UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

Financial Economics

Principal Component Analysis and the Cross-Sectional Variation of Returns

Authors: Armin Ramovic & Mikael Åkerman

Bachelor thesis 15 hp

Tutor: Marcin Zamojski

Spring term 2021

Abstract

We utilize Principal Component Analysis (PCA), a dimensionality reduction technique, on a set of 142 risk factors, including macroeconomic factors, proposed in financial literature to construct factor models with high explanatory powers when analysing the cross-sectional variation of portfolio returns. We apply a Fama and Macbeth (1973) two-pass regression to estimate risk premia commanded by our principal components. We perform a static PCA, which is what we call the conventional application of PCA, and a rolling window PCA, where the data is split into overlapping windows and where the PCA constructs principal components separately in each window. This allows us to construct two sets of factor models: one set from the static PCA, and one set from the rolling window PCA. The benchmark model for our research is the Fama and French (2015) five-factor asset pricing model. Our results suggest that both sets of factor models outperform the benchmark model in capturing the cross-sectional variation of returns. Furthermore, we find that the addition of macroeconomic factors adds explanatory power to our models. Finally, we find that the factor models from the rolling window PCA do not outperform the factor models from the static PCA.

Keywords: Principal Component Analysis, PCA, principal components, cross-sectional variation of returns, risk premia, asset pricing, dimensionality reduction, risk factors, machine learning

Acknowledgement

We would like to express our sincere gratitude toward our tutor Marcin Zamojski for his everlasting patience and continous support. Thank you for guiding us throughout the exciting field of machine learning and asset pricing, and for having evoked such an interest for these fields.

Table of Content

1. Introduction

There are many risk factors in financial research that seek to describe the cross-sectional variation of stock returns, such as the systematic risk factor beta from the Capital Asset Pricing Model (Treynor, 1962; Sharpe, 1964; Lintner, 1965a, b; Mossin, 1966), known as CAPM. In this paper, we look at 142 risk factors proposed in financial literature. It is possible that many of these factors have little to no contribution in explaining the cross-section of returns, either because of multiple hypothesis testing or because they duplicate the cross-sectional variation obtained in other factors. The aim of the essay is to mitigate this problem by applying Principal Component Analysis, or PCA for short, to reduce the dimensionality of the problem, such that we have a smaller number of factors called principal components. Our setting for the research replicates the one conducted by Kozak et al. (2018). We apply a Fama and Macbeth (1973) two-step regression to estimate risk premia commanded by our principal components. From this we obtain models to which we compare their explanatory powers to our benchmark model, namely the Fama and French (2015) five-factor asset pricing model.

We know from Kozak et al. (2018) that the principal component factors do in fact explain more of the cross-sectional variation of portfolio returns than the benchmark model. From our results we can confirm these findings. Also, we find that some of the principal components are significant. In our paper we look at whether the principal components that are extracted dynamically from our rolling window PCA specification can contribute more than the principal components extracted from what we refer to as the static PCA specification. Furthermore, we extend the models by including macroeconomic factors and looking at the impact of these non-investable factors. The inclusion of macroeconomic components is supported in the research conducted by Bianchi and McAlinn (2020). They apply macroeconomic variables to study the prediction of returns and found that the variables can contribute to predictions. Lastly, we look at more portfolios than in the research setting of Kozak et al.

From earlier research, as we will discuss in detail in our literature review section, we know that there are many ways to tackle the dimensionality reduction challenge in asset pricing. Feng et al. (2020) applies a double-selection LASSO regression, a variable selection and regularization technique, to reduce potential redundant factors in their high-dimensional setting. Kelly et al. (2018) expands on PCA by proposing a new modelling approach which is referred to as

Instrumented Principal Component Analysis, IPCA, to explain the cross-section of returns. Kozak et al., as well as our research, applies PCA to tackle the high-dimensionality challenge of asset pricing.

There is an important distinction to be made between our approach and the approach of Kozak et al. Our approach, and much of recent machine learning literature (Tsai et al. (2011); Huerta et al. (2013); Freyberger et al. (2017)), focus on estimating risk premia, while Kozak et al focus on estimation of risk prices. When estimating risk premia, we look at to what extent a risk factor is associated with variation in expected returns. When estimating risk prices, it is the extent to which the risk factor helps price assets by contributing to variation in the stochastic discount factor that is of interest (Kozak et al., 2018).

With all this in consideration, our research questions are as follows:

- Do the principal component factors extracted from the rolling window PCA explain more of the cross-sectional variation of our portfolio returns than the principal components extracted from the static PCA specification?
- Does the inclusion of macroeconomic factors increase the explanatory power of the rolling window PCA and static PCA models?

The hypothesis for the empirical analysis is that the principal component factors extracted from the rolling window PCA does in fact explain more of the cross-sectional variation of our portfolio returns than the principal components extracted from the static PCA specification. By allowing the components and the weights of the factors to vary over time, we should have a more comprehensive overview of the components' effect on stock returns and therefore be able to construct a factor model with higher explanatory power than the static PCA model. We also believe that the inclusion of macroeconomic factors will increase the explanatory power of the rolling window PCA models and the static PCA models. Even though we cannot invest in the macroeconomic factors, they do affect risk premia and stock returns and, therefore, will contribute to and correlate with many of the components in our models.

Our research contributes to an emerging literature that applies machine learning techniques to deal with the high-dimensionality challenge in asset pricing. In our literature review, we will look more closely on some of the literature that covers the topic of dimensionality. By conducting PCA in moving windows with time-varying components and macroeconomic factors, we believe that we add some discussion and perspectives regarding the topic of asset pricing. Furthermore, our contribution extends to the economic interpretation of the principal components and their effect on risk premia.

The remainder of the paper is structured as follows: in the next section we present a literature review where we provide an overview of current knowledge and position ourselves in relation to other researchers. We then look at the applied methodology, the design process, for our research. Here we will look at our approach when gathering data and techniques applied to transform and analyse the data. Then, we conduct our analysis on the results yielded from the static PCA specification. After this, we analyse the results yielded from the rolling window PCA specification. We then account for the robustness of our models by analysing how a separation of components yield different results. Finally, we draw a conclusion and discuss our findings and how further research can improve on our research.

2. Literature Review

In this section we discuss the theoretical assumptions that this paper builds upon. We will first look at the benchmark model as well as other factor models that attempts to capture the cross-sectional variation of stock returns. We then look at the topic of dimensionality reduction, and account for three papers which have had large influences on our research. Finally, we look at a proven method to estimate risk premia and discuss why we choose to apply it to our research.

2.1 Factor Models

Before we venture deeper into the discussion of the factor models, it is important to understand what the factors themselves provide for our research as well as previous research. The factors are used to establish time-series performances and observe whether the they yield abnormal rates of return, also referred to as alpha. The factors are also used to explain the cross-sectional differences in asset pricing. These dimensions allow the researcher to observe and establish certain patterns for certain factors in certain models. The high-dimensionality nature of asset pricing does not, however, allow the researcher to establish all patterns for all factors in all models. The analysis of the plethora of factors that are proposed in asset pricing literature are dependent on the researchers empirical setting, such as their applied methodology, reconstruction of certain factors, and their aim of the research. This means that different research settings will yield different results. There is one common element in many of these results, which is that the high-dimensional set of factors in asset pricing includes variables that add little to no contribution when explaining the cross-sectional variation of stock returns.

Our benchmark model, the Fama and French (2015) five-factor asset pricing model, is designed to capture profitability, size, value, and investment patterns in average stock returns and explain the cross-sectional variation of these returns. Building upon its predecessor, the five-factor model expands on the Fama and French (1993) three-factor model by constructing two additional factors referred to as profitability and investment. Profitability builds on the assumption that companies with higher future earnings have higher returns in the stock market, while investment suggests that companies which invest in major growth prospects are likely to experience losses in the stock market. Besides these two factors, the five-factor model and its predecessor include market risk, the outperformance of small versus big companies, known as SMB, and the outperformance of companies with high book-to-market ratios versus companies with small book-to-market ratios, known as HML. According to Fama and French's estimation, the five-factor model explains between 71 and 94 percent of the variation of expected returns for the portfolios they examine, where the results show that the highest expected returns in stock market are attained by companies that are small, profitable and value companies with no major growth prospects. The Fama and French five-factor model is chosen as the benchmark model because it outperforms the Fama and French three-factor model in capturing anomalies and explaining the variation of stock returns.

Besides the benchmark model, there are other models which expand on the Fama and French three-factor model in order to capture the evidence presented from the emergence of new factors. One such model is Carhart's (1997) four-factor model, which includes a momentum factor for asset pricing of stocks in addition to market risk, size, and value. The momentum factor, PR1YR, is constructed as the equal-weight average of firms with the highest 30 percent eleven-month returns lagged one month. It is important to note that the momentum factor was not discovered by Carhart, but rather documented by Jegadeesh and Titman (1993). Carhart, with the four-factor model specification, showed nevertheless that the momentum factor improved the explanatory power of the Fama and French three-factor model when explaining mutual funds' performance.

Another model is the *q*-factor model, introduced by Hou, Xue, and Zhang (2015). The *q*-factor model states that risk premia are captured by the market factor, a size factor, an investment factor, and a profitability factor. The contributions of the *q*-factor model are that it can largely summarize the cross section of average stock returns, and therefore capture many of the anomalies that proves challenging for many models. To evaluate the empirical performance of the *q*-factor model, the authors look at nearly 80 variables that cover all major categories of anomalies. Their examination of these variables shows that many claims in the literature regarding anomalies seem exaggerated and that the *q*-factor model outperforms the Fama and French three-factor model in capturing many of the significant anomalies.

2.2 Dimensionality Reduction

The topic of dimensionality reduction is not a new subject in financial literature, relatively speaking. In Stock and Watsons' (2002) research they conduct PCA in a high-dimensional macroeconomic setting to show that forecasts of a single series based on principal components of a large number of predictors can contribute to substantial improvement in forecasts beyond conventional models using a small number of variables.

Since this paper looks at the cross-sectional performance of low-dimensional datasets derived from a high-dimensional set of risk factors, the fundamental assumptions of this papers' theoretical framework build upon the research conducted by Kozak et al. (2018). In this paper, they use machine learning tools which are designed for high-dimensional settings to summarize the cross-section of expected returns in a low-dimensional factor model. From a primary dataset of 50 risk factors, they construct a stochastic discount factor, also referred to as the pricing kernel (Cochrane, 2001), that summarizes the joint explanatory power of the risk factors in explaining the cross-sectional variation of stock returns. Their results suggest that the quest to summarize the cross-section of returns with factor models containing only a few risk factors is futile as there is not enough redundancy among their high-dimensional dataset for such a factor model to adequately price the cross-section.

The topic of dimensionality reduction has further been covered by Kelly et al. (2018), where their research proposes a different modelling approach which expands on using PCA to estimate the factors and betas, and also expands on estimation of betas and alphas via regressions. They refer to this method as Instrumented Principal Component Analysis, or IPCA. From a dataset of 36 risk factors, they apply ICPA to find a low-dimension factor model which, according to the results, successfully explains cross-sectional variation in average returns. Furthermore, the results suggest that their factor model outperforms the Fama and French five-factor model.

Another paper that tries to tackle the high-dimensionality challenge with an alternative approach to ours is the paper from Feng et al. (2020), which looks at the marginal importance of any factor in a low-dimensional dataset in pricing the cross-section of expected returns relative to a high-dimensional dataset. Since this high-dimensional dataset consists of potentially hundreds of factors, the issue of reducing any potential redundant factor leads the authors to apply the double-selection LASSO method of Belloni et al. (2014). Then, by applying the Fama and Macbeth (1973) two-pass regressions, the paper tests the explanatory power of the low-dimensional dataset on asset prices in a high-dimensional setting. Their results suggest that by applying their tests over time, only a small number of risk factors proposed in the literature would be deemed significant.

Feng et al.'s approach of applying regularization techniques for variable selection could be an interesting approach to apply when conducting future research on dimensionality reduction. Park (2013) in his research proposed a methodology where PCA and Lasso-regression, a regularization technique, is combined. The first step entails utilizing PCA to extract principal components. The second step is to regress each component by Lasso-regression method. Each new principal component is computed as the linear combination of the original data set using the scaled estimated Lasso-regression coefficient as the coefficient of the combination. The results from Park's research shows that the methodology leads to easily interpretable principal components with more 0 coefficients by the properties of the Lasso-regression model. The reason for this is that the estimator of the regression of each component on the original data set is the corresponding eigenvector. We will cover the topic of eigenvectors in our methodology section.

2.3 Fama and Macbeth Estimation for Risk Premia

The Fama and Macbeth (1973) two-pass regression is a common method to estimate risk premia for our asset pricing models. We delve deeper into the technical aspects and specification of the regression in our methodology section. In this section, we will account for why the Fama and Macbeth regressions are the chosen method to estimate the risk premium for our principal components. Simply put, Fama and Macbeth regressions are a practical way of testing how the components describe portfolio returns.

