Erasmus University Rotterdam

Erasmus School of Economics

Thesis MSc Econometrics and Management Science

Using Instrumented Principal Component Analysis to Explain Delta-Hedged Options Returns

Author

Raymon Ligtenberg (503776)

Instructor

Dr. M. Grith

Second Assessor

Dr. X. Xiao

July 5, 2020

Abstract

This paper examines the key drivers of delta-hedged option returns by employing the Instrumented principal component analysis (IPCA) on a panel dataset of 1776 companies listed on American exchanges from 2005 to 2014. The basic IPCA constructed in Kelly et al. (2017) is improved by implementing three regularization techniques to obtain a more parsimonious model and two robust methods to account for the problem of heteroskedasticity and outliers in the data. Investment, profit to total assets, return on assets and earnings per share were statistically significant in explaining returns. Where the first two characteristics were positively associated with returns, the latter two were negatively related. Parsimonious models yielded better fit out-of-sample while robust methods did not outperform the OLS method in-sample nor out-of-sample. Even though returns are on average negative, tangency portfolios based on out-of-sample IPCA strategies yielded positive returns and exhibited low downside based on value at risk.

Keywords— IPCA, option returns, regularized and robust models

Contents

1 Introduction

Since the issuance of the first stock of the United East Indian Company (VOC) in the early 1600s, people wanted to invest to increase their amount of wealth. When more stocks became available people needed to decide which stock will yield the highest return and therefore needed to know which characteristics of a company drive stock returns. It took almost 350 years before the amount of research on the drivers of stock returns exploded. The first real breakthrough was the one-factor model constructed by Sharpe (1964) and Lintner (1975) that included the market risk factor, which became the well-known Capital Asset Pricing Model (CAPM). The CAPM stated that return on assets depends on their exposure to market risk. This model got extended twice by Fama and French. Their first extension included a size and a value factor as in Fama and French (1992). Two decades later Fama and French (2015) added a profitability and an investment factor to their model which now included five factors that all had some explanation why returns differ across assets. Afterward, the five-factor model became the standard and has been extended by a variety of factors of which the most common are momentum and reversal research by Jegadeesh and Titman (1993) and Jegadeesh (1990) respectively. Altogether, in the last 40 years a large number of factors were introduced to explain expected stock returns.

In contrast, limited research has been performed on the characteristics that drive delta-hedged option returns (from now on *returns* in contrast to *stock returns*) even though the volume of options being traded has increased significantly over recent decades. A possible reason for the limited amount of research according to Horenstein et al. (2018) is that options are merely viewed as a leveraged manner to buy a stock. However, the limited risk profile this type of hedged securities exhibit, simply due to its construction, makes it an exceptional investment for low-risk type investors. To ascertain that the investor obtains the highest return possible given his risk profile, a framework needs to be utilized that explains the main drivers of returns.

To determine which characteristics influence returns, we construct a type of option return that is normally considered in this type of research, namely monthly delta-hedged option returns. This type of return represents the risk premium related to disadvantageous variance risk which is constructed similar to Cao and Han (2013) and Horenstein et al. (2018) by buying a call and delta-hedged daily with the underlying stock during a one-month period. This dataset consists of 1776 companies listed on the main America exchanges including the NYSE and NASDAQ between 2005 and 2014.

The importance of different characteristics on expected returns was investigated by applying the principal component analysis (PCA) in Horenstein et al. (2018). This statistical method transforms a return matrix of observations that are possibly correlated into a set of linearly uncorrelated factors. The disadvantage of this method is that due to this statistical transformation the latent factors lack economic meaning and no information of the non-return characteristics is used. Therefore, the authors investigated how these orthogonal latent factors are related to a set of long-short portfolios based on characteristics. However, this conversion analysis requires a priori correct specification of the factors, which is exactly what needs to be researched. Furthermore, the PCA is a static model and therefore can not take into account the time-varying nature of the relationship between characteristics and return. For instance, Duesenberry (1965) found that econometric relationships often exhibit structural instability. Moreover, Fama and French (1993) discovered that there is a difference in returns for companies with different characteristics such as size, growth rates, and book-to-market value. Since companies evolve over time from small to large firms or vice versa, it would be logical to incorporate time-varying parameters. Furthermore, a static model is unsuitable for a conditional asset pricing model which is important for the typical investor, since they seek to optimize their portfolio and corresponding return every month.

Cao and Han (2013) investigated the relationship between returns and characteristics by using a Fama-MacBeth procedure defined in Fama and MacBeth (1973). The Fama-MacBeth follows a two-step procedure where first the effect of a variable on the cross-section of option returns is determined, the corresponding beta of the characteristic is then subsequently averaged such that the significance can be tested. The disadvantage of this method is that a large number of parameters need to be estimated.

Kelly et al. (2019) discuss in their paper about stock returns the instrumented principal component analysis (IPCA), this is a factor model that explains stock returns by time-varying parameters and implementing time-variant characteristics. With the IPCA the two disadvantages of the PCA are tackled in the following way. The first weakness, lack of economic meaning, is addressed by modeling the factor loadings as a function of large set characteristics that have been proven to be important in literature. Hence this framework contributes to our understanding of which set of instruments drives returns without imposing that the factor structure is known beforehand. Furthermore, by allowing the factor loadings to be time-variant, the second disadvantage of the PCA is solved and yields two benefits. First, the framework provides the opportunity to incorporate time-varying parameters and therefore yields a model that better matches the dynamic behavior of firm characteristics over time. Second, the structure allows for the estimation of a conditional factor model.

The IPCA can be evaluated as a generalization of the Fama-MacBeth model that uses L factors to construct a relationship between L characteristics and the option return. However, when a reduced specification is used where $K < L$ factors are used, a substantially more parsimonious model is obtained that solves the problem of the large set of estimators in the Fama-MacBeth.

Although the IPCA model is not yet applied to any other type of returns than stock returns, we strongly believe this model is also suitable for option returns because both option and stock returns share commonalities and use a very similar dataset of characteristics. First, the dataset used in this paper and in Kelly et al. (2019) consists of panel data with a finite amount of characteristics, second both dependent variables are heavy-tailed monthly returns with a mean of around 0. Furthermore, Horenstein et al. (2018) and Cao et al. (2017) showed that there do exist characteristics that drive returns. Hence, we conclude that since the two datasets share important commonalities and literature confirms the importance of instruments in modeling option returns, IPCA exhibits the properties needed to be a correct method for investigating our research question: 'What characteristics influence delta-hedged option returns'.

Apart from understanding the relationship between the characteristics and returns, our second interest lies in enhancing the IPCA model by using shrinkage techniques and tackling the unfavorable consequences of heteroskedasticity and extreme tails. We will research whether shrinkage techniques yield better out-of-sample fit than the basic variant and if these models determine the same characteristics to be important in the construction of the loadings and be statistically significant. Moreover, we expect that adjusting the ordinary least squares (OLS) to a robust counterpart that satisfies all assumptions such as the Huber loss or weighted least squares (WLS) results in better out-of-sample estimates.

To incorporate the wishes of a low-risk investor, not only a good out-of-sample fit is needed, but also a trading strategy that yields a high return to risk with a limited downside. Trading strategies are based either on a tangency portfolio or an anomaly portfolio. The tangency portfolio is derived from an optimal combination of risk factors obtained from the IPCA models which achieves the highest Sharpe ratio, whereas the anomaly portfolio is constructed on the anomaly loadings of the unrestricted IPCA model. The high return to risk is examined by the Sharpe ratio whereas the downside is investigated by the Value at Risk (VaR) metric.

The main results we found are that investments, profit to total assets, return on assets, and earnings per share are statistically significant instruments in explaining variation in returns. The first two characteristics were positively related to returns, whereas the latter two were negatively related. Moreover, the IPCA showed to have superior performance compared to the PCA, possibly caused by the inclusion of time-varying non-return characteristics. Supervised machine learning techniques showed a slight positive effect out-of-sample, whereas robust methods did not improve upon the IPCA. The tangency portfolio exhibited a positive return which increased even further when a ridge or lasso technique was applied and the VaR analysis established the belief that the downside was minimized to around 0. Therefore, trading strategies based on delta-hedged option returns are interesting for low-risk investors.

The main contribution to research is the extension of the application of IPCA on a new asset class, delta-hedged option returns. Moreover, existent methods that have been previously combined with certain factor models, but have never been integrated with IPCA, are constructed in this paper resulting in a major advancement in theoretical research on IPCA models. These extensions include regularization and robust techniques to cope with the problem of a large set of parameters or not satisfied assumptions of the OLS respectively. Moreover, the out-of-sample investigation of the IPCA in Kelly et al. (2019) is extended with the VaR analysis which facilitates the evaluation of the downside risk of the model.

Apart from the paper's theoretical contributions, it also has an exceptional practical benefit for investors. The investor can calculate the expected return of its option based on the firm's present characteristics as well as use the instruments that have shown to be statistically significant in explaining returns in their strategy to achieve a high Sharpe ratio with a low downside risk.

The remaining sections of the paper are outlined as follows. Section 2 presents the IPCA method and its extensions, establishes a test to determine the importance of variables, investigates the performance of the model, discusses forecasting methods and the VaR. Section 3 considers the data related to returns and characteristics. Section 4 shows the results of the IPCA and its extensions on the model and characteristic level. Section 5 concludes with a summary of the methods and results.

2 Methodology

This section starts with discussing which methods we apply to establish the relationship between characteristics and returns. First, we start with the PCA which is the time-invariant no instruments counterpart of the IPCA. Second, when we understand how this model works, we can include these two aforementioned features and construct the IPCA. Third, we use shrinkage and selection methods to derive a more parsimonious model that might improve forecasts. Fourth, the OLS method is adjusted for outliers by means of a Huber loss function and tries to tackle the problem of heteroskedasticity by replacing OLS by WLS.

When we have constructed all models, we create asset pricing tests to determine which characteristic is important in explaining returns. These tests consist of a likelihood ratio test for the calculation of p-value and F-values of the characteristics and marginal R^2. Subsequently, we evaluate how well all the characteristics together explain the variation in returns and expected returns by utilizing the R^2 metrics. To evaluate if a profitable trading strategy with low-downside can be constructed by utilizing the models, out-of-sample fit, Sharpe ratios, and VaR are examined. The section is completed by a robustness check that considers whether the relationships between the characteristics are similar for different models.

2.1 PCA

To assess the performance of our main model, the IPCA, we should naturally contrast this model to its static counterpart that does not include any instruments which is the PCA. To construct the PCA we use characteristic-managed portfolio returns ($\boldsymbol{x}_t$) introduced in Kelly et al. (2019) as dependent variable that is defined as

$$\boldsymbol{x}_{t+1} := \frac{1}{N_t} \boldsymbol{Z}_t^{\mathrm{T}} \boldsymbol{r}_{t+1}, \tag{1}$$

where $\boldsymbol{x}_{t+1} : \mathbb{R}^{(L\times 1)}$, N_t, $\boldsymbol{Z}_t : \mathbb{R}^{L\times N_t}$ and $\boldsymbol{r}_{t+1} : \mathbb{R}^{N_t\times 1}$ are respectively the return on a characteristic-managed portfolio at time $t+1$, the number of options at time t, all options characteristics at time t and the individual return at time $t+1$. The characteristic-managed portfolio return is the return on a portfolio weighted by the characteristics and normalized by $\frac{1}{N_t}$ such that the portfolio volatility is not affected by the number of companies available at time t. Even though our data is unbalanced, we only need option returns that are available for two consecutive periods. However, if we would have considered individual returns for the PCA we could only include options that are available every time period because otherwise the dependent variable's dimension would differ monthly and therefore a principal component would not exist. Hence the well-known static model constructed similarly as in Connor and Korajczyk (1988) with characteristic-managed portfolio returns is

$$\boldsymbol{x}_t = \boldsymbol{\beta} \boldsymbol{f}_t + \boldsymbol{\epsilon}_t, \tag{2}$$

where $\boldsymbol{\beta} : \mathbb{R}^{L\times K}$, $\boldsymbol{f}_t : \mathbb{R}^{K\times 1}$ and $\boldsymbol{\epsilon}_t$ are the loadings, the statistical orthogonal factors, and the errors respectively.

To find the optimal parameters $\boldsymbol{\beta}$ and $\boldsymbol{f}_t$ we need to minimize:

$$\min_{\boldsymbol{\beta},\boldsymbol{f}_t} \sum_{t=1}^{T} (\boldsymbol{x}_t - \boldsymbol{\beta}\boldsymbol{f}_t)^{\mathrm{T}} (\boldsymbol{x}_t - \boldsymbol{\beta}\boldsymbol{f}_t). \tag{3}$$

Taking the first-order conditions to $\boldsymbol{f}_t$ yields:

$$\boldsymbol{\beta}^{\mathrm{T}}(\boldsymbol{x}_t - \boldsymbol{\beta}\boldsymbol{f}_t) = \mathbf{0} \iff \boldsymbol{f}_t = (\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\beta})^{-1}\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_t. \tag{4}$$

Subsequently, we substitute $\boldsymbol{f}_t$ into Equation (3) and obtain the concentrated form of the objective function for $\boldsymbol{\beta}$

$$\max_{\boldsymbol{\beta}} \operatorname{tr}\Big(\sum_{t=1}^{T} (\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\beta})^{-1}\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}_t\boldsymbol{x}_t^{\mathrm{T}}\boldsymbol{\beta}\Big). \tag{5}$$

In this particular case, the PCA solution for $\boldsymbol{\beta}$ are the first K eigenvectors of the portfolio returns second moment $\sum_{t=1}^{T} \boldsymbol{x}_t\boldsymbol{x}_t^{\mathrm{T}}$.

The PCA framework can be enhanced by making the loadings of the PCA dependent on time-varying characteristics which boils down to the IPCA.

2.2 IPCA - Theoretical framework

We start the theoretical framework from the ground up, by investigating the very origin of the factor models. Subsequently, we make the loadings dependent on the information about the company and discuss all the features of the model. Lastly, we discuss the requirements needed for successful estimation.

To find a suitable factor model for option returns we utilize a framework similar to Kelly et al. (2019). We start by examining the Euler equation with the stochastic discount factor for investment returns. Only the no-arbitrage assumption is necessary to ensure the existence of a stochastic discount factor which satisfies for any asset return the following equation

$$E(m_{t+1}r_{i,t+1}|I_t) = 0 \iff E(m_{t+1}r_{i,t+1}|I_t) = \underbrace{\frac{cov_t(m_{t+1}r_{i,t+1})}{var_t(m_{t+1})}}_{\beta_{i,t}}\underbrace{\Big(-\frac{var_t(m_{t+1})}{E_t(m_{t+1})}\Big)}_{\lambda_t}, \tag{6}$$

where m_{t+1} is the discount factor at time $t+1$, $r_{i,t+1}$ is the excess return for company i at time $t+1$, and I_t indicates the information set at time t. We define the first part as the loading ($\beta_{i,t}$) which can be viewed as the amount of exposure to the systematic risk factors. Moreover, the second part can be seen as the price of risk (λ_t) related to the aforementioned factors. Given that m_{t+1} depends on a linear function of f_{t+1} Equation (6) is transformed into a factor model with as dependent variable excess returns in line with the methods used in Ross (1976) and Hansen and Richard (1987)

$$r_{i,t+1} = \alpha_{i,t} + \boldsymbol{\beta}_{i,t}^{\mathrm{T}}\boldsymbol{f}_{t+1} + \epsilon_{i,t}, \tag{7}$$

where $\boldsymbol{\beta}_{i,t} : \mathbb{R}^{1\times K}$ is the time-varying and option specific counterpart of the constant loading $\boldsymbol{\beta}$ and the constant and error term are respectively $\alpha_{i,t}$ and $\epsilon_{i,t}$. Moreover we assume $E(\epsilon_{i,t}|I_t) = 0$, $E(\epsilon_{i,t}\boldsymbol{f}_{i,t}|I_t) = \mathbf{0}_{K\times 1}$, $E(\boldsymbol{f}_{t+1}|I_t) = \boldsymbol{\lambda}_t$ and $\alpha_{i,t} = 0$ for all i and t. It follows that with Equation (7) the first flexibility feature of the factor loading, time-varying $\boldsymbol{\beta}_{i,t}$, is met. Now we implement the second feature by modeling the loading $\boldsymbol{\beta}_{i,t}$ as a function of non-return instruments/characteristics ($\boldsymbol{z}_{i,t}$) and obtain the IPCA model specification as defined in Kelly et al. (2019):

$$r_{i,t+1} = \alpha_{i,t} + \boldsymbol{\beta}_{i,t}\boldsymbol{f}_{t+1} + \epsilon_{i,t+1} \text{ where } \epsilon_{i,t+1} \sim \text{WN}(0, \sigma_i^2), \tag{8}$$

where $\alpha_{i,t} = \boldsymbol{z}_{it}^T\boldsymbol{\Gamma}_\alpha + \eta_{i,t}$, $\boldsymbol{\beta}_{i,t} = \boldsymbol{z}_{it}^T\boldsymbol{\Gamma}_\beta + \boldsymbol{u}_{i,t}$ and $\epsilon_{i,t+1} = \eta_{i,t} + \boldsymbol{u}_{i,t}\boldsymbol{f}_{t+1}$.

As Equation (8) describes, the relationship between return i at time $t+1$ ($r_{i,t+1}$) is determined by an intercept ($\alpha_{i,t}$) and a slope coefficient for K factors ($\boldsymbol{\beta}_{i,t} : \mathbb{R}^{1\times K}$) that both depend on L characteristics ($\boldsymbol{z}_{i,t} : \mathbb{R}^{L\times 1}$) and their corresponding weights ($\boldsymbol{\Gamma}_\alpha : \mathbb{R}^{L\times 1}$ and $\boldsymbol{\Gamma}_\beta : \mathbb{R}^{L\times K}$). Moreover, the contemporaneous common risk factor ($\boldsymbol{f}_{t+1} : \mathbb{R}^{K\times 1}$) can be seen as a proxy of sorted portfolio returns dependent on the loadings $\boldsymbol{\beta}_{i,t}$. Besides these estimated parameters the unexplained part is the error for option i at time $t+1$ ($\epsilon_{i,t+1}$) that is a function of errors in the loadings $\eta_{i,t}$ and $\boldsymbol{u}_{i,t} : \mathbb{R}^{1\times K}$. With $\boldsymbol{u}_{i,t}$ we acknowledge that risk exposure of the assets might not be fully retrievable by the instruments, a similar idea holds for $\eta_{i,t}$ which allows for idiosyncratic mispricing unassociated with both the instruments and the risk factor.

As described before, the IPCA is a factor model consisting of varying loadings that depend on characteristics. Including instruments lead to several important benefits. First, the instrument contributes to the model's efficiency even if the instruments are time-invariant as was discovered by Fan et al. (2016). Second, because the characteristics are time-varying, this yields time-variant loadings which leads to a dynamic factor model. With a dynamic factor model one can estimate a conditional return model. Lastly, and most importantly, the inclusion of characteristics helps in identifying what the real drivers of returns are and therefore can answer the main research question: 'Which characteristics explain option returns'.

The linear relation between the characteristics and the dynamic loadings is found by estimating the matrix $\tilde{\boldsymbol{\Gamma}} := [\boldsymbol{\Gamma}_\alpha, \boldsymbol{\Gamma}_\beta] : \mathbb{R}^{L\times(K+1)}$. In previous empirical research, a substantial amount of characteristics have shown to exhibit some degree of power in explaining the variation across returns. Therefore, a matrix is constructed that serves the purpose of reducing the high amount of characteristics for a substantial number of companies to a small matrix. If these characteristics are informative but noisy about the dynamic loadings, they are still able to improve the model since the noise can be averaged out and the signal separated and identified. While $\tilde{\boldsymbol{\Gamma}}$ is time-invariant and equal for every company, it dynamically makes portfolios based on the time-varying characteristics that are firm-specific. For instance, when a firm is small at the beginning of the sample and becomes large over time, the value of the dynamic betas that depend on the time-variant characteristic and constant $\boldsymbol{\Gamma}$ changes. This causes the firm to have a different predicted return in the beginning of the sample compared to the end of the sample, hence the risk-return identity can still be met.

Note that constructing sorted portfolios dynamically is also possible with the method from Horenstein et al. (2018). However, when a portfolio is constructed based on several characteristics and the number of characteristics explodes, their portfolio construction relying on test assets becomes infeasible.[1]

When the return characteristic relation is not driven by compensation for exposure to the latent risk factors, the characteristic does not exhibit risk and allocates the compensation to the intercept $\alpha_{i,t}$. Hence, this contributes to the possibility that $\alpha_{i,t} \neq 0$ and therefore to the possibility of the existence of an anomaly. The estimation of $\alpha_{i,t}$ is conducted by finding a linear combination of the characteristics and the matrix $\boldsymbol{\Gamma}_\alpha$ that is best able to describe the conditional returns while simultaneously controlling for the instruments in $\boldsymbol{\beta}_{i,t}$. If the characteristics align differently to the intercept than to the systematic risk factor, there is anomalous compensation for the delta-hedged option return beyond systematic risk.

The risk factors $\boldsymbol{f}_{t+1}$ capture the variation in returns that can not be arbitraged away. These can be viewed as returns on the dynamic portfolios constructed by $\boldsymbol{\beta}_{i,t}$. While we allow the number of characteristics to be large, the amount of risk factors is to be chosen small.

Several requirements need to be met for successful estimation of the parameters in IPCA according to Kelly et al. (2017) these are as follows

- First, the instrument should be orthogonal to the errors, that is $\mathbb{E}(\boldsymbol{z}_{i,t}\epsilon_{i,t}) = \mathbf{0}_{L\times 1}$. This assumption can be met by assuming that $\mathbb{E}(\boldsymbol{z}_{i,t}\eta_{i,t}) = \mathbf{0}_{L\times 1}$ and $\mathbb{E}(\boldsymbol{z}_{i,t}^{\mathrm{T}}\boldsymbol{u}_{i,t}\boldsymbol{f}_{t+1}) = \mathbf{0}_{L\times 1}$. This condition is comparable to the instrumental variable exclusion restriction that is often made in similar settings and ensures consistency.
- Second, several moments of the characteristics, returns, and a combination of the two need

[1] For instance, when the single sort method with 63 characteristics is used 63 portfolios are needed, while for double sort 2^{63} portfolios are needed.

to exist. These are $\mathbb{E}(\|\boldsymbol{f}_t\boldsymbol{f}_t\|^2))$, $\mathbb{E}(\|\boldsymbol{z}_{i,t}\epsilon_{i,t}\|^2))$, $\mathbb{E}(\|\boldsymbol{z}_{i,t}\boldsymbol{z}_{i,t}^{\mathrm{T}}\|^2)$, $\mathbb{E}(\boldsymbol{\Omega}_t^{z,\epsilon})$, $\mathbb{E}(\|\boldsymbol{z}_{i,t}\boldsymbol{z}_{i,t}^{\mathrm{T}}\|^2\|\boldsymbol{f}_t\boldsymbol{f}_t\|^2)$, and their conditional at time t variants of these moments, where $\boldsymbol{\Omega}_t^{z,\epsilon}$ is the covariance between the characteristics and the error. This assumption is important since it guarantees that $\boldsymbol{\Gamma}_\beta \boldsymbol{Z}_t^{\mathrm{T}} \boldsymbol{Z}_t \boldsymbol{\Gamma}_\beta$ is nonsingular and bounded which is needed in the first-order condition of the IPCA, where instruments data for N_t companies is $\boldsymbol{Z}_t : \mathbb{R}^{L\times N_t}$.

- Third, the parameter space of $\tilde{\boldsymbol{\Gamma}}$ is compact and away from rank deficient, which is the $\det(\tilde{\boldsymbol{\Gamma}}^{\mathrm{T}}\tilde{\boldsymbol{\Gamma}}) > 0$. This is required in the identification assumption of the IPCA.
- Fourth, $\boldsymbol{z}_{i,t}$ is bounded and $\det(\boldsymbol{\Omega}_{z,z}) > 0$.
- The last assumption is the central limit theorem, where we assume
$$\frac{1}{\sqrt{\sum_{t=1}^{T} N_t}} \sum_{i,t} vec(\boldsymbol{z}_{i,t}^{\mathrm{T}} e_{i,t} \tilde{\boldsymbol{f}}_t^{\mathrm{T}}) \xrightarrow{d} N(0, \Omega_{zef}), \tag{9}$$
where $\Omega_{zef} = var(vec(\boldsymbol{z}_{i,t}^{\mathrm{T}} e_{i,t} \tilde{\boldsymbol{f}}_t^{\mathrm{T}}))$.

2.3 Basic IPCA model

As discussed before our main model is the IPCA model as described by Equation (8), in this subsection we will cover two versions of this model. The first version is the restricted model, this model assumes that the $\alpha_{i,t} = 0$ in Equation (8). This means that there do not exist characteristics with a risk-free return, hence no anomalies. The second model that is discussed is the unrestricted model, here we assume that $\alpha_{i,t}$ in Equation (8) is not restricted to 0, and therefore this model allows for the existence of anomalies.

2.3.1 Estimation of the restricted IPCA ($\boldsymbol{\Gamma}_{\boldsymbol{\alpha}} = \boldsymbol{0}_{L\times 1}$)

In this part, we investigate the restricted model, and therefore set $\alpha_{i,t} = 0\ \forall i,t$ in Equation (8). We have to estimate two parameters $\boldsymbol{\Gamma}_\beta$ and $\boldsymbol{f}_{t+1}$ in the equation
$$r_{i,t+1} = \boldsymbol{z}_{i,t}^{\mathrm{T}} \boldsymbol{\Gamma}_\beta \boldsymbol{f}_{t+1} + \epsilon_{i,t+1}^*. \tag{10}$$
In matrix form for N_t companies at time t it becomes
$$\boldsymbol{r}_{t+1} = \boldsymbol{Z}_t \boldsymbol{\Gamma}_\beta \boldsymbol{f}_{t+1} + \boldsymbol{\epsilon}_{t+1}^*, \tag{11}$$
where $\boldsymbol{r}_{t+1} : \mathbb{R}^{(N_t\times 1)}$, $\boldsymbol{Z}_t : \mathbb{R}^{(N_t\times L)}$, $\boldsymbol{\Gamma}_\beta : \mathbb{R}^{(L\times K)}$, $\boldsymbol{f}_{t+1} : \mathbb{R}^{(K\times 1)}$ and $\boldsymbol{\epsilon}_{t+1}^* : \mathbb{R}^{(N_t\times 1)}$. Subsequently, we try to minimize sum of squared errors as shown below
$$\mathcal{L}^{ols}(\boldsymbol{r}_{t+1}, \boldsymbol{Z}_t, \boldsymbol{\Gamma}_\beta, \boldsymbol{f}_{t+1}) = \min_{\boldsymbol{\Gamma}_\beta, \{\boldsymbol{f}_{t+1}\}} \frac{1}{2} \sum_{t=1}^{T-1} (\boldsymbol{r}_{t+1} - \boldsymbol{Z}_t \Gamma_\beta \boldsymbol{f}_{t+1}))^T (\boldsymbol{r}_{t+1} - \boldsymbol{Z}_t \Gamma_\beta \boldsymbol{f}_{t+1}). \tag{12}$$
Equation (12) searches for $\boldsymbol{\Gamma}_\beta$ and $\{\boldsymbol{f}_{t+1}\}_{t=1,\ldots,T-1}$ given a sample of $\{r_{i,t}\}_{t=2,\ldots,T}$, $\{\boldsymbol{z}_{i,t}\}_{t=1,\ldots,T-1}$ that minimize the sum of squared errors over all i and t.

The first-order condition of $\mathcal{L}^{ols}(\boldsymbol{r}_i, \boldsymbol{Z}_t, \hat{\boldsymbol{\Gamma}}_\beta, \boldsymbol{f}_{t+1})$ with respect to $\boldsymbol{f}_{t+1}$ and given $\hat{\boldsymbol{\Gamma}}_\beta$ is

$$\hat{\boldsymbol{f}}_{t+1} = (\hat{\boldsymbol{\Gamma}}_\beta^{\mathrm{T}} \boldsymbol{Z}_t^{\mathrm{T}} \boldsymbol{Z}_t \hat{\boldsymbol{\Gamma}}_\beta)^{-1} \hat{\boldsymbol{\Gamma}}_\beta^{\mathrm{T}} \boldsymbol{Z}_t^{\mathrm{T}} \boldsymbol{r}_{t+1} \ \forall t \tag{13}$$

and for known $\{\hat{\boldsymbol{f}}_{t+1}\}_{t=1,\dots,T-1}$, we can transform Equation (12) to

$$\mathcal{L}^{ols}(\boldsymbol{r}_{t+1}, \boldsymbol{Z}_t, \boldsymbol{\Gamma}_\beta, \hat{\boldsymbol{f}}_{t+1}) = \min_{\boldsymbol{\Gamma}_\beta} \frac{1}{2} \sum_{t=1}^{T-1} (\boldsymbol{r}_{t+1} - \boldsymbol{Z}_t \otimes \hat{\boldsymbol{f}}_{t+1}^{\mathrm{T}} vec(\boldsymbol{\Gamma}_\beta^{\mathrm{T}}))^T (\boldsymbol{r}_{t+1} - \boldsymbol{Z}_t \otimes \hat{\boldsymbol{f}}_{t+1}^{\mathrm{T}} vec(\boldsymbol{\Gamma}_\beta^{\mathrm{T}})). \tag{14}$$

Subsequently, we take the first-order condition with respect to $vec(\boldsymbol{\Gamma}_\beta^{\mathrm{T}})$

$$\begin{aligned} vec(\boldsymbol{\Gamma}_\beta^{\mathrm{T}}) = (\sum_{t=1}^{T-1} [\boldsymbol{Z_t} \otimes \hat{\boldsymbol{f}}_{t+1}^{\mathrm{T}}]^{\mathrm{T}} [\boldsymbol{Z}_t \otimes \hat{\boldsymbol{f}}_{t+1}^{\mathrm{T}}])^{-1} (\sum_{t=1}^{T-1} [\boldsymbol{Z_t} \otimes \hat{\boldsymbol{f}}_{t+1}^{\mathrm{T}}]^{\mathrm{T}} \boldsymbol{r}_{t+1}) \overset{2}{=} \\ (\sum_{t=1}^{T-1} \boldsymbol{Z}_t^{\mathrm{T}} \boldsymbol{Z}_t \otimes \hat{\boldsymbol{f}}_{t+1} \hat{\boldsymbol{f}}_{t+1}^{\mathrm{T}})^{-1} (\sum_{t=1}^{T-1} [\boldsymbol{Z}_t^{\mathrm{T}} \boldsymbol{r}_{t+1} \otimes \hat{\boldsymbol{f}}_t]). \end{aligned} \tag{15}$$

[3] Since the companies in our sample do not have an observation every month, we have an unbalanced sample and hence the matrix multiplication of $\boldsymbol{Z}_t^T$ and $\boldsymbol{r}_{t+1}$ is not always possible. The reason for missing observations are: the company is not listed on an exchange every month during our entire sample period due to for instance an initial public offering, bankruptcy, taking private during our sample period or there did not exist any option that met our requirements, as defined in Section 3. Kelly et al. (2019) showed that the IPCA problem can be approximately solved by using a singular value decomposition (SVD) on characteristic-managed portfolio returns defined in Equation (1) instead of the raw returns.

We can simplify the first-order conditions resulting from Equation (12) by using $\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{Z}_t^{\mathrm{T}} \boldsymbol{Z}_t$ instead of $\boldsymbol{Z}_t^{\mathrm{T}} \boldsymbol{Z}_t \ \forall t$. When this approach is used $\boldsymbol{\Gamma}_\beta$ is equal to the K first eigenvectors of matrix $\sum_{t=1}^{T-1} \boldsymbol{x}_t \boldsymbol{x}_t$. And $\boldsymbol{f}_{t+1}$ would be simplified to the first K principal components of the matrix with managed portfolios for the entire sample $\boldsymbol{X} : \mathbb{R}^{(L \times T)}$ as defined in $\boldsymbol{X} := [\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_T]$. Although these values for $\boldsymbol{\Gamma}_\beta$ and $\boldsymbol{f}_{t+1}$ are not the exact solution of Equation (12) they offer a good starting point ($\boldsymbol{\Gamma}_\beta^{(0)}$ and $\boldsymbol{f}_{t+1}^{(0)}$), especially when the values of the characteristics do not differ substantially over the time period which means for all t $\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{Z}_t^{\mathrm{T}} \boldsymbol{Z}_t \approx \boldsymbol{Z}_t^{\mathrm{T}} \boldsymbol{Z}_t$. To be certain that our algorithm does not converge to a local minimum, we add a small random error to the previously defined starting points and subsequently iterate n times between Equation (13) and (15) until the difference between $\boldsymbol{\Gamma}_\beta^{(n)}$, $\boldsymbol{\Gamma}_\beta^{(n-1)}$ and $\boldsymbol{f}_{t+1}^{(n)}$, $\boldsymbol{f}_{t+1}^{(n-1)}$ are negligible for each element of the vector or matrix. When this process is repeated several times with slightly different starting values and still converges to the same optimal $\boldsymbol{\Gamma}_\beta^{(n)}$ and $\boldsymbol{f}_{t+1}^{(n)}$, we can be more certain that these parameters are indeed part of the global optima than in the scenario where we did not consider this randomness.

[3]by Kronecker product rules (1) the transpose property rule $(\boldsymbol{A} \otimes \boldsymbol{B})^{\mathrm{T}} = \boldsymbol{A}^{\mathrm{T}} \otimes \boldsymbol{B}^{\mathrm{T}}$ and (2) by mixed-product property $(\boldsymbol{A}^{\mathrm{T}} \otimes \boldsymbol{B}^{\mathrm{T}})(\boldsymbol{A} \otimes \boldsymbol{B}) = \boldsymbol{A}^{\mathrm{T}} \boldsymbol{A} \otimes \boldsymbol{B}^{\mathrm{T}} \boldsymbol{B}$ where the mixed property implies $(\boldsymbol{A} \otimes \boldsymbol{B})\boldsymbol{C} = \boldsymbol{AC} \otimes \boldsymbol{B}$ for matrices $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$

Rotational unidentification is a known issue in latent factor models as described in Kelly et al. (2019), for instance for any nonsingular $\boldsymbol{R} : \mathbb{R}^{(K\times K)}$ the combination of $\boldsymbol{\Gamma}_\beta \boldsymbol{R}$ and $\boldsymbol{R}^{-1}\boldsymbol{f}_{t+1}$ is also a solution of Equation (12). Hence, without any restrictions on $\boldsymbol{\Gamma}_\beta$ and $\boldsymbol{f}_{t+1}$ the IPCA algorithm has severe difficulty finding the optimal solution because at every iteration another 'optimal' solution is found. Therefore we follow Kelly et al. (2019) and Kelly et al. (2017) and restrict the parameters at every iteration such that $\boldsymbol{R}$ does not affect our solution and hence the factors do not jump around. The unidentifaction problem is solved by imposing the following restrictions $\boldsymbol{\Gamma}_\beta^{\mathrm{T}}\boldsymbol{\Gamma}_\beta = \boldsymbol{I}_k$, $E(\boldsymbol{f}_t) \geq 0$ and $cov(\boldsymbol{f}_t) = \boldsymbol{A}$ where $a_{i,j} = 0$ if $i \neq j$ and $a_{i,i} > a_{i+1,i+1}$. Note that these assumptions do not affect the sum of errors or change the economic meaning of the parameters.

2.3.2 Estimation of unrestricted model IPCA ($\boldsymbol{\Gamma_\alpha} \neq \mathbf{0}_{L\times 1}$)

In the unrestricted model we allow for the possibility that characteristics explain part of the return without bearing systematic risk, hence this model exhibits an anomaly feature. This translates to the unrestricted model having an intercept and is defined as follows

$$\boldsymbol{r}_{t+1} = \boldsymbol{Z}_t^{\mathrm{T}}\boldsymbol{\Gamma}_\alpha + \boldsymbol{Z}_t^{\mathrm{T}}\boldsymbol{\Gamma}_\beta \boldsymbol{f}_{t+1} + \boldsymbol{\epsilon}^*_{t+1}. \tag{16}$$

First we define $[\boldsymbol{\Gamma}_\alpha, \boldsymbol{\Gamma}_\beta] := \tilde{\boldsymbol{\Gamma}} : \mathbb{R}^{L\times(K+1)}$ and $[\mathbf{1}, \boldsymbol{f}_{t+1}] := \tilde{\boldsymbol{f}}_{t+1} : \mathbb{R}^{(K+1)\times 1}$. Subsequently, by taking first-order conditions the optimal $\boldsymbol{f}_{t+1}$ given $\hat{\boldsymbol{\Gamma}}$ is determined by

$$\boldsymbol{f}_{t+1} = (\hat{\boldsymbol{\Gamma}}_\beta^{\mathrm{T}}\boldsymbol{Z}_t^{\mathrm{T}}\boldsymbol{Z}_t\hat{\boldsymbol{\Gamma}}_\beta)^{-1}\hat{\boldsymbol{\Gamma}}_\beta^{\mathrm{T}}\boldsymbol{Z}_t^{\mathrm{T}}(\boldsymbol{r}_{t+1} - \boldsymbol{Z}_t\hat{\boldsymbol{\Gamma}}_\alpha). \tag{17}$$

Logically, the constant part of $\tilde{\boldsymbol{f}}_{t+1}$ is unaffected and remains 1 throughout the whole iteration procedure. Moreover, in order to calculate the optimal $\tilde{\boldsymbol{\Gamma}}$, we slightly adjust Equation (15) by replacing $\boldsymbol{\Gamma}_\beta$ with $\tilde{\boldsymbol{\Gamma}}$ and $\boldsymbol{f}_{t+1}$ with $\tilde{\boldsymbol{f}}_{t+1}$. Note that $\tilde{\boldsymbol{\Gamma}}$ will always have at least two columns of which the first one will be related to $\boldsymbol{\Gamma}_\alpha$ and the remainder of the columns will be related to $\boldsymbol{\Gamma}_\beta$. Apart from the identification restrictions imposed by the unrestricted model, we need additional restrictions to ensure a unique solution, hence we restrict $\boldsymbol{\Gamma}_\alpha$ to be orthogonal to $\boldsymbol{\Gamma}_\beta$ that is $\boldsymbol{\Gamma}_\alpha^{\mathrm{T}}\boldsymbol{\Gamma}_\beta = \mathbf{0}_{1\times K}$. To achieve this result we conduct the following steps. We use the estimated pair of gammas from the unrestricted version of Equation (15) to find the new $\boldsymbol{\Gamma}_\alpha = (\boldsymbol{I}_L - \boldsymbol{\Gamma}_\beta\boldsymbol{\Gamma}_\beta^{\mathrm{T}})\boldsymbol{\Gamma}_\alpha$ and the new $\boldsymbol{f}_t = \boldsymbol{f}_t + \boldsymbol{\Gamma}_\beta^{\mathrm{T}}\boldsymbol{\Gamma}_\alpha$. When these equations are used, the risk loadings can describe as much as possible of the options mean returns. Moreover, the intercept gets only assigned the orthogonal residual left over from the total return prediction of the instruments.

2.4 Regularized IPCA models

As described in Kelly et al. (2019), IPCA has a dimension reduction feature. For instance, when one risk factor is added, the number of estimated parameters increases with $L+T$ and when one characteristic is added K additional parameters need to be estimated. While this seems high, it is substantially lower than PCA which increases with $(T+N)$ for every additional factor. For a model that consists of 63 characteristics, 4 factors and a period of 112 months, we still obtain many parameters (N_p) of $N_p = LK + TK = 63 \times 4 + 112 \times 4 = 700$ parameters, of which a considerable amount is not necessary due to low statistical significance or magnitude of the characteristic.

In order to find a model with a more parsimonious structure, several supervised machine learning techniques can be used. The two most famous techniques are the ridge and least absolute shrinkage and selection operator (lasso) regression. Although both techniques share some commonalities such as an objective function that penalizes the sum of the coefficients, they also differ in several ways such as the fierceness of their restrictions. The ridge regression is discussed first, subsequently we explain the lasso regression. The last function that is considered is the elastic net which is a combination of the ridge and lasso.

Note that one of the common practices according to Marquardt and Snee (1975) for the ridge regression and Tibshirani (1996) for the lasso is the standardization of the instruments. This involves two components, (1) centering which sets the mean to 0, and (2) scaling which transforms the values of the variables such that all variables have equal variances. Centering has the advantage that it removes unnecessary ill-conditioning and hence reduces variance inflation. Scaling makes interpretation easier since the magnitude of the coefficients does not depend on the scale of the characteristic.

2.4.1 Ridge regression

The ridge regression constructed by Tikhonov et al. (1963) penalizes the number of regression coefficients by using the square of the magnitude of these coefficients as the penalty. Adding the penalty $\phi^{ridge}(\lambda) = \frac{1}{2}\lambda \sum_{l=1}^{L} \sum_{l=1}^{K} \gamma_{\beta,l,k}^2$ with ridge parameter $\lambda \in [0,1]$ to the objective of the restricted IPCA as in Equation (12) yields:

$$\mathcal{L}^{ridge}(\boldsymbol{r}_t, \boldsymbol{Z}_t, \boldsymbol{\Gamma}_\beta, \boldsymbol{f}_{t+1}, \lambda) = \min_{\boldsymbol{\Gamma}, \{\boldsymbol{f}_{t+1}\}} \frac{1}{2} \sum_{t=1}^{T-1} (\boldsymbol{r}_{t+1} - \boldsymbol{Z}_t \boldsymbol{\Gamma}_\beta \boldsymbol{f}_{t+1})^T (\boldsymbol{r}_{t+1} - \boldsymbol{Z}_t \boldsymbol{\Gamma}_\beta \boldsymbol{f}_{t+1}) + \frac{1}{2}\lambda \sum_{l=1}^{L} \sum_{l=1}^{K} \gamma_{\beta,l,k}^2. \tag{18}$$

Note that we do not penalize the $\boldsymbol{f}_t$ factor, because of two reasons. First, these values correspond to the return on the 'sorted' portfolio and hence is not our aim to shrink to zero. Second, we are interested in understanding the most important drivers of returns, which can be inspected by the magnitude and statistical significance of the $\boldsymbol{\Gamma}_\beta$ coefficients. The ridge parameter λ determines the weight that is given to the penalty function, a higher value for λ corresponds to more weight and therefore more coefficients close to zero. The coefficients of $\boldsymbol{\Gamma}_\beta$ do not become 0, and hence there is model shrinkage but no model selection. The minimization of Equation (18) yields the optimal $\boldsymbol{\Gamma}_\beta^{ridge}$ defined in the equation below

$$Vec({\boldsymbol{\Gamma}_\beta}^{ridge^{\mathrm{T}}}) = [(\sum_{t=1}^{T-1}(\boldsymbol{Z}_t^{\mathrm{T}}\boldsymbol{Z}_t) \otimes \hat{\boldsymbol{f}}_{t+1}\hat{\boldsymbol{f}}_{t+1}^{\mathrm{T}}) + \lambda \boldsymbol{I}_{LK}]^{-1}(\sum_{t=1}^{T-1}[\boldsymbol{Z}_t \otimes \hat{\boldsymbol{f}}_{t+1}^{\mathrm{T}}]^{\mathrm{T}}\boldsymbol{r}_{t+1}), \tag{19}$$

where $\boldsymbol{I}_{LK} : \mathbb{R}^{LK \times LK}$ is the identity matrix of dimension LK.

The previously defined orthonormality identification restriction on $\boldsymbol{\Gamma}_\beta$ was imposed to find a unique set of $\boldsymbol{\Gamma}_\beta$ and $\boldsymbol{f}_{t+1}$. However, with the ridge regression, we want to achieve that the sum of squared elements of $\boldsymbol{\Gamma}_\beta$ is lower than in the scenario when no penalty is considered. Hence, we relax the assumption on $\boldsymbol{\Gamma}_\beta$ to merely orthogonality whereas we restrict the factors to be orthonormal instead of orthogonal with decreasing elements. This method ensures that we still find a unique solution.

When the optimal set of $\boldsymbol{\Gamma}_\beta^{ridge}(\lambda)$ and $\boldsymbol{f}_t^{ridge}(\lambda)$ is found, we normalize them such that it is easier to compare the $\boldsymbol{\Gamma}_\beta(\lambda)$ for different λ. This normalization is conducted by the following procedure. First, we compute the variance-covariance matrix of the factors for the standard IPCA case ($\lambda = 0$) based on Equation (13) and their identification assumptions and define it as $\boldsymbol{R} = \boldsymbol{f}\boldsymbol{f}^{\mathrm{T}} : \mathbb{R}^{K \times K}$ which is a positive definite matrix. Second, we use this $\boldsymbol{R}$ matrix to uncover the normalized weights ($\boldsymbol{NW}(\lambda) : R^{K \times K}$) defined as $\boldsymbol{NW}(\lambda)) = (\boldsymbol{R}(\boldsymbol{f}^{ridge}(\lambda)\boldsymbol{f}^{ridge^{\mathrm{T}}}(\lambda))^{-1})^{\frac{1}{2}}$. Third, calculate the normalized ridge parameters $\boldsymbol{\Gamma}_\beta^{ridge,n}(\lambda) = \boldsymbol{\Gamma}_\beta^{ridge}(\lambda)\boldsymbol{NW}(\lambda)$ and $\boldsymbol{f}_t^{ridge,n}(\lambda) = \boldsymbol{NW}(\lambda)^{-1}\boldsymbol{f}_t^{ridge}(\lambda)$.[4] Note that this does not affect the predicted returns in any way since the product of $\boldsymbol{NW}$ and inverse of $\boldsymbol{NW}$ yield an identity matrix and hence is only used for comparison purposes. To conclude, we have now achieved that the normalized risk factors based on the ridge regression have an equal variance to the standard case, whereas the coefficients in the $\boldsymbol{\Gamma}_\beta^{ridge,n}(\lambda)$ matrix are smaller in total.

[4] $\boldsymbol{A}^{\frac{1}{2}}$ for matrix $\boldsymbol{A} : \mathbb{R}^{n \times n}$ is calculated utilizing a Cholesky decomposition $\boldsymbol{A} = \boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}$

2.4.2 Lasso operator

The lasso regression considers a penalty function with an absolute value of the coefficients instead of a square. This results in the possibility that the coefficients become equal to zero, instead of near-zero as in the ridge regression. Therefore, the lasso decides which variables should be removed from the model and provides an additional possibility to infer which characteristics are most important in explaining the variation in returns. The penalty function is defined as $\phi^{lasso}(\lambda) = \lambda \sum_{l=1}^{L} \sum_{l=1}^{K} |\gamma_{\beta,l,k}|$ which is added to IPCA objective Equation (12) and therefore the lasso objective equation becomes

$$\mathcal{L}^{lasso}(\boldsymbol{r}_t, \boldsymbol{Z}_t, \boldsymbol{\Gamma}_\beta, \boldsymbol{f}_{t+1}, \lambda) = \min_{\boldsymbol{\Gamma},\{\boldsymbol{f}_{t+1}\}} \sum_{t=1}^{T-1} (\boldsymbol{r}_{t+1} - \boldsymbol{Z}_t \boldsymbol{\Gamma}_\beta \boldsymbol{f}_{t+1}))^T (\boldsymbol{r}_{t+1} - \boldsymbol{Z}_t \boldsymbol{\Gamma}_\beta \boldsymbol{f}_{t+1}) + \lambda \sum_{l=1}^{L} \sum_{l=1}^{K} |\gamma_{\beta,l,k}|. \tag{20}$$

When λ is relatively large, this objective function will set a large number of parameters $\gamma_{\beta,l,k}$ to zero instead of almost zero as in the ridge regression. Hence, this model will lead to a reduced form of the IPCA which we call sparse instrumented principal component analysis (SIPCA) which is the advanced version of the sparse principal component analysis (SPCA) constructed by Zou et al. (2006). The downside of the lasso method according to Owen (2007) is that when a set of strongly correlated characteristics with large effects is used in the model, it is inclined to set a high portion or all except one of the characteristics to zero. Moreover, according to Tibshirani (1996) the lasso regression is outperformed by the ridge regression when the number of observations exceeds the number of coefficients to be estimated and when there exist high correlations between the characteristics. With the lasso we mainly want to use the variable selection feature, and therefore can impose the same identification assumptions as normally.

To derive the $\boldsymbol{\Gamma}_\beta$ and ease implementation we transform Equation (20) into

$$\mathcal{L}^{lasso}(\boldsymbol{r}_t, \boldsymbol{Z}_t, \boldsymbol{\Gamma}_\beta, \boldsymbol{f}_{t+1}, \lambda) = \min_{\boldsymbol{\Gamma},\{\boldsymbol{f}_{t+1}\}} \frac{1}{2} \sum_{t=1}^{T-1} (\boldsymbol{r}_{t+1} - \boldsymbol{Z}_t \otimes \boldsymbol{f}_{t+1}^{\mathrm{T}} vec(\boldsymbol{\Gamma}_\beta^{\mathrm{T}}))^{\mathrm{T}} (\boldsymbol{r}_{t+1} - \boldsymbol{Z}_t \otimes \boldsymbol{f}_{t+1}^{\mathrm{T}} vec(\boldsymbol{\Gamma}_\beta^{\mathrm{T}})) + \lambda \sum_{j=1}^{LK} |\gamma_{\beta,j}|. \tag{21}$$

When both functions are convex and continuous we can according to Rockafellar (1970) use Moreau-Rockafellar theorem and split the combined derivative into two parts $\partial(\mathcal{L}^{lasso}) = \partial\mathcal{L}^{ols} + \partial\phi^{lasso}(\lambda)$. Since both parts of Equation (21) are convex and continuous everywhere except at 0 we can implement their theorem for all values excluding 0. Hence we need to treat $\gamma_{\beta,l,k} = 0$ differently. Since the coordinate descent step procedure described in Friedman et al. (2010) is commonly used in the practice of finding optimal parameters in a lasso setting, we also use this method. We initially focus on the first part of Equation (21) which notation is slightly changed ($\boldsymbol{C}_t = \boldsymbol{Z}_t \otimes \boldsymbol{f}_{t+1}^{\mathrm{T}} : \mathbb{R}^{N_t \times LK}$) to obtain neater equations as in

$$\mathcal{L}(\boldsymbol{r}_t, \boldsymbol{z}_t, \boldsymbol{\Gamma}_\beta, \boldsymbol{f}_{t+1}) = \frac{1}{2}\sum_{t=1}^{T-1}(\boldsymbol{r}_{t+1} - \boldsymbol{C}_t vec(\boldsymbol{\Gamma}_\beta^{\mathrm{T}}))^{\mathrm{T}}(\boldsymbol{r}_{t+1} - \boldsymbol{C}_t vec(\boldsymbol{\Gamma}_\beta^{\mathrm{T}}))$$

$$= \frac{1}{2}\sum_{t=1}^{T-1}(\boldsymbol{r}_{t+1} - (\boldsymbol{c}_{t,1}, \boldsymbol{c}_{t,2}, \ldots, \boldsymbol{c}_{t,LK})(\gamma_{\beta,1}, \gamma_{\beta,2}, \ldots, \gamma_{\beta,LK})^{\mathrm{T}})^{\mathrm{T}}(\boldsymbol{r}_{t+1} - (\boldsymbol{c}_{t,1}, \boldsymbol{c}_{t,2}, \ldots, \boldsymbol{c}_{t,LK})(\gamma_{\beta,1}, \gamma_{\beta,2}, \ldots, \gamma_{\beta,LK})^{\mathrm{T}})$$

$$= \frac{1}{2}\sum_{t=1}^{T-1}(\boldsymbol{r}_{t+1} - (\boldsymbol{c}_{t,1}\gamma_{\beta,1}, \boldsymbol{c}_{t,2}\gamma_{\beta,2}, \ldots, \boldsymbol{c}_{t,LK}\gamma_{\beta,LK}))^{\mathrm{T}}(\boldsymbol{r}_{t+1} - (\boldsymbol{c}_{t,1}\gamma_{\beta,1}, \boldsymbol{c}_{t,2}\gamma_{\beta,2}, \ldots, \boldsymbol{c}_{t,LK}\gamma_{\beta,LK}))$$

$$= \frac{1}{2}\sum_{t=1}^{T-1}(\boldsymbol{r}_{t+1} - \sum_{j=1}^{LK}(\boldsymbol{c}_{t,j}\gamma_{\beta,j}))^{\mathrm{T}}(\boldsymbol{r}_{t+1} - \sum_{j=1}^{LK}(\boldsymbol{c}_{t,j}\gamma_{\beta,j})).$$

In order to find the optimal $\gamma_{\beta,lk}$ we take first-order conditions of $\mathcal{L}(r_{i,t}, \boldsymbol{z}_{i,t}, \boldsymbol{\Gamma}_\beta, \boldsymbol{f}_{t+1})$ to $\gamma_{\beta,lk}$

$$\frac{\partial\mathcal{L}(r_{i,t}, \boldsymbol{z}_{i,t}, \boldsymbol{\Gamma}_\beta, \boldsymbol{f}_{t+1})}{\partial\gamma_{\beta,lk}} = -\sum_{t=1}^{T-1}\boldsymbol{c}_{t,k}^{\mathrm{T}}(\boldsymbol{r}_{t+1} - \sum_{j=1}^{LK}\boldsymbol{c}_{t,j}\gamma_{\beta,j}) = -\sum_{t=1}^{T-1}\boldsymbol{c}_{t,k}^{\mathrm{T}}(\boldsymbol{r}_{t+1} - \sum_{j\neq lk}\boldsymbol{c}_{t,j}\gamma_{\beta,j} - \boldsymbol{c}_{t,k}\gamma_{\beta,lk})$$

$$= -\sum_{t=1}^{T-1}\boldsymbol{c}_{t,k}^{\mathrm{T}}(\boldsymbol{r}_{t+1} - \sum_{j\neq lk}\boldsymbol{c}_{t,j}\gamma_{\beta,j}) + \gamma_{\beta,lk}\sum_{t=1}^{T-1}\boldsymbol{c}_{t,k}^{\mathrm{T}}\boldsymbol{c}_{t,k} = -p_k + \gamma_{\beta,lk}z_k, \tag{22}$$

where $p_k = \sum_{t=1}^{T-1}\boldsymbol{c}_{t,k}^{\mathrm{T}}(\boldsymbol{r}_{t+1} - \sum_{j\neq lk}\boldsymbol{c}_{t,j}\gamma_{\beta,j})$ and $z_k = \sum_{t=1}^{T-1}\boldsymbol{c}_{t,k}^{\mathrm{T}}\boldsymbol{c}_{t,k}$ are used to ease notation. Subsequently, we add the partial derivative of the OLS and the penalty and set it to 0

$$\frac{\partial\mathcal{L}^{lasso}}{\partial\gamma_{\beta,lk}} = \frac{\partial\mathcal{L}^{ols}}{\partial\gamma_{\beta,lk}} + \frac{\partial\mathcal{L}^{penalty}}{\partial\gamma_{\beta,lk}} = 0 \iff -p_k + \gamma_{\beta,lk}z_k + \partial_{\gamma_{\beta,lk}}\lambda|\gamma_{\beta,lk}| = 0. \tag{23}$$

This results in three different equations depending on the value of $\gamma_{\beta,lk}$, including special case $\gamma_{\beta,lk} = 0$

$$0 = \begin{cases} -p_k + \gamma_{\beta,lk}z_k - \lambda & \text{if} \quad \gamma_{\beta,lk} < 0 \\ |-p_k - \lambda, -p_k + \lambda| & \text{if} \quad \gamma_{\beta,lk} = 0 \\ -p_k + \gamma_{\beta,lk}z_k + \lambda & \text{if} \quad \gamma_{\beta,lk} > 0 \end{cases}$$

We take into account that the closed interval of the second scenario should contain 0 to ensure that $\gamma_{\beta,lk}$ is a global minimum. Therefore the system of equations is transformed to:

$$\frac{1}{z_k}(p_k, \lambda) = \begin{cases} \frac{p_k+\lambda}{z_k} & \text{if} \quad p_k < -\lambda \\ 0 & \text{if} \quad -\lambda \leq p_k \leq \lambda \\ \frac{p_k-\lambda}{z_k} & \text{if} \quad p_k > \lambda \end{cases}$$

Since all necessary equations are defined, we can perform the following iterative procedure that gives us $\gamma_{\beta,lk}$

Algorithm 1: Coordinate descent step algorithm.

- For $k = 1, \ldots, KL$;
- Calculate $p_k = \sum_{t=1}^{T-1}\boldsymbol{c}_{t,k}^{\mathrm{T}}(\boldsymbol{r}_{t+1} - \sum_{j\neq lk}\boldsymbol{c}_{t,j}\gamma_{\beta,j})$;
- Calculate $z_k = \sum_{t=1}^{T-1}\boldsymbol{c}_{t,k}^{\mathrm{T}}\boldsymbol{c}_{t,k}$;
- Set $\gamma_{\beta,lk} = \frac{1}{z_k}(p_k, \lambda)$.

Subsequently, we take the vector of $\{\gamma_{\beta,i}\}_{i=1,\ldots,LK}$ as given and compute the vector $\boldsymbol{f}_t$ using the normal first-order condition Equation (13) and repeat the algorithm until convergence. Lastly, note that similarly to the normal IPCA, this algorithm only contains companies that have an observation for two consecutive periods.

2.4.3 Elastic net

A hybrid version of the lasso and ridge is the elastic net method. This method overcomes some of the downsides of the lasso and the ridge method. The penalty function used in the elastic net estimation leads to variable selection and coefficient shrinkage and hence allows for highly correlated instruments. The penalty of this method according to Zou and Hastie (2005) is defined as

$$\phi^{en}(\gamma, \lambda, \rho) = \lambda\rho \sum_{j=1}^{LK} |\gamma_{\beta,j}| + \frac{1}{2}\lambda(1-\rho) \sum_{j=1}^{LK} \gamma_{\beta,j}^2, \tag{24}$$

where hyperparameter $\rho \in [0, 1]$ determines the percentage that is allocated to the lasso penalty and the remainder $(1-\rho)$ is assigned to the ridge. Similarly as in the ridge and lasso, the penalty is added to the basic IPCA model objective Equation (12) and therefore becomes

$$\mathcal{L}^{lasso}(\boldsymbol{r}_t, \boldsymbol{Z}_t, \boldsymbol{\Gamma}_\beta, \boldsymbol{f}_{t+1}, \lambda) = \min_{\boldsymbol{\Gamma}, \{\boldsymbol{f}_{t+1}\}} \sum_{t=1}^{T-1} (\boldsymbol{r}_{t+1} - \boldsymbol{Z}_t \boldsymbol{\Gamma}_\beta \boldsymbol{f}_{t+1}))^T (\boldsymbol{r}_{t+1} - \boldsymbol{Z}_t \boldsymbol{\Gamma}_\beta \boldsymbol{f}_{t+1}) + \lambda\rho \sum_{j=1}^{LK} |\gamma_{\beta,j}| + \frac{1}{2}\lambda(1-\rho) \sum_{j=1}^{LK} \gamma_{\beta,j}^2. \tag{25}$$

To find the optimal $\boldsymbol{\Gamma}_\beta$ in an elastic net setting, a similar framework as the lasso in Section 2.4.2 that is slightly adjusted for the different penalty function can be applied. Hence we split the elastic net objective equation into the OLS part and the penalty function. The first part as shown in Equation (22) yielded $\frac{\partial \mathcal{L}^{OLS}}{\partial \gamma_l k} = -p_{lk} + \gamma_{lk} z_{lk}$, whereas the second part is now adjusted to $\frac{\partial \phi^{en}}{\partial \gamma_{lk}} = \partial_{\gamma_{lk}} \lambda\rho|\gamma_{lk}| + \lambda(1-\rho)\gamma_{lk}$. Together this leads to the following system of equations

$$0 = \begin{cases} -p_{lk} + \gamma_{lk} z_{lk} - \lambda\rho + \lambda(1-\rho)\gamma_{lk} & \text{if} \quad \gamma_{lk} < 0 \\ |-p_{lk} - \lambda\rho, -p_{lk} + \lambda\rho| & \text{if} \quad \gamma_{lk} = 0 \\ -p_{lk} + \gamma_{lk} z_{lk} + \lambda\rho + \lambda(1-\rho)\gamma_{lk} & \text{if} \quad \gamma_{lk} > 0 \end{cases} \tag{26}$$

And by adjusting the equation it follows that

$$\gamma_{lk} = \begin{cases} \frac{p_{lk} + \lambda\rho}{z_{lk} + \lambda(1-\rho)} & \text{if} \quad p_{lk} < 0 \text{ and } \lambda\rho < |p_{lk}| \\ 0 & \text{if} \quad \lambda\rho \geq |p_{lk}| \\ \frac{p_{lk} - \lambda\rho}{z_{lk} + \lambda(1-\rho)} & \text{if} \quad p_{lk} > 0 \text{ and } \lambda\rho < |p_{lk}| \end{cases} \tag{27}$$

This expression is similar to Donoho and Johnstone (1994) apart from the fact that they normalized z_k to 1. Algorithm 1 from Section 2.4.2 that uses Equation (27) in the fourth step is applied to find the complete matrix $\boldsymbol{\Gamma}_\beta$.

2.5 Robust IPCA models

One of the assumptions we made in Equation (8) was that the errors of the model were white noise $\epsilon_{i,t+1} \sim \text{WN}(0,\sigma_i)$ for every asset i. White noise errors are (1) serially uncorrelated and

(2) normally distributed $N(0, \sigma_i)$. The Durbin Watson test for panel data (d_{pd}) can be applied to test autocorrelation in the errors. This statistics investigates how well the lagged error value ($\epsilon_{i,t}$) explains $\epsilon_{i,t+1}$ that is

$$d_{pd} = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T_i - 1} (\epsilon_{i,t+1} - \epsilon_{i,t})^2}{\sum_{i=1}^{N} \sum_{t=1}^{T_i} \epsilon_{i,t+1}^2} \tag{28}$$

where $d_{pd} \in [0, 4]$. A d_{pd} around 0, 2 or 4 indicates respectively strong positive, no, or strong negative correlation. To understand whether assumption (2) holds, we investigate the Normal Q-Q plots of the restricted IPCA model by individual options and of the whole dataset.

Although an OLS framework has favorable properties, wrong results can be obtained when the assumptions are not satisfied. For instance, OLS is very sensitive to outliers in the variables and given that our data consists of returns that are often heavy-tailed we should consider a robust regression. A Huber loss function can tackle the problem of outliers whereas the heteroskedasticity problem can be solved by WLS. Another method to solve these problems is using white standard errors instead of regular standard errors in the calculation of the p-values and F-values of the characteristics. We proceed first with the robust regression and explain the second method in Section 2.6.

2.5.1 Huber loss function

In order to tackle the problem of extreme observations, a modified OLS model with Huber loss function is constructed by Huber (1992). The Huber loss framework from Gu et al. (2019) is applied on the regular IPCA as in Equation (12) that together yields:

$$\mathcal{L}^{Huber}(r_{i,t}, \boldsymbol{z}_{i,t}, \boldsymbol{\Gamma}_\beta, \boldsymbol{f}_{t+1}, \xi) = \min_{\boldsymbol{\Gamma}_\beta, \{\boldsymbol{f}_{t+1}\}} \sum_i \sum_t H(r_{i,t+1} - \boldsymbol{z}_{i,t}^{\mathrm{T}} \boldsymbol{\Gamma}_\beta \boldsymbol{f}_{t+1}, \xi), \tag{29}$$

where $\mathcal{L}^{Huber}(r_{i,t}, \boldsymbol{z}_{i,t}, \boldsymbol{\Gamma}_\beta, \boldsymbol{f}_{t+1}, \xi) : \mathbb{R} \times \mathbb{R}^{1\times L} \times \mathbb{R}^{L\times K} \times \mathbb{R}^{K\times 1} \times \mathbb{R} \mapsto \mathbb{R}$ and $H(.) : \mathbb{R} \times \mathbb{R} \mapsto \mathbb{R}$ is the Huber Loss function defined below

$$H(x, \xi) := \begin{cases} \frac{1}{2}x^2 & \text{if} \quad |x| \leq \xi \\ \xi|x| - \frac{1}{2}\xi^2 & \text{if} \quad |x| > \xi \end{cases}$$

where ξ is a tuning constant that is set to a value such that a high-efficiency level is achieved while simultaneously aiming for protection against outliers. The protection against outliers is achieved because the Huber loss is a combination of a squared loss function and absolute loss function for small and large errors respectively. Now we know the properties of the Huber loss function we can relate it to the general M-estimation case that is used to obtain a robust regression by Fox and Weisberg (2002). The minimization equation used in their M-estimation method is defined by

$$\mathcal{L}^{M}(y_i, \boldsymbol{x}_i, \boldsymbol{b}) = \sum_{i=1}^{n} \rho(y_i - \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{b}), \tag{30}$$

where y_i, $\boldsymbol{x}_i : \mathbb{R}^{L\times 1}$, $\boldsymbol{b} : \mathbb{R}^{L\times 1}$ are the dependent variable, covariates and coefficients respectively. The function $\rho() : \mathbb{R} \times \mathbb{R}^{L\times 1} \times \mathbb{R}^{L\times 1} \mapsto \mathbb{R}$ needs to satisfy the following properties:

- Always non-negative $\rho(e) > 0$;
- Equal to zero when its argument is zero, $p(0) = 0$;
- Symmetric $\rho(e) = \rho(-e)$;
- Monotone in $|e_i|$, $p(e_i) \geq p(e_j)$ for $|e_i| > |e_j|$.

All these properties are satisfied for the Huber Loss function, therefore we can implement the technique of Fox and Weisberg (2002) and generalize it to panel data instead of cross-sectional data. The substitution of our variables into their framework yields equation

$$\sum_{i=1}^{N_t} \rho(r_{i,t+1} - \boldsymbol{z}_{i,t}^{\mathrm{T}} \boldsymbol{\Gamma}_\beta \boldsymbol{f}_{t+1}). \tag{31}$$

This equation serves the purpose of calculating the risk factor $\boldsymbol{f}_{t+1}$ which can be transformed to an equation that is easier for the computation of the loadings $\boldsymbol{\Gamma}_\beta$ as in

$$\sum_{i=1}^{T-1} \sum_{i=1}^{N_t} \rho(r_{i,t+1} - \boldsymbol{z}_{i,t}^{\mathrm{T}} \otimes \boldsymbol{f}_{t+1}^{\mathrm{T}} vec(\boldsymbol{\Gamma}_\beta^{\mathrm{T}})). \tag{32}$$

Subsequently, we derive the influence curve which is the derivative of $p(e)$ defined by $\psi(e) = p(e)'$. When the influence curve is applied to the previously defined equations the first-order conditions that give the optimal values of the parameters are obtained:

$$\frac{\partial \rho(r_{i,t+1}, \boldsymbol{z_{i,t}}, \boldsymbol{\Gamma}_\beta, \boldsymbol{f}_{t+1})}{\partial \boldsymbol{f}_{t+1}} = \sum_{i=1}^{N_t} \boldsymbol{\Gamma}_\beta^{\mathrm{T}} \boldsymbol{z}_{i,t} \psi(r_{i,t+1} - \boldsymbol{z}_{i,t}^{\mathrm{T}} \boldsymbol{\Gamma}_\beta \boldsymbol{f}_{t+1}) = \mathbf{0}_{K\times 1} \tag{33}$$

and

$$\frac{\partial \rho(r_{i,t+1}, \boldsymbol{z_{i,t}}, \boldsymbol{\Gamma}_\beta, \boldsymbol{f}_{t+1})}{\partial vec(\boldsymbol{\Gamma}_\beta^{\mathrm{T}})} = \sum_{i=1}^{T-1} \sum_{i=1}^{N_t} \boldsymbol{z}_{i,t} \otimes \boldsymbol{f}_{t+1} \psi(r_{i,t+1} - \boldsymbol{z}_{i,t}^{\mathrm{T}} \otimes \boldsymbol{f}_{t+1}^{\mathrm{T}} vec(\boldsymbol{\Gamma}_\beta^{\mathrm{T}})) = \mathbf{0}_{LK\times 1}. \tag{34}$$

To ease the calculations, the influence curve is transformed to a weight function, that is $w_{i,t} := \frac{\psi(e_{i,t})}{e_{i,t}}$ which can take the values:

$$w_{i,t} := \begin{cases} 1 & \text{if} \quad |e_{i,t}| \leq \xi \\ \frac{\xi}{|e_{i,t}|} & \text{if} \quad |e_{i,t}| > \xi \end{cases} \tag{35}$$

The second case describes that if the error is larger than the tuning constant, less weight is assigned to this observation. Similar to Fox and Weisberg (2002) we use a tuning constant that depends on the data that is equal to $1.345\hat{\sigma}$ where $\hat{\sigma} = \text{median}(\frac{|e_{i,t}|}{0.6745})$. The benefit of this procedure according to Jiang et al. (2019) is that a higher efficiency level can be achieved.