Risk premia can also be retrieved by Ordinary Least Squares, or OLS for short. The difference between the Fama and Macbeth regressions and the OLS is that the Fama and Macbeth specification uses the estimated betas retrieved from the time-series regression to perform a cross-sectional regression for each observation date. The estimated premium is then the average of the estimate from each cross-sectional regression. OLS, in contrast, performs a single cross-sectional regression if the betas are observable to calculate risk premia (Kan & Zhang, 1999). Since betas are not observable in practice, we decide to apply the Fama and Macbeth two-step regression to estimate our betas and to retrieve the risk premia commanded by our principal components.

3. Methodology

3.1 Data

In our empirical analysis we use a set of standard portfolios of U.S. equities. We choose to analyse U.S. equities because of the quality of the data and the availability of data for our chosen timeframe.

In total, we collect 142 risk factors. In Appendix A we report a list of these factors. The factors are collected at a monthly frequency for the period from November 1973 to December 2019. We use Kenneth French's data library to obtain the following factors: excess market return, size, value, momentum, investment, and profitability. We download 55 factors from Serhiy Kozak's database, some of which are duplicates from Kenneth French's dataset. The duplicates pose no problem for PCA, as PCA only includes the variables that capture unique variation of the dataset. We include the following factors from the AQR data library: Betting-Against-Beta and HML Devil. We use Federal Reserve Economic Data, FRED for short, to obtain the following eight macroeconomic variables: Growth in Money Supply, Lagged Excess Returns from the S&P500 index, Relative T-Bill Rate, Chicago FED National Activity Index, Total Consumption Growth, Consumption Growth Rate in Credit Goods, Monthly Growth in Industry Production, and Consumer Price Index based Inflation. The ninth macroeconomic variable, CAY- consumption, aggregate wealth, and expected stock returns, is retrieved from Martin Lettau's database. We retrieve all available financial ratios from the Compustat and CRSP merged data library. These ratios are constructed by data including only common stocks of companies listed on the NYSE, AMEX, or NASDAQ and are a set of commonly used ratios by academic researchers.

Before the empirical analysis, each factor in the dataset is standardized. The standardization is performed to transform the factors that use different measurements or scales. This means that the factors are assigned equal importance after the standardization. When factors with different measurements are not standardized, the principal components can be dominated by one single factor. When all factors contributing to the component is measured in the same scale, we can make interpretations based on the weights from the factors.

3.2 Test Assets

We focus on a range of portfolios rather than individual assets as characteristic sorted portfolios are less prone to missing data issues, have more stable betas, and higher signal-to-noise ratios. (Feng et al., 2020).

We use a total of 46 portfolios as test assets. We start from a set of 36 portfolios, which are constructed monthly by intersecting two portfolios formed on size and three portfolios formed on the following characteristics: book-to-market ratio, operating profitability, investment, long-term reversal on prior return, short-term reversal on prior return, momentum on prior return. We also add the standard ten industry portfolios, which are constructed by assigning each NYSE, AMEX, and NASDAQ stock to a value weighted industry portfolio. This set of test assets captures a large cross-section of anomalies and exposures to different factors and are all available from Kenneth French’s website.

3.3 Computational Methods

3.3.1 Principal Component Analysis

Principal Component Analysis, or PCA, is a technique of multivariate analysis that allows us to summarize the information content of a high-dimensional data set by means of a smaller set of uncorrelated variables known as principal components. Principal components are constructed as linear combinations of the initial factors. These combinations are structured in a way where the first few components retain most of the variation present in all the original variables. In other words, PCA attempts to put as much information from the original variables as possible in the first principal component, and then attempts to put as much information in the second principal component and so on. This continues until the total principal components have been calculated, where the number of components used is decided by the researcher. (Jolliffe, 2002).

There are different methods when choosing the number of components to be retrieved for the research. The researcher can try to achieve a specific number of variance explained by the components, discard components with eigenvalues below a certain set figure, or visually inspect the eigenvalues in a plot to see when eigenvalues start to become substantially small.

3.3.1.1 Static PCA

Conventionally, PCA is applied as a static method. We therefore refer to it as the static PCA to emphasize the structural differences between it and the PCA conducted in rolling windows, which we refer to as rolling window PCA. To retrieve static principal components, we compute a covariance matrix for the standardized dataset:

$$C = \left(cov\left(x_{ij}, y_{ij} z_{ij}\right)\right).$$

This is an example of a covariance matrix with three factors x, y, and z, with each entry in the matrix resulting from the calculation of the covariance between two separate dimensions (Smith, 2002). The eigenvectors, which are the directions of the axes where there is the most variance and whose directions remain unchanged when a linear transformation is applied to them. Eigenvectors can only be found on square matrices, such as the covariance matrix, and not all matrices have eigenvectors. Each eigenvector is paired with an eigenvalue, which describes the amount of variance carried in each principal component.

If we let C be our covariance matrix, a square matrix, ν a non-zero vector and λ a scalar that satisfies

$$C\nu = \lambda\nu,$$

then λ is called eigenvalue associated with eigenvector ν of matrix C. The eigenvalues of C are roots of the characteristic equation

$$\det(C - \lambda I) = 0,$$

where det is the determinant, the scalar value which is a function of the entries in C, and I is an identity matrix, a square matrix with only ones in the main diagonal and zeros elsewhere

$$I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

By simplifying $(C - \lambda I)$, we can find the determinant and solve for λ, so that we get the eigenvalues of C. This allows us to calculate the eigenvectors corresponding to the retrieved eigenvalues. Then we rank the eigenvectors based on its eigenvalues, highest to lowest, to get the principal components in order of significance. (Andrew, 1973; Jolliffe, 2002).

With the help of a scree plot, a line plot of the eigenvalues of the principal components, we determine the total number of components that will be used for our regression analysis. The determination is done by examining the scree plot to find a kink, where the next component adds a substantially smaller amount of the total variation in the dataset. We then select all the components up until this kink and apply these in our analysis.

For our analysis, we apply one PCA for our financial factors and one separate PCA for our macroeconomic variables. The reasoning behind this is that we wish to study how the loadings of the components change with and without the extension of macroeconomic variables. The choice of separating the components is also a way to increase the interpretability of the components. By separating, we can analyse the impact of macroeconomic factors on the components. The reason we wish to analyse the impact of these factors is because even though it is not possible to invest in a portfolio of macroeconomic factors, these factors can still affect risk premia.

3.3.1.2 Rotation of components

For the sake of interpretability, we rotate the factors using the Varimax criterion to perform our rotation. By applying the Varimax criterion, components are orthogonally rotated so that a single component is marked by a restricted set of factors with high contributions while all the remaining factors have zero or negligible contributions to this component. This is referred to as "simple structure." The Varimax maximizes loadings of variance on each component. This procedure is performed by searching for a linear combination that maximizes:

$$v = \sum(q_{j,l}^2 - \bar{q_{j,l}^2})^2,$$

where $q_{j,l}^2$ is the squared loadings of variable j on factor l, and $\bar{q_{j,l}^2}$ is the mean of the squared loadings. After the rotation, we observe that there are only a few factors which have significant contributions to the components. This allows us to create meaningful interpretation of relationships of components to the original data set. (Rösler & Manzey, 1981).

3.3.1.3 Rolling Window PCA

The PCA with rolling window uses a similar procedure as the static version with one exception. The analysis introduces time variation in factor weights. The data is split into different windows of five years, or 60 months, length. Within each window we conduct a conventional PCA according to the static PCA method. Each window has a set of components with unique weights of factors within them. The design of the rolling window rolling forward one month at each step gives us a set of principal components for each month.

When the PCA is conducted in one window there is no certainty that it will be the same in the next window. To match the components, we assume that the components are the same over time. This is a limitation of our approach which future research can improve upon. This means that component one in the first window will be matched with component one in the second window. We perform correlation tests to control if the components have "flipped signs. When a component is "flipped" it means that its sign is changed. It also means that the signs of the weights from the factors in the components are changed. If we notice that the signs have changed, we "flip" the components before we merge them. We then merge the components by using the first window of 60 months and then add the estimated components from the next month for each step in the rolling regression.

3.3.2 Fama-Macbeth Regression

Following the selection of the principal components, we estimate factor risk premia. This is done by applying the Fama and MacBeth (1973) two-pass regressions. The first step of the approach is time-series estimation of the beta factor for each portfolio *i*:

$$R_{i,t} = a_i + \beta_{i,pc1} pc1 + \cdots + \beta_{i,pc_k} pc_k + \epsilon_{i,t},$$

where R is the return of portfolio i and the right-hand variables are the risk factors, i.e. the principal components. In our case the estimation is done in moving window of 60 months.

In the second step, we run a cross-sectional regression at each period of time T over all portfolios *i*:

$$R_{i,t} = \widehat{\Upsilon}_t \lambda_t + a_{it} + \delta$$

where $\widehat{\Upsilon}_t$ is the estimated beta factor of the portfolio, λ_t is the risk premium of our risk premia β_i, a_{it} is the pricing error, which is the part of the return $R_{i,t}$ unexplained by our risk premia β_i, δ is the intercept, i.e. the mean of the pricing error. We include the intercept as some of our factors are not investable.

This results in estimates for λ_t and α_{it} for each period of time *T*. Fama and Macbeth (1973) suggest that we estimate $\hat{\lambda}$ and $\widehat{a_i}$ as the average of these cross-sectional regression estimates:

$$\hat{\lambda} = \frac{1}{T}\sum_{t=1}^{T} \lambda_t\,, \qquad \hat{a}_i = \frac{1}{T}\sum_{t=1}^{T} \alpha_{it}$$

$\hat{\lambda}_t$ is the estimate of the risk-premiums and $\hat{a}_i$ is the estimated pricing error for each cross-sectional regression. If there is a linear relationship between our risk premiums β_i and portfolio returns in period *t*, then by applying a cross-sectional regression at each period *t* of time *T* we should retrieve statistically significant risk premiums $\hat{\lambda}_t$.

4. Static PCA

In this section we look at the results yielded by the static PCA specification from our empirical analysis. We will first discuss our approach of selecting the number of components applied to our analysis. Then we look at the outputs retrieved from the unrotated components and the analysis regarding these outputs. We then move on to interpret and analyse the outputs yielded from the rotated components.

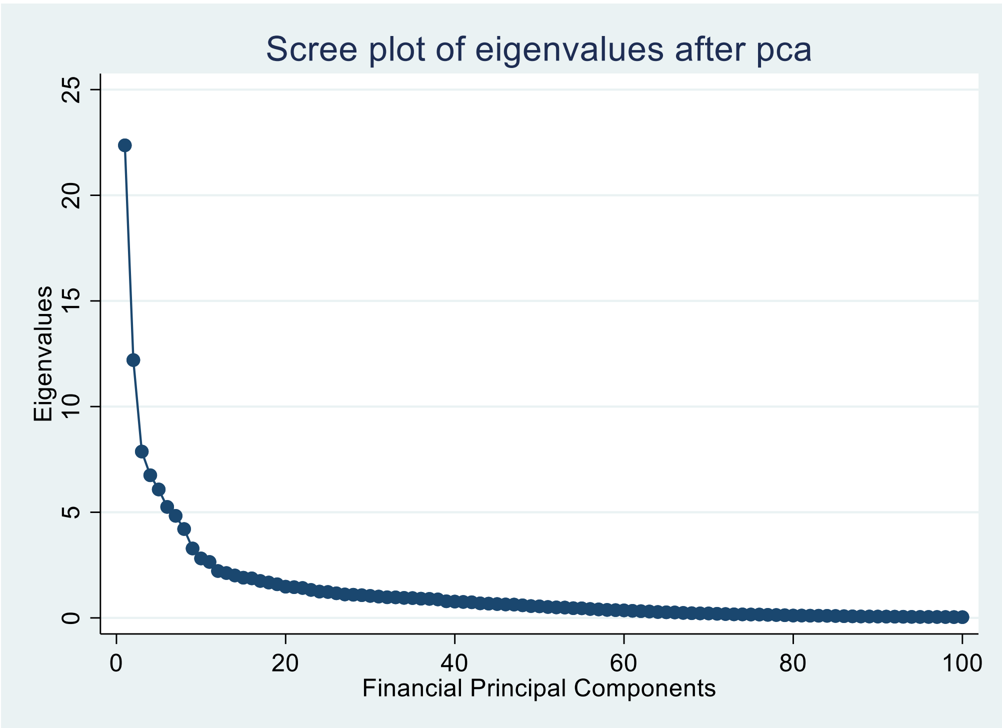
4.1 Unrotated Components

The analysis will start off by looking at the outputs from the unrotated components. However, the interpretation of the principal components will focus on the rotated components. This is related to our discussion regarding rotations in our methodology section.

4.1.1 Component Analysis

To determine the number of financial components to extract, we look at the following scree plot:

Figure 1. Scree plot of eigenvalues after PCA on financial factors



The scree plot shows us that there is a kink between the eighth and ninth component. We decide to extract ten components because there is some distance between the ninth and tenth component that we deem significant. As we mentioned in our methodology discussion, there is no correct number of components one should extract. Rather, the researcher decides based on the aim of the research and their understanding of the marginal contribution that omitted components would bestow upon the research. For example, Kozak et al. (2018) describes that the characteristic-sorted portfolios could be explained by only two components. This then captures two thirds of the variation in their test, but it can also be described by ten components which captures almost all of the variation in their test.

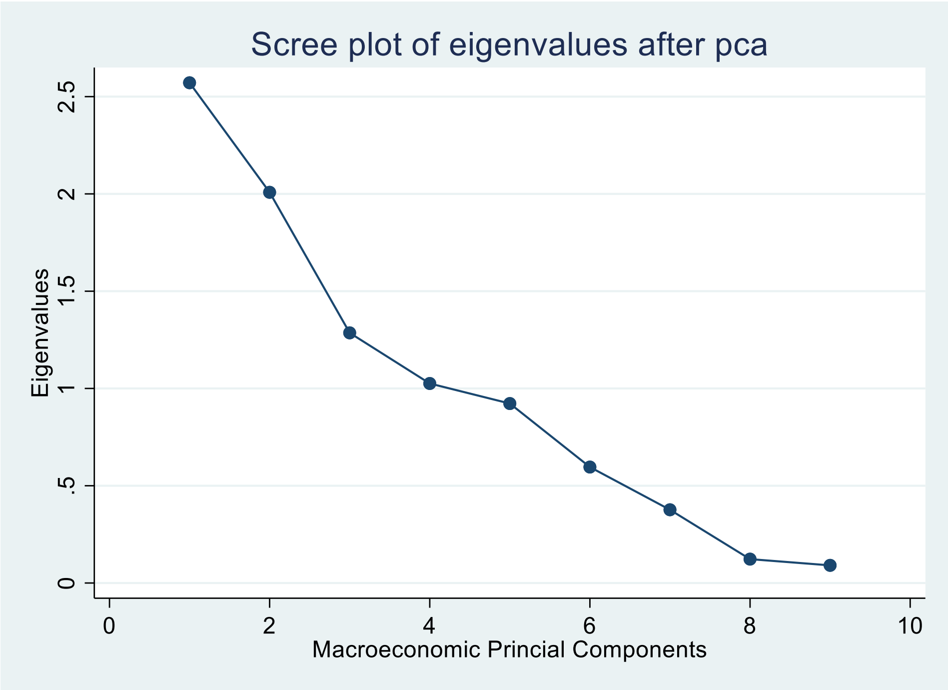
We consider that the marginal contributions of the components computed after the tenth component are not significant for our research. This could be a topic for future research, as the difference in extracted principal components can yield different results. In Table 1 and 2 we illustrate eigenvalues attained from each component and the cumulative variation retained with our selected components

Table 1. Eigenvalues and cumulative variation - Financial components

Component	Eigenvalue	Difference	Proportion	Cumulative
Comp1	22.363	10.162	0.168	0.168
Comp2	12.201	4.329	0.092	0.260
Comp3	7.873	1.122	0.059	0.319
Comp4	6.751	0.672	0.051	0.370
Comp5	6.079	0.827	0.046	0.416
Comp6	5.253	0.422	0.040	0.455
Comp7	4.831	0.624	0.036	0.491
Comp8	4.207	0.921	0.032	0.523
Comp9	3.285	0.471	0.025	0.548
Comp10	2.814	0.166	0.021	0.569

To determine the number of macroeconomic components to extract, we look at the following scree plot:

Figure 2. Scree plot of eigenvalues after PCA on macroeconomic factors



The scree plot shows us that there is a kink between the fifth and sixth macroeconomic component. We decide to extract five components. The gap between the fifth and sixth component is significantly large. This means that we consider the sixth component and beyond to have insignificant contribution to our research. Table 2 shows that five macroeconomic components captures approximately 87% of the variation of the macroeconomic factors.

Table 2. Eigenvalues and cumulative variation - Macroeconomic components

Component	Eigenvalue	Difference	Proportion	Cumulative
Comp11	2.571	0.563	0.286	0.286
Comp12	2.008	0.722	0.223	0.509
Comp13	1.286	0.260	0.143	0.652
Comp14	1.026	0.103	0.114	0.766
Comp15	0.923	0.326	0.102	0.868

4.1.2 The Benchmark Model and unrotated Static Factor Models

Table 3. First stage mean adjusted R-Squared

First Stage Mean Adjusted R-Squared	
Fama-French 5-factor model	0,900
Unrotated Static PCA 10-factor model	0,827
Unrotated Static PCA 10+5-factor model	0,833

We will now look at the explanatory power of the benchmark model and our unrotated static factor models after conducting the first stage in the Fama and Macbeth regressions, the time-series regression. The explanatory power of the models is measured by the adjusted R-squared value. The adjusted R-squared value reflects how much variation the factors and components capture in the portfolio returns.

The values in Table 3 are the mean adjusted R-squared values. The values observed are in line with Fama and French's (2015) reported values and indicate that the five-factor model performs as intended in our test. What is of interest to us, however, is the second stage of the Fama and Macbeth regressions, the cross-sectional regression, and to observe how the benchmark model compares to our static factor models.

Table 4. The benchmark model and the unrotated static factor models

Fama-French 5-factor model				Unrotated Static PCA 10-factor model				Unrotated Static PCA 10+5-factor model		
Factors FF5	Coef,	St,Err,	p-value	Factors	Coef,	St,Err,	p-value	Coef,	St,Err,	p-value
Mkt-Rf	-0,032	0,06	0,592	PC 1	0,203	0,231	0,382	-0,226	0,234	0,335
SMB	0,048	0,044	0,275	PC 2	0,555(***)	0,102	0	-0,696(***)	0,095	0
HML	-0,002	0,047	0,969	PC 3	-0,007	0,137	0,961	0,016	0,14	0,909
RMW	0,094 (*)	0,05	0,062	PC 4	0,018	0,134	0,892	0,047	0,128	0,712
CMA	0,07	0,054	0,192	PC 5	-0,032	0,11	0,772	0,209(**)	0,106	0,048
				PC 6	0,147	0,172	0,392	0,024	0,174	0,89
				PC 7	0,454(***)	0,14	0,001	0,487(***)	0,147	0,001
				PC 8	0,025	0,179	0,891	0,031	0,172	0,856
				PC 9	-0,352(***)	0,16	0,029	-0,315(**)	0,131	0,016
				PC10	0,148	0,097	0,13	0,195(**)	0,094	0,039
				PC 11				-0,35(***)	0,079	0
				PC 12				-0,311(***)	0,116	0,008
				PC 13				-0,009	0,061	0,889
				PC 14				-0,046	0,068	0,496
				PC 15				-0,026	0,077	0,735
Constant	0,012(***)	0,002	0	Constant	0,008(***)	0,002	0	0,009(***)	0,001	0
R-squared			0,5616			0,7321	0,7321			0,8229

*** p<0,01, ** p< 0,05, * p< 0,1,

Table 4 shows the output after the second stage of Fama and Macbeth regressions from the benchmark as well as the unrotated outputs from our static factor models. Here we list the coefficients and their respective significance levels for the benchmark model and our two unrotated static factor models. From the static factor model, we report an adjusted R-squared value of 0.7321. We do see an increase in the adjusted r-squared value when extending the static factor model with macroeconomic components. From the extended static factor model, we report an adjusted R-squared value of 0.8229. Furthermore, the inclusion of macroeconomic components raises the components' significance levels. This indicates that the macroeconomic factors correlate with the financial components and affect the risk premia or risk discount commanded by the components.

We also notice that the benchmark model yields an adjusted R-squared value of 0.5616. This is a lower value than expected and lower than the static factor models. This is an interesting aspect of the Fama and French five-factor model, as we can clearly see in Table 3 that the model better explains the time-series patterns of average returns than the static factor models. However, our results suggest that benchmark model does not explain the cross-sectional differences of the performance of our assets better than the static factor models, which indicates that our principal components are better equipped to capture the cross-sectional variation of returns. One reason for this could be the difference in the set of portfolios applied to the analysis. In their research, Fama and French (2015) looks at 25 characteristic-sorted portfolios. We know from Lewellen et al. (2010) and Kozak et al. (2018) that these portfolio returns have such a strong factor structure that the portfolio returns are close to being linear combinations of the HML and SMB factors. This means that the five-factor model does not necessarily capture descriptions of average returns as well when faced with a high-dimensional setting.

Furthermore, we notice that the benchmark model's factors have a low significance level when compared to the model with unrotated components. The only factor in the benchmark model that is significant is the Robust-Minus-Weak factor, which measures profitability. The unrotated components of the static PCA yield higher relative significance when compared to the benchmark model, and three out of the ten components are significant.

From the analysis of the outputs from Table 4, we can conclude that the unrotated components in our models are more significant than the factors in the benchmark model. The static factor models have higher explanatory power and benefit from the inclusion of macroeconomic components.

Examining adjusted value of the R-squared, it is important to note that this metric punishes the application of more factors in the statistical test (Jaggia & Kelly, 2016). We drop factors that are not able to capture unique variation in the data set that is used, thereby adding only information of use in capturing the variance of our dataset. When the adjusted R-squared value is higher, it indicates that the expanded static model has smaller differences between the observed data and the fitted value compared to the other models, and therefore has higher explanatory power compared to the other models. The expanded static model can therefore explain a larger part of the cross-section of stock returns. The results are in line with the findings of Kozak et.al (2018), where they find that sparse factor models offer lower R-squared values, explaining less of the cross-sectional variation of returns, when compared to the principal component factors, which display higher R-squared values and explains a larger amount of the cross-sectional variation of returns than the benchmark model.

4.2 Rotated Components

Before we interpret the rotated components, we wish to present an observation we make after rotating the components.

Table 5. Unrotated and rotated static factor models with macroeconomic components

Unrotated Static PCA 10+5-factor model				Rotated Static PCA 10+5-factor model			
Factors	Coef,	St,Err,	p-value	Factors	Coef,	St,Err,	p-value
PC 1	-0,226	0,234	0,335	PC 1	0,299	0,217	0,17
PC 2	-0,696(***)	0,095	0	PC 2	0,488 (***)	0,076	0
PC 3	0,016	0,14	0,909	PC 3	0,395 (***)	0,11	0
PC 4	0,047	0,128	0,712	PC 4	0,182	0,142	0,199
PC 5	0,209(**)	0,106	0,048	PC 5	0,235 (*)	0,13	0,071
PC 6	0,024	0,174	0,89	PC 6	0,240 (*)	0,131	0,066
PC 7	0,487(***)	0,147	0,001	PC 7	-0,204	0,192	0,29
PC 8	0,031	0,172	0,856	PC 8	0,166	0,186	0,373
PC 9	-0,315(**)	0,131	0,016	PC 9	-0,021	0,103	0,836
PC10	0,195(**)	0,094	0,039	PC10	-0,509 (***)	0,13	0
PC 11	-0,35(***)	0,079	0	PC 11	0,218 (***)	0,065	0,001
PC 12	-0,311(***)	0,116	0,008	PC 12	-0,418 (***)	0,123	0,001
PC 13	-0,009	0,061	0,889	PC 13	-0,006	0,066	0,926
PC 14	-0,046	0,068	0,496	PC 14	0,005	0,078	0,948
PC 15	-0,026	0,077	0,735	PC 15	-0,017	0,065	0,799
Constant	0,009(***)	0,001	0	Constant	0,009(***)	0,001	0
R-squared			0,8229				0,8229

*** p<0,01, ** p< 0,05, * p< 0,1,

Studying the output from the rotated components relative to the unrotated components we can observe that there are some changes of the component's coefficients and significance levels. These changes are expected since the simple structure of the components, because of the Varimax rotation, alters how the components are constructed. One thing to note is that the components in the unrotated model are not same as the components in the rotated model. PC 1 in the unrotated model does not have the same loadings as PC 1 in the rotated model. This is because of the Varimax rotation and the simple structure we retrieve from the rotation.

Furthermore, Kozak et al. (2018) propose that future research investigates the economic interpretation on a selected number of components. The choice of rotation in our research are done for the purpose of interpretability. In the next section we interpret the rotated components. In Table 6, we name the components based on factor contributions.