We have obtained a formulation for the weight that is contributed to each observation given a tuning constant and an error, subsequently we need to define the weight adjusted equations for the parameters and transform them to obtain an expression for $\boldsymbol{\Gamma}_\beta$ and $\boldsymbol{f}_{t+1}$

$$\hat{\boldsymbol{\Gamma}}_\beta^{\mathrm{T}} \boldsymbol{Z}_t^{\mathrm{T}} \boldsymbol{W}_t (\boldsymbol{r}_{t+1} - \boldsymbol{Z}_t \hat{\boldsymbol{\Gamma}}_\beta \boldsymbol{f}_{t+1}) = \mathbf{0}_{K\times 1}$$
$$\iff \hat{\boldsymbol{\Gamma}}_\beta^{\mathrm{T}} \boldsymbol{Z}_t^{\mathrm{T}} \boldsymbol{W}_t \boldsymbol{r}_{t+1} = \hat{\boldsymbol{\Gamma}}_\beta^{\mathrm{T}} \boldsymbol{Z}_t^{\mathrm{T}} \boldsymbol{W}_t \boldsymbol{Z}_t \hat{\boldsymbol{\Gamma}}_\beta \boldsymbol{f}_{t+1} \iff \boldsymbol{f}_{t+1} = (\hat{\boldsymbol{\Gamma}}_\beta^{\mathrm{T}} \boldsymbol{Z}_t^{\mathrm{T}} \boldsymbol{W}_t \boldsymbol{Z}_t \hat{\boldsymbol{\Gamma}}_\beta)^{-1} \hat{\boldsymbol{\Gamma}}_\beta^{\mathrm{T}} \boldsymbol{Z}_t^{\mathrm{T}} \boldsymbol{W}_t \boldsymbol{r}_{t+1}$$

and

$$
\begin{aligned}
&\sum_{i=1}^{T-1} \boldsymbol{Z}_t^{\mathrm{T}} \otimes \hat{\boldsymbol{f}}_{t+1} \boldsymbol{W}_t(\boldsymbol{r}_{t+1} - \boldsymbol{Z}_t \otimes \hat{\boldsymbol{f}}_{t+1}^{\mathrm{T}} vec(\boldsymbol{\Gamma}_\beta^{\mathrm{T}})) = \mathbf{0}_{LK\times 1} \\
\Longleftrightarrow &\sum_{i=1}^{T-1} \boldsymbol{Z}_t^{\mathrm{T}} \otimes \hat{\boldsymbol{f}}_{t+1} \boldsymbol{W}_t \boldsymbol{r}_{t+1} = \sum_{i=1}^{T-1} \boldsymbol{Z}_t^{\mathrm{T}} \otimes \hat{\boldsymbol{f}}_{t+1} \boldsymbol{W}_t \boldsymbol{Z}_t \otimes \hat{\boldsymbol{f}}_{t+1}^{\mathrm{T}} vec(\boldsymbol{\Gamma}_\beta^{\mathrm{T}}) \\
\Longleftrightarrow\ & vec(\boldsymbol{\Gamma}_\beta^{\mathrm{T}}) = (\sum_{i=1}^{T-1} \boldsymbol{Z}_t^{\mathrm{T}} \otimes \hat{\boldsymbol{f}}_{t+1} \boldsymbol{W}_t \boldsymbol{Z}_t \otimes \hat{\boldsymbol{f}}_{t+1}^{\mathrm{T}})^{-1} \sum_{i=1}^{T-1} \boldsymbol{Z}_t^{\mathrm{T}} \otimes \hat{\boldsymbol{f}}_{t+1} \boldsymbol{W}_t \boldsymbol{r}_{t+1} \\
\Longleftrightarrow\ & vec(\boldsymbol{\Gamma}_\beta^{\mathrm{T}}) = (\sum_{i=1}^{T-1} \boldsymbol{Z}_t^{\mathrm{T}} \boldsymbol{W}_t \boldsymbol{Z}_t \otimes \hat{\boldsymbol{f}}_{t+1} \hat{\boldsymbol{f}}_{t+1}^{\mathrm{T}})^{-1} \sum_{i=1}^{T-1} \boldsymbol{Z}_t^{\mathrm{T}} \boldsymbol{W}_t \boldsymbol{r}_{t+1} \otimes \hat{\boldsymbol{f}}_{t+1}.
\end{aligned}
$$

As can be seen in the Equation (35) the estimated parameter $w_{i,t}$ depends on the errors while the system of equations above show that errors depend on the estimated parameters the weight matrix $\boldsymbol{W}_t : \mathbb{R}^{N_t \times N_t}$ with weights $w_{i,t}$ on its diagonal, $\boldsymbol{\Gamma}_\beta$ and $\boldsymbol{f}_{t+1}$. Therefore an iterative algorithm, the iteratively reweighted least-squares (IRLS) is required for the calculation of $w_{i,t}$, $\boldsymbol{\Gamma}_\beta$ and $\boldsymbol{f}_{t+1}$ as shown below:

Algorithm 2: Huber modified OLS.

We start with the optimal values of $\boldsymbol{\Gamma}_\beta^*$ and $\boldsymbol{f}_{t+1}^*$ obtained by running the IPCA based on the normal OLS as in Equation (12).

1. Calculate residual $e_{i,t}^{(j)}$ and the corresponding weight $w_{i,t}^{(j)}(e) = w(e_{i,t}^{(j)})$;
2. Solve the following two equations

$$
\boldsymbol{f}_{t+1} = (\hat{\boldsymbol{\Gamma}}_\beta^{\mathrm{T}} \boldsymbol{Z}_t^{\mathrm{T}} \boldsymbol{W}_t \boldsymbol{Z}_t \hat{\boldsymbol{\Gamma}}_\beta)^{-1} \hat{\boldsymbol{\Gamma}}_\beta^{\mathrm{T}} \boldsymbol{Z}_t^{\mathrm{T}} \boldsymbol{W}_t \boldsymbol{r}_{t+1} \ \forall t \tag{36}
$$

$$
vec(\boldsymbol{\Gamma}_\beta^{\mathrm{T}}) = (\sum_{i=1}^{T-1} \boldsymbol{Z}_t^{\mathrm{T}} \boldsymbol{W}_t \boldsymbol{Z}_t \otimes \hat{\boldsymbol{f}}_{t+1} \hat{\boldsymbol{f}}_{t+1}^{\mathrm{T}})^{-1} \sum_{i=1}^{T-1} \boldsymbol{Z}_t^{\mathrm{T}} \boldsymbol{W}_t \boldsymbol{r}_{t+1} \otimes \hat{\boldsymbol{f}}_{t+1}; \tag{37}
$$

3. We iterate this procedure until convergence of both parameters is achieved.

2.5.2 Weighted least squares

Where the Huber loss function robustness feature is mainly focused on solving the problem of outliers, the WLS main focus is on heteroskedasticity. The whole structure as previously established in the Huber loss section can be applied with a minor alteration made to the weights. For the first iteration, the errors based on the normal OLS IPCA model as in Equation (12) are used subsequently, we follow the steps defined below

Algorithm 3: WLS.

1. For every t the variance-covariance matrix $\boldsymbol{\Sigma}_t$ of the errors is calculated;
2. The weight matrix is set equal to the inverse of the variance-covariance matrix $\boldsymbol{W}_t = \boldsymbol{\Sigma}_t^{-1}$;
3. Equations (36) and (37) are used to calculate the parameters;

4. Based on these new parameters the errors are calculated and the algorithm is started from the beginning until convergence is reached.

This algorithm is constructed in a similar way as the feasible weighted least squares (FWLS) which is a version of the WLS where the covariance of the errors is unknown. For large samples, the FWLS is asymptotically more efficient than OLS when the covariance of the error term is consistently estimated. The FWLS has according to Greene (2003) two main disadvantages. First, when either the sample is not sufficiently large or the covariance is not consistently estimated, it can be less efficient than OLS. Second, the weights are unknown and therefore when the weights are estimated based on a small sample size this can lead to bad regression analysis. However, we have to note that according to Kelly et al. (2017) IPCA exhibits due to its parsimonious parameterization of $\tilde{\mathbf{\Gamma}}$ a convergence rate of $\sqrt{NT}$, while PCA has a convergence rate of $\sqrt{T}$ because of its company-by-company time-series regression. Therefore, a smaller sample for the IPCA WLS variant is needed than when the PCA WLS variant is used.

2.6 Asset pricing tests

The main goal of this paper is about finding the drivers of option returns. We distinguish between two cases, the first case investigates which characteristics are statistically significant in explaining returns and are not related to risk factors while the second case examines which characteristics are statistically significant in the explanation of returns and are related to risk factors. First, we will test whether there exist instruments that are not related to risk factors but contribute to explaining the returns. Formally this means the existence of characteristics that cause the hypothesis $\mathbf{\Gamma}_\alpha = \mathbf{0}_{L\times 1}$ to be rejected. Moreover, when we have established the belief that $\mathbf{\Gamma}_\alpha \neq \mathbf{0}_{L\times 1}$, we will investigate which characteristic or set of characteristics is most important in rejecting the hypothesis.
Second, we investigate which characteristics have high explanation power in the variation in returns and are related to risk factors. This means we test the joint hypothesis of $\boldsymbol{\gamma}_{\beta,l} = \mathbf{0}_{K\times 1}$ and calculate their corresponding F-values. Lastly, we inspect the significance level of every characteristic for every individual factor by making use of the hypothesis $\gamma_{\beta,l,k} = 0$, where $\boldsymbol{\gamma}_{\beta,l} = [\gamma_{\beta,l,1}, \ldots, \gamma_{\beta,l,K}]$.

2.6.1 Testing the significance of $\mathbf{\Gamma}_{\boldsymbol{\alpha}}$

In this section, we discuss the framework of testing whether there exists an anomaly where characteristics are not associated with risk factors but do contribute in explaining the variation in returns. If this scenario holds, the restricted model is incorrectly specified and yields a worse fit compared to the unrestricted model. We start by investigating whether hypothesis $\mathbf{\Gamma}_\alpha = \mathbf{0}_{L\times 1}$ is true. If this hypothesis is rejected we subsequently want to determine which characteristic or set of characteristics is most responsible for rejecting $\mathbf{\Gamma}_\alpha = \mathbf{0}_{L\times 1}$, hence inspecting $\gamma_{\alpha,l} = 0$. It follows that our null hypothesis is H_0: $\mathbf{\Gamma}_\alpha = \mathbf{0}_{L\times 1}$ and the alternative hypothesis is $H_1 : \mathbf{\Gamma}_\alpha \neq \mathbf{0}_{L\times 1}$. Note that testing $\mathbf{\Gamma}_\alpha$ is different from jointly testing $\alpha_{i,t}$ since $\alpha_{i,t} = \boldsymbol{z}_{i,t}^{\mathrm{T}}\mathbf{\Gamma}_\alpha + \eta_{i,t}$, includes an error term $\eta_{i,t}$. Hence, the hypothesis test allows for mispricing only if the mispricing is not related to the characteristics $\boldsymbol{z}_{i,t}$.

The following procedure that is in line with Kelly et al. (2019) is used to determine whether the null hypothesis is rejected. First, we estimate the unrestricted model as in Equation (16) that includes the possibility of an anomaly and store the values of $\hat{\mathbf{\Gamma}}_\alpha$ $\hat{\mathbf{\Gamma}}_\beta$ and $\{\hat{\boldsymbol{f}}_{t+1}\}_{t=1}^{T-1}$. To compare the models under different hypotheses, Wald statistics are often used. The Wald statistic is an approximation of the likelihood ratio and is almost equal to the t-test or an F-test when the number of observations is large. Second, we use a Wald-type test statistic to determine the significance levels of the intercept coefficients $\mathbf{\Gamma}_\alpha$. This test is defined as the product of the estimated gammas of the intercepts, that is

$$W_\alpha = \hat{\mathbf{\Gamma}}_\alpha^{\mathrm{T}}\hat{\mathbf{\Gamma}}_\alpha.$$

Third, we utilize the bootstrap method to obtain a sample of $\mathbf{\Gamma}_\alpha$. The bootstrap method is conducted using managed portfolios $\boldsymbol{x}_t$ instead of raw returns $\boldsymbol{r}_t$, because it is less computationally expensive to resample L instead of N_t values while still being correctly specified. The reason is that for calculating the parameters $\tilde{\mathbf{\Gamma}}$ and $\boldsymbol{f}_{t+1}$ managed portfolio returns, $\boldsymbol{Z}_t^{\mathrm{T}}\boldsymbol{r}_{t+1}$, are used. However, managed portfolio returns cannot be used for WLS and Huber because their weight matrix depends on the errors and vice versa as shown in Algorithm 2 and 3. Thus, the bootstrap method needs to be adjusted which is explained at the end of Section 2.6.1.

In the first step of the bootstrap method, we obtain the errors of the unrestricted model which are the differences between the actual return and the return predicted by the model $\boldsymbol{r}_{t+1} - (\boldsymbol{Z}_t\mathbf{\Gamma}_\alpha + \boldsymbol{Z}_t\mathbf{\Gamma}_\beta\boldsymbol{f}_{t+1}) = \boldsymbol{\epsilon}_{t+1} \iff \boldsymbol{Z}_t^{\mathrm{T}}\boldsymbol{r}_{t+1} - (\boldsymbol{Z}_t^{\mathrm{T}}\boldsymbol{Z}_t\mathbf{\Gamma}_\alpha + \boldsymbol{Z}_t^{\mathrm{T}}\boldsymbol{Z}_t\mathbf{\Gamma}_\beta\boldsymbol{f}_{t+1}) = \boldsymbol{Z}_t^{\mathrm{T}}\boldsymbol{\epsilon}_{t+1} = \boldsymbol{\nu}_{t+1}$, where we define $\boldsymbol{\nu}_{t+1} : \mathrm{R}^{L\times 1}$ to be the residuals of the managed portfolio. The benefit of rewriting this equation is that in this way we can resample the managed portfolios returns $\boldsymbol{x}_{t+1}$ by adjusting the error $\boldsymbol{\nu}_{t+1}$.

The second step involves obtaining the bootstrapped portfolio returns based on the unrestricted estimates and error $\boldsymbol{\nu}_{t+1}$ but excluding $\mathbf{\Gamma}_\alpha$ as defined in the following equation

$$\boldsymbol{x}_{t+1}^{(b)} = (\boldsymbol{Z}_t^{\mathrm{T}}\boldsymbol{Z}_t)\hat{\mathbf{\Gamma}}_\beta\hat{\boldsymbol{f}}_{t+1} + \tilde{\boldsymbol{\nu}}_{t+1}^{(b)}, \tag{38}$$

where $\tilde{\boldsymbol{\nu}}_{t+1}^{(b)} = s_{t+1}^{(b)} \hat{\boldsymbol{\nu}}_{\iota_{t+1}^{(b)}}$. Moreover, $s_{t+1}^{(b)}$ is the value from a Student's t-distribution with 5 degrees of freedom and unit standard deviation (t_5) for bootstrap number b and $\hat{\boldsymbol{\nu}}_{\iota_{t+1}^{(b)}}$ is the managed portfolios error at time $\iota_{t+1}^{(b)}$. The variable $\iota_{t+1}^{(b)}$ is a resampled vector consisting of a sequence from 1 to $T-1$ which indicates the time index of the error. Moreover, every time index is sampled without replacement and therefore can only be taken once. Since our dataset consists of return data that according to Schwert and Seguin (1990) often suffers from heteroskedasticity, we used a Student's t-distribution instead of a normal distribution for $s_{t+1}^{(b)}$ which according to Goncalves and Kilian (2004) tackles the lack of efficiency problem in heteroskedastic data.

When all the $\boldsymbol{\Gamma}_\alpha^{(b)}$ $\forall$ $b \in \{1, \ldots, 1000\}$ are obtained the Wald-like statistics for these bootstrapped values are computed $W_\alpha^{(b)} = \boldsymbol{\Gamma}^{(b)\mathrm{T}}_\alpha \boldsymbol{\Gamma}_\alpha^{(b)}$. Next, the corresponding F-value is computed by $\frac{1}{1000} \sum_{b=1}^{1000} \mathbb{1}(W_\alpha^{(b)} > W_\alpha)$, where $\mathbb{1}(A)$ is an indicator value which is 1 if the statement A is true and 0 otherwise. When the F-value indicates that $\boldsymbol{\Gamma}_\alpha \neq \mathbf{0}_{L\times 1}$, we can subsequently determine which characteristic is most responsible for rejecting the null hypothesis by inspecting $\gamma_{\alpha,l}$. The process of testing this hypothesis is explained in greater detail in Section 2.6.3.

As discussed before we use this procedure for all models except the Huber loss model and the WLS. These two models use the same framework but some slight adjustments are made to the equations to incorporate the special features of these two loss functions. First, raw returns instead of portfolio returns are used when calculating the returns. Because in Equations (36) and (37) $\boldsymbol{Z}_t^{\mathrm{T}} \boldsymbol{W}_t \boldsymbol{r}_{t+1}$ is used instead of $\boldsymbol{Z}_t^{\mathrm{T}} \boldsymbol{r}_{t+1}$ where the former depends on the weight matrix that depends on an error term. Since a change in $\boldsymbol{W}_t$ causes the parameters to change we need to use raw returns even though this slows down our algorithm substantially. Second, a normal distribution is used instead of a t-distribution for the WLS, since the weight matrix already accounts for heteroskedasticity.

2.6.2 Testing the significance of $\boldsymbol{\gamma}_{\beta,l}$

Apart from knowing whether there exists an anomaly as discussed in Section 2.6.1, it is also interesting to understand which characteristics are statistically significant in the explanation of returns. This analysis on the statistical significance level of one characteristic is conducted while controlling for the other instruments. Moreover, a similar framework as for $\boldsymbol{\Gamma}_\alpha$ is used to perform inference on $\boldsymbol{\gamma}_{\beta,l}$ from Equation (15). Our null hypothesis is that all coefficients related to the characteristic l are zero, hence our null hypothesis is $H_0 : [\boldsymbol{\gamma}_{\beta,1}, \boldsymbol{\gamma}_{\beta,2}, \ldots, \boldsymbol{\gamma}_{\beta,l-1}, \mathbf{0}_{1\times K}, \boldsymbol{\gamma}_{\beta,l+1}, \ldots, \boldsymbol{\gamma}_{\beta,L}]^{\mathrm{T}}$ and our alternative hypothesis is $H_1 : [\boldsymbol{\gamma}_{\beta,1}, \boldsymbol{\gamma}_{\beta,2}, \ldots, \boldsymbol{\gamma}_{\beta,l-1}, \boldsymbol{\gamma}_{\beta,l}, \boldsymbol{\gamma}_{\beta,l+1}, \ldots, \boldsymbol{\gamma}_{\beta,L}]^{\mathrm{T}}$, where $\boldsymbol{\gamma}_{\beta,l} \neq$

$\mathbf{0}_{1\times K}$.

In a similar fashion as in Section 2.6.1, we first estimate the model under the alternative hypothesis and store parameters $\mathbf{\Gamma}_\beta$ and $\{\boldsymbol{f}_{t+1}\}_{t=0}^{T-1}$ and residuals $\{\hat{\boldsymbol{d}}_t\}_{t=1}^{T}$. Subsequently, we calculate a Wald-type statistic $W_{\beta,l} = \hat{\gamma}_{\beta,l}\hat{\gamma}_{\beta,l}^{\mathrm{T}}$. Next, we implement the same bootstrapping method as performed previously to conduct inference, that is when determining the significance of characteristic l we use the estimated $\mathbf{\Gamma}_\beta$ under the alternative hypothesis and set the lth block to zero and obtain $\tilde{\mathbf{\Gamma}}_\beta = [\boldsymbol{\gamma}_{\beta,1}, \hat{\boldsymbol{\gamma}}_{\beta,2}, \ldots, \hat{\boldsymbol{\gamma}}_{\beta,l-1}, \mathbf{0}_{1\times K}, \hat{\boldsymbol{\gamma}}_{\beta,l+1}, \ldots, \hat{\boldsymbol{\gamma}}_{\beta,L}]^{\mathrm{T}}$. Based on these parameters, we calculate the managed portfolios $\tilde{\boldsymbol{x}}_{t+1}^{b} = \boldsymbol{Z}_t\tilde{\mathbf{\Gamma}}_\beta\hat{\boldsymbol{f}}_{t+1} + \tilde{\boldsymbol{d}}_t^{\,b}$. These managed portfolios can subsequently be used in Equation (13) and (15) to recalculate $\mathbf{\Gamma}_\beta$. Lastly, the Wald-like statistic for $\boldsymbol{\gamma}_{\beta,l}^{(b)}$ is calculated for every bootstrap, combining all these b bootstrapped statistics and comparing it to the $W_{\beta,l}$ similarly as with $\mathbf{\Gamma}_\alpha$ gives us the F-value of instrument l.

Another manner of assessing the importance of a certain characteristic is by using the marginal R^2 method as discussed in Kelly et al. (2019). This method involves calculating the difference of total model R^2 and the same model where the coefficients of $\mathbf{\Gamma}_\beta$ corresponding to characteristic l are set to 0 which results in R^2_{-l}. Therefore, the marginal $R^2_{m,l}$ for characteristic l is defined as

$$R^2_{m,l} := R^2 - R^2_{-l}. \tag{39}$$

2.6.3 Testing the significance of $\gamma_{\alpha,l}$ and $\gamma_{\beta,l,k}$

For the determination of $\gamma_{\alpha,l} \neq 0$, we perform the same testing procedure as for $\boldsymbol{\gamma}_{\beta,l}$ but with a small alteration. We consider the constant vector $\mathbf{\Gamma}_\alpha$ instead of $\mathbf{\Gamma}_\beta$, hence $\boldsymbol{\gamma}_{\beta,l} = \mathbf{0}_{1\times K}$ becomes $\gamma_{\alpha,l} = 0$. When one compares this to the renowned 'GRS-test' as in Gibbons et al. (1989) there are several noticeable differences. First, with the GRS-test one knows which asset is mispriced, whereas this test shows which characteristic contributes (most) to the mispricing and hence which characteristic yields a statistically significant return unrelated to the risk factors. Second, the α_i in the GRS are the intercepts obtained after the regression and are therefore a residual, while our unrestricted IPCA model is a model on its own. Hence, when we reject the null hypothesis of $\mathbf{\Gamma}_\alpha = \mathbf{0}_{L\times 1}$ we know that that the unrestricted model is better specified than the restricted version.

Kelly et al. (2019) showed that different factors give importance to different characteristics in terms of the magnitude of their loadings in $\mathbf{\Gamma}_\beta$, which led us to believe that there might also be differences in terms of significance. Hence, the hypothesis of $\gamma_{\beta,l,k} = 0$ is evaluated. We proceed by utilizing a similar framework as before. First, we estimate the model based on $\gamma_{\beta,l,k} \neq 0$, afterward we calculate the corresponding Wald statistic $W_{\beta,l,k} = \gamma^2_{\beta,l,k}$. Subsequently, we set $\gamma_{\beta,l,k} = 0$, compute the managed portfolios and based on these recalculate $\mathbf{\Gamma}_\beta$ of which $\gamma^{(b)}_{\beta,l,k}$ can be extracted. Lastly, we use the same test to investigate the significance level of characteristic l for factor k.

2.7 In-sample performance

Apart from the significance of the individual instruments, it is also of crucial importance to evaluate how well the characteristics together are able to explain returns, which naturally leads to the coefficient of determination (R^2). With the R^2 statistic we can compare and contrast models based on how well the combination of characteristics ($\boldsymbol{z}_{i,t}$), factor loadings $[\mathbf{\Gamma}_\alpha, \mathbf{\Gamma}_\beta]$, and risk factor ($\boldsymbol{f}_{t+1}$) explain the variation in returns. This paper uses two types of R^2 in line with Kelly et al. (2019) to determine and compare the performance of the models. Hence the 'total R^2' is defined by

$$R^2_{total} := 1 - \frac{\sum_{i,t}(\boldsymbol{r}_{i,t+1} - \boldsymbol{z}^{\mathrm{T}}_{i,t}(\hat{\mathbf{\Gamma}}_\alpha + \hat{\mathbf{\Gamma}}_\beta \hat{\boldsymbol{f}}_{t+1}))^2}{\sum_{i,t} \boldsymbol{r}^2_{i,t+1}}. \tag{40}$$

This type of R^2 evaluates how much of the variance in the returns is explained by the dynamic conditional loadings ($\boldsymbol{z}^{\mathrm{T}}_{i,t}\mathbf{\Gamma}$) together with contemporaneous risk factors ($\boldsymbol{f}_t$).
The 'predictive R^2' is the second measure that we consider and is defined as

$$R^2_{predictive} := 1 - \frac{\sum_{i,t}(\boldsymbol{r}_{i,t+1} - \boldsymbol{z}^{\mathrm{T}}_{i,t}(\hat{\mathbf{\Gamma}}_\alpha + \hat{\mathbf{\Gamma}}_\beta \hat{\lambda}))^2}{\sum_{i,t} \boldsymbol{r}^2_{i,t+1}}, \tag{41}$$

where the constant risk price $\boldsymbol{\lambda} = E(\boldsymbol{f}_{t+1})$ is the unconditional expectation of the contemporaneous time series $\boldsymbol{f}_{t+1}$. With this statistic we can investigate how well the estimated conditional expected returns explain the variation in realized returns. Note that dynamics of the returns in IPCA are driven by two dependent components, $\boldsymbol{z}^{\mathrm{T}}_{i,t}\tilde{\mathbf{\Gamma}}$ and $\boldsymbol{f}_{t+1}$, hence no separate identification of risk price dynamics is possible. Therefore, when we use the constant risk price, predictions are solely based on dynamic instrumented loadings.

To obtain the portfolio fit, one simply replaces $\boldsymbol{r}_{i,t+1}$ with $\boldsymbol{x}_{i,t+1}$ and $\boldsymbol{z}^{\mathrm{T}}_{i,t}$ with $\boldsymbol{z}_{i,t}\boldsymbol{z}^{\mathrm{T}}_{i,t}$

2.8 Out-of-sample performance

This section discusses the out-of-sample fit of the models, the Sharpe ratio based on the construction of tangency portfolios and anomaly portfolios, and the downside risk utilizing the VaR.

2.8.1 Out-of-sample fit

The main part of this paper discusses the importance of discovering which set of characteristics contributes to the explanation of delta-hedged option returns in-sample. However, in order to find a trading strategy that yields high profits while maintaining a low-risk profile we need to understand how well our model performs out-of-sample. To perform the forecasts a recursive backward-looking estimation procedure similar to Kelly et al. (2019) will be conducted, that is only information up to time t is used to make the forecast for time $t+1$. Hence, we estimate the dynamic loading as usual with information up to time t based on Equation (15) or their regularized or robust counterparts and obtain $\hat{\boldsymbol{\Gamma}}_{\beta,t}$. Subsequently, we estimate the out-of-sample realized factor return $\hat{\boldsymbol{f}}_{t+1|t}$ where for the basic and regularized models Equation (13) changes to

$$\hat{\boldsymbol{f}}_{t+1|t} = (\hat{\boldsymbol{\Gamma}}_{\beta|t}^{\mathrm{T}} \boldsymbol{Z}_t^{\mathrm{T}} \boldsymbol{Z}_t \hat{\boldsymbol{\Gamma}}_{\beta|t})^{-1} \hat{\boldsymbol{\Gamma}}_{\beta|t}^{\mathrm{T}} \boldsymbol{Z}_t^{\mathrm{T}} \boldsymbol{r}_{t+1}. \tag{42}$$

The $\hat{\boldsymbol{f}}_{t+1|t}$ is the return at time $t+1$ based on the portfolio construction at time t, since all the terms apart from $\boldsymbol{r}_{t+1}$ are based on information at time t. One can view this portfolio construction and return calculation in a similar way as the portfolio sorts technique that was used by for instance Jegadeesh and Titman (1993). The assets are first sorted in a portfolio and subsequently in the post-formation period the return of this sorted portfolio is calculated. For the robust methods, Equation (36) is adjusted in a similar fashion as above and therefore includes the weight matrix.

To assess the predictability of the out-of-sample returns one computes the out-of-sample total R^2 by replacing $\boldsymbol{Z}_t \hat{\boldsymbol{\Gamma}}_\beta$ in Equation (40) that computes the in-sample R^2 by $\hat{\boldsymbol{r}}_{t+1} = \boldsymbol{Z}_t \hat{\boldsymbol{\Gamma}}_{\beta|t} \hat{\boldsymbol{f}}_{t+1|t}$. Similarly, the out-of-sample predictive R^2 uses the price of risk $\boldsymbol{\lambda}_t : \mathbb{R}^{K \times 1}$ instead of $\boldsymbol{\lambda}$ in Equation (41), where $\boldsymbol{\lambda}_t = \frac{1}{t} \sum_{i=1}^{t} \boldsymbol{f}_i$. Because $\boldsymbol{\lambda}_t$ does not depend on any information beyond time t, we can truly say this coefficient of determination is based on out-of-sample predictions.

Since several researchers showed that out-of-sample performance can be improved when a parsimonious model is used, we investigate if this is also true for our models. Hence, we compare the basic IPCA model including all characteristics with two more parsimonious models. The first model only uses characteristics that have been proven to be statistically significant in the in-sample estimation, whereas the second model only uses instruments that have not been set to 0 by the lasso in-sample estimation.

2.8.2 Sharpe ratio based on tangency and anomaly portfolio

Apart from using the factors for comparing the out-of-sample fit of the different models, these can also be utilized in the construction of portfolios for trading strategies. Therefore we can evaluate which model yields the highest return out-of-sample and should be implemented by the investor. The first type of trading strategy is based on tangency portfolios constructed by the out-of-sample risk factors of the model based on the following algorithm

Algorithm 4: Tangency portfolio return algorithm.

- Denote $\tilde{\boldsymbol{t}} = \{\tilde{t}_1, \tilde{t}_2, \ldots, \tilde{T}\}$ as the out-of-sample dates that are equidistant.
- Calculate $\boldsymbol{\Gamma}_{\beta|\tilde{t}_j}$ based on information up to and including $\tilde{t}_j$ and compute out-of-sample $\hat{\boldsymbol{f}}_{\tilde{t_{j+1}}|\tilde{t}_j}$ based on Equation (42) or adjusted with the weight matrix $\boldsymbol{W}_t$.
- Calculate the tangency weight ($\boldsymbol{\pi}_{\tilde{t}_j} : \mathbb{R}^K$) defined as $\boldsymbol{\pi}_{\tilde{t}_j} = \frac{1}{\boldsymbol{\iota}^{\mathrm{T}} \tilde{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{f}}_{\tilde{t}_j}} \tilde{\boldsymbol{\mu}}_{\hat{\boldsymbol{f}}_{\tilde{t}_j}}} \tilde{\boldsymbol{\Sigma}}^{-1}_{\hat{\boldsymbol{f}}_{\tilde{t}_j}} \tilde{\boldsymbol{\mu}}_{\hat{\boldsymbol{f}}_{\tilde{t}_j}}$ where $\tilde{\boldsymbol{\mu}}_{\hat{\boldsymbol{f}}_{\tilde{t}_j}} = \frac{1}{\tilde{t}_j} \sum_{i=\tilde{t}_1}^{\tilde{t}_j} \boldsymbol{f}_i$ and $\tilde{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{f}}_{\tilde{t}_j}} = \boldsymbol{f}_{1:\tilde{t}_j} \boldsymbol{f}^{\mathrm{T}}_{1:\tilde{t}_j}$. The construction of the risk factors only depends on information up to time $\tilde{t}_j$ and standardize the tangency weights such that these sum to 1.
- We form the tangency portfolio return based on Brandt et al. (2009) where the investor chooses weights at time $\tilde{t}_j$ according to the tangency formulation and obtains a post-formation return at $\tilde{t}_{j+1}$ of $\boldsymbol{r}_{\tilde{t}_{j+1}}$ defined as $\boldsymbol{r}_{\tilde{t}_{j+1}} = \boldsymbol{\pi}_{\tilde{t}_j} \hat{\boldsymbol{f}}_{\tilde{t}_{j+1}|\tilde{t}_j}$.

Another interesting portfolio is the anomaly portfolio, which is based on the loadings of the anomaly parameter $\boldsymbol{\Gamma}_\alpha$ and therefore utilizes all information associated to return without bearing risk exposure. We construct anomaly weights based on information up to time t for companies that have observations in two consecutive periods similar as Kelly et al. (2019) that is

$$\boldsymbol{\pi}^{an}_t = \boldsymbol{Z}_t (\boldsymbol{Z}_t^{\mathrm{T}} \boldsymbol{Z}_t)^{-1} \boldsymbol{\Gamma}_{\alpha|t}, \tag{43}$$

where $\boldsymbol{\pi}^{an}_t : \mathbb{R}^{N_t \times 1}$ are the weights of the anomaly portfolio. Subsequently, we multiply this with the return at time $t+1$ for the firms to get the anomaly portfolio return ($r_{t+1,p}$)

$$r_{t+1,p} = \boldsymbol{\pi}^{an^{\mathrm{T}}}_t \boldsymbol{r}_{t+1}. \tag{44}$$

This portfolio assigns more weight to companies in the portfolio that have a large value for characteristic l that has a positive $\gamma_{\alpha l}$. This means that weights are given based on how high the expected return is beyond risk-based compensation for a characteristic.