4.2.1 Interpretation of Ten Financial Components

Here we interpret the factor loadings in the ten rotated financial principal components. The weights are available in Appendix B, including weights of factors in the macroeconomic components. A close inspection of the weights allows us to interpret the principal components. The standardization ensures that the weights and our interpretation of the weights are meaningful. The interpretation is done by looking at which factors contribute most to each respective component, i.e. what factors have the largest weights, in absolute values, for each component. We start off the interpretations by first discussing the components with factor weights which are relatively less complicated in their interpretation, and then move on to the components with factor weights which are more complicated to interpret. What we mean by this is that, for some components, the factors which contribute the most are related to the same category of risk factors, such as market excess returns or liquidity ratios, which makes it less complicated to interpret, while others receive contribution from a variety of factors, making interpretation more complicated.

One of the less complicated components to interpret is the third component, which we label as "*Liquidity.*" The factors with the highest positive contribution are associated with companies' short-term liquidity. The factors having a negative contribution to the component are cash flow to debt ratios. The weights are interpreted as that companies with large cash holdings and high short-term debt will increase the coefficient and companies with small cash holdings and low short-term debt will decrease the coefficient.

The fourth component, "*Market Return*", is mostly connected to market returns in various forms. These include the equal- and value-weighted industry excess returns as well as equal- and value-weighted excess market returns. This component is contributed by factors sorted in portfolios. We thereby interpret the component as representing market return premium.

We interpret the fifth component as "*Profitability*." The factors contributing most to the fifth component is mainly related to profitability through returns on assets and returns to equity. The sixth component, "*Momentum*", receives most contribution from different momentum-based factors. The Factor having most contribution is "Value-Momentum-Profitability" factor, which sums ranks of book-to-market, momentum, and profitability. This factor is followed by differently constructed factors which capture momentum of different horizons. The size factor contributes most to the component in negative terms. This represents going long in a portfolio with a high vale-momentum-profitability characteristic and going short in a portfolio with companies that have a high size factor.

Component number seven is referred to as "*Margin*." This component gets most contribution from several profit margin measures, such as net profit margin or cash flow margin. Higher profit margin ratios are viewed as having a positive effect on excess returns, which would attain a premium based on the margin.

We label the eighth component as "*Leverage*." The factors that have the highest positive contribution to this component are mainly debt to invested capital and sales to invested capital. The factor with the largest negative contribution is the common equity to invested capital. Component ten, which we interpret as "*Earnings*", is the final component among the most easily interpretable. This component gets the most contribution from price to earnings and interest coverage. Higher interest coverage might indicate untapped growth potential if a company does not invest in new projects. If the companies' have stocks that are overvalued relative to the companies' earnings, this will show in a negative coefficient in the test. If the companies are undervalued relative to earnings this will show a positive coefficient.

We now look at the more complicated components to interpret. The first component is labelled as "*Value Premium*." We notice that the risk factors with the most contribution, in absolute values, to this component focus on the relationship of companies' book value of equity to their market value of equity. Investing in a portfolio that is constructed from companies which are valued higher and that have more leverage is expected to yield a risk premium. Based on the factor weights for this component, we see that the component is mostly affected by value premium.

We interpret the second component as "*Profitability & Value*." Some of the factors contributing most, in absolute values, to the second component are related to profitability through stock price to cash flow and stock price to sales and to value premium through debt to assets. Observing the rotation matrix, this component is particularly hard to interpret as it is seemingly being affected by leverage ratios. These leverage ratios can indicate whether the companies are exposed to value stocks or growth stocks and connected to value premium. Since this component and the first component are both related to value premium, we test the factors by creating the variance inflation factor, which measures the correlation and the strength of the correlations to control for high correlation between the components and multicollinearity (O'Brien, 2007). We then conclude that the components "*Value Premium*" and "*Profitability &*

Value" are not highly correlated. That is because there are several other factors that are contributing enough to both components to offset the potentially high correlation between these components.

Lastly, we refer to component nine as "*Pricing*." This component contains several factors of the Fama and French five-factor asset pricing model, albeit in different forms. The component gets a substantial contribution from the Small-Minus-Big factor and the gross profitability factor. This component is also contributed by the value profitability factor. We interpret component nine as reflecting the five-factor asset pricing models factors for estimating variations in the cross-section of returns.

4.2.2 Interpretation of Five Macroeconomic Components

Here we account for the factor loadings in the five macroeconomic components. The rotated components score matrix containing the weights of factors in each component is found in Appendix B. The interpretations of these components proved less complicated than for the ten factor components.

We denote first macroeconomic principal component in our dataset as “*Market Premium*.” The largest factor loadings are associated with market risk premia. The factor with highest positive contribution on the component is lagged excess return of the S&P 500 index. The components with the highest negative contribution are the risk-free rate and inflation factors. We can draw comparisons between this component and the fourth factor component “*Market Return*” as they both are largely contributed by the market risk premia. We once again test the factors by creating the variance inflation factor and find that they are not highly correlated. The reason for this is the same as for “*Value Premium*” and “*Profitability & Value*”: both components have several other factors that are contributing enough to offset potential high correlation between these components.

The second macroeconomic component is referred to as “*Business Cycle*.” The component is contributed mainly by an index, the Chicago Fed National Activity Index (CFNAI), which is designed to measure overall economic activity and inflationary pressure. The third macroeconomic component is referred to as “*Consumption*.” The component gets most contribution from changes in consumption related to asset wealth, labour income, and inflation. We interpret the fourth macroeconomic component as “*Inflation*.” The component is mainly contributed by the growth in real money supply. Finally, the fifth macroeconomic component is referred to as “*Lending*”, as the component is mainly contributed by the growth in consumption of credit goods.

4.2.3 Statistical Significance

Table 6. The rotated static factor models

	Rotated Static PCA 10-factor model			Rotated Static PCA 10+5-factor model		
Factors	Coef,	St,Err,	p-value	Coef,	St,Err,	p-value
Value Premium	0,281	0,215	0,193	0,299	0,217	0,17
Profitability & Value	0,507 (***)	0,082	0	0,488 (***)	0,076	0
Liquidity	0,122	0,115	0,289	0,395 (***)	0,11	0
Market Return	0,146	0,145	0,315	0,182	0,142	0,199
Profitability	0,181	0,135	0,181	0,235 (*)	0,13	0,071
Momentum	0,194	0,13	0,137	0,240 (*)	0,131	0,066
Margin	-0,066	0,183	0,718	-0,204	0,192	0,29
Leverage	0,198	0,195	0,311	0,166	0,186	0,373
Debt	0,012	0,105	0,907	-0,021	0,103	0,836
Earnings	-0,488 (***)	0,156	0,002	-0,509 (***)	0,13	0
Market Premium				0,218 (***)	0,065	0,001
Business Cycle				-0,418 (***)	0,123	0,001
Consumption				-0,006	0,066	0,926
Inflation				0,005	0,078	0,948
Lending				-0,017	0,065	0,799
Constant	0,008(***)	0,002	0	0,009(***)	0,001	0
R-squared			0,7321			0,8229

*** p<0,01, ** p< 0,05, * p< 0,1,

In Table 6, we notice that the rotated static ten-factor model has two significant components, namely "*Profitability & Value*" and "*Earnings*", while the rest of the components have insignificant risk premia. The significant components are interpreted as a premium or a discount, *Profitability & Value* premium and *Earnings* discount. By investing in companies with high market capitalization, the investor can expect a "*Profitability & Value*" premium. The analysis of *"Earnings"* component indicates that investing in companies with high price to earnings and interest coverage ratio, the investor can expect an "*Earnings*" discount.

We also see some clear changes to the components level of significance when our static regression model is extended with macroeconomic components. One notable change is that the component "*Liquidity*" goes from having an insignificant risk premium in the factor model without macroeconomic factors to a significant risk premium when the model is extended. This is because the macroeconomic factors in the in the macroeconomic components have a substantial effect on how companies utilize their cash holdings and how companies structure their debt financing. For example, inflation uncertainty has been found to reduce the number of investment projects financed by issuing debt and may also reduce the number of capital investment projects that the firm undertakes because it increases interest rate uncertainty

(Hatzinikolaou et al., 2002). By investing in companies with large cash holdings and high short-term debt, the investor can expect a “*Liquidity*” discount in returns.

We also see that “*Momentum*” and “*Profitability*” have higher significance when extending the model. The effect of the macroeconomic variables is not as drastic for the remaining set of principal components. We do observe that all ten principal components' significance level from the non-extended model obtain lower p-values. This does not mean that all components become significant.

On the extended model, we conclude that “*Market Premium*” and “*Seasonality*” are significant macroeconomic components. The analysis of the macroeconomic components is distinguished from the other factors. When investing in companies’ or portfolios with high correlation with the *“Market premium”* component, the investor can expect a premium. Analysing the *“Seasonality”* component, we conclude that investing in companies’ or portfolios with high seasonal dependence or correlation with business cycles, the investor can expect a *“Seasonality”* discount.

We draw comparisons between the components connected to market risk premium, namely “*Market Return*” and “*Market Premium*.” We see that “*Market Return*” is not significant. This is a surprising result as we, before conducting the PCA, hypothesized that it would be significant since the factors with the largest contributions are associated with excess market returns. We did establish earlier that the “*Market Return*” component is ‘contaminated’ by other factors which are not excess market returns.

5. Rolling Window PCA

Table 7. Unrotated static factor model and rolling window factor model

Unrotated Static PCA 10+5-factor model				Rolling Window PCA 10+5-factor model			
Factors	Coef,	St,Err,	p-value	**Factors**	Coef,	St,Err,	p-value
PC 1	-0,226	0,234	0,335	PC 1	0,807 (*)	0,476	0,091
PC 2	-0,696(***)	0,095	0	PC 2	-0,15	0,42	0,721
PC 3	0,016	0,14	0,909	PC 3	0,192	0,38	0,613
PC 4	0,047	0,128	0,712	PC 4	0,075	0,325	0,818
PC 5	0,209(**)	0,106	0,048	PC 5	0,063	0,291	0,828
PC 6	0,024	0,174	0,89	PC 6	-0,137	0,297	0,644
PC 7	0,487(***)	0,147	0,001	PC 7	-0,035	0,266	0,895
PC 8	0,031	0,172	0,856	PC 8	-0,445 (*)	0,241	0,066
PC 9	-0,315(**)	0,131	0,016	PC 9	0,335	0,223	0,135
PC10	0,195(**)	0,094	0,039	PC10	-0,504 (**)	0,214	0,019
PC 11	-0,35(***)	0,079	0	PC 11	-0,044	0,192	0,818
PC 12	-0,311(***)	0,116	0,008	PC 12	-0,495 (***)	0,185	0,008
PC 13	-0,009	0,061	0,889	PC 13	-0,069	0,119	0,563
PC 14	-0,046	0,068	0,496	PC 14	-0,02	0,095	0,837
PC 15	-0,026	0,077	0,735	PC 15	-0,04	0,101	0,69
Constant	0,009(***)	0,001	0	Constant	0,012(***)	0,002	0
R-squared			0,8229				0,7841

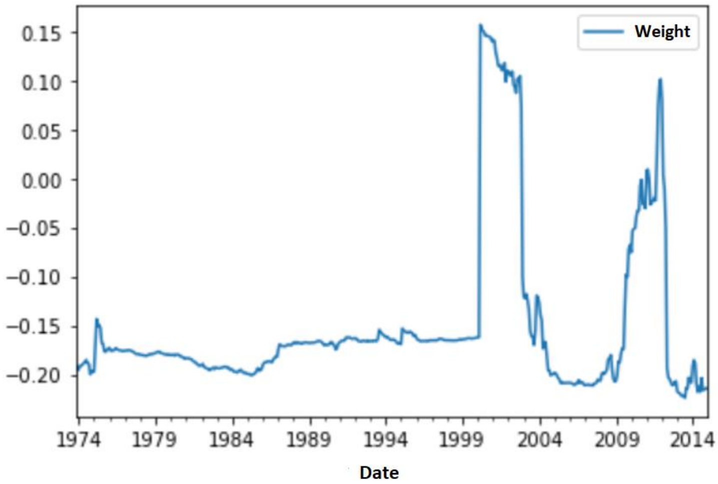
*** p<0,01, ** p< 0,05, * p< 0,1,

The Rolling Window PCA is conducted in separate windows, constructing a set of components for each window. Since the PCA is conducted on each window, the components can be of different values in each window. We mitigate this by matching the components by correlation to control for changes in signs of factor loadings over time. The matching procedure is done by absolute correlation. We opt to use the same number of components as in our Static PCA for transparency and comparability. As the windows consist of different loadings of our factors, the number of components might not be as suitable for the rolling window factor model specification as for the static factor model specification. This can be a topic for future research, where the set of components are based on the data yielded by rolling window PCA rather than static PCA.

We study the weights of the components from the rolling window PCA to see whether the loadings of factors in the components are stable or volatile over time. The loadings of factors affect the beta and risk premium commanded by the principal component, and therefore affect the components' significance level.