To investigate whether a risk factor, tangency portfolio, or anomaly portfolio based on delta-hedged option returns is an attractive asset class we need to take both risk and return into account. A solution that is suitable for this problem is the unconditional Sharpe ratio because it incorporates the mean and the standard deviation of the portfolio's return. The Sharpe ratio for factor i is defined similarly to Lettau and Pelger (2018) as

$$Sh(f_i) := \frac{\mu_{fi}}{\sigma_{fi}}, \tag{45}$$

where $\mu_{fi} = \mathbb{E}(f_i) = \frac{1}{T}\sum_{i=1}^{T-1} = f_{t+1,i}$ and $\Sigma_{f1} = var(f_1) = \frac{1}{T}\sum_{i=1}^{T-1} f_{t+1,i}f_{t+1,i}$. When the factor is replaced by the tangency or anomaly portfolio return, one obtains the Sharpe ratio of these portfolios.

2.8.3 VaR

Section 2.8.2 showed a method to construct the tangency portfolio and the anomaly portfolio and its corresponding return and Sharpe ratio. However, an average investor not only bases its portfolio on return and variation in risk but is also interested in a measure that evaluates its risk in extreme events such as crises. This naturally leads to the utilization of the Value at Risk (VaR) measure, as described in Jorion (2000), for the tangency portfolio return. To calculate the VaR, a sample of returns is needed. This sample can be obtained by applying a similar bootstrap method as was conducted for the p- and F-values in Section 2.6.1. Hence, we add a random shock to the observed value of the managed portfolio return as shown in the equation below

$$\tilde{\boldsymbol{x}}_{t+1}^{(b)} = \boldsymbol{Z}_t^{\mathrm{T}} \boldsymbol{r}_{t+1} + \boldsymbol{Z}_t^{\mathrm{T}} \boldsymbol{\epsilon}_{t+1}^{*}. \tag{46}$$

We calculate the bootstrapped managed portfolio returns ($\tilde{\boldsymbol{x}}_{t+1}^{(b)}$) for all the out-of-sample dates for a large number of bootstrap iterations based on Equation (46). Subsequently, we can calculate the out-of-sample factor returns using these bootstrapped returns in Equation (42). The last step involves the application of Algorithm 4 which yields the tangency portfolio returns. When all the bootstrapped tangency portfolio returns are collected, a distribution of the returns can be obtained out of which the VaR can be computed, which is defined as

$$VaR_{1-\alpha}(\boldsymbol{r}_{t+1}) := \underset{x \in R}{inf}\{x : Pr(\boldsymbol{r}_{t+1} \leq x) \geq 1 - \alpha\}. \tag{47}$$

According to Bayer (2018), one key problem with forecasting VaR is the dependence on data in the model's performance. In financial markets that are stable a more parsimonious model is plausible to outperform a very parameterized model, whereas in financial crises the opposite might hold. Therefore, we will investigate the discrepancies in model performance between regularized models and the normal IPCA model.

2.9 Robustness

Since our main investigation of the relationships between the characteristics and returns heavily depends on data, it is critical to ascertain the credibility of the coefficients and their specification.

This is especially important in forecasting since the inclusion or exclusion of one period represents a larger part of the out-of-sample period than one period in the in-sample estimation period. Therefore, we use a modelling dataset consisting of either 75 or 90 observations. Furthermore, to mitigate the threat of model misspecification, a wide array of models is used to evaluate the relationship between instruments and return. Especially, the lasso and elastic net are able to investigate misspecification because most models include a different set of characteristics.

3 Data

In this section the data that is used to model the delta-hedged option returns is discussed. First, we discuss how this type of return can be calculated and discuss their summary statistics. Second, we consider the data of the characteristics. Finally, we will explain which features influenced our decisions to consider some observations as outliers and therefore were deleted from our sample.

3.1 Delta-hedged option returns

For the calculation of the delta-hedged option returns, we used near the money option with a maturity of approximately one month. These option returns had the following features: price (C_t), the delta of the option (Δ_t), the corresponding stock price from 1776 companies listed on major American exchanges including the NYSE and NASDAQ (S_t), and the risk-free rate of a ten year US treasury rate (rf_t) from Optionmetrics which share similar features with Horenstein et al. (2018). This dataset ranges from January 2005 till June 2014 and consists of 61,120 monthly delta-hedged call option returns.

To get a dataset that does not heavily depend on outliers and unrealistic data we applied a filtering procedure similar to Horenstein et al. (2018). First, illiquid options that had the following features were removed: zero trade volume, a bid or ask price of zero or an average bid and ask price below 0.125. Second, since our paper focuses on the volatility and risk premium and not on the early exercise premium we exclude returns of call options that would have received a dividend in that particular month. As a result, our American type call options can be viewed as European type call options. Third, options with a higher ask than bid are excluded. Fourth, only near the money options are used, that is the moneyness ranges between 0.8 and 1.2. Lastly, options with volatility lower than 0.1 or higher than 1 are excluded.

The process of calculating these types of returns based on the framework of Horenstein et al. (2018) will be discussed below. The gain of delta-hedged option i at time t with maturity $t+\tau$ ($\Pi^i_{t,t+\tau}$) consists of the gain of the option $C^i_{t+\tau} - C^i_t$ in excess of the delta hedges to the changes in the stock price (dS) and the risk-free rate earned by this portfolio

$$\Pi^i_{t,t+\tau} = C^i_{t+\tau} - C^i_t - \int_t^{t+\tau} \Delta^i_u dS_u - \int_t^{t+\tau} rf(C^i_u - S^i_u \frac{\partial C_u}{\partial S^i_u})du. \tag{48}$$

To use the delta-hedged gain as the dependent variable in Equation (8), we need to discretize Equation (48) into

$$\Pi^i_{t,t+\tau} = C^i_{t+\tau} - C^i_t - \sum_{n=0}^{N-1} \Delta_{t_n}(S_{t_{n+1}} - S_{t_n}) - \sum_{n=0}^{N-1} \frac{d_n rf_{t_n}}{365}(C^i_{t_n} - \Delta_{t_n} S_{t_n}), \tag{49}$$

where Δ_{t_n} is the delta of the option at time t_n, rf_{t_n} is the risk-free rate at time t_n and d_n are the number of days between t_n and t_{n+1}. Hence, we need to construct an algorithm that dynamically hedges C^i N times. Since there are large discrepancies between gains due to a variety of prices, we divided the delta-hedged gain by the amount invested which results in the delta-hedged return, that is $r_{i,t} = \frac{\Pi^i_{t,t+\tau}}{\Delta_t S_t - O_t}$.

When applying Equation (49), we obtain a delta-hedged return for every unique option that has at least two observations. Ideally, we would like to have options that start at the beginning of the month and lasts till the end of the month, which makes the returns more comparable since they do not depend on the timing and time period. Hence, when an option starts at the end of month t and has several observations in month $t+1$, the return will be calculated from the first observation in month $t+1$ till the last observation in month $t+1$. For the same purpose, we discard observations that are not in the 3- to 5-week range. Subsequently, for every month and every unique company that has more than one return that satisfies the restrictions, we keep the observation that is closest to at-the-money at time of maturity (t_N) with at least a remainder of 20 days of maturity at time t_N.

Table 1 below shows detailed information about the option returns.

Table 1: Characteristics of monthly delta-hedged option returns.

	Call options		
	gain	return	moneyness
min	-16.8785	-0.2837	0.8002
1st Qu	-0.3091	-0.0265	0.9745
Median	-0.1181	-0.0097	0.9975
Mean	-0.1070	-0.0104	0.9980
3rd Qu	0.0633	0.0041	1.0193
Max	19.93	1.1630	1.1996

Note: This table displays the summary statistics of the main features of the option returns. The amount of options returns calls by year are respectively 2005: 7040 2006: 8001 2007: 8753 2008: 7767 2009: 7712 2010: 8473 2011: 8231 2012: 7957 2013: 8241 2014: 3777.

Similar to Horenstein et al. (2018) and Cao et al. (2017) our mean of delta-hedged option returns are negative and seem to have fat tails of which the last feature is common in return data. A dependent variable that is non-normal can cause the errors to be non-normally distributed.

Although the estimates of $\boldsymbol{\beta}_{i,t}$ in Equation (8) by OLS remain unbiased with heteroskedasticity, inference is often problematic and it lacks efficiency because the homoskedasticity assumption in Gauss-Markov Theorem does not hold. Therefore we test if the errors are normally distributed by evaluating the Normal Q-Q plot of the options. Moreover, the errors are tested on serially correlation by inspecting the Durbin-Watson statistics as in Equation (28). When the errors are serially correlated the OLS method does not yield a minimum variance estimator and the coefficients are biased.

Figure 5 in Appendix B.1 shows that for all 9 randomly chosen companies the errors, based on the restricted basic 4-factor IPCA model as in Equation (11), are heavy-tailed and therefore we reject that the errors for all 1776 companies follow a normal distribution. Figure 6 in Appendix B.1 depicts that when all assets are considered simultaneously extreme heavy tails are shown. The d_{pd} value for the restricted and unrestricted model with 1 to 4 risk factors ranges from 1.69 to 1.95 which is around 2, therefore almost no autocorrelation between the errors exists. Combining these two observations leads us to conclude that the errors are not normally distributed but do not exhibit autocorrelation. Therefore, the two methods that account for heteroskedasticity, the wild bootstrap and WLS are appropriate.

3.2 Characteristics

As was previously discussed in Section 2.2 and explained by Kelly et al. (2019), it is beneficial to have a large number of characteristics that are informative even though some might be noisy or even spurious, because the noise is possibly averaged out when aggregating the instruments. Thus, we include a large number of characteristics that have been proven to be important in either equity returns or option returns.

Vasquez and Xiao (2018) found that default risk based on default probability or credit ratings is monotonically negatively related to delta-hedged equity option returns. Hence, it would be logical to include characteristics that are associated with the riskiness of a security. For instance, debt-to-market capitalization can be included as instrument in the IPCA since it is commonly viewed as proxy for riskiness. Furthermore, we use 5 characteristics that Cao et al. (2017) showed to be important in explaining option gains such as stock reversal, momentum, the size of the company, profitability, and cash-to-asset ratio. They found that option gains are positively related to reversal, momentum, size, and profitability and negatively related to cash-to-asset.

Moreover, we include variables that are related to trading frictions such as idiosyncratic volatility and liquidity of the underlying or the option. Cao and Han (2013) found that option returns are negatively monotonically related to idiosyncratic volatility in the underlying security because options with high idiosyncratic volatility are harder to hedge and market makers demand higher compensation for offering these type of options. Furthermore, they found that less liquid options yield lower returns which is also due to the difficulty of hedging these types of options.

These aforementioned characteristics are also used in the analysis on stock returns by Kim et al. (2019). We complement this set of instruments with the remaining characteristics mentioned in this paper. In total, a dataset consisting of 62 characteristics and one constant with a similar time period as the option returns is considered which is sourced from Kim et al. (2019).

To give structure to this high number of characteristics, these are divided into groups similar to Kim et al. (2019): (1) Past returns, (2) Investment, (3) Profitability, (4) Intangibles, (5) Value, and (6) Trading frictions. The instruments belonging to these categories are listed and explained in detail in Table 8 in Appendix A.

To reduce the effects of outliers and implement the common practice for parsimonious models discussed in Section 2.4, we standardize the characteristics in the following manner similar to Kelly et al. (2019). First, we calculate the rank of certain characteristic for option i at time t ($z_{i,t,l}$) based on sorting the set of characteristics $\boldsymbol{z}_{t,l}$ from high to low. Second, we transform the ranks by dividing by the number of available options N_t at time t and subtract a half. This procedure results in that all values of characteristics l are in the set $[-0.5, 0.5]$. We perform the same method for all time periods $t \in \{1, \ldots, T\}$ and all L characteristics.

Note that in contrast to Horenstein et al. (2018), we do not include an index as characteristics. Hence, we stay in line with Kelly et al. (2019) to only include a constant in the characteristics $\boldsymbol{Z}_t$ used in Equation (11) to incorporate the shared variation of all the returns at time t. The reason for not including both a constant and an index in the model is that the optimal parameters cannot be found because the inverse of the first part of Equation (13), due to $\boldsymbol{Z}_t$, is undefined. [5]

3.3 Determination of the number of factors

It is important to have a model that performs well both in-sample as well as out-of-sample. It is shown in Ledolter and Abraham (1981) that a non-parsimonious model yields higher forecasting

[5]Consider matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ where $\boldsymbol{A}$ has no full rank and hence has a determinant (det) of 0. Then by det($\boldsymbol{AB}$)=det($\boldsymbol{A}$)det($\boldsymbol{B}$) it follows det($\boldsymbol{AB}$)=0 and hence the inverse of matrix $\boldsymbol{AB}$ is undefined.

errors. Moreover, Box et al. (1970) preached for a parsimonious model because it is easier to understand and because every parameter exhibits an estimation error. Hence, we strive for a parsimonious model and therefore need to assess how many risk factors to include.

The most common and simple method was developed by Cattell (1966) where one visually investigates a scree plot of the eigenvalues of the covariance of the managed portfolio returns $\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}$, where $\boldsymbol{X}$ is the matrix based on Equation (1) consisting of all periods. When performing this method on Figure 7 in Appendix B.1, it is clear that the importance of the factors shrinks substantially between 1 and 2, and that the factors after the 4th factor do almost not contribute to the total eigenvalue. Because this visual investigation is subjective and not clear in all situations, Ahn et al. (2018) constructed the Eigenvalue Ratio (ER) test which is a more mathematical and objective method. The ER test finds the optimal number of factors based on changes in eigenvalues of managed portfolios. We define the kth eigenvalue of matrix $\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}} : \mathbb{R}^{LXL}$ by $\mu_{L\tilde{T},k} = \lambda_k(\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}})$ and the optimal number of factors by $\tilde{k}_{ER} = \underset{k<kmax}{\operatorname{argmax}} \frac{\mu_{L\tilde{T},k}}{\mu_{L\tilde{T},k+1}}$. Figure 8 in Appendix B.1 shows that there is a steep decline from the third factor onwards, hence our focus lies on the first three factors and the fourth factor serves as a robustness check.

Where we constructed a method for the determination of the number of factors, Kelly et al. (2019) did not use such a method. They simply explained that the number of factors is kept small, with a maximum of 6 factors. Their results showed that there is only a slight increase in R^2 when 5 or 6 factors are used.

4 Results

The results section is split into two parts, where the first part discusses the basic IPCA model as in Section 2.3 and the second discuss the IPCA extensions of regularized and robust models from Section 2.4 and 2.5 respectively. In the basic IPCA and the regularized models we will first try to answer the main research question: 'What characteristics influence delta-hedged option returns'. The investigation is structured as follows. First, we inspect the coefficients, significance levels, and individual R^2 for the basic model. Second, we examine the existence of an anomaly by researching whether there are instruments that are not related to the risk factors but contribute in explaining returns. Third, we investigate the in-sample and out-of-sample performance based on the R^2, Sharpe ratios, and VaR from Section 2.7 and 2.8. Lastly, we compare the results of the basic IPCA model with their extensions which allows us to answer the research question whether regularized or more robust methods perform better than basic models.

4.1 Basic IPCA

The features of the basic IPCA are discussed in this part including the coefficients and significance levels of the characteristics as well as the fit and performance of the trading strategies all based on Equations (11) and (16).

4.1.1 Coefficients and significance

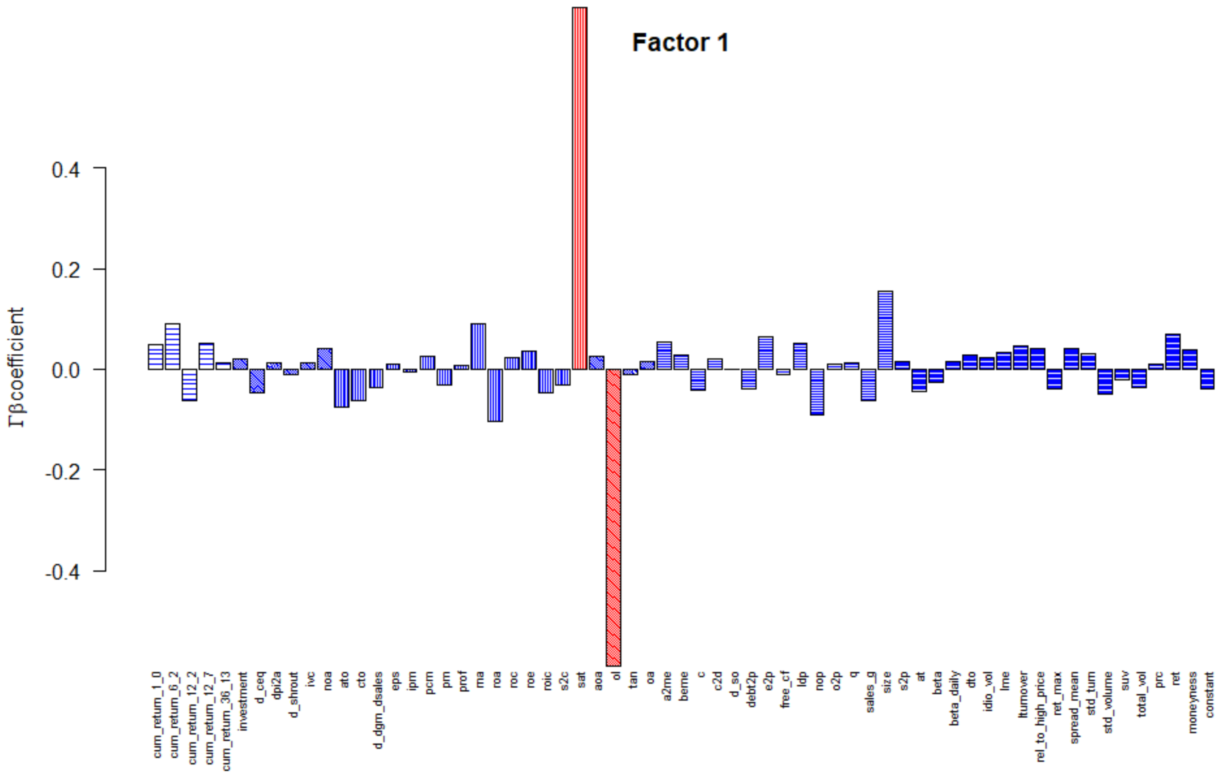


Figure 1: Estimates of $\mathbf{\Gamma}_\beta$ for the restricted 4-factor model.

Note: The estimates of $\mathbf{\Gamma}_\beta$ related to the second risk factor of the restricted 4-factor model as defined in Equation (11) are shown. A red bar indicates a value that is in absolute terms larger than 0.2 and different patterns are used for different categories.

Figure 1 for the first risk factor and Figures 9 till 11 in Appendix for the remaining three factors show the loadings $\mathbf{\Gamma}_\beta$ of the characteristics based on the restricted 4-factor model, which refers to the IPCA model that imposes $\mathbf{\Gamma}_\alpha = \mathbf{0}_{L\times 1}$ as described in Equation (11). The first factor exhibits the largest loadings in terms of magnitude on sales-to-total assets *(sat)*, costs of goods sold and selling, general and administrative expenses to total assets *(ol)*. Since we have imposed the expected value of $\boldsymbol{f}_t$ to be non-negative, companies with large *sat* have large positive betas on factor 1 and therefore a higher return.[6] The same idea holds for *ol* but then the opposite relationship holds. Hence the first risk factor is related to both total assets and profit (sales -

[6]This imposition on $\boldsymbol{f}_t$ does not restrict the economic meaning but serves to find a unique solution as discussed in Section 2.3.1.

costs). Higher profits increase the option return which affirms the result in Cao et al. (2017), while total assets yield an indifferent effect.

Factor 2 exhibits large magnitudes for size *(size)*, cash-to-total assets *(c)*, and sales-to-cash *(S2C)*, where the first one is positive and the latter two are negative. Hence, it seems that cash, size, total assets, and sales are important variables in the constitution of this factor. The signs of the first two variables suggest that a larger size or an increase in total assets lead to higher returns possibly because larger companies have a lower chance of default and therefore yield higher returns as Vasquez and Xiao (2018) suggested. Furthermore, an increase in cash leads to either lower or higher returns. Sales also seem to have a negative effect on returns. Lastly, the negative sign of cash-to-total assets confirms the result in Cao et al. (2017) that higher cash-to-total assets yields lower option returns.

Factor 3 has a large magnitude for market capitalization *(lme)*, return on assets *(roa)*, and sales-to-lagged net operating assets *(ato)* next to the constant. All these characteristics have a positive sign, hence higher returns, sales and market capitalization lead to higher option returns, while larger assets or net operating assets lead to lower option returns. It follows that market capitalization as a proxy for size yields the same positive result as described in Cao et al. (2017).

The last factor has large loadings on *lme* and total assets *(at)* where the first is negative and the second positive. Hence, this suggests that a value strategy that is mainly based on taking a long position in companies with high book-to-market value and a short position in companies with a low value leads to higher returns.

The coefficients of the unrestricted IPCA model are shown in Figure 12 to 16 in Appendix C.1.1. When considering the unrestricted model, we are mainly interested in the coefficients and the significance of the anomaly loading $\mathbf{\Gamma}_\alpha$. The loading informs us how much weight is given to a characteristic related to this factor, hence a larger value corresponds to a larger anomaly effect. We have imposed the anomaly factor to have a periodic return of 1 in Equation (16) which does not change the economic meaning but serves to pin down a unique solution. Therefore, the weights are substantially lower on the anomaly factor than on the other risk factors that have an expected return of around 0.01. For the loadings of the anomaly factor we find a large positive magnitude for *lme*, *roa*, *sat*, which shows that a positive value for market capitalization, assets, or return on assets yield high anomaly return. In contrast, a large negative magnitude is found for the standard deviation of volume *(std_volume)* which shows more volatile trade volumes reduce the return beyond systematic risk. Higher volatility corresponds to investors being uncertain about the value of the stocks and are therefore perceived as riskier. Lastly, for the loadings on the risk factors we observe quantitatively different but similar results as for the unrestricted model for characteristics: *sat*, *ol*, *size*, and *lme*, where sales-to-assets and costs-to-assets are respectively positive and negative, the instrument size and market capitalization vary by factor. Therefore, the risk factors loadings of the restricted and unrestricted model share commonalities.

Next to the magnitude of the relationship, it is also important to investigate which characteristics are statistically significant in explaining returns based on their F-values and p-values and are perceived as major characteristics in explaining returns based on the marginal R^2. We start with the importance of the characteristic for all the risk factors, that is F-values and marginal R^2. Subsequently, we attempt to identify whether there exist characteristics that are statistically significant for an individual factor based on their p-values, and hence contribute in explaining variation in returns. Lastly, we inspect if the hypothesis $\mathbf{\Gamma}_\alpha = \mathbf{0}_{L\times 1}$ holds. In the scenario that the hypothesis is rejected we further investigate which characteristic is most responsible for this rejection.

Table 2: P-values of the characteristics of the restricted model.

Characteristic/k	1	2	3	4	Characteristic/k	1	2	3	4
Past returns					**Value**				
cum_return_1_0	0.98	0.32	0.88	0.92	a2me	0.52	0.69	0.55	0.76
cum_return_6_2	0.96	0.69	0.86	0.90	beme	0.81	0.66	0.44	0.40
cum_return_12_2	0.69	0.72	0.94	0.98	c	0.80	0.32	**0.07**	0.18
cum_return_12_7	0.62	0.76	0.96	0.96	c2d	0.46	0.60	0.74	0.80
cum_return_36_13	0.94	0.23	0.76	0.82	d_so	0.58	0.74	0.82	0.78
					debt2p	0.64	0.80	0.74	0.86
Investment					e2p	0.52	0.78	0.92	0.88
investment	**0.02**	**0.00**	**0.00**	**0.00**	free_cf	0.90	1.00	0.60	0.46
d_ceq	0.90	0.54	0.76	0.85	ldp	0.39	0.12	0.24	0.40
dpi2a	**0.10**	0.22	0.27	0.48	nop	0.82	0.96	0.77	0.73
d_shrout	0.82	0.92	0.99	0.78	o2p	0.60	0.74	0.87	0.80
ivc	0.70	0.83	0.86	0.94	q	0.64	0.88	0.92	0.94
noa	**0.09**	0.32	0.40	0.55	sales_g	0.94	0.42	0.66	0.78
					size	0.57	0.68	0.60	0.80
Profitability					s2p	0.86	0.98	0.98	0.98
ato	0.12	0.32	**0.01**	**0.00**					
cto	0.82	0.64	0.88	0.96	**Trading frictions**				
d_dgm_dsales	0.78	0.42	0.98	0.20	at	0.77	0.32	0.40	0.73
eps	**0.07**	0.31	0.26	0.20	beta	0.28	0.18	0.14	0.22
ipm	0.35	0.40	0.60	0.15	beta_daily	0.40	0.26	0.34	0.55
pcm	0.68	0.88	0.30	0.13	dto	0.98	0.98	0.50	0.61
pm	0.68	0.97	0.69	0.40	idio_vol	0.70	0.98	0.98	0.91
prof	0.96	0.98	0.92	0.36	lme	0.84	0.52	0.74	0.14
rna	0.98	0.76	**0.10**	0.14	lturnover	0.16	0.13	0.30	0.26
roa	0.30	**0.01**	**0.08**	0.12	rel_to_high_price	0.88	0.90	0.88	0.87
roc	0.70	0.60	0.80	0.24	ret_max	0.72	0.49	0.85	0.94
roe	0.47	0.16	0.32	0.30	spread_mean	0.44	0.16	0.55	0.56
roic	0.66	0.59	0.89	0.86	std_turn	**0.04**	**0.10**	0.22	0.30
s2c	0.77	0.30	0.18	0.29	std_volume	**0.08**	0.26	0.41	0.45
sat	**0.10**	0.16	**0.06**	**0.04**	suv	1.00	0.96	0.91	0.29
					total_vol	0.34	0.42	0.77	0.60
Intangibles					prc	0.38	0.38	0.33	0.38
aoa	0.11	**0.02**	0.16	0.27	ret	0.47	0.30	0.38	0.64
ol	**0.04**	**0.10**	**0.06**	0.14	moneyness	0.73	0.18	0.38	0.41
tan	0.48	0.48	0.03	**0.10**					
oa	0.41	0.67	0.92	0.78	constant	**0.00**	**0.00**	**0.00**	**0.00**

Note: The p-values for the characteristics of $\mathbf{\Gamma}_\beta$ displayed are based on the restricted model with k systematic risk factors as defined in Equation (11). The p-values are based on the equations in Section 2.6.2. Moreover, a bold value indicates that the significance level is below 10%.

Table 2 shows that the following characteristics are statistically significant at the 10% level for at least two of the four models: investment, *(ato)*, *(roa)*, *(sat)*, *(ol)*, standard deviation of turnover *(std turn)*, and the constant. Hence, the characteristics are widely spread out over the different groups, but profitability seems to be the most important category. The importance of this group is because the pricing of returns is related to the risk profile of the company which is an important predictor for returns as was suggested by Cao et al. (2017). Moreover, instruments that are also significant for equity returns such as characteristics related to profit and return to assets are significant for delta-hedged option returns too. Although according to Jegadeesh and Titman (1993) and Jegadeesh (1990) momentum and reversal characteristics seem to be important in the explanation of equity returns, this is not the case for delta-hedged option returns. Where according to Hong and Stein (1999) strategies on momentum and reversal work for equities because these are based on investors trading in the historical direction of the stock, investors have probably not as widely implemented these strategies for delta-hedged option returns and therefore fail to work. Another explanation is that the momentum and reversal feature of stock returns do not translate to option returns.

Lastly, statistical significance and magnitude of the coefficient characteristics seem to be related with *ato*, *roa*, *sat*, and *ol* being both statistically significant as well as having a large effect on the constitution of the risk factor.

As was discussed in Section 2.6.2 the marginal $R^2_{m,l}$ as defined in Equation (39) can be used to identify the importance of an instrument l in explaining returns.

Table 3: Marginal decrease in R^2 for the restricted 4-factor model.

Characteristic	$R^2_{m,l}$	Stat. sign.	Characteristic	$R^2_{m,l}$	Stat. sign.	Characteristic	$R^2_{m,l}$	stat. sign.
constant	0.1211	*	pcm	0.0007		s2c	0.0006	
ato	0.0015	*	tan	0.0007	*	prc	0.0005	
rna	0.0010	**	beta	0.0006		ldp	0.0004	
ol	0.0009	**	ret	0.0007		roa	0.0004	**
sat	0.0009	*	c	0.0006	**	std_volume	0.0003	**
investment	0.0008	*	eps	0.0006	**	std_turn	0.0003	**

Note: The $R^2_{m,l}$ are based on the restricted 4-factor model as in Equation (11). Note that the sum of all $R^2_{m,l}$ is not necessary R^2 because of the correlation of the variables. When variables x and y are strongly correlated, excluding x results in that y can explain more. Moreover, * indicates that the characteristic was statistically significant at the 10% level for the restricted 4-factor model, whereas ** indicates that it was statistically significant for another restricted model that did not include 4 factors.

Table 3 depicts 18 variables that solely explain most of the variation in the returns in terms of marginal $R^2_{m,l}$ for the restricted 4-factor model based on Equation (11). It confirms that the constant is very important in explaining expected returns, hence the returns of different companies at the same time share a significant commonality unrelated to any of the other 62 characteristics. But we also have to note that the constant is not related to any other variable, while all the other characteristics are to a certain degree related.[7] The two most important categories based on the marginal $R^2_{m,l}$ metric are the profitability and trade friction category which contribute respectively seven and five characteristics to the 18 most important instruments. This confirms the previous results that profitability and trading frictions are the most important groups in terms of their ability to explain variation in returns. Moreover, Table 3 shows the robustness for a high number of characteristics because they contributed substantially to the increase in R^2 and were statistically significant in restricted IPCA models. Similarly, Figures 1, 9, and Table 3 show that sales-to-cash (*s2c*) has a substantial influence on the constitution of the risk factors in terms of their magnitude and on the R^2.

Table 2 based on the restricted IPCA model defined as in Equation (11) shows that once the IPCA model is extended with another risk factor, the characteristics might change from significant to insignificant or vice versa. A possible reason is that the instrument is statistically significant for one factor and insignificant for the other and the combination turns out to be insignificant. Hence, where we previously looked at the importance of characteristics for the constitution of all the risk factors, which was based on magnitude, we now examine how significant a characteristic is for one risk factor to explain the returns. Table 9 in Appendix C.1.1 depicts the p-values and F-values of the characteristics. The F-value can be observed as a weighted function of the p-values related to the individual risk factors, hence these F-values often lay between the lowest p-value and the highest p-value of the individual factors. Characteristics that have been statistically significant for at least five out of the combined set of ten p-values and four F-values are: *investment*, *roa*, *sat*, *eps*, and *ol*. The characteristics investment and sat have a positive association with return and the remainder a negative relationship with return as shown in Table 10 in Appendix C.1.1. Moreover, we also conclude that different risk factors of a model have different instruments that are statistically significant contributors in the explanation of returns.

Lastly, we discuss the p-values of the anomaly factor of the unrestricted models. Table 11 in Appendix C.1.1 shows that a substantial amount of characteristics are not related to the risk factor but do explain variation in the returns via the anomaly factor, hence the hypothesis of a non-existent anomaly is rejected. Especially variables in trading frictions and profitability are

[7] This means that when excluding characteristic a which is correlated with characteristic b does not decrease the R^2 as much because the variation that was previously explained by omitted characteristic a is now partly explained by b.

often statistically significant. Instruments *noa*, *EPS*, *sat*, *ldp*, *lme*, *std_volume*, *prc*, and *ret* and the constant are statistically significant for all the models at the 5% level. Moreover, we observe that on average the p-values slightly increase when a factor is added and are not robust which leads to the belief that part of the anomaly return is actually related to risk.

4.1.2 In-sample fit

This subsection considers how well the characteristics explain the variation in returns and expected returns in-sample.

Table 4: Total and predicted R^2 in-sample of the restricted and unrestricted model.

	Total R^2				Predicted R^2			
	Individual returns ($\boldsymbol{r}_t$)		Portfolio returns ($\boldsymbol{x}_t$)		Individual returns ($\boldsymbol{r}_t$)		Portfolio returns ($\boldsymbol{x}_t$)	
k	$\boldsymbol{\Gamma}_\alpha = 0$	$\boldsymbol{\Gamma}_\alpha \neq 0$	$\boldsymbol{\Gamma}_\alpha = 0$	$\boldsymbol{\Gamma}_\alpha \neq 0$	$\boldsymbol{\Gamma}_\alpha = 0$	$\boldsymbol{\Gamma}_\alpha \neq 0$	$\boldsymbol{\Gamma}_\alpha = 0$	$\boldsymbol{\Gamma}_\alpha \neq 0$
1	0.1901	0.2045	0.8927	0.9144	0.0720	0.0810	0.3487	0.3655
2	0.2133	0.2202	0.9283	0.9310	0.0794	0.0808	0.3627	0.3649
3	0.2273	0.2315	0.9430	0.9445	0.0797	0.0807	0.3635	0.3644
4	0.2366	0.2407	0.9487	0.9494	0.0797	0.0807	0.3629	0.3645

Note: The R^2 displayed are based on the restricted $\boldsymbol{\Gamma}_\alpha = 0$ and unrestricted $\boldsymbol{\Gamma}_\alpha \neq 0$ IPCA model with k systematic factors as defined in Equations (11) and (16). The total R^2 and predictive R^2 are calculated based on Equation (40) and (41) respectively.