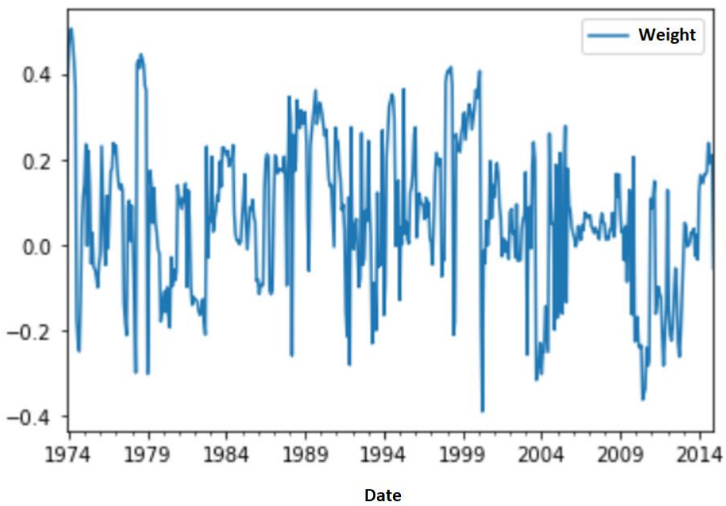
Regarding the financial components, only the first principal component is observed to have relatively stable factor loadings. While its factor loadings vary over time, they have a relatively stable region of loadings. Figure 3 presents how one of the main contributing factors to component one, the factor High-Minus-Low, varies over time.

Figure 3. Factor loadings in component 1 – High-Minus-Low



Looking at Table 7, we see that component eight and ten are significant. When observing the factor loadings of these components we notice that the main contributing factors have volatile loadings over time. In Figure 4 we present the loadings of one of the main contributing factors to component ten, the factor HML-Devil. We notice that the loadings change signs frequently over time.

Figure 4. Factor loadings in component 10 - HML-Devil



The pattern we draw from our data is that the initial financial components provide relatively stable loadings over time, while subsequent financial components present more volatile factor loadings over time. We observe that the macroeconomic components have more stable factor loadings than the financial components. They seem to provide the same pattern, where the first macroeconomic component provides relatively stable factor loadings while the subsequent components have more volatile loadings. Component eleven is the most stable macroeconomic component. While component twelve is more volatile, the loadings do not change signs as frequently as the other components. This could explain the relative high coefficient of component twelve and its significance level.

In conclusion, the weights of the components vary over time and changes sign frequently enough to affect the components coefficients. Loadings that have positive contribution in one window and then negative contribution in the next window will affect the significance of the components by smaller coefficients, in absolute values. This is because the betas will move closer to zero as the factor loadings keep switching signs.

When discussing the possible outputs of our testing, we predicted an increasing significance of the components from the rolling window PCA compared to the static PCA. The reasoning behind this prediction is that by allowing the components weights of factors to vary in time, we would retrieve components that are more accurately reflecting the variation of the correlation between factors in time. The significance of the test when analysing the components in rolling window PCA with fifteen factors yields an unexpected result. Compared to static PCA, the rolling window PCA yields four components which have significant risk premia while the static PCA yields seven components with significant risk premia. Studying the changes in weights we also see that many of our components, including some that are significant, have unstable weights. This indicates that there is noise in our data and that the correlations procedure is not enough to control for the errors in our variables.

For the rolling window PCA we used a rolling window of 60 months. It is possible that outcomes vary with different specifications, which is something that we do not control for in this paper. Furthermore, there are different interpretations on what number of components to use when performing a PCA. The choice of the number of components are made by the researcher. Therefore, we cannot compare the number of components that we apply with other models as this will depend on the amount of variation the research seeks to capture, and the data used for the research.

We conclude that our specification of the rolling window PCA for our dataset with the same number of components as retrieved from the static PCA does not improve the explanatory power compared to the static factor models. The results suggest that further research is needed to develop the specification for the rolling window PCA. Adjusting the number of financial components and adjusting the windows for the rolling window PCA used to construct the rolling window factor models will yield different results. Further research is also proposed concerning the creation of the components in rolling window. We suggest that the components are matched by both absolute correlation testing for flipped signs and correlation with other components to see if this can better match the components.

6. Robustness Test - Separation of Components

In this section we explore the potential of separating the components. The collected factors we use are different to earlier research. The financial ratios in the setting of Kozak et.al (2018) are sorted in portfolios. The financial ratios we use are not compiled in portfolios and the macroeconomic factors are not investable. We do not separate the financial ratios from the portfolio components in our research as PCA should be able to differentiate between these variables. We first describe component selection analysis. We then perform Fama and Macbeth regressions, and finally, we discuss implications of the separation of the components. We do not repeat the analysis of the macroeconomic components as these do not differ from our previously conducted analysis.

Figure 5. Scree plot of portfolio factors

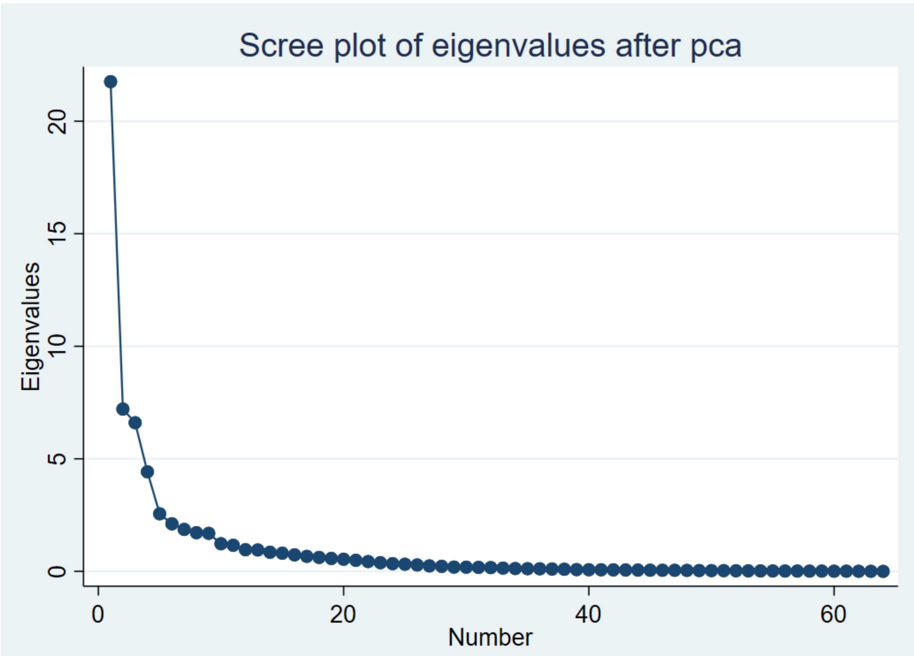


Table 8. Cumulative R-squared captured for the portfolio components

Component	Eigenvalue	Difference	Proportion	Cumulative
Comp1	21.757	14.547	0.340	0.340
Comp2	7.210	0.607	0.113	0.453
Comp3	6.603	2.180	0.103	0.556
Comp4	4.423	1.867	0.069	0.625
Comp5	2.557	0.444	0.040	0.665

When analysing the financial ratios and portfolio factors, we choose to conduct the analysis in the same manner as for the static PCA. This means that we select the number of components that we apply to the factor models by analysing a scree plot. The kink in the scree plot indicate that the sixth component do not capture a substantial amount of variation from the factors. We therefore decide to select five components for our analysis. Table 8 and Table 9 describes captured variation by the components we select.

Figure 6. Scree plot of financial ratios

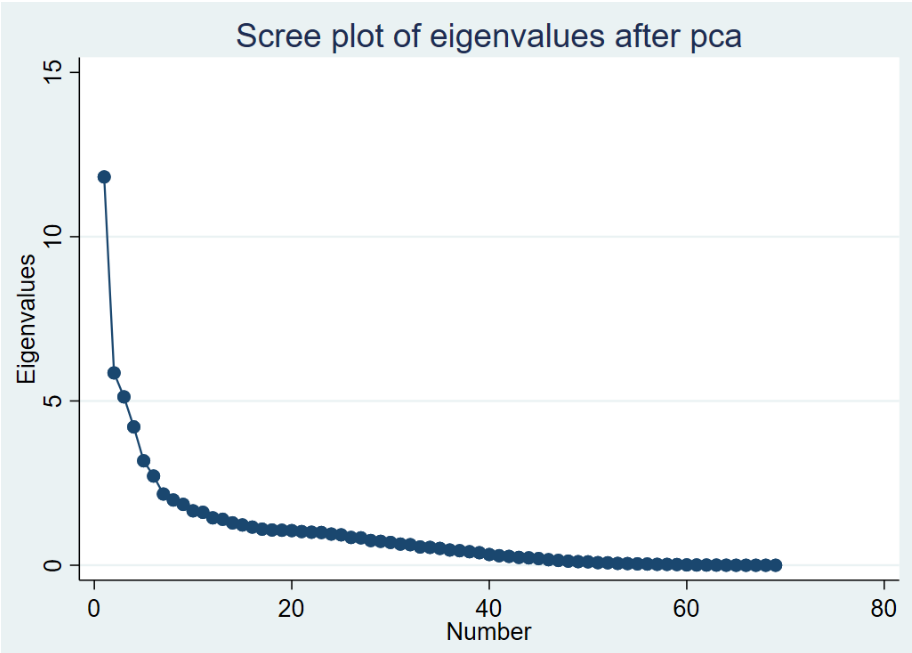


Table 9. Cumulative R-squared for financial ratio components

Component	Eigenvalue	Difference	Proportion	Cumulative
Comp1	11.819	5.965	0.171	0.171
Comp2	5.854	0.729	0.085	0.256
Comp3	5.125	0.914	0.074	0.330
Comp4	4.211	1.032	0.061	0.391
Comp5	3.179	0.464	0.046	0.438
Comp6	2.715	0.547	0.039	0.477
Comp7	2.168	0.182	0.031	0.508

The analysis of the scree plot for the financial ratio components show us a kink between components six and seven. We interpret this as that adding component eight will not add a significant amount of variation from the set of factors. The analysis indicates that we should retrieve seven financial ratio components. Studying Table 9, we see that seven ratio components captures in total 0.508 of the total variation.

Table 10. Output Unrotated separate components Financial portfolios and ratios

Separated 5 Portfolio Components				Separated 5 Portfolio and 7 Ratio Components				Unrotated Static PCA 10-factor model			
Factors	Coef,	St,Err,	p-value	Factors	Coef,	St,Err,	p-value	Factors	Coef,	St,Err,	p-value
PC 1	-0,273	0,223	0,222	PC 1	-0,305	0,225	0,175	PC 1	0,203	0,231	0,382
PC 2	0,256(*)	0,132	0,053	PC 2	0,18	0,127	0,159	PC 2	0,555(***)	0,102	0
PC 3	0,018	0,137	0,895	PC 3	0,016	0,132	0,903	PC 3	-0,007	0,137	0,961
PC 4	0,375(***)	0,11	0,001	PC 4	0,358(***)	0,105	0,001	PC 4	0,018	0,134	0,892
PC 5	0,135	0,094	0,153	PC 5	0,12	0,087	0,169	PC 5	-0,032	0,11	0,772
				PC 6	-0,792(***)	0,1	0	PC 6	0,147	0,172	0,392
				PC 7	-0,134	0,108	0,214	PC 7	0,454(***)	0,14	0,001
				PC 8	0,185	0,213	0,386	PC 8	0,025	0,179	0,891
				PC 9	0,366(*)	0,199	0,067	PC 9	-0,352(***)	0,16	0,029
				PC10	-0,421(***)	0,149	0,005	PC10	0,148	0,097	0,13
				PC 11	0,216(*)	0,122	0,077				
				PC 12	-0,173	0,126	0,171				
Constant	0,009(***)	0,002	0	Constant	0,011(***)	0,002	0	Constant	0,008(***)	0,002	0
R-squared			0,5682				0,7867				0,7321

*** p<0,01, ** p< 0,05, * p< 0,1,

When comparing the separated model, extended with both portfolio and ratio components, with the ten unrotated components from our previous section, we see higher significance and a higher R-squared value from the separated model. This is not a surprising result as these factor models contain more components and should therefore capture more of the cross-sectional variation of returns. We conclude that by separating the financial components we can construct factor models with higher explanatory power. The results do not differ substantially from our previous factor models. The results indicate that even without separating the financial components, our unseparated models are still able to capture a large amount of the cross-sectional variation of portfolio returns. Furthermore, from Table 10 we observe that the financial ratios have a significant effect on the extended models' explanatory power and can therefore conclude that the financial ratios correlate with the portfolio components. The explanatory power from our ratios are supported by the research conducted by Bianchi and McAlinn (2020). One added element that the separation of factors contributes with, besides higher explanatory power, is a more accurate component analysis. The separation allows us to, more accurately, pinpoint the contributions to components from financial factors with portfolios versus financial ratios without portfolios. This is something future research should consider when constructing the original high-dimensional dataset.

7. Conclusion & Discussion

For our empirical analysis we construct factor models by utilizing PCA, a dimensionality reduction technique, on a set of 142 risk factors proposed in financial literature and apply a Fama and Macbeth two-pass regression to estimate risk premia commanded by our principal components. We perform a static PCA, which is what we call the conventional application of PCA, and a rolling window PCA, where the data is split into overlapping windows and where the PCA constructs principal components separately in each window. From the static and rolling window PCA we can construct six factor models: unrotated static PCA factor model, unrotated extended static PCA factor model, rotated static PCA factor model, rotated extended PCA factor model, rolling window PCA factor model, and the extended rolling window PCA factor model. The extended models refer to the addition of five macroeconomic components to the models. Lastly, we perform a robustness control with separation of financial factors sorted in portfolios and WRDS ratios when performing static PCA. This is to control if the unseparated factors increase the difficulty of interpretation of components and if our models have less explanatory power relative to the models with separated factors.

The benchmark model for our research is the Fama and French five-factor asset pricing model. We know from Kozak et al. (2018) that the principal component factors obtained from what we refer to as static PCA can explain more of the variation in the cross-section of stock returns than the benchmark model. One of our research questions revolves around whether the components extracted from the rolling window specification can explain more of this variation than the components extracted from static PCA. The results suggest, surprisingly, that the components from the rolling window specification does not explain more of this variation compared to the static specification. The added complexity and potential accuracy of the rolling window factor model did not improve its explanatory power. Our other research questions look at whether the inclusion of macroeconomic components increases the explanatory power of the rolling window PCA and the static PCA models. The results show that the addition of macroeconomic components increases the explanatory power for both the rolling window and static factor models.

When interpreting the rotated components, we find that some components are less complicated to interpret than others. For the more complicated-to-interpret components, the contribution of the risk factors is where the complexity stems from. Different risk factors from different categories have nearly identical level of contribution for some of the components, giving room for flexibility in the interpretations.

Our research shows that the use of PCA is a viable option when facing the high-dimensionality challenge in asset pricing. Further research regarding the rolling window PCA factor model is proposed. While the model showcases higher explanatory power than the benchmark model, our assumption was that the model would provide higher explanatory power than the extended static PCA factor models. Different specification and computation of the model may improve its explanatory power when benchmarked against the static PCA factor models. Rather than using the components retrieved from static PCA, rolling window PCA could be conducted on the original dataset of 142 variables where the researcher can then decide the number of components applied to the analysis. This can potentially increase the explanatory powers of the rolling window factor models. We also suggest improving the matching process of the components from different windows to reduce noise in the variables.

Furthermore, there are two suggested routes for future research concerning the construction of the original dataset. The first suggestion is a separation of factors which are sorted in portfolios, and the financial ratios, for further interpretability possibilities. Our results suggest that separation can improve the accuracy of component analysis. The second suggestion is that future research construct portfolios of the financial ratios. Because of a limited timeframe, we were not able to sort out portfolios based on the ratios.

Bibliography

Andrew, A.L. (1973). "Eigenvectors of certain matrices." *Linear Algebra and its Applications*. Vol 7, no. 2, pp. 151-162.

Belloni, A., Chernozhukov V and Hansen, C. (2014). "High-Dimensional Methods and Inference on Structural and Treatment Effects." *Journal of Economic Perspectives*. Vol 28, No. 2, pp. 29-50.

Bianchi, D., and McAlinn, K. (2020). "Divide and Conquer: Financial Ratios and Industry Returns Predictability." *Working paper*.

Carhart, M. (1997). "On persistence in Mutual Fund Performance." *Journal of Financial Economics*. Vol 52, no. 1, pp. 57-82

Cochrane, J. (2001). "Asset Pricing." New Jersey: Princeton University Press.

Fama, E., and French, K. (1993). "Common risk factors in the returns on stocks and bonds." *Journal of Finance*. Vol 33, no.1, pp. 3-56

Fama, E., and French, K. (2015). "A five-factor asset pricing model". *Journal of Financial Economics*. Vol 116, no.1, pp. 1-22.

Fama, E., and Macbeth, J.D. (1973). "Risk, return, and equilibrium: Empirical tests." *Journal of Political Economy*. Vol 81, no.3, pp. 607-636.

Feng, G., Giglio, S., and Xiu, D. (2020). "Taming the Factor Zoo: A Test of New Factors". *The Journal of Finance*. Vol 75, no.3, pp. 1327-1370.

Freyberger, J., Neuhierl, A., and Weber, M. (2017). "Dissecting characteristics non-parametrically." Technical report, University of Chicago.

Hatzinikolaou, D., Katsimbris, G., Noulas, A. (2002). "Inflation uncertainty and capital structure: Evidence from a pooled sample of the Dow-Jones industrial firms." *International Review of Economics and Finance*. Vol 11, no.1, pp. 45-55.

He, Z., Kelly, B., Manela, A. (2017). "Intermediary asset pricing: New evidence from many asset classes." *Journal of Financial Economics*. Vol 126, no.1, pp. 1-35.

Hou, K., Xue, C., & Zhang, L. (2015). “Digesting Anomalies: An Investment Approach.” *Review of Financial Studies*. Vol 28, no.3, pp. 650-705.

Huerta, R., Corbacho, F., and Elkan, C. (2013). “Nonlinear support vector machines can systematically identify stocks with high and low future returns.” *Algorithmic Finance*. Vol 2, no. 1, pp. 45-58.

Jaggia, S., and Kelly, A. (2016). “Business Statistics: Communicating with Numbers.” 2nd ed. New York: McGraw-Hill.

Jegadeesh, N., and Titman, S. (1993). “Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency.” *The Journal of Finance*. Vol 48, No. 1, pp. 65-91

Jolliffe, I.T. (2002). “Principal Component Analysis.” 2nd ed. New York: Springer. *Journal of Financial Economics*. Vol. 135, no. 2, pp 271-292.

Kelly, B., Pruitt, S., Su, Y. (2018). “Characteristics Are Covariances: A Unified Model of Risk and Return.” *Journal of Financial Economics*. Vol 134, No. 3, pp. 501-524.

Kozak, S., Nagel, S., and Santosh, S. (2020) “Shrinking the Cross-Section.” *Journal of Financial Economics*. Vol 135, no.2, pp 271-292

Lewellen, J., Nagel, S., and Shanken, J. (2010). “A skeptical appraisal of asset-pricing tests.” *Journal of Financial Economics*. Vol 96, no.2, pp. 175-194.

Lintner, John. 1965a. “The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets.” *Review of Economics and Statistics*. Vol 47, no.1, pp. 13-37.

Lintner, John. 1965b. “Security Prices, Risk and Maximal Gains from Diversification.” *Journal of Finance*. Vol 20, no.4, pp. 587-615.

Mossin, Jan. 1966. “Equilibrium in a Capital Asset Market.” *Econometrica*. Vol 34, no.4, pp. 768-83.

O’Brien, R.M. (2007). “A Caution Regarding Rules of Thumb for Variance Inflation Factors.” *Quality & Quantity*. Vol 41, no. 5, pp. 673–690.

Park, C. (2013). “Simple principal component analysis using Lasso.” *Journal of the Korean Data and Information Science Society*. Vol 24, no. 3, pp. 533-541.

Rösler, F., and Manzey, D. (1981). “Principal Components and varimax-rotated components in event-related potential research: Some remarks on their interpretation.” *Biological Psychology*. Vol 13, pp. 3-26.

Sharpe, William F. (1966). “Mutual Fund Performance.” *Journal of Business*. Vol 39, no.1, pp. 119-38.

Smith, L. (2002). “A tutorial on Principal Components Analysis.” OUCS-2002-12. University of Otago Department of Computer Science, New Zealand.

Stock, J.H., and Watson, M.W. (2002). “Forecasting Using Principal Components from a Large Number of Predictors.” *Journal of American Statistical Association*. Vol 97, No. 460, pp. 1167-1179.

Treynor, J.L. (1962). “Toward a Theory of Market Value of Risky Assets.” Unpublished manuscript. Final version in Asset Pricing and Portfolio Performance, 1999, Robert A. Korajczyk, ed., London: Risk Books, pp. 15-22.

Tsai, C-F., Lin, Y-C., Yen, D., and Chen, Y-M. (2011). “Predicting stock returns by classifies ensembles.” *Applied Soft Computing*. Vol 11, no. 2, pp. 2452-2459.

Appendix A

The first 72 factors are listed with the year of publication and corresponding reference. When noted “Mean value per month” the data collected for each month is averaged for all the observations. Data that is Backfilled is retrieved in Quarterly form and then backfilled to match our tests, the backfill is performed by filling the gap in for example January, February, March and April with the same value as Quarterly reported in April for the whole period.

No.	Factor	Reference	Alteration/Note:
1	Excess Market Return	Kozak (2020)	Mean value per month
2	Equal Weighted Excess Return	Kozak (2020)	Mean value per month
3	Size	Fama and French (1993).	Mean value per month
4	Value (Annual)	Fama and French (1993).	Mean value per month
5	Gross Profitability	Novy Marx (2013a).	Mean value per month
6	Vale-Profitability	Novy Marx (2013b).	Mean value per month
7	Piotroski's F-score	Piotroski (2000).	Mean value per month
8	Debt Issuance	Spiess and Affleck-Graves (1999).	Mean value per month
9	Share Repurchases	Ikenberry et al. (1995).	Mean value per month
10	Share Issuance	Pontiff and Woodgate (2008).	Mean value per month
11	Accruals	Sloan (1996).	Mean value per month
12	Asset Growth	Cooper et al. (2008).	Mean value per month
13	Gross Margins	Novy Marx (2013a).	Mean value per month

14	Dividend Yield	Naranjo et al. (1998).	Mean value per month
15	Earnings/Price	Basu (1977).	Mean value per month
16	Cash Flow / Market value of Equity	Lakonishok et al. (1994).	Mean value per month
17	Net operating assets	Hirshleifer et al. (2004).	Mean value per month
18	Investment	Chen et al. (2011);	Mean value per month
19	Investment-to-capital	Xing (2008).	Mean value per month
20	Investment Growth	Xing (2008).	Mean value per month
21	Sales Growth	Lakonishok et al. (1994).	Mean value per month
22	Leverage	Follows Bhandari (1988).	Mean value per month
23	Return on Assets	Chen et al. (2011).	Mean value per month
24	Return on Equity	Haugen and Baker (1996).	Mean value per month
25	Sales to Price	Barbee Jr et al. (1996).	Mean value per month
26	Growth in LTNOA	Fairfield et al. (2003).	Mean value per month
27	Momentum (6 month)	Jagadeesh and Titman (1993).	Mean value per month
28	Industry Momentum	Moskowitz and Grinblatt (1999).	Mean value per month
29	Value-Momentum	Novy Marx (2013b).	Mean value per month
30	Value-Momentum-Profitability	Novy Marx (2013b).	Mean value per month
31	Short Interest	Dechow et al. (1998).	Mean value per month
32	Momentum (1 year)	Jagadeesh and Titman (1993).	Mean value per month

33	Momentum-Reversal	Jagadeesh and Titman (1993).	Mean value per month
34	Long-Term Reversals	DeBondt and Thaler (1985).	Mean value per month
35	Value	Asness and Frazzini (2013).	Mean value per month
36	Share Issuance (monthly)	Pontiff and Woodgate (2008).	Mean value per month
37	PEAD (SUE)	Foster et al. (1984).	Mean value per month
38	Return on Book Equity	Chen et al. (2011).	Mean value per month
39	Return on Market Equity	Chen et al. (2011).	Mean value per month
40	Return on Assets	Chen et al. (2011).	Mean value per month
41	Short-term Reversal	Jegadeesh (1990).	Mean value per month
42	Idiosyncratic Volatility	Ang et al. (2006).	Mean value per month
43	Beta Arbitrage	Cooper et al. (2008).	Mean value per month
44	Seasonality	Heston and Sadka (2008).	Mean value per month
45	Industry Relative Reversals	Da et al (2013)	Mean value per month
46	Industry Relative Reversals(Low Volatility)	Da et al (2013)	Mean value per month
47	Composite Issuance	Daniel and Titman (2006)	Mean value per month
48	Price	Blume Husic (1973)	Mean value per month
49	Firm Age	Barry and Brown (1984)	Mean value per month
50	Share Volume	Datar et al. (1998)	Mean value per month
51	Industry Momentum-Reversal	Moskowitz and Grinblatt (1999).	Mean value per month

52	Asset Turnover	Soliman (2008).	Mean value per month
53	Cash Flow duration	Dechow et al. (2004).	Mean value per month
54	Exchange Switch	Kozak (2020)	Mean value per month
55	Initial Public Offering	Kozak (2020)	Mean value per month
56	Return equally weighted NYSE index	Chen, Roll and Ross (1986)	
57	Return value weighted NYSE index	Chen, Roll and Ross (1986)	
58	HML Devil	Asness and Frazzini (2013)	
59	Betting against Beta	Frazzini and Pedersen (2014)	
60	Growth in Moneysupply	Maio, Silva (2020)	
61	SPEX-lagged excess returns S&P500	Chen, Ludvigson (2009)	
62	RREL-relative t-bill rate	Chen, Ludvigson (2009)	
63	CFNAI-chicago fed national activity index	Michael W. Mccracken (2007)	
64	CG-total consumption growth	Maio, Silva (2020)	
65	C2G-consumption growth rate in credit goods	Maio, Silva (2020)	
66	IP-monthly growth in industrial production	Michael W. Mccracken (2007)	
67	PI-pce based inflation	Michael W. Mccracken (2007)	
68	Cay-Consumption, aggregate wealt, excess return	Lettau, Ludvigson (2002)	Backfilled
69	Excess Market Return	Fama and French (1993).	
70	Small Minus Big	Fama and French (1993).	
71	High Minus Low	Fama and French (1993).	
72	Conservative Minus aggressive	Fama and French (2015)	
73	Robust Minus Weak	Fama and French (2015)	

WRDS Industry Financial Ratios

Continuing the factor list we here list the factors retrieved from Compustat. These WRDS Industry Financial Ratios are a collection of financial ratios commonly used by academic researchers. The list contains 69 ratios covering several categories. The categories are Profitability, Liquidity, Financial Soundness/Solvency, Efficiency, Capitalization, Valuation and others. The ratios are provided through the Financial Ratios Suite by WRDS.