Table 4 shows that when the number of risk factors increases the total R^2 of both, restricted and unrestricted, as well as raw returns and portfolio returns increase. However, this increase in model fit is at a decreasing rate which is a logical consequence of the identification assumption made in Section 2.3 where the covariance matrix of the factors has descending diagonal elements. This means that the first factor has the highest variance and therefore the highest explanation power. Moreover, it shows that if at least 3 factors are included, almost all variation in returns is explained by the IPCA model since the increase in total R^2 from the 3- to 4-factor model is low.

Moreover, the fit of the unrestricted model performs better for all model specifications, because the unrestricted model has one additional factor. Even though this factor is constant it helps to explain the variation of the returns. The higher total R^2 of the unrestricted model can be logically explained because the hypothesis $\boldsymbol{\Gamma}_\alpha = \mathbf{0}_{L\times 1}$ is rejected. However, the constant factor of the unrestricted model does not perform as well as the factors of the restricted model, as one can observe when comparing the total R^2 of the restricted model with $k+1$ risk factors to unrestricted model with k risk factors and one constant factor. The difference in fit between the restricted and unrestricted model is negatively related to the number of risk factors, this might be partly caused by an increase in the p-value for the $\boldsymbol{\Gamma}_\alpha = \mathbf{0}_{L\times 1}$ hypothesis. Lastly, the restricted 4-factor model captures 98% of the unrestricted model, hence suggesting the minor contribution of anomaly loading on the 4-factor model.

Our IPCA method is based on OLS and therefore targets the total R^2 directly. This essentially means that the risk factors are constructed based on the criteria that they need to explain as much as possible of the variation in returns. Since the method is established on returns and not expected returns, the predicted R^2 is not optimized. Table 4 suggests that the IPCA exhibits risk factors that can explain risk compensation across the options.
The total and predicted R^2 of the characteristics-managed portfolio returns are high which is in line with the findings of Kelly et al. (2019). This suggests that the managed portfolios that function as test assets exhibit the power to explain systematic risk and differences in expected returns.

Table 12 in Appendix C.1.2 shows that in-sample IPCA performs better than PCA, this difference is most substantial when few factors are considered. However, when 4 factors are considered the effect of time-varying parameters and implementation of instruments seems to be marginal.

4.1.3 Out-of-sample analysis

In order to investigate if a trading strategy constructed based on the basic IPCA exhibits good fit and is lucrative, we investigate the R^2, the Sharpe ratio, and the VaR. The usual structure of splitting the dataset into two parts is used, where the first part is the modeling dataset and the second the forecasting dataset. In the modeling dataset the parameters are estimated, while the forecasting dataset serves as out-of-sample observations. The modeling dataset consists of 75 or 90 observations which corresponds to around 65 to 80% of the total number of observations and increases with one observation every period until we reach the end of the forecast period $T = 112$. The estimation accuracy in terms of R^2 for both restricted and unrestricted models are shown in Table 5.

Table 5: Out-of-sample R^2 for the basic IPCA model.

	Individual returns ($\boldsymbol{r}_t$)				Portfolio returns ($\boldsymbol{x}_t$)			
	Total R^2		Predicted R^2		Total R^2		Predicted R^2	
k	$\boldsymbol{\Gamma}_\alpha = \mathbf{0}$	$\boldsymbol{\Gamma}_\alpha \neq \mathbf{0}$	$\boldsymbol{\Gamma}_\alpha = \mathbf{0}$	$\boldsymbol{\Gamma}_\alpha \neq \mathbf{0}$	$\boldsymbol{\Gamma}_\alpha = \mathbf{0}$	$\boldsymbol{\Gamma}_\alpha \neq \mathbf{0}$	$\boldsymbol{\Gamma}_\alpha = \mathbf{0}$	$\boldsymbol{\Gamma}_\alpha \neq \mathbf{0}$
1	0.1338	0.0843	0.0750	0.0837	0.8517	0.5279	0.4531	0.4923
2	0.1466	0.0809	0.0823	0.0834	0.8926	0.4715	0.4829	0.4900
3	0.1536	0.0766	0.0827	0.0833	0.9009	0.4277	0.4851	0.4899
4	0.1567	0.0788	0.0832	0.0834	0.9062	0.4208	0.4881	0.4897

Note: The R^2 displayed are based on the restricted ($\boldsymbol{\Gamma}_\alpha = \mathbf{0}$) or unrestricted ($\boldsymbol{\Gamma}_\alpha \neq \mathbf{0}$) IPCA model with k systematic factors. The total R^2 and predictive R^2 are calculated based on Equation (40) and (41) that are slightly adjusted as described in Section 2.8. Moreover, t=90 is used as the first out-of-sample date.

It follows from Tables 4 and 5 that the IPCA model not only performs strong in-sample but also out-of-sample since the decrease in R^2 is not too substantial. Therefore, our model exhibits the feature that it is good at explaining the most important drivers of returns in-sample as well as out-of-sample. Furthermore, the table shows that the R^2 out-of-sample increases with a decreasing rate when more factors are used as was expected by imposing that every factor should explain less than the previous factor and also confirms what Figure 8 in Appendix B.1 suggested. Especially the increase from the first to the third factor seems important. Moreover, the predicted R^2 that is based on $\boldsymbol{\lambda}_t$ seems to be relatively high compared to the predicted R^2 in-sample, suggesting a good out-of-sample fit for expected returns. The reason for this might be a combination of two reasons where the former is related to the risk factors and the latter to the risk loadings. First, the in-sample $\boldsymbol{\lambda}$ is based on information at time $t+1$ but also at information later than $t+1$ which seems to be less useful than or even contradictory to past information. Second, similarly as for the risk factors, $\boldsymbol{\Gamma}_{\beta,t}$ might be better estimated by not including future information. However, the second effect is definitely less important, because otherwise the total R^2 of the out-of-sample would also be higher than the total R^2 in-sample.

Moreover, the out-of-sample statistics in Table 13 in Appendix C.1.3 show that the IPCA almost always outperforms the PCA which is similar to the findings for stock returns in Kelly et al. (2019). Note that for estimation of managed portfolios the same amount of parameters are used for IPCA as PCA, whereas normally for raw returns we would expect that the IPCA's dimension reduction feature would cause IPCA to outperform the PCA more substantially.

Table 6: Sharpe ratios for the basic IPCA model.

k	f_1	f_2	f_3	f_4	f_{tang}	f_α
1	1.13	x	x	x	1.13	0.30
2	1.60	0.46	x	x	1.76	0.35
3	1.36	0.63	0.59	x	1.74	0.30
4	0.85	0.91	0.27	0.50	1.94	0.38

Note: All the out-of-sample monthly Sharpe ratios are calculated based on Equation (45). f_i stands for factor i, f_{tang} for the tangency portfolio based on algorithm 4 and f_α is based on Equation (43) and (44). All factors are based on the restricted model with k factors as in Equation (11) apart from the anomaly factor that is based on the unrestricted model as in Equation (16). Moreover, t=90 is used as the first out-of-sample date.

Table 6 depicts high Sharpe ratios for the tangency portfolio and for the individual factors implying that the IPCA is able to capture a large portion of the co-movement between the options and can successfully align the expected returns with the factor loadings. Moreover, these Sharpe ratios seem to be in line with the results of Cao et al. (2017), who found monthly Sharpe ratios for delta-hedged option returns between 0.01 and 2 for portfolios sorted on different characteristics.

Even though the mean of the returns is negative, the Sharpe ratio of the tangency portfolio is positive, because it exhibits a diversification feature and exploits the idiosyncratic risk in the portfolio. Moreover, the Sharpe ratio increases when a factor is added because the tangency portfolio has more freedom in choosing the optimal weights when more factors can be used. As expected, the tangency Sharpe ratio always exceeds the maximum of their individual factors, because an optimal weight for the factors is chosen based on risk and return.

The table also shows that the Sharpe ratios based on the IPCA pure-alpha portfolio are smaller than based on the tangency portfolio. Hence, it suggests that arbitrage possibilities show low potential. Note that this superfluity of the arbitrage factor is confirmed by Table 5 that shows that the total R^2 out-of-sample is lower than the restricted variant.

To assess how the IPCA performs in extreme events such as crises, a VaR is considered.

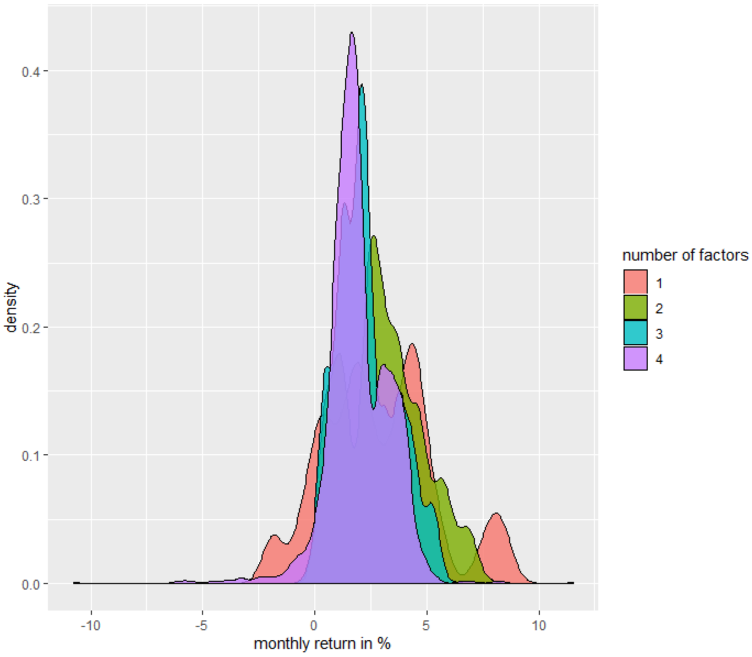


Figure 2: VaR of the restricted basic IPCA model.

Note: The VaR is calculated based on Equation (46) that uses the tangency portfolio returns as defined in Algorithm 4 for the restricted IPCA model.

Figure 2 shows that on average the returns of the tangency portfolio seem promising especially considering that the average option return is negative. When more risk factors are included, a lower maximum monthly return can be obtained but the downside has decreased too.

We conclude that a trading strategy based on the basic IPCA model that uses tangency portfolios seems favorable given the limited downside and the high unconditional Sharpe ratio. However, since no transaction costs are incorporated in the tangency returns, the actual returns will be lower especially when the tangency portfolio depends on characteristics that change frequently.

4.2 Regularized IPCA model

This subsection considers the more parsimonious model of the basic IPCA by utilizing the ridge, lasso, and elastic net as defined in Equations (18), (20) and (25).

4.2.1 Coefficients and significance

In the previous subsection we have established which instruments are important in constructing the risk loadings and explaining the variation in returns. To investigate how robust these observations are we compare them with regularized models that have a selection feature.

Tables 14 till 17 in Appendix C.2.1 show the minimum, mean, and maximum of the coefficients based on the lasso model for 5 different penalty parameters λ. For most models and most instruments the sign of the minimum is not different than that of the maximum, therefore indicating robustness. However, this robustness decreases when more factors are taken into consideration because the model has more possibilities to reach a high R^2. Moreover, we observe that the value and trading frictions categories seem to have the largest impact on the constitution of the loadings. This is as expected due to the nature of this category and is in line with Cao et al. (2017). The most important instruments seem to be price (*prc*), *size*, book to market (*beme*), Price to 52-week high price ($rel\_to\_high\_price$) and beta. Where *price*, *size*, *beme*, and $rel\_to\_high\_price$ have a positive relationship with returns, the relationship between beta and return is unclear. Moreover, Table 18 in Appendix C.2.1 shows how often an instrument is set to 0 by the lasso and therefore is not selected to be important in the constitution of the loadings. We observe next to the aforementioned characteristics that *roa*, *at*, *lturnover*, ret_max, and $total\_vol$ also seem to be important, since these are infrequently set to 0.

We are particularly interested if the most important characteristics that constitute the portfolios of the 4-factor model depicted in Figure 1 and 9 to 11 in Appendix C.1.1 are robust. Table 17 shows that when a λ unequal to zero is considered, a substantial number of characteristics that have turned out to be important in the constitution of the factors in the basic IPCA model switch sign and have a lower weight in absolute value terms. Hence, we conclude that the portfolio constitution can substantially differ between model specifications.

When comparing the 5 characteristics ($investment, roa, sat, eps$, and *ol*) that were often statistically significant in Table 9 in Appendix C.1.1 with the robustness check from the lasso as in Tables 14 till 17 in Appendix C.2.1, we find that these 5 characteristics are not entirely stable showing sometimes different signs for different model specifications. However, the mean of the coefficient based on the robustness analysis does often correspond with the signs of the 5 aforementioned characteristics based on the basic IPCA model as shown in Table 10 in Appendix C.1.1.

4.2.2 In-sample fit

In this part we discuss the in-sample fit of the three special cases of the elastic net. First, we discuss the elastic variant with hyperparameter $\rho = 0$, which is essentially a ridge regression. Subsequently, we deal with the special case $\rho = 1$ which is a lasso and lastly we discuss the elastic net scenario where the hyperparameter is between 0 and 1.

Ridge

In the ridge regression we try to find the optimum in the trade-off between variance and biasedness. When the assumptions of the OLS are satisfied then by the Gauss-Markov theorem the covariance of the ridge estimator $\mathbf{\Gamma}_\beta$ is reduced compared to the OLS variant, while the bias is increased. Figure 3 shows that indeed compared to the normal OLS ($\lambda = 0$), the covariance of the estimators decreases. When λ increases, the values of the parameters are more and more centered around 0 which is as expected since parameters that are not zero are fiercer punished in the objective equation.

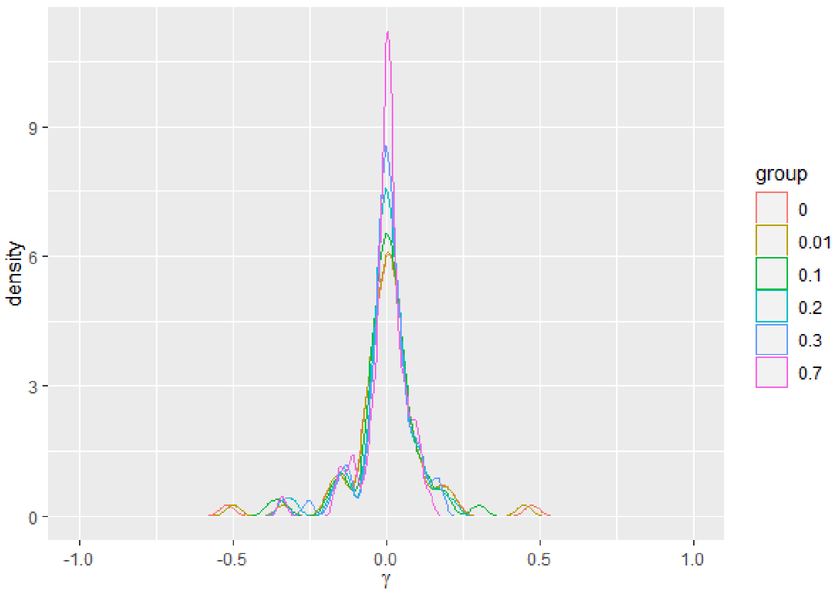


Figure 3: Coefficients $\mathbf{\Gamma}_\beta$ distribution of the restricted ridge 1-factor model.

Note: The coefficients of the restricted ridge 1-factor model are based on Equation (19). The distribution focuses more on values around 0 when λ increases. The third quartile of the absolute value shrinks from 0.042 to 0.03 when λ increases from 0.01 to 0.7.

To assess the quality of the model and the effect of the penalty function we assess the in-sample R^2. Table 19 in Appendix C.2.2 shows that a similar pattern is visible for the ridge as the basic model. That is when the number of factors increases, the fit increases and that the predicted R^2 is around a third of the total R^2. When λ increases both the total and predicted R^2 decreases for raw and portfolio returns. The basic model always outperforms the ridge model on total R^2, but the difference decreases when more factors are considered. The reason is that a model with a large number of parameters benefits more, however still not sufficiently, from shrinkage.

Lasso and elastic net

When we evaluated the properties of lasso and elastic net, we found that due to the nature of the penalty, the coefficients of $\mathbf{\Gamma}_\beta$ are in theory set to 0 when they do not contribute sufficiently to the reduction of squared errors.

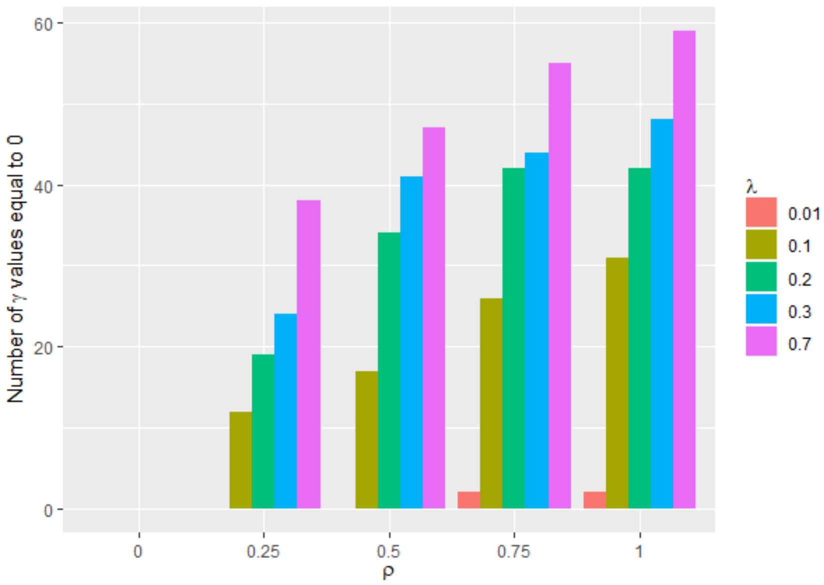


Figure 4: Amount of values of the γ vector equal to zero for elastic net restricted 1-factor IPCA model. *Note*: The coefficients of the elastic net restricted 1-factor IPCA model are based on Equation (27). The distribution focuses more on values around 0 when λ and/or ρ increase.

Figure 4 shows empirically that when ρ increases and hence more weight is attributed to the lasso, the number of elements of $\mathbf{\Gamma}_\beta$ that are set equal to 0 increases. This shows that the lasso has a more severe penalty function than the ridge. Moreover, the graph shows that an increase in λ also results in an increase of values that are set to 0 because a larger weight is attributed to the penalty function in the objective equation.

Table 20 in Appendix C.2.2 depicts the normal pattern for the in-sample total R^2, where an increase in penalty weight λ decreases the R^2. Moreover, it shows that when ρ increases, which means that more weight is attributed to setting the parameters equal to 0 instead of near 0, the total R^2 decreases. In contrast to the linear relationship in the total R^2, the predicted R^2 shows a pattern where the highest R^2 is on the diagonal high ρ and low λ till low ρ and high λ. This result can be interpreted as that the existence of a constraint is beneficial when it is not too fierce (high ρ and λ) nor too weak (low ρ and λ). Hence, the elastic net often outperforms the lasso and the ridge. Table 21 in Appendix C.2.2 shows for the 2-factor model a relatively similar pattern where shrinkage and selection should not be too firm.

4.2.3 Out-of-sample analysis

Empirical research by Bayer (2018) has shown that out-of-sample estimation based on regularized models often perform better because the effect of overfitting that causes poor performance out-of-sample is reduced. Moreover, he found that in a quiet period, ridge regression performs better in terms of VaR when no variable selection is needed, whereas in a period of financial crises, lasso and elastic net perform better due to the selection feature.

Ridge

Table 22 in Appendix C.2.3 shows that occasionally a slightly better fit is achieved when the parameters are shrinked for the 2 and 4-factor model. A possible reason is that these models might exhibit larger parameter instability and therefore a shrinkage of the parameters has a beneficial effect.

Apart from fit, we also examine the Sharpe ratio of the ridge and contrast it to previous methods. Table 23 in Appendix C.2.3 shows that an increase in λ generally increases the Sharpe ratio until the penalty parameter reaches 0.2 and afterward decreases. Hence, a value of around 0.2 for the ridge parameter is most beneficial when we consider a trading strategy that both accounts for return and risk.

A trading strategy should not only yield a good mean-variance ratio but also have a limited downside in risky periods such as crisis which is especially important for low-risk investors. Figure 17 in Appendix B.3 shows a similar pattern as the basic IPCA where more factors yield more stable returns and a lower risk profile. The downside risk decreases when a higher λ is considered. However, to achieve a higher upward potential for the 3-factor model, the penalty parameter should be adjusted to a value between 0.1 and 0.3.

Lasso and reduced model based on p-values

To obtain a parsimonious model that according to Ledolter and Abraham (1981) can increase out-of-sample fit a two-step method is used based on two distinct methods. The first method selects the characteristics based on p-values that were statistically significant (lower than 0.1) in the restricted IPCA and uses this reduced set of characteristics to calculate the out-of-sample fit R^2. Whereas the second method uses the characteristics that were not set to 0 by the in-sample estimation of the lasso, in the out-of-sample estimation.

Table 24 in Appendix C.2.3 shows that a regularization based on lasso often improves the fit out-of-sample for restricted models including more than 1 factor and λ between 0.01 and 0.7. Moreover, the difference becomes larger when more factors are considered because then the basic IPCA model includes too many parameters and hence overfits. The unrestricted model benefits even more from a severely reduced model specification, possibly explained by the larger set of parameters, due to the inclusion of the anomaly factor, that normally needs to be estimated. The fit of the portfolio returns as depicted in Table 25 in Appendix C.2.3 also benefits from a reduced model, this is probably because the test assets that are being used as dependent variable in the reduced formation are based on fewer characteristics that exhibit variation that can be more easily explained by the same set of characteristics. In other words, characteristics that exhibit variation that is hard to explain and therefore construct test assets that are harder to explain are removed from the estimation model. Since these observations hold for different time periods the results are robust.

In terms of the Sharpe ratio, the reduced models sometimes outperformed the basic variant as depicted in Table 26 in Appendix C.2.3. When few factors are considered a lasso model based on a low penalty parameter occasionally outperforms the basic model whereas for 3 or 4 risk factors a higher penalty parameter is needed for the lasso to outperform the basic. This is a logical consequence of the beneficial feature of the lasso that performs well when there are too many parameters that need to be estimated.

Figure 18 in Appendix B.3 shows that when a higher penalty parameter λ and therefore fewer characteristics are considered, the downside risk is often lower but also decreases the upward potential.

4.3 Robust IPCA model

This subsection discusses the robust version of the basic IPCA based on OLS by using either a Huber loss function to tackle the problem of outliers or WLS to solve the heteroskedasticity problem as defined in Algorithm 2 and 3.

4.3.1 In-sample fit

Table 7: R^2 based on robust methods.

	Total R^2						Predicted R^2					
	Individual returns ($\boldsymbol{r}_t$)			Portfolio returns ($\boldsymbol{x}_t$)			Individual returns ($\boldsymbol{r}_t$)			Portfolio returns ($\boldsymbol{x}_t$)		
k	OLS	Huber	WLS	OLS	Huber	WLS	OLS	Huber	WLS	OLS	Huber	WLS
1	0.1901	0.1277	0.1709	0.8927	0.6483	0.8255	0.0720	0.0414	0.0732	0.3487	0.1826	0.3531
2	0.2133	0.1426	0.1907	0.9283	0.6838	0.8826	0.0794	0.0437	0.0781	0.3627	0.1879	0.3616
3	0.2273	0.1529	0.2060	0.9430	0.7160	0.9128	0.0797	0.0476	0.0772	0.3635	0.2007	0.3609
4	0.2366	0.1570	0.2127	0.9487	0.7239	0.9210	0.0797	0.0481	0.0780	0.3629	0.2026	0.3608

Note: The R^2 displayed are based on the restricted OLS, Huber, and WLS IPCA model with k systematic factors as defined in Equation (11) and Algorithms 2 and 3. The total R^2 and predictive R^2 are calculated based on Equation (40) and (41) respectively.

Table 7 shows that the R^2 is almost always lower for the IPCA based on the two robust methods, Huber and WLS, than the basic IPCA based on OLS. The only exception is where the WLS 1-factor model has a higher predicted R^2 than the other models. According to Greene (2003), the reason for the poorer fit might be due to the estimation of the weights. Because when there exists substantial uncertainty about the weights, the wrong observation might have received a large weight which results in low efficiency and hence a low R^2.

4.3.2 Out-of-sample analysis

We established that the in-sample performance of robust methods does seldom contribute to an improved fit, which was as expected. However, research showed that a more robust method can increase the out-of-sample fit. Table 27 in Appendix C.3.1 shows that the basic model beats almost always the Huber and the WLS model. Hence, either the outliers and heteroskedasticity in the data do not substantially affect the performance or the weights that were estimated in the robust methods were instable. The latter is a possible reason because the model is applied on empirical data that does not have a substantial number of observations every month.

Since both start dates show lower Sharpe ratios for WLS and Huber than the basic OLS indicated by Table 28 in Appendix C.3.1, the results are robust and investors are better of by not implementing these robust methods in their trading strategy. The reason for the lower Sharpe ratio is probably the instability of the weights as explained before.

4.4 Comparison with stock returns

It is interesting to examine whether option returns can be better described by IPCA than stock returns and whether the same relationships between returns and instruments are established. The comparison in association between characteristics and returns is based on three metrics: the signs, p-values and F-values of the instruments, whereas the model's performance of the two asset classes is contrasted based on R^2 and Sharpe ratio. In this comparison we use Kelly et al. (2019) as reference material. Note that they used data on 12,813 firms with 36 characteristics ranging from July 1962 to May 2014. Hence, there are noticeable differences in the dataset at first glance. However, we have to bear in mind that we have seen that different time-periods do not necessarily lead to different sample fits, and that a substantial number of variables are either correlated or do not contribute to improving the sample fit as was shown by the two-step method. Hence, the differences in the sample period and the number of characteristics might not have a large effect on the output.

The loadings of the market capitalization, value of the assets, and market beta were described as important for stock returns and option returns. However, we did not find a contribution of the momentum and reversal characteristic, a possible cause is that momentum for stocks does not affect option returns or the non-existence of trading strategies that exploit momentum and reversal instruments.

The statistical significance of the instruments for option returns was often in-line with stock returns, apart from the past return category. Characteristics that were important for option returns but not for equity returns were profits or their individual parts revenue and costs as well as cash. A possible explanation is that one can argue that a higher profit and cash position yield a lower default probability which has been important in the explanation of option returns according to Cao et al. (2017).

The values for the total and predicted R^2 in-sample and out-of-sample seemed to exhibit similar patterns: these metrics increase with factors, where the unrestricted model outperformed the restricted model and values that were relatively similar. However, a noticeable difference existed in predicted R^2. The predicted R^2 for option returns seemed to be substantially higher, a reason can be that due to the hedging feature the options risk factor f_t is more stable, which causes the price of risk λ and risk factor f_t to be relatively similar.

Lastly, the yearly Sharpe ratios of the tangency returns based on stock returns were similar to the monthly counterpart. When we would convert the monthly Sharpe ratios of the options to yearly, we would obtain substantially higher Sharpe ratios. This indicates that a trading strategy based on options might be favorable next to or as a substitution of stocks.

5 Further research

Similar to Kelly et al. (2019) and Horenstein et al. (2018) we do not include transaction costs, hence optimal trading strategies that involve high monthly turnover might not be optimal when accounted for transaction costs. Since the most important category for the constitution of loadings is trading frictions which is probably positively related to turnover this might have an unfavorable effect on the tangency portfolio returns and therefore on the Sharpe ratio. Furthermore, Cao et al. (2017) found that characteristics behave in the same way for delta-hedged call option returns as for put option returns. It might be interesting to examine if this also holds for the IPCA model.

6 Conclusion

The first part answers the key research question of the paper: which characteristics explain delta-hedged option returns. Subsequently, the second part consists of the comparison between different models and methods based on in-sample and out-of-sample fit, Sharpe ratios, and VaR.

The IPCA restricted model found that based on regularization techniques the price of the stock, size of the company, book-to-market value, price relative to the 52-week high price, and beta have a large magnitude for the factor loading and therefore are important in the constitution of the risk factors f_t portfolios. The first four instruments are positively related to returns, whereas the relationship between the latter characteristic and returns is unstable. In contrast, the sorted anomaly portfolio is mainly driven by market capitalization and standard deviation of volume. The first instrument is positive suggesting that larger companies have higher option returns beyond what is related to the risk exposure, which is similar to Cao et al. (2017). The latter characteristic is negative indicating that lower volume increases return beyond the risk exposure, possibly caused by the difficulty of hedging an option when it is traded infrequently.

Moreover, the characteristics that best describe returns based on their p and F-values are investment, sales to total assets (sat), return on assets, earnings per share, and costs to total assets (ol). Where the first two instruments are positively related to returns and the latter three negatively. One can view sat and ol as profit to total assets which has a positive relation with returns.

Most of the variables that showed to be important in the portfolio constitution or explanation of variation in returns were also important in the marginal R^2.

The in-sample total R^2 of the IPCA ranged from 0.19 for the 1-factor restricted model to 0.24

for the 4-factor unrestricted model. Almost all models indicated that to explain returns around 3 to 4 factors are needed, because the increase in the in-sample R^2 when an additional factor is added decreases substantially from the three to four-factor model. Moreover, the anomaly factor decreases in importance when more factors are considered.

The regularized models were often not able to beat the basic version, hence selection and shrinkage methods in-sample are not necessary. Moreover, the robust methods also underperformed the basic version, possibly due to the uncertainty of the weights.

The out-of-sample R^2 showed that the IPCA is able to explain a substantial amount of variation in returns out-of-sample because there is only a minor decrease compared to the in-sample R^2. The out-of-sample monthly Sharpe ratios seemed promising having relatively high values compared to the equity returns yearly Sharpe ratio variant in Kelly et al. (2019) and option return models based on sorted portfolios Cao et al. (2017). Moreover, we found that the tangency portfolio has higher Sharpe ratios when more factors are considered due to its diversification feature. The downside risk based on VaR of the IPCA is limited especially when more than 1 factor is considered.

Shrinking the coefficients to 0 with the ridge method increased the fit when the penalty parameter λ was not too high. Moreover, a higher Sharpe ratio on the tangency portfolio can be obtained when λ is around 0.2 and the downside risk compared to the basic variant can be decreased when a λ between 0.1 and 0.3 is considered. Hence, the shrinkage method which shrinks the coefficients slightly is useful for both explaining the variation in return and constructing a trading strategy.

Furthermore, selection methods based on the two-step procedure also proved their worth with a substantially better fit when λ is between 0.01 and 0.7. This is especially evident when either 3 or 4 factors are considered since the basic IPCA finds it difficult to estimate a large set of parameters due to the instability such a large set causes to their coefficients. Apart from fit, Sharpe ratios also increased when either few factors and a small penalty parameter or 3 to 4 factors with a large penalty parameter were considered. Hence, a more parsimonious model can increase out-of-sample fit as well as offer a better trading strategy when the number of factors and the value of the penalty parameter are carefully chosen.

Robust methods did not increase fit out-of-sample nor yield a higher Sharpe ratio possibly caused by a similar reason as in the in-sample estimation. We conclude that even though some OLS assumption are not satisfied, accounting for these with the WLS or Huber Loss function did not increase performance.

References

S. C. Ahn, A. R. Horenstein, and N. Wang. Beta matrix and common factors in stock returns. *Journal of Financial and Quantitative Analysis*, 53(3):1417–1440, 2018.

S. Bayer. Combining value-at-risk forecasts using penalized quantile regressions. *Econometrics and statistics*, 8:56–77, 2018.

G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung. *Time series analysis: forecasting and control*. John Wiley and Sons, 1970.

M. W. Brandt, P. Santa-Clara, and R. Valkanov. Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns. *The Review of Financial Studies*, 22(9): 3411–3447, 2009.

J. Cao and B. Han. Cross section of option returns and idiosyncratic stock volatility. *Journal of Financial Economics*, 108(1):231–249, 2013.

J. Cao, B. Han, Q. Tong, and X. Zhan. Option return predictability. *Rotman School of Management Working Paper*, (2698267), 2017.

R. B. Cattell. The scree test for the number of factors. *Multivariate behavioral research*, 1(2): 245–276, 1966.

G. Connor and R. A. Korajczyk. Risk and return in an equilibrium apt: Application of a new test methodology. *Journal of Financial Economics (JFE)*, 21(2), 1988.

D. L. Donoho and J. M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *biometrika*, 81(3):425–455, 1994.

J. S. Duesenberry. *The Brookings quarterly econometric model of the United States*. Rand McNally, 1965.

E. F. Fama and K. R. French. The cross-section of expected stock returns. *the Journal of Finance*, 47(2):427–465, 1992.

E. F. Fama and K. R. French. Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, 33(1):3–56, 1993.

E. F. Fama and K. R. French. A five-factor asset pricing model. *Journal of financial economics*, 116(1):1–22, 2015.

E. F. Fama and J. D. MacBeth. Risk, return, and equilibrium: Empirical tests. *Journal of political economy*, 81(3):607–636, 1973.

J. Fan, Y. Liao, and W. Wang. Projected principal component analysis in factor models. *Annals of statistics*, 44(1):219, 2016.

J. Fox and S. Weisberg. Robust regression. *An R and S-Plus companion to applied regression*, page 91, 2002.

J. Friedman, T. Hastie, and R. Tibshirani. Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software*, 33(1):1, 2010.

M. R. Gibbons, S. A. Ross, and J. Shanken. A test of the efficiency of a given portfolio. *Econometrica: Journal of the Econometric Society*, pages 1121–1152, 1989.

S. Goncalves and L. Kilian. Bootstrapping autoregressions with conditional heteroskedasticity of unknown form. *Journal of Econometrics*, 123(1):89–120, 2004. URL `https://EconPapers.repec.org/RePEc:eee:econom:v:123:y:2004:i:1:p:89-120`.

W. H. Greene. *Econometric analysis*. Pearson Education India, 2003.

S. Gu, B. T. Kelly, and D. Xiu. Autoencoder asset pricing models. *Available at SSRN*, 2019.

L. P. Hansen and S. F. Richard. The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models. *Econometrica: Journal of the Econometric Society*, pages 587–613, 1987.