No.	**Factor**	**Alteration/Note:**
74	Equal Weighted Industry Return	Average across industries
75	Value Weighted Industry Return	Average across industries
76	Dividend Payout Ratio	Average across industries
77	Trailing P/E to Growth	Average across industries
78	Book/Market	Average across industries
79	Shillers Cyclically Adjusted P/E Ratio	Average across industries
80	Dividend Yield	Average across industries
81	Enterprise Value Multiple	Average across industries
82	Price/Cash flow	Average across industries
83	P/E (Diluted, Excl. EI)	Average across industries
84	P/E (Diluted, Incl. EI)	Average across industries
85	Price/Sales	Average across industries
86	Price/Book	Average across industries
87	Effective Tax Rate	Average across industries
88	Gross Profit/Total Assets	Average across industries
89	After-tax Return on Average Common Equity	Average across industries
90	After-tax Return on Total Stockholders Equity	Average across industries
91	After-tax Return on Invested Capital	Average across industries
92	Gross Profit Margin	Average across industries
93	Net Profit Margin	Average across industries
94	Operating Profit Margin After Depreciation	Average across industries
95	Operating Profit Margin Before Depreciation	Average across industries

96	Pre-tax Return on Total Earning Assets	Average across industries
97	Pre-tax return on Net Operating Assets	Average across industries
98	Pre-tax Profit Margin	Average across industries
99	Return on Assets	Average across industries
100	Return on Capital Employed	Average across industries
101	Return on Equity	Average across industries
102	Capitalization Ratio	Average across industries
103	Common Equity/Invested Capital	Average across industries
104	Long-term Debt/Invested Capital	Average across industries
105	Total Debt/Invested Capital	Average across industries
106	Inventory/Current Assets	Average across industries
107	Receivables/Current Assets	Average across industries
108	Free Cash Flow/Operating Cash Flow	Average across industries
109	Operating CF/Current Liabilities	Average across industries
110	Cash Flow/Total Debt	Average across industries
111	Cash Balance/Total Liabilities	Average across industries
112	Cash Flow Margin	Average across industries
113	Short-Term Debt/Total Debt	Average across industries
114	Profit Before Depreciation/Current Liabilities	Average across industries
115	Current Liabilities/Total Liabilities	Average across industries
116	Total Debt/EBITDA	Average across industries
117	Long-term Debt/Book Equity	Average across industries
118	Interest/Average Long-term Debt	Average across industries
119	Interest/Average Total Debt	Average across industries
120	Long-term Debt/Total Liabilities	Average across industries
121	Total Liabilities/Total Tangible Assets	Average across industries
122	Total Debt/Equity	Average across industries
123	Total Debt/Total Assets	Average across industries
124	Total Debt/Total Assets (Double)	Average across industries
125	Total Debt/Capital	Average across industries
126	After-tax Interest Coverage	Average across industries

127	Interest Coverage Ratio	Average across industries
128	Cash Conversion Cycle	Average across industries
129	Cash Ratio	Average across industries
130	Current Ratio	Average across industries
131	Quick Ratio	Average across industries
132	Asset Turnover	Average across industries
133	Inventory Turnover	Average across industries
134	Payables Turnover	Average across industries
135	Receivables Turnover	Average across industries
136	Sales/Stockholders Equity	Average across industries
137	Sales/Invested Capital	Average across industries
138	Sales/Working Capital	Average across industries
139	Accruals/Average Assets	Average across industries
140	Research and Development/Sales	Average across industries
141	Avertising Expenses/Sales	Average across industries
142	Labor Expenses/Sales	Average across industries

Appendix B

Appendix B list the rotated components matrix, containing weights of factors in the rotated components. The first part of the list contains the ten rotated principal Components for the factors excluding macroeconomic factors.

Variable	Comp 1	Comp 2	Comp 3	Comp 4	Comp 5	Comp 6	Comp 7	Comp 8	Comp 9	Comp 10	Unexplained
Operating CF/Current Liabilities	-0,011	-0,084	-0,329	-0,033	-0,01	-0,016	-0,009	0,005	0,077	0,056	0,181
Cash Flow/Total Debt	-0,008	0,017	-0,305	-0,028	0	-0,01	-0,003	-0,005	0,061	0,04	0,211
Profit before depreciation/ Current Liabilities	-0,003	0,099	-0,24	-0,029	0,002	-0,015	0,01	-0,017	0,06	0,081	0,25
Long term debt/Total liabilites	0,003	0,071	-0,199	0,071	0	0,045	0,019	-0,029	-0,189	-0,12	0,377
Return on net operating assets	0	0,165	-0,189	-0,019	0,006	0,001	0,048	-0,021	0,04	0,117	0,142
Dividend Yield	0,014	0,192	-0,154	-0,028	0,007	-0,004	0,017	0,012	-0,052	0,036	0,236
Price/Cash Flow	-0,004	-0,253	-0,121	0,012	0,003	0	0,01	-0,049	0,059	0,162	0,355
Gross Profti/Total Assets	0,011	0,233	-0,096	-0,024	0,011	0,001	0,027	-0,014	0,028	0,109	0,17
Inventory/Current Assets	-0,005	0,235	-0,095	0,026	-0,009	0,021	0,019	0,002	-0,065	0,049	0,164
Return on Total earnings	0,015	-0,118	-0,085	-0,009	0,024	0,002	0,016	0,029	-0,002	0,006	0,845
Asset Turnover	0,001	0,26	-0,081	0,002	0,002	0,013	0,024	-0,006	-0,012	0,083	0,075
SMB	-0,038	0,032	-0,056	0,095	-0,089	0,056	0	-0,023	0,289	-0,09	0,357
Total Debt/Total Assets	0,01	0,096	-0,051	0,079	0,002	0,033	-0,001	0,003	-0,202	-0,168	0,617
Price/Book	-0,005	-0,049	-0,046	-0,005	-0,002	0,001	0,004	0,313	0,021	0,035	0,513
Book/Market	-0,012	0,024	-0,043	0,022	-0,005	-0,003	0,007	0	-0,008	-0,003	0,968
Research and Development/ Sales	0,004	-0,168	-0,033	-0,039	0,017	-0,019	0,003	0,074	0,082	0,021	0,646
Return on Invested Capital	-0,023	0,002	-0,031	0,003	-0,066	-0,035	-0,002	-0,004	0,063	-0,095	0,913
Enterprise value Multiple	0,003	-0,003	-0,031	0,038	0,013	0,023	0,006	0,013	-0,075	-0,065	0,957
Net operating assets	0,024	0,004	-0,028	-0,008	0,098	-0,088	-0,003	-0,022	0,068	0,117	0,81
Return on Average	0,004	-0,012	-0,027	-0,033	0,007	0,004	0,016	-0,015	0,04	-0,028	0,979

Common Equity											
Return on Total Stockholders Equity	-0,015	0,013	-0,026	-0,008	-0,01	0,011	0,015	-0,03	0,026	0,077	0,951
Sales/Working Capital	0,017	-0,015	-0,025	-0,032	-0,05	0,019	0,005	0,002	0,014	-0,105	0,922
After-Tax Return on Average Common Equity	-0,004	0,007	-0,023	0,004	0	0,019	-0,002	0,007	-0,025	0,035	0,983
Dividend Growth	0,069	-0,04	-0,022	0,098	0,135	0,022	-0,015	0,017	-0,11	-0,02	0,728
Equal Weighted Industry Return	-0,025	0,001	-0,022	0,311	-0,048	-0,013	-0,001	0,005	0,106	-0,01	0,095
Idiosyncratic Volatility	0,117	0,024	-0,017	0,1	-0,124	-0,002	-0,004	0,009	0,134	-0,029	0,074
Share Repurchases	-0,141	-0,005	-0,017	-0,019	0,154	-0,089	-0,011	0,005	-0,031	-0,009	0,179
Long-Term Debt/Invested Capital	0,001	-0,03	-0,015	0,002	0,005	0,001	0	0,456	-0,009	-0,01	0,051
PEAD	0,082	0,01	-0,015	-0,003	0,118	0,206	-0,006	-0,006	-0,11	0,072	0,512
Total Debt/Total Assets	0,014	-0,246	-0,014	0,037	0,001	-0,003	-0,009	0,01	-0,064	-0,098	0,356
Exchange Switch	0,029	0,011	-0,014	0,061	0,077	-0,029	-0,012	0,014	-0,008	0,052	0,898
Sales/Invested Capital	-0,001	0,067	-0,013	-0,015	0,001	-0,003	0,007	0,435	0,024	0,021	0,162
Equal Weighted Excess Return	-0,028	0,011	-0,011	0,041	0,015	-0,002	0,06	0,063	0,101	-0,005	0,899
Piotroski’s F-score	-0,072	0,009	-0,01	-0,05	0,218	-0,059	-0,015	-0,011	-0,033	-0,016	0,321
Momentum-Reversal	0,154	-0,02	-0,01	-0,02	0,029	0,081	0	0,011	-0,115	0,076	0,539
Asset Turnover	0,005	0,044	-0,009	0,031	0,222	-0,009	0,007	-0,005	0,32	-0,031	0,286
Asset Growth	0,193	-0,006	-0,009	0,064	0,047	-0,041	0,013	0	-0,005	0,005	0,167
Cash Flow/Market Value of Equity	-0,211	0	-0,009	0,008	-0,019	0,022	0,005	0,015	-0,016	0,022	0,154
Earnings/Price	-0,186	-0,006	-0,008	-0,004	0,061	0,008	-0,001	0,016	-0,062	-0,007	0,152
Return on Assets (Annual)	0,084	0,024	-0,007	-0,072	0,327	-0,046	0,002	0,011	0,12	-0,054	0,137
Share Volume	0,128	0,028	-0,007	0,09	-0,11	0,017	0,001	0,016	0,108	-0,033	0,123
Sales Growth	0,176	0,015	-0,007	0,057	0,019	0	-0,002	0,005	0,029	-0,006	0,258
Return on Market Equity	-0,158	-0,011	-0,007	0,028	0,149	0,011	-0,007	-0,007	-0,097	0,037	0,159
Gross Profit Margin	-0,002	0,238	-0,006	0,03	-0,001	0,019	-0,002	-0,039	-0,032	-0,06	0,4
Momentum (1 year)	0,055	0,004	-0,006	-0,01	0,025	0,337	-0,02	-0,014	-0,007	0,025	0,17
Composite Issuance	0,117	0,005	-0,005	0,115	-0,074	-0,008	-0,011	0,005	0,084	-0,011	0,259