H. Hong and J. C. Stein. A unified theory of underreaction, momentum trading, and overreaction in asset markets. *The Journal of finance*, 54(6):2143–2184, 1999.

A. Horenstein, A. Vasquez, and X. Xiao. Common factors in equity option returns. *SSRN Electronic Journal*, 01 2018. doi: 10.2139/ssrn.3290363.

P. J. Huber. Robust estimation of a location parameter. In *Breakthroughs in statistics*, pages 492–518. Springer, 1992.

N. Jegadeesh. Evidence of predictable behavior of security returns. *The Journal of finance*, 45 (3):881–898, 1990.

N. Jegadeesh and S. Titman. Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, 48(1):65–91, 1993. doi: 10.1111/j.1540-6261.1993.tb04702.x. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1993.tb04702.x`.

Y. Jiang, Y.-G. Wang, L. Fu, and X. Wang. Robust estimation using modified huber's functions with new tails. *Technometrics*, 61(1):111–122, 2019. doi: 10.1080/00401706.2018.1470037. URL `https://doi.org/10.1080/00401706.2018.1470037`.

P. Jorion. Value at risk. 2000.

B. Kelly, S. Pruitt, and Y. Su. Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics*, 1 2019. ISSN 0304-405X. doi: 10.1016/j.jfineco.2019.05.001.

B. T. Kelly, S. Pruitt, and Y. Su. Instrumented principal component analysis. *Available at SSRN 2983919*, 2017.

S. Kim, R. A. Korajczyk, and A. Neuhierl. Arbitrage portfolios. 2019.

J. Ledolter and B. Abraham. Parsimony and its importance in time series forecasting. *Technometrics*, 23(4):411–414, 1981.

M. Lettau and M. Pelger. Factors that fit the time series and cross-section of stock returns. Working Paper 24858, National Bureau of Economic Research, July 2018. URL `http://www.nber.org/papers/w24858`.

J. Lintner. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. In *Stochastic optimization models in finance*, pages 131–155. Elsevier, 1975.

D. W. Marquardt and R. D. Snee. Ridge regression in practice. *The American Statistician*, 29 (1):3–20, 1975.

A. B. Owen. A robust hybrid of lasso and ridge regression. *Contemporary Mathematics*, 443(7): 59–72, 2007.

R. T. Rockafellar. *Convex analysis*. Number 28. Princeton university press, 1970.

S. A. Ross. The arbitrage theory of capital asset pricing. In *Handbook of the fundamentals of financial decision making: Part I*, pages 11–30. World Scientific, 1976.

G. W. Schwert and P. J. Seguin. Heteroskedasticity in stock returns. *the Journal of Finance*, 45 (4):1129–1155, 1990.

W. F. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance*, 19(3):425–442, 1964.

R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.

A. N. Tikhonov, A. Tikhonov, and A. TIKHONOV. Solution of incorrectly formulated problems and the regularization method. 1963.

A. Vasquez and X. Xiao. Default risk and option returns. *Working Paper*, 2018.

H. Zou and T. Hastie. Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, 67(2):301–320, 2005.

H. Zou, T. Hastie, and R. Tibshirani. Sparse principal component analysis. *Journal of computational and graphical statistics*, 15(2):265–286, 2006.

Appendices

A Characteristics

Table 8: Characteristics grouped in categories - abbreviation and their meaning.

Past-returns
(1) r_{2_1} return 1 month before prediction
(2) r_{6_2} return from 6 to 2 months before prediction
(3) r_{12_2} return from 12 to 2 months before prediction
(4) r_{12_7} return from 12 to 7 months before prediction
(5) r_{36_13} return from 36 to 13 months before prediction

Investment
(6) *Investment* % change in AT
(7) Δ *CEQ* % change in BE
(8) Δ *PI2A* Change in PP&E and inventory over lagged AT
(9) Δ *Shrout*, % changes in shares outstanding
(10) *IVC* Change in inventory over average AT
(11) *NOA* Net-operating assets over lagged AT

Profitability
(12) *ATO* Sales to lagged net operating assets
(13) *CTO* Sales to lagged total assets
(14) Δ(ΔGM-ΔSales *Delta*% change in gross margin and % change in sales
(15) *EPS* earnings per share
(16) *IPM* Pre-tax income over sales
(17) *PCM* Sales minus costs of goods sold of sales
(18) *PM* OI after depreciation over sales
(19) *Prof* Gross profitability over BE
(20) *RNA* OI after depreciation to lagged net operating assets
(21) *ROA* Income before extraordinary items to lagged AT
(22) *ROC* Size + longterm debt - total assets to cash
(23) *ROE* income before extraordinary items to lagged BE
(24) *ROIC* Return on invested capital
(25) *S2C* Sales to cash
(26) *SAT* Sales to total assets

Intangibles
(27) *AOA* Absolute value of operating accruals
(28) *OL* Costs of goods sold + SG&A tot total assets
(29) *Tan* Tangibility
(30) *OA* Operating accruals

Value
(31) *A2ME* Total assets to Size
(32) *BEME* Book to Market ratio
(33) *C* Cash to AT
(34) *C2D* Cash to total liabilities
(35) Δ *SO* Log change in split-adjusted shares outstanding
(36) Debt2P Total debt to Size
(37) *E2P* Income before extraordinary items to Size
(38) *Free CF* Free cash flow to BE
(39) *LDP* Trailing 12-months dividends to price
(40) *NOP* Net payouts to Size
(41) *O2P* Operating payouts to market cap
(42) *Q* Tobin's Q
(43) *S2P* Sales to price
(44) *sales_g* Growth of sales

Trading frictions
(45) *AT* Total assets
(46) *Beta* Correlation x ratio of vols
(47) *Beta daily* CAPM beta using daily returns
(48) *DTO* De-trended Turnover - market Turnover
(49) *Idio vol* Idiosyncratic vol of FF 3 factor model
(50) *LME* Price times shares outstanding
(51) *Size* Lagged size variable
(52) *Lturnover* Last months volume to shares outstanding
(53) *Rel_to_high* price Price to 52 week high price
(54) *Ret_max* Maximum daily return
(55) *Spread* Average daily bid-ask spread
(56) *Std. turnover* Standard deviation of daily turnover
(57) *Std. volume* Standard deviation of daily volume
(58) *SUV* Standard unexplained volume
(59) *Total vol* Standard deviation of daily return
(60) *price*, Price of the stock
(61) *return*, Return on the stock
(62) *moneyness*, Option price to strike

Note: The variables are based on the paper of Kim et al. (2019). This paper discusses how the instruments are exactly constructed in more detail.

B Figures

B.1 Preliminary analysis

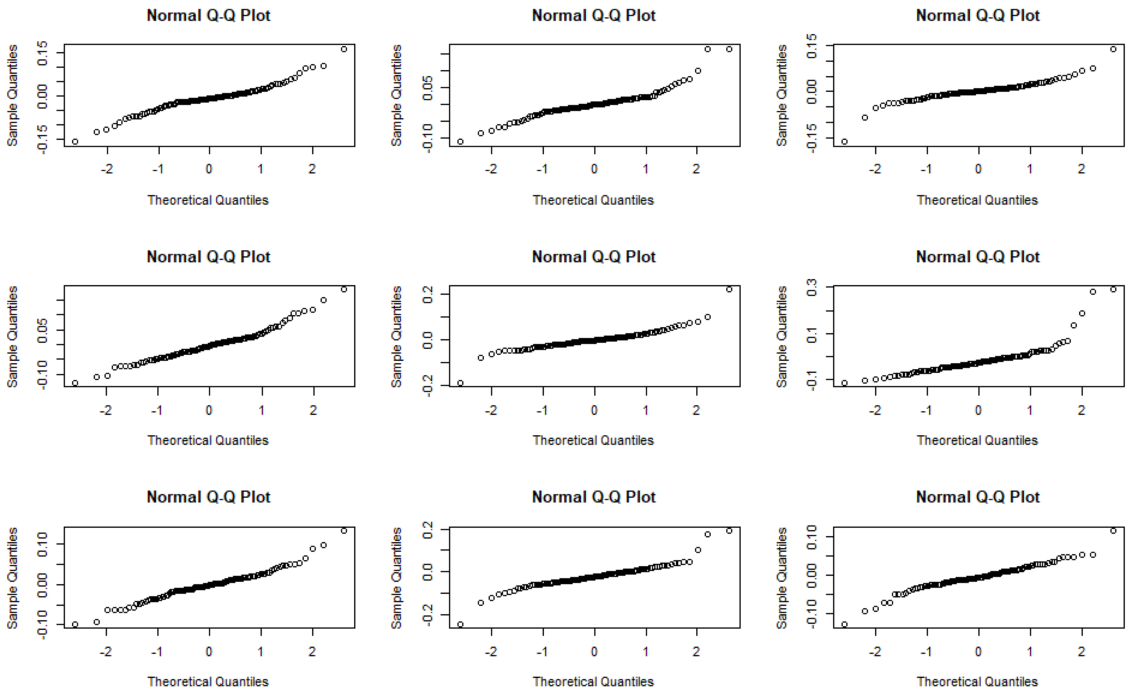


Figure 5: Errors of restricted 4-factor model of 9 different assets with more than 100 observations.

Note: The Q-Q method explained in Section 3.1 is used to assess whether the errors of the restricted 4-factor model as defined in Equation (11) are normally distributed for 9 different options.
Since the line is not linear at the minimum and maximum, we reject the assumption that the errors are normally distributed.

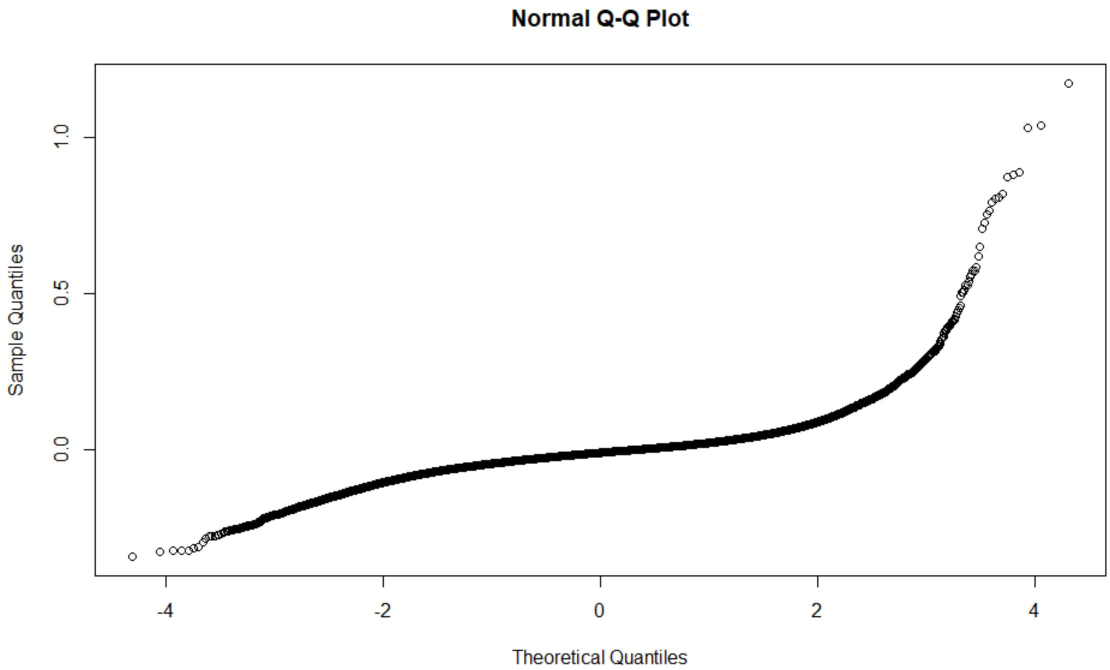
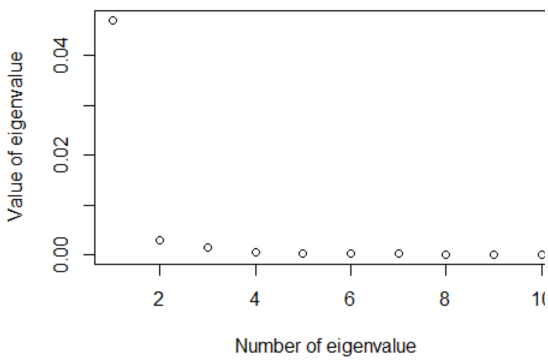


Figure 6: Errors of restricted 4 factor model for all assets.

Note: The Q-Q method explained in Section 3.1 is used to assess whether the errors of the restricted 4-factor model as defined in Equation (11) are normally distributed for all options.
Since the line is not linear at the minimum and maximum, we reject the assumption that the errors are normally distributed.



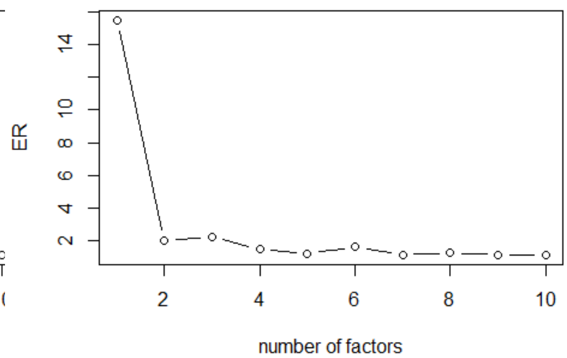


Figure 7: Value of the eigenvalue. **Figure 8:** Eigenvalue ratio.

Note: The Eigenvalue ratio method explained in Section 3.3 is used to assess how many factors to include. The eigenvalues are based on the covariance matrix of the managed portfolio returns.
Since there is a substantial decline from factor 3 to 4 and factor 4 to 5, a maximum of 4 factors are included. Where the fourth factor serves the purpose of last check.

B.2 Basic IPCA

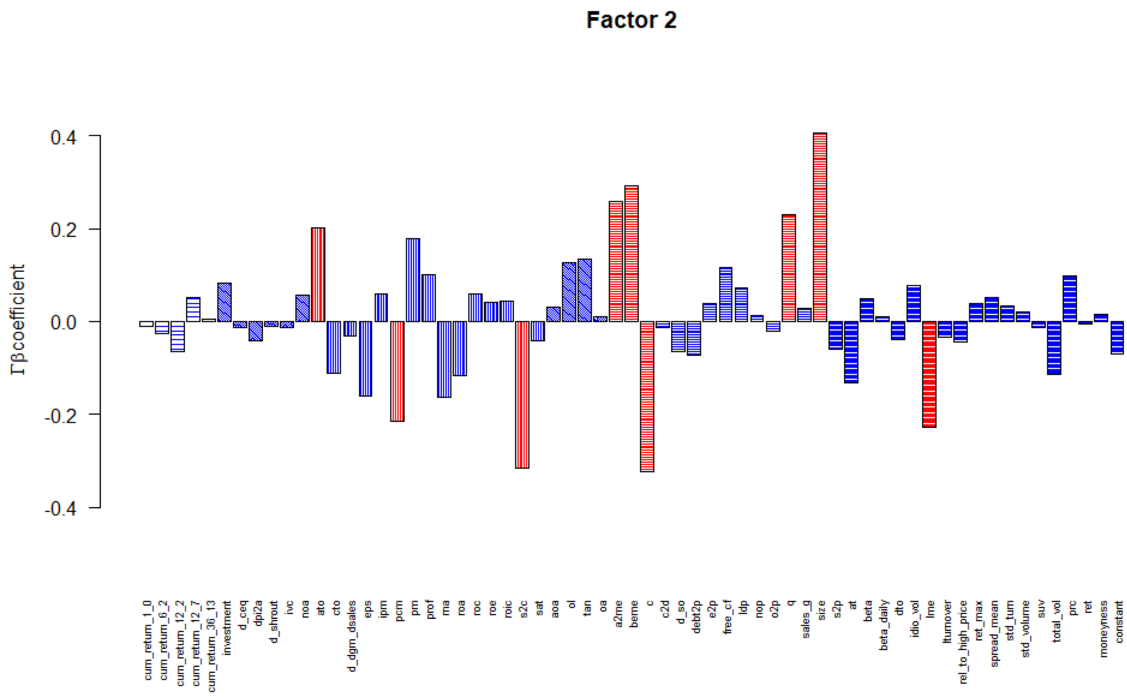


Figure 9: Estimates of $\mathbf{\Gamma}_\beta$ for the restricted 4-factor model.

Note: The estimates of $\mathbf{\Gamma}_\beta$ related to the second risk factor of the restricted 4-factor model as defined in Equation (11) are shown. A red bar indicates a value that is in absolute terms larger than 0.2 and different patterns are used for different categories.

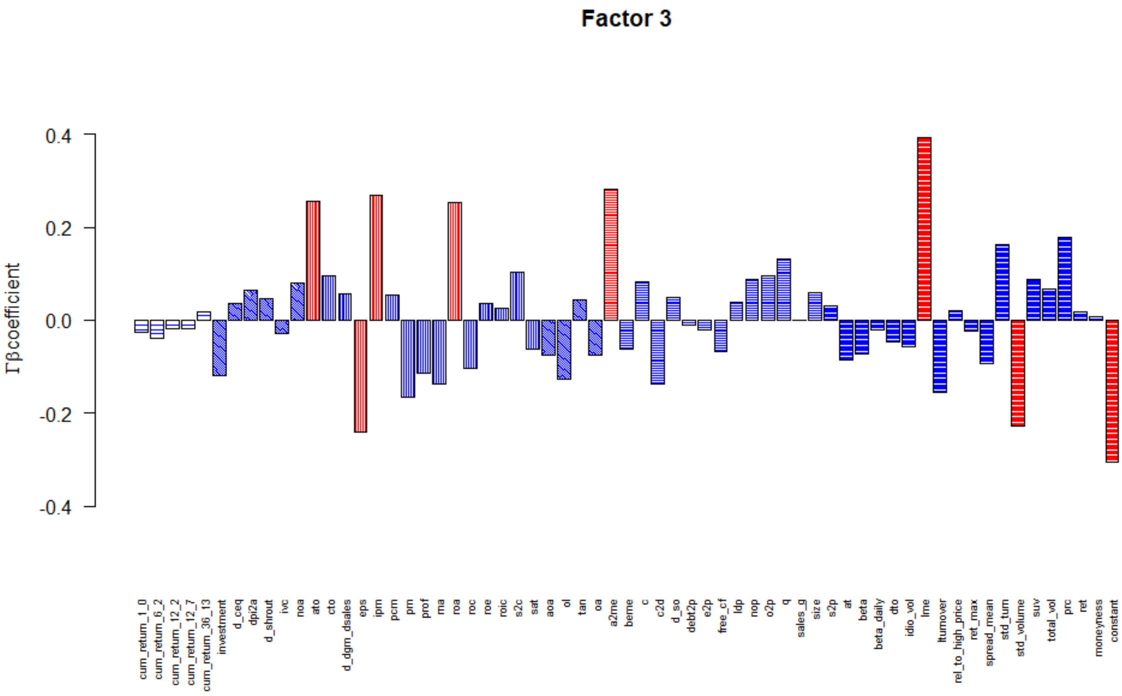


Figure 10: Estimates of $\mathbf{\Gamma}_\beta$ for the restricted 4-factor model.

Note: The estimates of $\mathbf{\Gamma}_\beta$ related to the third risk factor of the restricted 4-factor model as defined in Equation (11) are shown. A red bar indicates a value that is in absolute terms larger than 0.2 and different patterns are used for different categories.

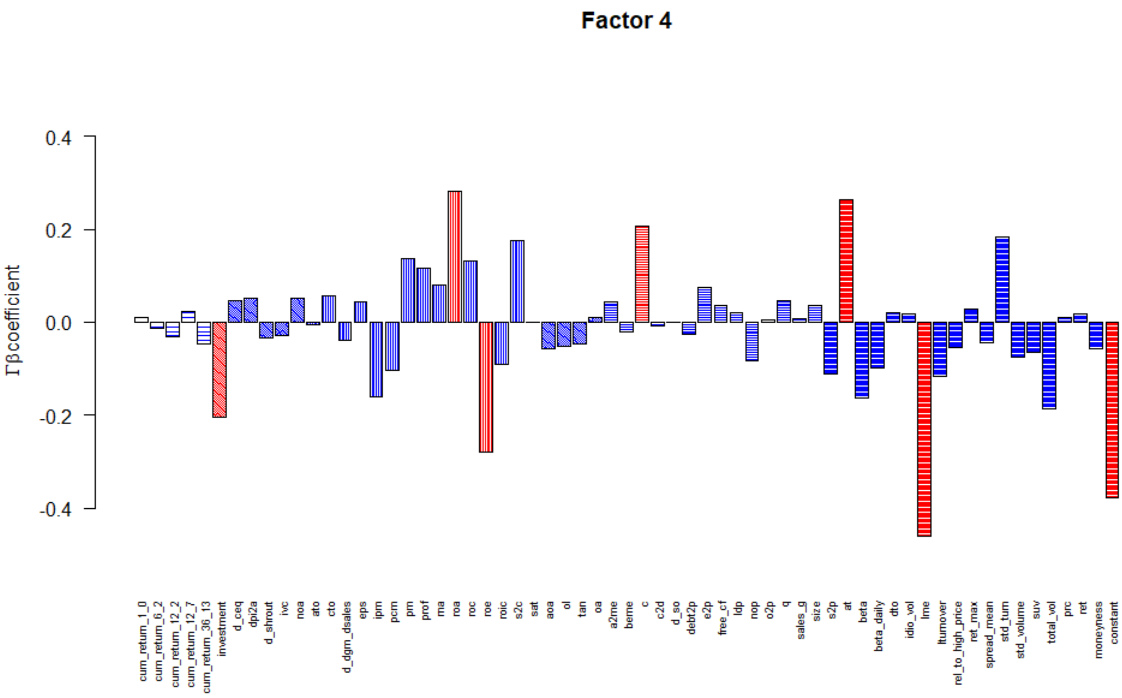


Figure 11: Estimates of $\mathbf{\Gamma}_\beta$ for the restricted 4-factor model.

Note: The estimates of $\mathbf{\Gamma}_\beta$ related to the fourth risk factor of the restricted 4-factor model as defined in Equation (11) are shown. A red bar indicates a value that is in absolute terms larger than 0.2 and different patterns are used for different categories.

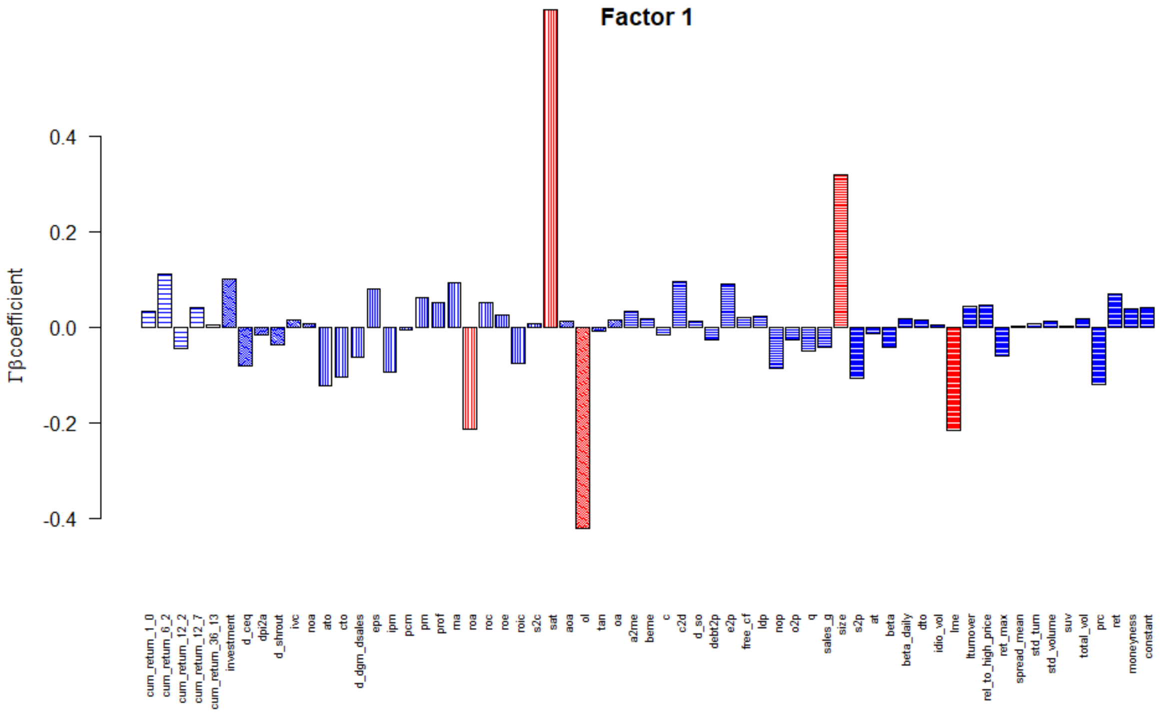


Figure 12: Estimates of $\mathbf{\Gamma}_\beta$ for the unrestricted 4-factor model.

Note: The estimates of $\mathbf{\Gamma}_\beta$ related to the first risk factor of the unrestricted 4-factor model as defined in Equation (16) are shown. A red bar indicates a value that is in absolute terms larger than 0.2 and different patterns are used for different categories.

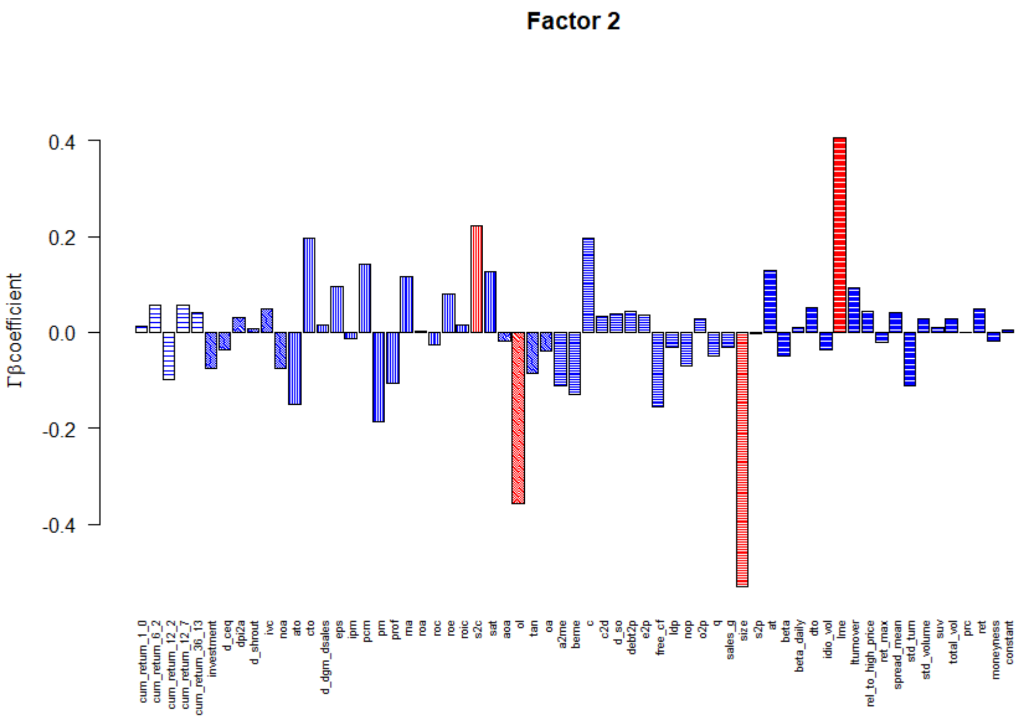


Figure 13: Estimates of $\mathbf{\Gamma}_\beta$ for the unrestricted 4-factor model.

Note: The estimates of $\mathbf{\Gamma}_\beta$ related to the second risk factor of the unrestricted 4-factor model as defined in Equation (16) are shown. A red bar indicates a value that is in absolute terms larger than 0.2 and different patterns are used for different categories.

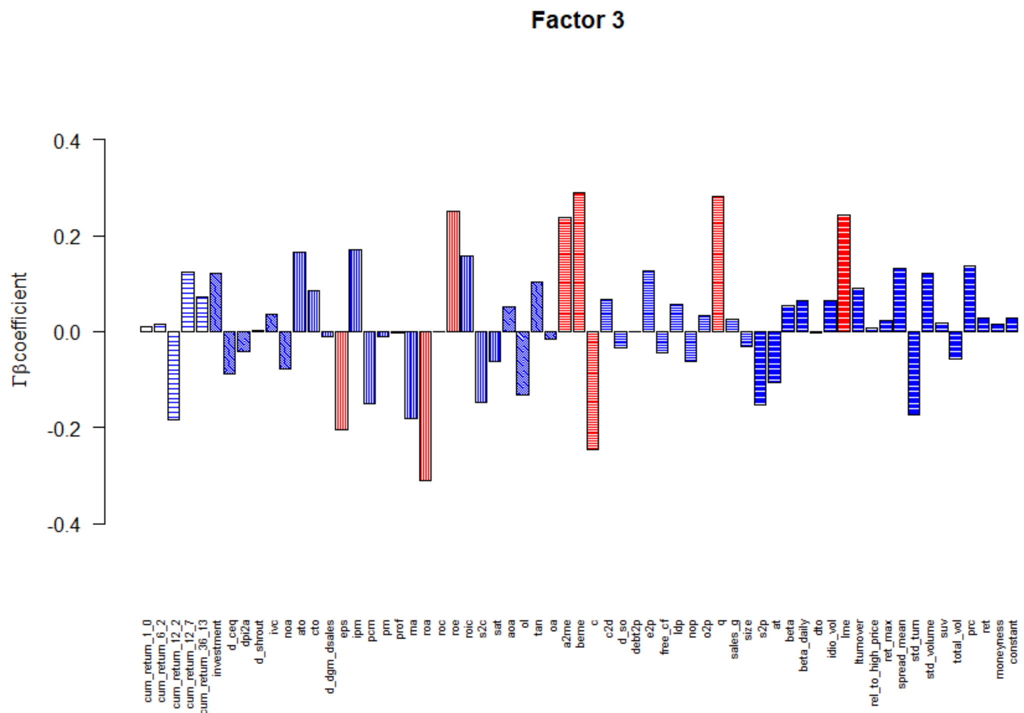


Figure 14: Estimates of $\mathbf{\Gamma}_\beta$ for the unrestricted 4-factor model.

Note: The estimates of $\mathbf{\Gamma}_\beta$ related to the third risk factor of the unrestricted 4-factor model as defined in Equation (16) are shown. A red bar indicates a value that is in absolute terms larger than 0.2 and different patterns are used for different categories.

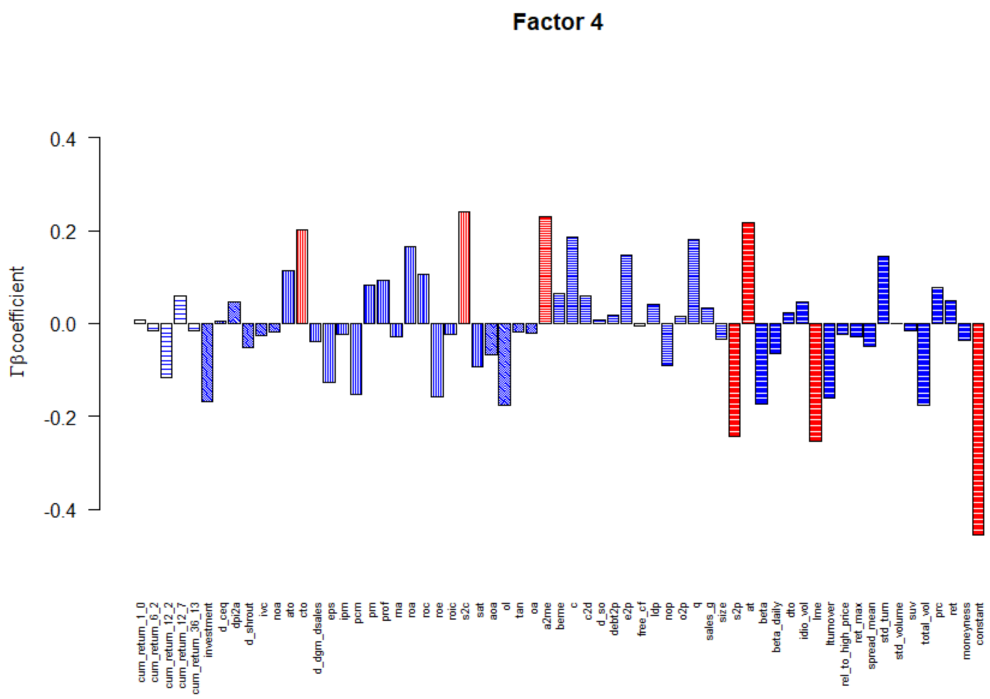


Figure 15: Estimates of $\mathbf{\Gamma}_\beta$ for the unrestricted 4-factor model.

Note: The estimates of $\mathbf{\Gamma}_\beta$ related to the fourth risk factor of the unrestricted 4-factor model as defined in Equation (16) are shown. A red bar indicates a value that is in absolute terms larger than 0.2 and different patterns are used for different categories.

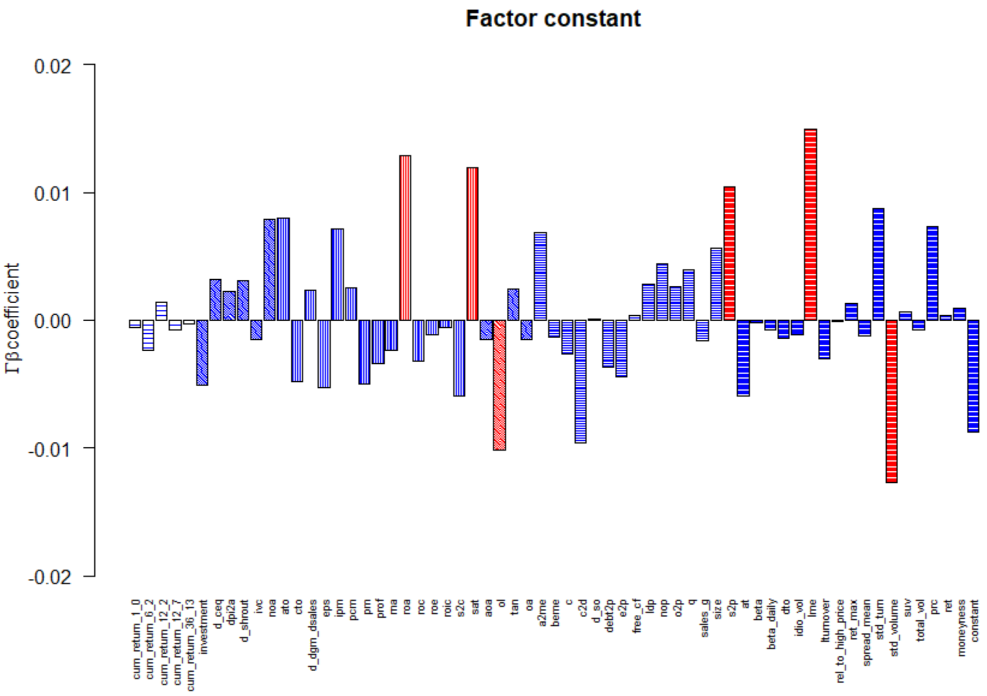


Figure 16: Estimates of $\mathbf{\Gamma}_\alpha$ for the unrestricted 4-factor model.

Note: The estimates of $\mathbf{\Gamma}_\alpha$ related to the anomaly factor of the unrestricted 4-factor model as defined in Equation (16) are shown. A red bar indicates a value that is in absolute terms larger than 0.2 and different patterns are used for different categories.

B.3 Regularized models

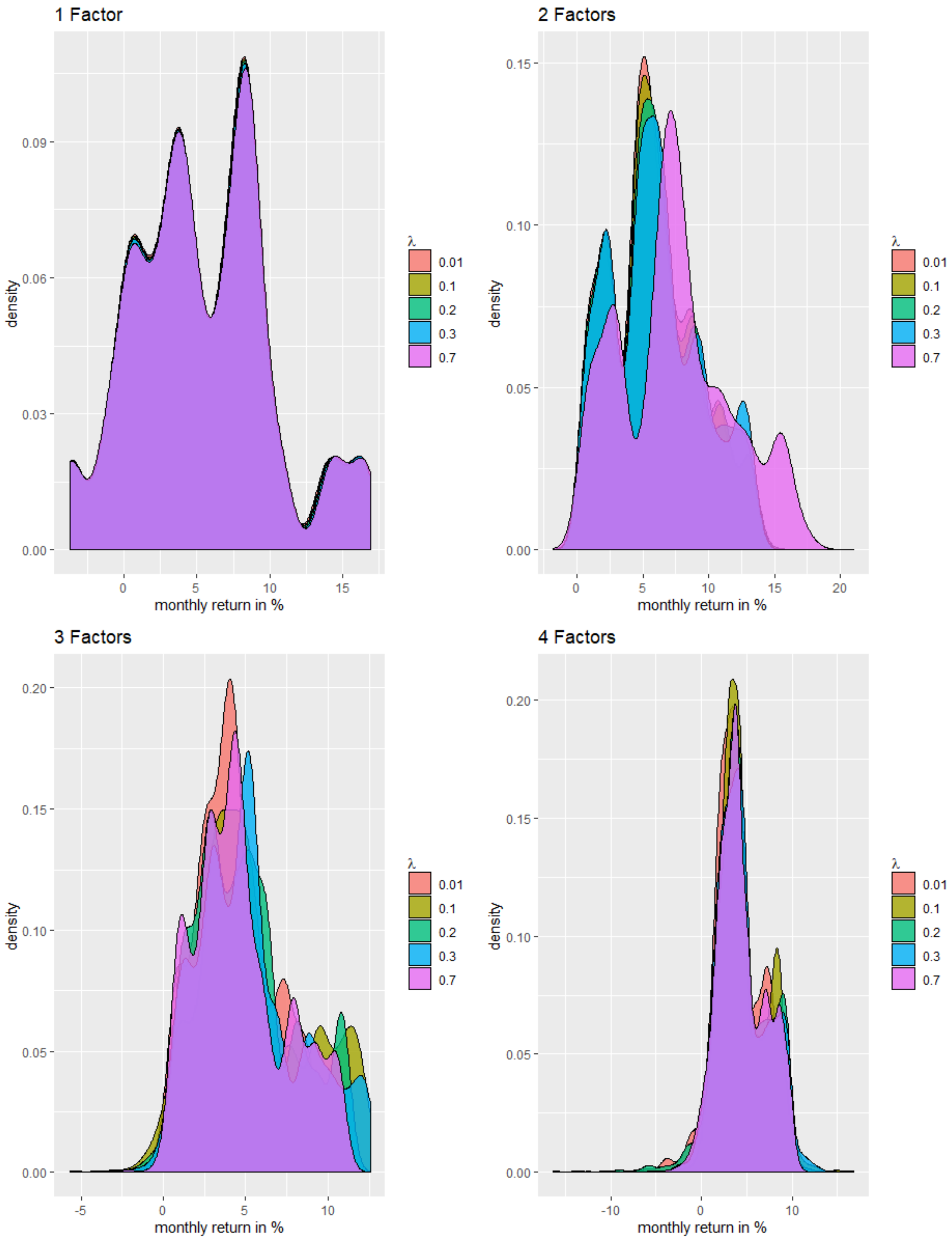


Figure 17: Distribution of the tangency portfolio returns based the restricted ridge model.

Note: The distribution of the monthly returns of the tangency portfolio based on the restricted ridge model are displayed. The returns are based on tangency portfolios defined in Algorithm 4 that is based on the ridge model as in Equation (18).

A higher penalty parameter λ corresponds to lower downside risk and more factors decreases the variation in monthly returns.

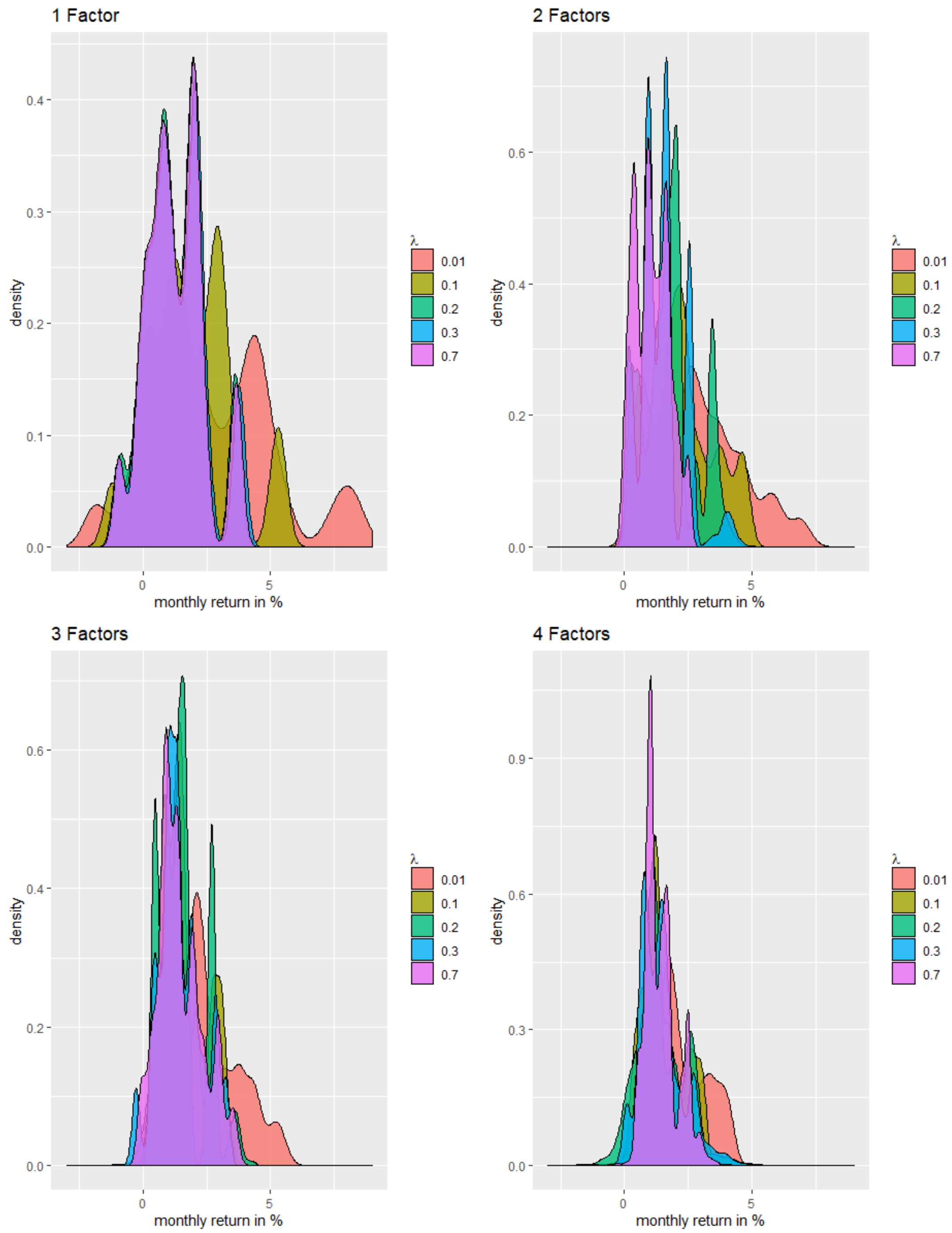


Figure 18: Distribution of the tangency portfolio returns based the reduced lasso factor model.

Note: The distribution of the monthly returns of the tangency portfolio based on the lasso model are displayed. The returns are calculated using tangency portfolios defined in Algorithm 4 that is based on the lasso model as in Algorithm 1.

A higher penalty parameter λ often corresponds to lower downside risk and more factors decreases the variation in monthly returns.

C Tables

C.1 Basic IPCA

C.1.1 Coefficients and significance

Table 9: P-value of the characteristics for $\mathbf{\Gamma}_\beta$ of the restricted model.

Characteristic / risk factor	F1	p2.1	p2.2	F2	p3.1	p3.2	p3.3	F3	p4.1	p4.2	p4.3	p4.4	F4
Past returns													
cum_return_1_0	0.98	0.41	0.22	0.32	0.39	0.86	0.96	0.88	0.22	0.82	0.59	0.87	0.92
cum_return_6_2	0.96	0.53	0.66	0.69	0.23	0.79	0.75	0.86	0.26	0.83	0.73	0.93	0.9
cum_return_12_2	0.69	0.42	0.83	0.72	0.59	0.73	0.88	0.94	0.6	0.69	0.9	0.88	0.98
cum_return_12_7	0.62	0.5	0.72	0.76	0.64	0.69	0.92	0.96	0.45	0.59	0.85	0.83	0.96
cum_return_36_13	0.94	0.23	0.28	0.23	0.55	0.62	0.56	0.76	0.73	0.87	0.72	0.42	0.82
Investment													
investment	0.02	0.76	0	0	0.8	0.32	0	0	0.7	0.25	0.11	0.01	0
d_ceq	0.9	0.56	0.48	0.54	0.41	0.8	0.42	0.76	0.33	0.81	0.57	0.55	0.85
dpi2a	0.1	0.58	0.1	0.22	0.44	0.64	0.16	0.27	0.76	0.38	0.2	0.46	0.48
d_shrout	0.82	0.72	0.76	0.92	0.94	0.76	0.97	0.99	0.78	0.82	0.36	0.67	0.78
ivc	0.7	0.82	0.54	0.83	0.68	0.74	0.57	0.86	0.68	0.76	0.51	0.68	0.94
noa	0.09	0.18	0.32	0.32	0.56	0.18	0.43	0.4	0.32	0.34	0.26	0.57	0.55
Profitability													
ato	0.12	0.14	0.38	0.32	0.19	0	0.24	0.01	0.16	0	0	0.98	0
cto	0.82	0.76	0.46	0.64	0.94	0.68	0.63	0.88	0.68	0.55	0.71	0.79	0.96
d_dgm_dsales	0.78	0.31	0.38	0.42	0.63	0.74	0.99	0.98	0.12	0.32	0.14	0.43	0.2
eps	0.07	0.1	0.66	0.31	0.98	0.07	0.62	0.26	0.89	0.08	0.03	0.73	0.2
ipm	0.35	0.15	0.81	0.4	0.64	0.22	0.9	0.6	0.96	0.62	0.1	0.28	0.15
pcm	0.68	0.76	0.72	0.88	0.2	0.1	0.79	0.3	0.66	0.02	0.57	0.48	0.13
pm	0.68	0.77	0.86	0.97	0.26	0.48	0.91	0.69	0.72	0.28	0.23	0.36	0.4
prof	0.96	0.86	0.96	0.98	0.6	0.72	0.98	0.92	0.92	0.24	0.26	0.24	0.36
rna	0.98	0.5	0.72	0.76	0.12	0.05	0.92	0.1	0.16	0.03	0.17	0.48	0.14
roa	0.3	0.66	0.01	0.01	0.78	0.65	0.05	0.08	0.37	0.43	0.1	0.09	0.12
roc	0.7	0.78	0.3	0.6	0.54	0.88	0.56	0.8	0.67	0.36	0.19	0.16	0.24
roe	0.47	0.48	0.08	0.16	0.62	0.5	0.12	0.32	0.7	0.71	0.8	0.1	0.3
roic	0.66	1	0.3	0.59	0.61	0.68	0.72	0.89	0.5	0.65	0.84	0.55	0.86
s2c	0.77	0.42	0.2	0.3	0.6	0.11	0.16	0.18	0.74	0.06	0.64	0.33	0.29
sat	0.1	0.08	0.96	0.16	0.02	0.7	0.79	0.06	0	0.92	0.83	1	0.04
Intangibles													
aoa	0.11	0.55	0.04	0.02	0.72	0.63	0.06	0.16	0.37	0.42	0.15	0.32	0.27
ol	0.04	0.12	0.38	0.1	0.01	0.82	0.62	0.06	0.01	0.58	0.57	0.84	0.14
tan	0.48	0.22	0.55	0.48	0.3	0	0.62	0.03	0.8	0	0.43	0.52	0.1
oa	0.41	0.97	0.5	0.67	0.9	0.9	0.63	0.92	0.68	0.85	0.18	0.9	0.78

Characteristic / risk factor	F1	p2.1	p2.2	F2	p3.1	p3.2	p3.3	F3	p4.1	p4.2	p4.3	p4.4	F4
Value													
a2me	0.52	0.42	0.98	0.69	0.96	0.2	0.48	0.55	0.9	0.39	0.1	0.72	0.76
beme	0.81	0.54	0.57	0.66	0.75	0.12	0.73	0.44	0.84	0.02	0.58	0.9	0.4
c	0.8	0.36	0.18	0.32	0.7	0.08	0.16	0.07	0.64	0.05	0.59	0.2	0.18
c2d	0.46	0.74	0.48	0.6	0.99	0.59	0.54	0.74	0.81	0.89	0.23	0.96	0.8
d_so	0.58	0.98	0.48	0.74	0.72	0.6	0.68	0.82	0.98	0.21	0.45	0.98	0.78
debt2p	0.64	0.39	0.98	0.8	0.94	0.34	0.86	0.74	0.42	0.31	0.9	0.86	0.86
e2p	0.52	0.5	0.91	0.78	0.5	0.78	0.82	0.92	0.38	0.65	0.78	0.56	0.88
free_cf	0.9	0.96	0.94	1	0.27	0.34	0.98	0.6	0.81	0.08	0.49	0.74	0.46
ldp	0.39	0.02	0.96	0.12	0.31	0.05	0.56	0.24	0.12	0.06	0.41	0.83	0.4
nop	0.82	0.7	0.98	0.96	0.3	0.58	0.96	0.77	0.2	0.88	0.52	0.46	0.73
o2p	0.6	0.43	0.91	0.74	0.42	0.84	0.74	0.87	0.86	0.78	0.42	0.95	0.8
q	0.64	0.68	0.98	0.88	0.96	0.41	0.58	0.92	0.98	0.45	0.54	0.72	0.94
sales_g	0.94	0.22	0.48	0.42	0.08	0.92	0.87	0.66	0.08	0.52	0.98	0.92	0.78
size	0.57	0.46	0.91	0.68	0.96	0.1	0.78	0.6	0.68	0.21	0.82	0.8	0.8
Trading frictions													
s2p	0.86	0.97	0.89	0.98	0.9	0.99	0.78	0.98	0.85	0.73	0.84	0.58	0.98
at	0.77	0.32	0.18	0.32	0.98	0.19	0.34	0.4	0.74	0.51	0.72	0.27	0.73
beta	0.28	0.82	0.09	0.18	0.34	0.59	0.06	0.14	0.56	0.38	0.41	0.12	0.22
beta_daily	0.4	0.67	0.12	0.26	0.5	0.74	0.21	0.34	0.64	0.82	0.79	0.23	0.55
dto	0.98	0.75	0.91	0.98	0.27	0.2	0.95	0.5	0.28	0.32	0.26	0.71	0.61
idio_vol	0.7	0.71	0.92	0.98	0.99	0.7	0.99	0.98	0.71	0.46	0.68	0.85	0.91
lme	0.84	0.56	0.22	0.52	0.5	0.91	0.31	0.74	0.88	0.47	0.05	0.01	0.14
lturnover	0.16	0.64	0.06	0.13	0.68	0.5	0.1	0.3	0.44	0.59	0.11	0.24	0.26
rel_to_high_price	0.88	0.63	0.93	0.9	0.38	0.79	0.86	0.88	0.44	0.48	0.73	0.59	0.87
ret_max	0.72	0.43	0.33	0.49	0.26	0.73	0.93	0.85	0.38	0.59	0.9	0.83	0.94
spread_mean	0.44	0.47	0.12	0.16	0.62	0.81	0.28	0.55	0.36	0.45	0.31	0.63	0.56
std_turn	0.04	0.25	0.14	0.11	0.52	0.6	0.16	0.22	0.66	0.71	0.24	0.25	0.3
std_volume	0.08	0.13	0.47	0.26	0.41	0.56	0.19	0.41	0.57	0.9	0.09	0.6	0.45
suv	1	0.76	0.93	0.96	0.91	0.57	0.98	0.91	0.48	0.84	0.11	0.3	0.29
total_vol	0.34	0.9	0.32	0.42	0.97	0.79	0.42	0.77	0.71	0.41	0.74	0.18	0.6
prc	0.38	0.2	0.46	0.38	0.77	0.11	0.52	0.33	0.88	0.24	0.08	0.94	0.38
ret	0.47	0.07	0.66	0.3	0.07	0.88	0.69	0.38	0.06	0.85	0.68	0.73	0.64
moneyness	0.73	0.19	0.2	0.18	0.23	0.26	0.26	0.38	0.11	0.63	0.95	0.31	0.41
constant	0	0	0	0	0	0	0	0	0	0	0	0	0

Note: The p-values of the characteristics are shown based on the restricted model with k systematic risk factors as defined in Equation (11). $pi.j$ indicates the p-value for risk factor i of a model with j risk factors, whereas the $F.i$ indicates the F-value based on all risk factors of the i-factor model.

Table 10: Coefficients of statistically significant characteristics of the restricted models.

characteristic / risk factor	f1	f2.1	f2.2	f3.1	f3.2	f3.3	f4.1	f4.2	f4.3	f4.4
constant	**-0.336**	**-0.159**	**-0.420**	**-0.043**	**-0.127**	**-0.478**	**-0.039**	**-0.070**	**-0.304**	**-0.377**
ato	0.108	0.104	0.066	-0.073	**0.246**	0.107	-0.075	**0.202**	**0.255**	-0.005
investment	**-0.132**	0.014	**-0.248**	-0.010	0.062	**-0.240**	0.020	0.083	-0.118	**-0.205**
aoa	-0.060	0.018	**-0.117**	0.009	0.020	**-0.106**	0.026	0.031	-0.074	-0.058
roa	0.164	-0.054	**0.383**	-0.036	-0.065	**0.363**	-0.103	-0.116	**0.253**	**0.282**
lturnover	-0.145	-0.025	**-0.158**	0.021	-0.054	**-0.153**	0.045	-0.033	-0.155	-0.118
beta	-0.069	-0.013	**-0.152**	-0.047	0.037	**-0.164**	-0.025	0.048	-0.073	-0.164
ldp	0.047	**0.091**	-0.003	0.029	**0.089**	0.033	0.051	**0.073**	0.039	0.021
c	0.046	-0.144	0.239	0.041	**-0.291**	0.253	-0.041	**-0.324**	0.082	0.207
dpi2a	**0.062**	0.019	**0.082**	0.028	-0.019	0.085	0.011	-0.042	0.063	0.051
tan	0.025	0.051	-0.031	-0.036	**0.139**	-0.028	-0.009	**0.134**	0.044	-0.046
ol	**-0.518**	-0.476	-0.209	**-0.605**	-0.053	-0.114	**-0.590**	0.128	-0.128	-0.051
std_volume	**-0.190**	-0.167	-0.106	-0.082	-0.070	-0.179	-0.051	0.020	**-0.228**	-0.076
rna	0.002	-0.043	0.030	0.092	**-0.183**	0.008	0.090	**-0.162**	-0.138	0.081
eps	**-0.155**	**-0.146**	-0.051	-0.003	**-0.216**	-0.057	0.011	**-0.161**	**-0.240**	0.044
roe	-0.069	0.060	**-0.201**	0.044	0.074	-0.205	0.035	0.042	0.035	**-0.279**
pcm	-0.035	-0.021	-0.035	0.090	**-0.148**	-0.028	0.025	**-0.215**	0.055	-0.103
ret	0.038	**0.073**	-0.023	**0.073**	0.007	0.025	**0.069**	-0.005	0.018	0.018
sat	**0.474**	**0.558**	0.007	**0.670**	0.106	-0.074	**0.719**	-0.041	-0.062	0.000
sales_g	-0.004	-0.052	0.038	**-0.061**	0.005	0.007	**-0.061**	0.029	0.001	0.008

Note: The coefficients of the restricted factor model with 1 to 4 risk factors as defined in Equation (11) are displayed. Only characteristics that were statistically significant for a risk factor for two different models were included. Subsequently we ranked the characteristics from low to high based on their average p-value. f$i.j$ indicates risk factor j of model with i factors. Values in bold denote that the characteristic was statistically significant for that specific risk factor. When all the bold values were positive or negative for a characteristic, the cell of characteristics name is indicated in green or red respectively.

Table 11: P-value of the characteristics for $\boldsymbol{\Gamma}_\alpha$ of the unrestricted model.

characteristic / k	1	2	3	4		1	2	3	4
Past returns					**Value**				
cum_return_1_0	0.82	0.58	0.32	0.5	a2me	0.03	0.04	0.04	0.02
cum_return_6_2	0.24	0.16	0.08	0.1	beme	0.6	0.54	0.27	0.62
cum_return_12_2	0.21	0.3	0.21	0.38	c	0.08	0.12	0.3	0.2
cum_return_12_7	0.18	0.16	0.36	0.71	c2d	0.02	0.04	0	0
cum_return_36_13	0.12	0.35	0.36	0.74	d_so	0.77	0.86	0.38	0.96
					debt2p	0.04	0.06	0.04	0
Investment					e2p	0.64	0.77	0.02	0.01
investment	0.9	0.49	0	0	free_cf	0.76	0.71	0.42	0.88
d_ceq	0.48	0.64	0.02	0.01	ldp	0	0	0.04	0
dpi2a	0.22	0.13	0.02	0.01	nop	0.66	0.92	0	0
d_shrout	0	0	0	0	o2p	0.12	0.1	0.17	0.02
ivc	0.86	0.9	0.16	0.11	q	0.34	0.24	0.42	0.18
noa	0	0	0	0	sales_g	0	0	0.04	0.05
					size	0	0.02	0	0.05
Profitability					s2p	0.18	0.1	0	0
ato	0.04	0.01	0	0					
cto	0.12	0.04	0	0.1	**Trading frictions**				
d_dgm_dsales	0.95	0.6	0	0	at	0.08	0.14	0.04	0.01
eps	0.04	0.02	0.08	0	beta	0.5	0.08	0.66	0.79
ipm	0.01	0.02	0	0	beta_daily	0.47	0.72	0.11	0.43
pcm	0.32	0.24	0.03	0.14	dto	0.12	0.34	0.01	0.05
pm	0.32	0.08	0.02	0.02	idio_vol	0.2	0.88	0.44	0.53
prof	0.26	0.34	0.06	0.01	lme	0.01	0	0.02	0
rna	1	0.89	0.62	0.09	lturnover	0.39	0.98	0.08	0.01
roa	0.7	0.32	0	0	rel_to_high_price	0.27	0.22	0.57	0.92
roc	0.56	0.35	0.24	0.02	ret_max	0.84	0.91	0.12	0.25
roe	0.28	0.38	0.12	0.59	spread_mean	0.57	0.77	0.04	0.28
roic	0.44	0.42	0.34	0.76	std_turn	0	0.01	0	0
s2c	0.04	0.03	0.01	0.01	std_volume	0	0	0	0
					suv	0.18	0.36	0.08	0.47
Intangibles					total_vol	0.29	0.64	0.7	0.73
sat	0	0	0	0	prc	0.01	0.01	0.01	0
aoa	0.96	0.67	0.04	0.05	ret	0	0	0.7	0.58
ol	0	0	0.03	0.01	moneyness	0.01	0	0.12	0.2
tan	0.07	0.02	0.06	0.02					
oa	0.81	0.44	0.32	0.21	constant	0	0	0	0

Note: The p-values for all 63 characteristics of $\boldsymbol{\Gamma}_\alpha$ are displayed based on the unrestricted model with k systematic risk factors as defined in Equation (16). A green cell indicates that this value is below 0.1 and therefore statistically significant at the 10 % level.

C.1.2 In-sample fit

Table 12: In-sample R^2 for managed portfolios ($\boldsymbol{x}_t$) comparison between IPCA and PCA.

	Total R^2 portfolio return ($\boldsymbol{x}_t$)			Predicted R^2 portfolio return ($\boldsymbol{x}_t$)		
k	$\boldsymbol{\Gamma}_\alpha = 0$	$\boldsymbol{\Gamma}_\alpha \neq 0$	PCA	$\boldsymbol{\Gamma}_\alpha = 0$	$\boldsymbol{\Gamma}\alpha \neq 0$	PCA
1	0.8927	0.9144	0.8502	0.3487	0.3655	0.3437
2	0.9283	0.931	0.905	0.3627	0.3649	0.3603
3	0.943	0.9445	0.9326	0.3635	0.3644	0.3612
4	0.9487	0.9494	0.9451	0.3629	0.3645	0.3618

Note: The R^2 displayed are based on the restricted $\boldsymbol{\Gamma}_\alpha = 0$ and unrestricted $\boldsymbol{\Gamma}_\alpha \neq 0$ IPCA model and restricted with k systematic factors in Equations (11), (16) and (2). The total R^2 and predictive R^2 are calculated based on Equation (40) and (41) respectively. For ease of comparison a colour scheme is used where a high R^2 is indicated in green and a low R^2 in red.

C.1.3 Out-of-sample analysis

Table 13: Out-of-sample R^2 for managed portfolio return ($\boldsymbol{x}_t$) comparison between IPCA and PCA.

	Total R^2 portfolio return ($\boldsymbol{x}_t$)			Predicted R^2 portfolio return ($\boldsymbol{x}_t$)		
k	$\boldsymbol{\Gamma}_\alpha = 0$	$\boldsymbol{\Gamma}_\alpha \neq 0$	PCA	$\boldsymbol{\Gamma}_\alpha = 0$	$\boldsymbol{\Gamma}\alpha \neq 0$	PCA
1	0.8517	0.5279	0.8104	0.4531	0.4923	0.4263
2	0.8926	0.4715	0.8956	0.4829	0.49	0.4772
3	0.9009	0.4277	0.9185	0.4851	0.4899	0.4758
4	0.9062	0.4208	0.9265	0.4881	0.4897	0.4761

Note: The R^2 displayed are based on the restricted ($\Gamma_\alpha = 0$) and unrestricted ($\Gamma_\alpha \neq 0$) IPCA model and PCA with k systematic factors. The total R^2 and predictive R^2 are calculated based on Equation (40) and (41) that are slightly adjusted as described in Section 2.8. For ease of comparison a colour scheme is used where a high R^2 is indicated in green and a low R^2 in red.

C.2 Regularized IPCA models

C.2.1 Coefficients and significance

Table 14: Coefficients of $\boldsymbol{\Gamma}_\beta$ of the lasso 1-factor restricted model.

	factor 1		
characteristics	Min.	Mean	Max
Past returns			
cum_return_1_0	-0.0009	-0.0002	0
cum_return_6_2	-0.0482	-0.0125	0
cum_return_12_2	-0.056	-0.0147	0
cum_return_12_7	-0.0203	-0.0041	0
cum_return_36_13	-0.0349	-0.0127	0
Investment			
investment	-0.0648	-0.0138	0
d_ceq	0	0.0005	0.0026
dpi2a	0	0.01	0.0498
d_shrout	-0.0283	-0.0057	0
ivc	-0.0137	-0.0027	0
noa	0	0.0439	0.1043
Profitability			
ato	0	0.0406	0.0774
cto	-0.046	-0.0092	0
d_dgm_dsales	0	0.0007	0.0035
eps	-0.093	-0.0213	0
ipm	0	0.0065	0.0323
pcm	0	0.0275	0.1058
pm	0	0.0051	0.0255
prof	0	0.0148	0.0456
rna	0	0.0106	0.0529
roa	0	0.0595	0.0776
roc	-0.0634	-0.0127	0
roe	0	0.0235	0.0461
roic	-0.0125	0.0023	0.024
s2c	-0.0023	0.0167	0.045
Intangibles			
sat	0	0.0021	0.0103
aoa	-0.05	-0.0117	0
ol	0	0.035	0.1033
tan	0	0	0
oa	0	0.0003	0.0013
Value			
a2me	0	0.001	0.005
beme	0	0.0144	0.0632
c	-0.0814	-0.0302	0
c2d	0	0.0116	0.0262
d_so	-0.0216	-0.0105	0
debt2p	0	0.0028	0.0138
e2p	-0.0177	-0.0035	0
free_cf	0	0.0008	0.0041
ldp	0	0.0435	0.0814
nop	0	0.0009	0.0033
o2p	0	0.0281	0.0629
q	-0.0184	-0.007	0
sales_g	-0.0384	-0.0078	0
Trading frictions			
size	0	0.1114	0.1824
s2p	0	0.0126	0.0413
at	-0.0995	-0.0199	0
beta	-0.1576	-0.1356	-0.0731
beta_daily	-0.0783	-0.0396	0
dto	0	0.0003	0.0017
idio_vol	0	0.0164	0.0822
lme	-0.0189	-0.0038	0
lturnover	-0.2311	-0.1714	0
rel_to_high_price	-0.0168	-0.0034	0
ret_max	-0.0642	-0.0482	0
spread_mean	-0.0379	-0.0076	0
std_turn	0	0.0349	0.1246
std_volume	-0.0428	-0.0096	0
suv	-0.0182	-0.0036	0
total_vol	-0.1478	-0.0545	0
prc	0.0123	0.3551	0.474
ret	0	0.0439	0.122
moneyness	-0.0364	-0.0073	0
constant	-0.9862	-0.8388	-0.7064

Note: The minimum, mean and maximum of the coefficients of the characteristics of the lasso restricted 1-factor model are shown based on Algorithm 1 with $\lambda = \{0.01, 0.1, 0.2, 0.3, 0.7\}$. Red indicates a low value and green indicates a high value.