Investment Growth	0,145	0,003	-0,005	0,053	0,037	-0,067	0,017	-0,005	0,013	0,035	0,485
Sales-to-Price	-0,23	0,007	-0,004	0,09	0,076	0,014	0,003	0,005	0,1	0,013	0,079
Value(monthly)	-0,193	0,004	-0,003	0,032	-0,034	-0,163	-0,001	0,002	-0,037	-0,007	0,061
Return on Assets	0,066	0,006	-0,003	-0,024	0,316	0,085	-0,009	-0,01	0,061	0,007	0,216
Investment	0,163	0,004	-0,003	0,025	0,048	-0,03	-0,002	0,005	0,055	0,026	0,408
Investment-to-Capital	0,17	0,015	-0,002	0,062	-0,047	-0,015	-0,002	0,011	0,085	-0,039	0,101
Value-Momentum	-0,108	-0,011	-0,002	-0,012	-0,04	0,328	-0,005	0,001	0,001	0,002	0,13
Share Issuance(monthly)	0,159	0,007	-0,002	0,018	-0,157	0,044	-0,001	0,005	0,001	0,022	0,143
Value (Annual)	-0,227	-0,002	-0,001	0,016	-0,061	0,022	-0,004	0,006	-0,018	0,009	0,065
Excess Return. Equal weighted portfolio all stock	-0,006	-0,029	-0,001	0,339	-0,013	0,013	-0,002	-0,004	0,04	0,025	0,093
Total Debt/EBITDA	0,005	0,028	0	0,061	0,013	0,049	0,005	0,017	-0,12	-0,069	0,908
Cash flow Duration	0,232	-0,01	0	-0,039	0,058	-0,026	0,005	-0,003	-0,086	-0,016	0,139
Cash Ratio	0,001	-0,037	0	0,023	0,019	0,007	-0,005	0,198	-0,034	0	0,797
Momentum(6 months)	0,019	-0,014	0,001	0,002	0,011	0,37	-0,005	-0,004	0,002	0	0,091
Share Issuance	0,163	0,005	0,001	0,025	-0,141	-0,009	0	0,006	-0,011	0,019	0,19
Industry Momentum-Reversal	-0,031	0,049	0,001	0,159	-0,006	0,24	0,004	0,012	-0,077	0,041	0,549
After-Tax Interest Coverage	0,019	-0,013	0,001	-0,05	-0,036	-0,037	-0,029	0,016	0,078	0,293	0,645
After-Tax Return on Invested Capital	0,019	-0,013	0,001	-0,05	-0,036	-0,037	-0,029	0,016	0,078	0,293	0,645
CMA	-0,181	0,026	0,002	-0,052	-0,103	0,05	0,006	0,006	0,046	-0,056	0,3
Common Equity/Invested Capital	-0,001	0,034	0,002	-0,002	-0,001	0,001	0,001	-0,452	0,006	0,022	0,056
Labour Expenses/Sales	-0,005	-0,038	0,002	0,038	-0,006	0,024	-0,002	0,008	-0,039	0,04	0,975
Beta Arbitrage	0,102	0,023	0,003	0,146	-0,084	-0,041	-0,006	0,017	0,095	-0,043	0,184
Industry Momentum	-0,017	-0,007	0,003	0,002	-0,006	0,348	0,006	0,008	0,015	-0,031	0,199
Cash Flow Margin	0,001	-0,002	0,003	-0,002	0	-0,003	0,438	0,002	0	-0,004	0,01
Net Profit Margin	0,001	-0,002	0,003	-0,002	-0,001	-0,002	0,438	0,002	0	-0,003	0,01
Operating Profit Margin After Depreciation	0,001	-0,002	0,003	-0,002	-0,001	-0,002	0,438	0,002	0	-0,003	0,01
Operating Profit Margin	0,001	-0,002	0,003	-0,002	-0,001	-0,002	0,438	0,002	0	-0,003	0,01

Before Depreciation											
Pre-Tax Profit Margin	0,001	-0,002	0,003	-0,002	-0,001	-0,002	0,438	0,002	0	-0,003	0,01
Return on Book Equity	-0,005	-0,018	0,003	0,051	0,325	0,109	-0,007	-0,007	-0,057	0,028	0,193
Free Cash flow/Operating Cash Flow	0,003	-0,013	0,003	0,022	-0,006	0,041	-0,013	0,018	-0,024	0,135	0,927
Gross Profitability	0,129	0,052	0,004	-0,03	0,14	-0,004	-0,001	0,01	0,26	-0,063	0,238
Return on Equity Annual	-0,011	-0,009	0,004	0,023	0,326	-0,021	0,005	0,016	-0,01	-0,045	0,21
Value-Momentum-Profitability	-0,061	0,009	0,004	0,017	0,048	0,343	-0,006	0,01	0,095	-0,027	0,143
Leverage	-0,22	-0,02	0,004	0,059	-0,035	0,018	-0,002	0,004	-0,08	0,021	0,093
Inventory Turnover	-0,016	-0,044	0,005	0,013	-0,024	-0,016	0,008	-0,01	0,029	-0,088	0,932
Accruals	-0,015	-0,014	0,005	0,102	0,199	0,006	0,019	-0,014	0,136	-0,06	0,587
Pre-Tax Return on Total Earning Assets	0,005	0,04	0,005	0,003	-0,002	-0,005	0,024	0,005	-0,003	0,015	0,98
Value Weighted Excess Return	-0,024	0,003	0,006	0,035	-0,002	-0,011	0,068	0,049	0,089	-0,008	0,914
HML-Devil	-0,032	0,047	0,006	-0,015	-0,034	-0,013	0,025	0,069	0,067	0	0,927
HML	-0,218	0,007	0,007	0,042	-0,094	0,003	-0,02	0,007	-0,003	-0,017	0,182
Industry Relative Reversals(Low volatility)	0,052	-0,096	0,008	-0,212	0,005	0,073	-0,002	0,002	0,048	-0,022	0,495
Short-Term Reversal	0,021	-0,022	0,008	-0,171	-0,022	0,132	0,02	0,01	0,09	0,032	0,609
Dividend Yield	-0,135	0,009	0,009	-0,148	-0,138	0,009	0,006	0,003	-0,052	-0,001	0,236
Firm Age	-0,159	-0,013	0,009	-0,043	0,106	-0,054	0,004	-0,004	-0,08	0,011	0,09
Long-Term Reversals	0,197	-0,02	0,009	-0,002	0,092	-0,025	0,018	0,012	-0,148	0,033	0,331
Value Weighted Excess Market Return	0,009	-0,039	0,011	0,338	0,005	0,014	-0,004	0,002	-0,014	0,045	0,139
Excess Market Return	0,009	-0,039	0,011	0,338	0,005	0,014	-0,004	0,002	-0,014	0,045	0,139
Industry Relative Reversals	0,026	-0,053	0,013	-0,203	-0,028	0,131	0,016	0,007	0,113	0,022	0,481
Cash Conversion Cycle	0,015	0,036	0,015	0,003	0,021	-0,05	0	0,005	0,04	-0,045	0,959
Short Interest	0,043	0,026	0,015	0,159	0,022	-0,004	0,002	-0,003	0,164	-0,021	0,472
Value weighted Industry Return	-0,017	-0,023	0,015	0,349	0,024	0,002	0	-0,005	-0,007	0,05	0,16
Debt Issuance	0,137	0,01	0,016	-0,046	-0,134	0,097	0,002	0,004	0,02	-0,065	0,339

Long term Debt/Book Equity	0,008	-0,013	0,019	0,002	-0,009	-0,017	0	-0,007	-0,014	-0,009	0,99
Shillers Cyclically Adjusted P/E Ratio	0,004	-0,029	0,02	0,002	-0,012	-0,005	-0,005	0,026	-0,036	-0,055	0,962
Total Liabilities/Total Tangible Assets	0,025	-0,036	0,02	0,009	-0,013	0,018	0,008	-0,01	-0,051	-0,031	0,957
Return on Equity	-0,013	0,044	0,022	0,061	0,002	0,027	0,008	-0,008	0,008	0,025	0,949
RMW	-0,058	-0,005	0,024	-0,002	0,288	0,002	0,004	0,012	-0,016	0,029	0,223
Betting against Beta	0,003	0,016	0,024	0,006	0,02	0,027	-0,017	-0,011	0,007	0,045	0,98
Value-Profitability	-0,164	0,058	0,028	0,026	0,069	0,051	-0,015	0,018	0,216	-0,086	0,362
Effective Tax Rate	0,005	0,226	0,028	-0,041	-0,01	-0,017	-0,027	0,026	0,059	-0,069	0,561
Total Debt/Invested Capital	0,004	0,048	0,028	-0,029	-0,013	-0,02	-0,006	0,001	0,022	-0,024	0,975
Price	0,108	-0,017	0,028	-0,123	0,032	0,265	0,006	0,017	-0,095	0,064	0,194
Interest/Average Total Debt	0,004	-0,085	0,028	-0,08	0,004	-0,056	0,004	-0,033	0,183	0,134	0,703
Growth in LTNOA	-0,058	0,005	0,032	-0,001	-0,074	0,032	-0,018	-0,017	0,087	-0,001	0,89
Seasonality	0,099	0,057	0,033	0,048	0,046	-0,038	-0,057	0,009	-0,119	-0,053	0,709
P/E (Diluted, Incl. EI)	-0,015	0,052	0,033	0,074	0,005	0,022	0,007	-0,002	-0,07	0,401	0,321
Gross Margins	0,193	0,015	0,034	-0,056	-0,018	-0,038	-0,029	0,008	-0,099	-0,089	0,362
Sales/Stockholders Equity	-0,005	-0,006	0,036	0,003	-0,028	-0,004	0,014	-0,007	-0,041	0,003	0,977
Total Debt/Invested Capital	-0,001	0,039	0,036	0,007	-0,004	0,005	0,001	0,459	-0,02	0,011	0,076
P/E (Diluted, Excl. EI)	-0,015	0,056	0,036	0,072	0,004	0,024	-0,001	0,003	-0,072	0,399	0,324
Size	0,028	-0,02	0,039	-0,05	0,076	-0,208	0,006	0,02	-0,256	0,045	0,28
Payables Turnover	0,006	0,021	0,04	0,041	0,022	0,029	0,004	-0,006	-0,047	0,032	0,965
Interest/Average Total Debt	0,011	-0,117	0,044	-0,101	0,013	-0,057	-0,008	-0,035	0,19	0,101	0,602
Accruals/Average Assets	-0,021	-0,199	0,049	-0,007	-0,02	-0,022	-0,079	0,071	0,064	-0,118	0,302
Price/Sales	0,005	-0,242	0,058	0,044	0,001	0,041	0,047	-0,05	-0,048	-0,07	0,291
Dividend Payout Ratio	-0,018	0,02	0,072	0,001	-0,031	0,003	-0,001	0,052	-0,015	0,054	0,933
Return on Capital Employed	0	0,119	0,074	0,026	-0,012	0,02	0,008	0,003	-0,066	0,005	0,835
Total debt/Equity	-0,001	0,02	0,083	0,011	-0,028	0,004	0,011	-0,008	-0,014	0,043	0,943
Trailing P/E to growth	-0,011	-0,165	0,11	0,044	-0,021	0,004	0,011	0,022	0,009	0,196	0,485
Advertising Expenses/Sales	-0,025	0,084	0,131	0,028	0,047	0,009	0,03	-0,013	-0,03	-0,284	0,493
Price/Book	0,014	-0,16	0,175	0,06	-0,01	0,045	0,051	-0,026	-0,051	-0,014	0,341

Cash Ratio	0,007	-0,012	0,2	0,017	0,029	-0,001	0,011	-0,006	0,028	0,037	0,666
Current Liabilities/Total Liabilities	-0,003	0,299	0,201	-0,037	0,002	-0,011	-0,022	0,014	0,079	0,032	0,161
Short term Debt/Total Debt	0,002	0,292	0,232	-0,028	0,001	-0,016	-0,034	0,037	0,036	-0,021	0,149
Quick Ratio	-0,002	-0,04	0,32	-0,018	-0,004	-0,01	0,016	-0,014	0,035	0,048	0,119
Current Ratio	-0,006	0,066	0,337	-0,01	-0,009	-0,002	0,03	-0,017	0,017	0,091	0,198
Cash Balance/Total Liabilities	-0,005	0,015	0,347	-0,013	0,004	-0,004	0,001	-0,013	0,029	0,041	0,088

The next part of this list are the weights of factors in the rotated components based on the Macroeconomic factors.

Variable	**Comp1**	**Comp2**	**Comp3**	**Comp4**	**Comp5**	**Unexplained**
Consumption Growth Rate in Credit Goods	0,011	0,014	0,016	-0,011	0,987	0,015
Monthly Growth in Industry Production	-0,047	0,658	-0,074	0,033	0,032	0,118
Chicago FED National Activity Index	-0,018	0,675	-0,017	0,011	0,024	0,072
Relative T-Bill Rate	0,601	-0,059	-0,06	0,079	0,019	0,1
Consumer Price Index based Inflation	0,418	-0,016	0,501	-0,095	0,004	0,22
Growth in Money Supply	0,017	0,011	0,021	0,982	-0,011	0,016
Lagged Excess Returns S&P500	-0,598	0,013	0,174	-0,034	-0,04	0,099
Consumption, Aggregate Wealth, Excess Returns	0,239	0,107	-0,714	-0,106	-0,072	0,19
Total Consumption Growth	0,217	0,31	0,446	-0,079	-0,132	0,357