Table 15: Coefficients of $\mathbf{\Gamma}_{\beta}$ of the lasso 2-factor restricted model.

	factor 1			factor 2		
Characteristics	Min.	Mean	Max.	Min.	Mean	Max.
Past returns						
cum_return_1_0	-0.0016	0.0289	0.0974	0	0	0
cum_return_6_2	-0.0269	0.0026	0.0424	0	0	0
cum_return_12_2	-0.0484	-0.0112	0.0055	0	0	0
cum_return_12_7	-0.0251	-0.0073	0	0	0	0
cum_return_36_13	-0.0024	0.0223	0.1138	0	0	0
investment						
investment	-0.0307	0.0322	0.1432	0	0	0
d_ceq	0	0.0012	0.0045	0	0	0
dpi2a	0	0.0098	0.0485	0	0	0
d_shrout	-0.0018	0.0168	0.086	0	0	0
ivc	-0.0149	0.0001	0.0155	0	0	0
noa	-0.0994	0.0112	0.0673	0	0.0006	0.003
Profitability						
ato	0.0094	0.047	0.0896	0	0.0008	0.0041
cto	-0.0313	-0.0005	0.0289	0	0	0
d_dgm_dsales	-0.0827	-0.0193	0	0	0	0
eps	-0.1256	-0.0468	0	0	0	0
ipm	0	0.0278	0.06	0	0	0
pcm	-0.0248	0.021	0.1034	0	0	0
pm	0	0.0328	0.0672	0	0	0
prof	0	0.0351	0.0708	0	0	0
rna	0	0.0102	0.0502	0	0	0
roa	-0.1139	-0.0019	0.0565	0	0.0096	0.0357
roc	-0.0532	-0.0108	0	0	0	0
roe	0.0036	0.0204	0.0462	0	0.0001	0.0004
roic	-0.0074	0.0035	0.0251	0	0.0077	0.0172
s2c	0	0.0033	0.0143	0	0.0096	0.0398
Intangibles						
sat	0	0.0025	0.0121	0	0	0
aoa	-0.0049	0.046	0.1355	0	0	0
ol	-0.1842	-0.0521	0.0322	0	0	0
tan	-0.0032	0.0449	0.1176	0	0	0
oa	-0.038	-0.0073	0.0015	0	0	0
Value						
a2me	-0.0029	0.0178	0.0612	0	0	0
beme	0	0.2069	0.397	0	0	0
c	-0.0909	-0.0459	-0.0012	-0.0157	-0.005	0
c2d	-0.0252	-0.001	0.0081	0	0.0001	0.0006
d_so	-0.0247	-0.0126	-0.0003	-0.0016	-0.0003	0
debt2p	-0.0046	0.0388	0.0958	0	0	0
e2p	-0.002	0.0102	0.0379	0	0.0043	0.0168
free_cf	-0.0261	-0.0052	0	0	0	0
ldp	0.0794	0.1683	0.2351	0	0.0138	0.0514
nop	0	0	0	0	0	0
o2p	0.0008	0.0246	0.061	0	0.0092	0.0362
q	-0.0401	-0.0214	0	0	0	0
sales_g	-0.0473	-0.0096	0	0	0	0
Trading frictions						
size	0.2237	0.3511	0.5302	0	0.3621	0.8731
s2p	0.0027	0.0171	0.0381	0	0.0003	0.0013
at	-0.0837	0.0125	0.1048	-0.0001	0.0213	0.0853
beta	-0.0988	0.0773	0.3115	-0.5148	-0.1678	-0.0109
beta_daily	-0.0291	0.0517	0.1524	0	0	0
dto	-0.0062	-0.0016	0	0	0	0
idio_vol	-0.083	-0.008	0.0428	-0.005	-0.0013	0
lme	-0.0077	0.0154	0.0491	-0.0005	0.0463	0.1247
lturnover	-0.1739	-0.0097	0.2441	-0.2132	-0.0981	-0.0192
rel_to_high_price	-0.0025	0.0236	0.1204	0	0	0
ret_max	-0.0387	-0.0027	0.0644	-0.1075	-0.04	-0.0043
spread_mean	-0.0032	0.05	0.169	0	0	0
std_turn	-0.1483	-0.0091	0.0928	0	0	0
std_volume	-0.0133	0.0177	0.0879	0	0	0
suv	-0.0008	0.0075	0.0383	0	0	0
total_vol	-0.0339	-0.0096	0.0174	-0.3992	-0.1211	-0.0037
prc	0.1825	0.4088	0.5495	0.2178	0.406	0.6449
ret	0.013	0.0885	0.161	0	0.001	0.0048
moneyness	-0.0062	0.051	0.1338	0	0	0
constant	-0.7916	-0.2416	0.4716	-0.9474	-0.1243	0.4871

Note: The minimum, mean and maximum of the coefficients of the characteristics of the restricted lasso 2-factor model are shown based on Algorithm 1 with $\lambda = \{0.01, 0.1, 0.2, 0.3, 0.7\}$. Red indicates a low value and green indicates a high value.

Table 16: Coefficients of $\mathbf{\Gamma}_\beta$ of the lasso 3-factor restricted model.

	factor 1			factor 2			factor 3		
Characteristics	Min.	Mean	Max.	Min.	Mean	Max.	Min.	Mean	Max.
Past returns									
cum_return_1_0	-0.0004	0.0361	0.0904	-0.0028	-0.0006	0	0	0	0
cum_return_6_2	-0.0137	0.0276	0.1176	0	0.0741	0.2479	0	0	0
cum_return_12_2	-0.0365	0.0281	0.1866	0	0.0267	0.1335	0	0	0
cum_return_12_7	-0.0302	-0.0103	0	0	0	0	0	0	0
cum_return_36_13	-0.0081	0.0243	0.1084	0	0	0.0001	0	0	0
investment									
investment	-0.1088	0.0115	0.1739	-0.0143	-0.0029	0	0	0	0
d_ceq	-0.007	0.0089	0.048	0	0	0	0	0	0
dpi2a	-0.0009	0.013	0.0528	0	0	0	0	0	0
d_shrout	-0.0022	0.0245	0.0646	0	0	0	0	0	0
ivc	-0.0221	0.0003	0.0279	0	0	0	0	0	0
noa	-0.1047	0.0272	0.089	-0.0112	0.0023	0.0218	0	0	0
Profitability									
ato	-0.0307	0.0344	0.0839	-0.0144	0.0042	0.0341	0	0	0
cto	-0.0332	-0.0015	0.0156	0	0	0	0	0	0
d_dgm_dsales	-0.0881	-0.0183	0.0014	0	0	0	0	0	0
eps	-0.1241	-0.0506	-0.0027	0	0.0012	0.0061	0	0	0
ipm	-0.0246	0.021	0.0637	-0.0029	-0.0006	0	0	0	0
pcm	-0.0722	0.0599	0.2135	-0.0058	-0.0012	0	0	0	0
pm	-0.091	0.0146	0.095	-0.0039	-0.0008	0	0	0	0
prof	0.0065	0.0555	0.0986	0	0.001	0.005	0	0	0
rna	-0.0051	0.017	0.0495	0	0	0	0	0	0
roa	-0.1268	0	0.0476	-0.0026	0.0274	0.0837	-0.032	0.0847	0.192
roc	-0.05	-0.0099	0.0017	0	0	0	0	0	0
roe	-0.1602	-0.0065	0.0648	-0.005	0.0012	0.0063	-0.001	0.0017	0.0095
roic	-0.0069	0.0053	0.0277	0	0.0029	0.0147	-0.0023	0.0039	0.022
s2c	-0.0004	0.0008	0.0036	-0.0004	0.0177	0.0745	-0.0119	0.1327	0.3685
Intangibles									
sat	-0.0003	0.0023	0.0083	0	0	0	0	0	0
aoa	-0.0079	0.0515	0.1331	-0.0385	-0.0077	0	0	0	0
ol	-0.2253	0.0003	0.2165	0	0.0048	0.0239	0	0	0
tan	-0.1424	0.016	0.1539	-0.0166	-0.0033	0	0	0	0
oa	-0.1439	-0.0288	0.0013	0	0	0	0	0	0
Value									
a2me	-0.0364	0.0086	0.0504	-0.0614	-0.006	0.0341	-0.1849	0.0799	0.5167
beme	0.0623	0.2132	0.3788	-0.1505	-0.03	0.0003	-0.0026	0.0052	0.0285
c	-0.1002	-0.0413	0.0034	-0.0597	-0.0099	0.0125	0	0	0
c2d	-0.0416	0.0111	0.0666	-0.0022	0.0026	0.0142	-0.0023	0.0038	0.0214
d_so	-0.0242	-0.0069	0.0217	-0.0134	-0.0019	0.0042	0	0	0
debt2p	-0.0051	0.0427	0.1034	-0.0359	-0.0072	0.0001	-0.0006	0.0013	0.0069
e2p	-0.0071	0.007	0.0351	-0.0005	0.0038	0.0184	0	0	0
free_cf	-0.0231	-0.0089	0	0	0	0	0	0	0
ldp	-0.1717	0.111	0.2952	-0.0828	0.0018	0.0877	-0.0077	0.0156	0.0859
nop	-0.0138	-0.0023	0.002	0	0.0001	0.0004	-0.0001	0.0001	0.0006
o2p	-0.0091	0.0191	0.0654	-0.0055	0.0107	0.0308	-0.0043	0.0072	0.0405
q	-0.0546	-0.0121	0.0486	0	0.0008	0.004	0	0	0
sales_g	-0.1718	-0.0397	0.0407	0	0.0002	0.0008	0	0	0
Trading frictions									
size	-0.096	0.2908	0.5376	-0.2342	0.1538	0.7838	-0.1403	0.3553	0.9545
s2p	-0.0087	0.03	0.0818	-0.0051	0.001	0.0096	0	0	0
at	-0.086	0.0146	0.1099	-0.0382	0.0129	0.1138	-0.3795	-0.0229	0.2425
beta	-0.0771	0.0898	0.2876	-0.1851	-0.0769	0.0023	-0.7686	-0.1742	0.053
beta_daily	-0.0364	0.0723	0.1527	-0.0512	-0.0146	0	-0.1407	0.0113	0.1681
dto	-0.0503	0.0189	0.1426	0	0	0	0	0	0
idio_vol	-0.0857	-0.0084	0.0525	0	0	0	0	0	0
lme	-0.002	0.0287	0.0878	-0.0141	0.0124	0.0866	0	0	0
lturnover	-0.1583	0.0207	0.2654	-0.229	-0.0866	0.0045	-0.3438	-0.046	0.1891
rel_to_high_price	0.0017	0.1103	0.4083	0	0.3871	0.9875	0	0	0
ret_max	-0.0369	0.0016	0.0635	-0.0634	-0.0242	0.001	-0.2478	-0.0587	0.0315
spread_mean	-0.0631	0.0349	0.1919	-0.0163	-0.0033	0	0	0	0
std_turn	-0.1392	-0.0124	0.0957	-0.0022	-0.0004	0	0	0	0
std_volume	-0.0034	0.0365	0.0921	-0.0085	-0.0017	0	0	0	0
suv	-0.0284	0.0069	0.0631	0	0	0	0	0	0
total_vol	-0.0278	-0.005	0.0095	-0.0504	-0.0166	0.0017	-0.3772	-0.0713	0.0452
prc	0.122	0.3804	0.5566	-0.048	0.314	0.5003	-0.0068	0.2067	0.6678
ret	0.0231	0.1317	0.3139	-0.0008	0.0213	0.0762	0	0	0
moneyness	-0.0054	0.0518	0.1315	-0.0382	-0.0076	0	0	0	0
constant	-0.7168	-0.025	0.4775	-0.8279	-0.2165	0.4639	-0.8094	-0.023	0.5578

Note: The minimum, mean and maximum of the coefficients of the characteristics of the restricted lasso 3-factor model are shown based on Algorithm 1 with $\lambda = \{0.01, 0.1, 0.2, 0.3, 0.7\}$. Red indicates a low value and green indicates a high value.

Table 17: Coefficients of Γ_β of the lasso 4-factor restricted model.

	factor 1			factor 2			factor 3			factor 4		
Characteristics	Min.	Mean	Max.	Min.	Mean	Max.	Min.	Mean	Max.	Min.	Mean	Max.
Past returns												
cum_return_1_0	-0.0002	0.0321	0.1007	-0.0012	0.0005	0.0046	0	0	0	0	0	0
cum_return_6_2	-0.0155	0.031	0.1158	-0.0016	0.0638	0.2172	-0.1429	0.0171	0.2083	0	0.1127	0.5634
cum_return_12_2	-0.0369	0.033	0.139	0	0.0289	0.1032	0	0	0	0	0	0
cum_return_12_7	-0.0829	0.0182	0.2208	0	0	0	0	0	0	0	0	0
cum_return_36_13	-0.0281	0.0206	0.1141	-0.0003	0.0001	0.0004	0	0	0	0	0	0
investment												
investment	-0.1145	-0.0019	0.1787	-0.0088	0.0031	0.0323	0	0	0	0	0	0
d_ceq	-0.0046	0.0088	0.0374	0	0	0	0	0	0	0	0	0
dpi2a	-0.0093	0.0138	0.0527	0	0	0	0	0	0	0	0	0
d_shrout	-0.1441	0.0571	0.3531	0	0	0	0	0	0	0	0	0
ivc	-0.031	-0.0046	0.0329	0	0	0	0	0	0	0	0	0
noa	-0.1173	0.032	0.1221	-0.007	0.0037	0.0239	-0.0013	0.0004	0.0035	0	0	0
Profitability												
ato	-0.0187	0.034	0.085	-0.0082	0.0062	0.0373	-0.002	0.0007	0.0054	0	0	0
cto	-0.0338	0.0029	0.0356	0	0	0	0	0	0	0	0	0
d_dgm_dsales	-0.1872	0.0114	0.3238	0	0	0	0	0	0	0	0	0
eps	-0.1238	-0.0492	-0.0029	-0.0022	0.0005	0.0033	0	0	0	0	0	0
ipm	-0.0295	0.0222	0.0621	-0.0015	-0.0002	0.001	0	0	0	0	0	0
pcm	-0.0867	0.0694	0.3141	-0.0039	-0.0006	0.0025	0	0	0	0	0	0
pm	-0.1048	0.0115	0.088	-0.0021	-0.0003	0.0014	0	0	0	0	0	0
prof	-0.1134	0.0486	0.1973	0	0.02	0.0686	0	0	0	0	0	0
rna	-0.002	0.0162	0.0493	0	0	0	0	0	0	0	0	0
roa	-0.1209	0.0015	0.0481	-0.005	0.0108	0.0577	-0.0031	0.0839	0.2359	0	0.0097	0.0484
roc	-0.0497	-0.0148	0.012	0	0	0	0	0	0	0	0	0
roe	-0.127	-0.0176	0.0662	-0.0026	0.0011	0.0076	-0.0004	0.0001	0.0011	0	0	0
roic	-0.0078	0.0068	0.0242	0	0	0	0	0	0	0	0	0
s2c	-0.0064	0.0001	0.0048	0	0.0037	0.0183	-0.001	0.0003	0.0027	0	0	0
Intangibles												
sat	-0.0066	0.0029	0.013	0	0	0	0	0	0	0	0	0
aoa	-0.0077	0.0298	0.1333	-0.0332	0.0117	0.1224	0	0	0	0	0	0
ol	-0.2365	0.0352	0.2909	-0.715	-0.0608	0.4553	0	0	0	0	0	0
tan	-0.1026	0.0006	0.1504	-0.0088	0.0031	0.0326	0	0	0	0	0	0
oa	-0.1525	-0.037	0.0042	0	0	0	0	0	0	0	0	0
Value												
a2me	-0.0391	-0.0033	0.0268	-0.0157	0.0366	0.1456	-0.565	-0.0224	0.5081	-0.3356	0.1006	0.578
beme	-0.0768	0.1319	0.3595	-0.1107	0.0836	0.3968	-0.7716	-0.1396	0.1245	-0.0237	0.1111	0.2289
c	-0.1559	-0.0436	0.1024	-0.0522	-0.0093	0.0076	-0.0076	-0.001	0.0028	0	0	0
c2d	-0.0418	0.0126	0.0626	-0.0015	0.0006	0.004	-0.0002	0.0001	0.0006	0	0	0
d_so	-0.029	-0.0067	0.0202	-0.0149	-0.0025	0.0031	-0.0022	-0.0003	0.0008	0	0	0
debt2p	-0.0228	0.0198	0.0714	-0.0316	0.0191	0.1163	-0.0045	0.016	0.0538	-0.024	0.0084	0.0488
e2p	-0.0356	0.0014	0.0351	0	0.0023	0.0115	-0.0006	0.0002	0.0017	0	0	0
free_cf	-0.0235	-0.0101	0	0	0	0	0	0	0	0	0	0
ldp	-0.2754	0.0592	0.2886	-0.0585	0.0305	0.1864	-0.0042	0.0014	0.0114	0	0	0
nop	-0.007	-0.0015	0.0009	0	0	0	0	0	0	0	0	0
o2p	-0.0885	0.0208	0.0828	-0.003	0.006	0.0324	-0.0017	0.0006	0.0047	0	0	0
q	-0.0514	-0.0114	0.0404	-0.0014	0.0003	0.0021	0	0	0	0	0	0
sales_g	-0.2124	-0.0453	0.0429	-0.0007	0.0002	0.0011	0	0	0	0	0	0
Trading frictions												
size	-0.0659	0.1764	0.4366	-0.1779	0.289	0.7962	-0.1239	0.2266	0.7283	-0.3523	0.1824	0.6998
s2p	-0.0131	0.0418	0.1174	-0.5012	-0.0961	0.0369	-0.0004	0.0002	0.0012	0	0	0
at	-0.0865	-0.0082	0.0601	-0.0519	0.0458	0.2115	-0.5053	-0.086	0.0476	-0.4868	-0.1586	0.0237
beta	-0.0732	0.0361	0.2473	-0.1145	-0.0042	0.2841	-0.3433	-0.0679	0.101	-0.4983	-0.0997	0
beta_daily	-0.0339	0.05	0.1736	-0.042	0.0155	0.1565	-0.0002	0	0.0003	0	0	0
dto	-0.0459	0.0304	0.1088	0	0	0	0	0	0	0	0	0
idio_vol	-0.0932	-0.0078	0.0487	0	0	0	0	0	0	0	0	0
lme	-0.0021	0.0316	0.1396	-0.025	0.0508	0.1775	-0.6581	-0.0545	0.4685	-0.6562	0.0151	0.4941
lturnover	-0.1607	-0.0193	0.261	-0.176	-0.0178	0.1464	-0.1288	-0.0297	0.017	0	0.0318	0.1588
rel_to_high_price	-0.0012	0.1155	0.313	-0.0073	0.2582	0.7356	-0.073	0.2102	0.9636	-0.0592	0.2253	0.469
ret_max	-0.0347	-0.009	0.0109	-0.0423	-0.0017	0.0654	-0.1957	-0.0465	0.0173	-0.2078	-0.0416	0
spread_mean	-0.087	0.0203	0.191	-0.0104	0.0036	0.0382	0	0	0	0	0	0
std_turn	-0.145	-0.0208	0.093	-0.0015	-0.0002	0.001	0	0	0	0	0	0
std_volume	-0.0063	0.0409	0.1131	-0.0058	0.0021	0.0215	0	0	0	0	0	0
suv	-0.0357	0.0067	0.0602	0	0	0	0	0	0	0	0	0
total_vol	-0.0263	-0.0033	0.0118	-0.0352	-0.0069	0.0008	-0.0051	-0.0006	0.0019	0	0	0
prc	0.0322	0.3622	0.5567	0.0314	0.2759	0.4972	-0.0999	0.1225	0.3089	-0.0198	0.2156	0.4831
ret	-0.0709	0.1428	0.4776	0.0001	0.0361	0.101	-0.0017	0.0006	0.0045	0	0	0
moneyness	-0.0201	0.0274	0.1222	-0.0296	0.0104	0.109	0	0	0	0	0	0
constant	-0.7235	-0.0798	0.3369	-0.7996	0.0424	0.4547	-0.7761	-0.1557	0.1277	-0.5905	-0.0573	0.2432

Note: The minimum, mean and maximum of the coefficients of the characteristics of the restricted lasso 4-factor model are shown based on Algorithm 1 with $\lambda = \{0.01, 0.1, 0.2, 0.3, 0.7\}$. Red indicates a low value and green indicates a high value.

Table 18: Percent of times the parameter of the characteristic is set to 0 in the lasso restricted model.

Characteristic/k	1	2	3	4	Characteristic/k	1	2	3	4
past returns					**value**				
cum_return_1_0	80	60	60	60	a2me	80	60	0	0
cum_return_6_2	60	60	40	20	beme	60	60	40	0
cum_return_12_2	60	60	60	60	c	40	20	40	40
cum_return_12_7	80	80	80	80	c2d	40	40	20	45
cum_return_36_13	60	80	60	60	d_so	40	40	40	45
investment					debt2p	80	60	40	0
investment	60	60	60	60	e2p	80	40	60	65
d_ceq	80	80	80	80	free_cf	80	80	80	80
dpi2a	80	80	80	80	ldp	20	20	20	40
d_shrout	80	80	80	80	nop	60	100	60	80
ivc	80	80	80	80	o2p	20	20	20	40
noa	40	40	40	40	q	60	60	60	60
profitability					sales_g	60	80	60	60
ato	40	40	40	40	size	20	20	0	0
cto	80	80	80	80	**trading friction**				
d_dgm_dsales	80	80	80	80	s2p	60	40	40	45
eps	60	60	60	60	at	80	20	0	0
ipm	80	60	60	60	beta	0	0	0	20
pcm	60	60	60	60	beta_daily	20	60	20	40
pm	80	60	60	60	dto	80	80	80	80
prof	40	60	60	60	idio_vol	80	60	80	80
rna	80	80	80	80	lme	80	20	40	0
roa	20	20	0	20	lturnover	20	0	0	20
roc	80	80	80	80	rel_to_high_price	80	80	40	0
roe	20	40	20	45	ret_max	20	0	0	20
roic	60	40	60	80	spread_mean	80	60	60	60
s2c	40	40	20	65	std_turn	60	60	60	60
sat	80	80	80	80	std_volume	60	60	60	60
intangibles					suv	80	80	80	80
aoa					total_vol	20	0	0	40
ol	60	60	60	60	prc	0	0	0	0
tan	100	60	60	60	ret	40	40	40	40
oa	80	80	80	85	moneyness	80	60	60	60
					constant	0	0	0	0

Note: Percent of times the parameter is set to zero by the lasso restricted model with k risk factors as defined in Algorithm 1 is displayed. A low percentage (lower or equal to 20% is indicated in green) therefore indicates that the characteristic is not often set to zero and therefore is important in the explanation of returns.

C.2.2 In-sample fit

Table 19: R^2 in-sample based on ridge restricted model.

	Total R^2 individual return ($\boldsymbol{r}_t$)						Predicted R^2 individual return ($\boldsymbol{r}_t$)				
k / λ	0	0.1	0.2	0.3	0.7	k / λ	0	0.1	0.2	0.3	0.7
1	0.1901	0.1864	0.186	0.1856	0.1844	1	0.072	0.0694	0.0689	0.0686	0.0677
2	0.2133	0.2089	0.2078	0.2067	0.2028	2	0.0794	0.0782	0.0777	0.0771	0.0751
3	0.2273	0.2202	0.2182	0.2162	0.2104	3	0.0797	0.0789	0.0781	0.0773	0.0749
4	0.2366	0.2263	0.2238	0.221	0.2133	4	0.0797	0.0788	0.0783	0.0776	0.075
	Total R^2 managed portfolio return ($\boldsymbol{x}_t$)						Predicted R^2 managed portfolio return ($\boldsymbol{x}_t$)				
k / λ	0	0.1	0.2	0.3	0.7	k / λ	0	0.1	0.2	0.3	0.7
1	0.8927	0.8775	0.8765	0.8757	0.8731	1	0.3487	0.3395	0.3386	0.3377	0.3346
2	0.9283	0.9232	0.9233	0.9228	0.9184	2	0.3627	0.3623	0.362	0.3613	0.3574
3	0.943	0.9396	0.9398	0.9393	0.9333	3	0.3635	0.3637	0.3631	0.3621	0.3576
4	0.9487	0.9448	0.9458	0.9469	0.9398	4	0.3629	0.3634	0.3632	0.3626	0.3578

Note: The in-sample R^2 for the restricted ridge model are depicted for different penalty parameters λ. The total R^2 and predictive R^2 are calculated based on Equation (40) and (41) respectively. For ease of comparison a colour scheme is used where a high R^2 is indicated in green and a low R^2 in red.

Table 20: R^2 in-sample for elastic net 1-factor model.

	Total R^2 individual return ($\boldsymbol{r}_t$)						Predicted R^2 individual return ($\boldsymbol{r}_t$)				
ρ / λ	0.01	0.1	0.2	0.3	0.7	ρ / λ	0.01	0.1	0.2	0.3	0.7
0	0.1897	0.1864	0.186	0.1856	0.1844	0	0.0703	0.0694	0.0689	0.0686	0.0677
0.25	0.1855	0.1854	0.1851	0.1848	0.1842	0.25	0.0704	0.0706	0.0706	0.0706	0.0704
0.5	0.1856	0.1851	0.1842	0.1839	0.1838	0.5	0.0705	0.0706	0.0702	0.0696	0.0688
0.75	0.1856	0.1847	0.1841	0.1838	0.1832	0.75	0.0705	0.0703	0.0694	0.0682	0.0669
1	0.1855	0.1843	0.1824	0.1789	0.1744	1	0.0706	0.0686	0.0652	0.0617	0.0583
	Total R^2 managed portfolio return ($\boldsymbol{x}_t$)						Predicted R^2 managed portfolio return ($\boldsymbol{x}_t$)				
ρ / λ	0.01	0.1	0.2	0.3	0.7	ρ / λ	0.01	0.1	0.2	0.3	0.7
0	0.8927	0.8775	0.8765	0.8757	0.8731	0	0.3482	0.3395	0.3386	0.3377	0.3346
0.25	0.8881	0.8879	0.8876	0.8874	0.8871	0.25	0.3484	0.3487	0.3488	0.3488	0.3486
0.5	0.8881	0.8876	0.8872	0.8872	0.8872	0.5	0.3486	0.3488	0.3484	0.3469	0.3449
0.75	0.8881	0.8875	0.8874	0.8869	0.8849	0.75	0.3487	0.3485	0.3466	0.3431	0.339
1	0.888	0.8876	0.8811	0.8679	0.8525	1	0.3491	0.3443	0.3332	0.3204	0.3083

Note: R^2 for the 1 factor restricted model in-sample based on elastic net. The total R^2 and predictive R^2 are calculated based on Equation (40) and (41) respectively. The total R^2 for raw returns decreases when λ increases and when ρ increases. The predicted R^2 mostly decreases when λ increases, however this does not hold for ρ is 0.25 or 0.5. It seems that for these values of ρ an optimum lies between λ 0.1 and 0.3. For ease of comparison a colour scheme is used where a high R^2 is indicated in green and a low R^2 in red.

Table 21: R^2 in-sample for elastic net 2-factor model.

Total R^2 individual return ($\boldsymbol{r}_t$)						Predicted R^2 individual return ($\boldsymbol{r}_t$)					
ρ / λ	0.01	0.1	0.2	0.3	0.7	ρ / λ	0.01	0.1	0.2	0.3	0.7
0	0.2053	0.2056	0.2053	0.2053	0.204	0	0.0758	0.0761	0.0758	0.0759	0.0758
0.25	0.2053	0.2056	0.2053	0.2053	0.2023	0.25	0.0758	0.0761	0.0758	0.0759	0.0755
0.5	0.2055	0.2056	0.2051	0.2042	0.2023	0.5	0.076	0.0761	0.076	0.0758	0.0755
0.75	0.2055	0.2046	0.2051	0.2042	0.1947	0.75	0.076	0.0757	0.076	0.0758	0.0728
1	0.2056	0.2046	0.204	0.2014	0.1947	1	0.0761	0.0757	0.0758	0.0753	0.0728
Total R^2 managed portfolio return ($\boldsymbol{x}_t$)						Predicted R^2 managed portfolio return ($\boldsymbol{x}_t$)					
ρ / λ	0.01	0.1	0.2	0.3	0.7	ρ / λ	0.01	0.1	0.2	0.3	0.7
0	0.9241	0.9243	0.9242	0.9242	0.9228	0	0.3596	0.3597	0.3597	0.3598	0.3595
0.25	0.9241	0.9242	0.9242	0.9242	0.9226	0.25	0.3596	0.3596	0.3597	0.3598	0.3596
0.5	0.9243	0.9242	0.9238	0.9227	0.9226	0.5	0.3597	0.3596	0.3599	0.3596	0.3596
0.75	0.9243	0.9231	0.9238	0.9227	0.9136	0.75	0.3597	0.3589	0.3599	0.3596	0.3548
1	0.9243	0.9231	0.9228	0.9221	0.9136	1	0.3597	0.3589	0.3595	0.3597	0.3548

Note: R^2 for the 2 factor restricted model in-sample based on elastic net. The total R^2 and predictive R^2 are calculated based on Equation (40) and (41) respectively. The total R^2 for raw returns almost always decreases when λ increases and when ρ increases. The predicted R^2 does not differ substantially between the different parameters, however we do observe that higher values are centered more around ρ 0.25 and 0.5 and that an increase in λ yields more often than not a decrease in predicted R^2. For ease of comparison a colour scheme is used where a high R^2 is indicated in green and a low R^2 in red.

C.2.3 Out-of-sample analysis

Table 22: R^2 out-of-sample based on ridge restricted model with λ.

	Total R^2					Predicted R^2				
	Individual returns ($\boldsymbol{r}_t$)					Individual returns ($\boldsymbol{r}_t$)				
k	Basic	$\lambda = 0.1$	$\lambda = 0.2$	$\lambda = 0.3$	$\lambda = 0.7$	Basic	$\lambda = 0.1$	$\lambda = 0.2$	$\lambda = 0.3$	$\lambda = 0.7$
1	0.1338	0.1337	0.1336	0.1336	0.1335	0.075	0.0749	0.0747	0.0746	0.0743
2	0.1466	0.1468	0.147	0.147	0.1472	0.0823	0.0825	0.0825	0.0825	0.0825
3	0.1536	0.1534	0.1532	0.1531	0.1528	0.0827	0.0829	0.0829	0.0829	0.0827
4	0.1567	0.1565	0.1571	0.1572	0.1565	0.0832	0.0832	0.0835	0.0837	0.0835
	Managed portfolio return ($\boldsymbol{x}_t$)					Managed portfolio return ($\boldsymbol{x}_t$)				
k	Basic	$\lambda = 0.1$	$\lambda = 0.2$	$\lambda = 0.3$	$\lambda = 0.7$	Basic	$\lambda = 0.1$	$\lambda = 0.2$	$\lambda = 0.3$	$\lambda = 0.7$
1	0.8517	0.8514	0.8512	0.851	0.8506	0.4531	0.4524	0.4519	0.4516	0.4506
2	0.8926	0.893	0.8933	0.8936	0.8942	0.4829	0.4825	0.4824	0.4824	0.4823
3	0.9009	0.9013	0.9015	0.9016	0.902	0.4851	0.4848	0.4846	0.4844	0.4839
4	0.9062	0.9075	0.9071	0.9101	0.9104	0.4881	0.4876	0.4881	0.4874	0.4867

Note: The out-of-sample R^2 for the restricted ridge model are depicted for different penalty parameters λ. The total R^2 and predictive R^2 are calculated based on Equation (40) and (41) that are slightly adjusted as described in Section 2.8. For ease of comparison a colour scheme is used where a high R^2 is indicated in green and a low R^2 in red.

Table 23: R^2 out-of-sample sharpe ratios of restricted ridge model t=90.

λ	k=1		k=2			k=3				k=4				
	tangency	f1	tangency	f1	f2	tangency	f1	f2	f3	tangency	f1	f2	f3	f5
0.01	1.1267	1.1267	1.7500	1.5991	0.4616	1.7173	1.4052	0.5726	0.5829	1.9257	1.0245	0.7437	0.6681	0.3121
0.1	1.1252	1.1252	1.7561	1.5779	0.4727	1.6988	1.5188	0.2036	0.5072	1.9383	1.2556	0.4859	0.5633	0.2637
0.2	1.1238	1.1238	1.7610	1.5674	0.4762	1.8022	1.2765	0.5042	0.3875	2.0051	1.2593	0.4647	0.3241	0.3821
0.3	1.1226	1.1226	1.7671	1.5740	0.4559	1.7325	1.1481	0.8506	0.0673	1.9572	1.1350	0.5324	0.5743	0.2389
0.7	1.1189	1.1189	1.8108	1.6646	0.0523	1.7381	1.1318	0.3288	0.6985	1.8092	1.1379	0.2863	0.2805	0.7292

Note: The out-of-sample Sharpe ratios of the tangency portfolios based on the restricted ridge model defined in Equation (18) are shown. The table shows that when more factors are considered the Sharpe ratio of the tangency portfolio often increases. Moreover, for the three and four factor model a λ of around 0.2 is optimal, whereas for 2 factors a high penalty parameter is better.

Table 24: Out-of-sample R^2 for individual returns ($\boldsymbol{r}_t$) comparison.

t=90	Restricted							Unrestricted						
	Total R^2 individual returns ($\boldsymbol{r}_t$)							Total R^2 individual returns ($\boldsymbol{r}_t$)						
k	Basic	P-value	λ=0.01	λ=0.1	λ=0.2	λ=0.3	λ=0.7	Basic	P-value	λ=0.01	λ=0.1	λ=0.2	λ=0.3	λ=0.7
1	0.1338	0.128	0.1338	0.1342	0.1327	0.1309	0.1273	0.0843	0.1298	0.0843	0.1175	0.1098	0.1317	0.1349
2	0.1466	0.141	0.1466	0.1476	0.1479	0.1457	0.1385	0.0809	0.1351	0.0796	0.1226	0.0701	0.103	0.1152
3	0.1536	0.1478	0.1536	0.1522	0.1531	0.1508	0.1498	0.0766	0.139	0.0766	0.1127	0.1397	0.1487	0.1481
4	0.1567	0.1541	0.1575	0.1562	0.1571	0.1574	0.1524	0.0788	0.1517	0.0783	0.1139	0.1429	0.1548	0.1479
	Predicted R^2 individual returns ($\boldsymbol{r}_t$)							Predicted R^2 individual returns ($\boldsymbol{r}_t$)						
k	Basic	P-value	λ=0.01	λ=0.1	λ=0.2	λ=0.3	λ=0.7	Basic	P-value	λ=0.01	λ=0.1	λ=0.2	λ=0.3	λ=0.7
1	0.075	0.068	0.075	0.0745	0.0732	0.0718	0.0703	0.0837	0.0732	0.084	0.0842	0.0843	0.0827	0.0768
2	0.0823	0.077	0.0823	0.0827	0.0827	0.0801	0.075	0.0834	0.0783	0.0835	0.0835	0.0844	0.0828	0.078
3	0.0827	0.0794	0.0827	0.0834	0.0839	0.0826	0.081	0.0833	0.0803	0.0833	0.084	0.0849	0.0831	0.0818
4	0.0832	0.0837	0.0831	0.0835	0.0842	0.082	0.0792	0.0834	0.0848	0.0834	0.0838	0.0847	0.082	0.0793

t=75	Restricted							Unrestricted						
	Total R^2 individual returns ($\boldsymbol{r}_t$)							Total R^2 individual returns ($\boldsymbol{r}_t$)						
k	Basic	P-value	λ=0.01	λ=0.1	λ=0.2	λ=0.3	λ=0.7	Basic	P-value	λ=0.01	λ=0.1	λ=0.2	λ=0.3	λ=0.7
1	0.1405	0.1384	0.1406	0.1421	0.1406	0.1401	0.1375	0.0897	0.1042	0.0896	0.1216	0.1118	0.1383	0.1439
2	0.1545	0.1487	0.1547	0.1568	0.154	0.1516	0.1454	0.0892	0.1152	0.0877	0.1316	0.0837	0.113	0.1248
3	0.1645	0.1618	0.1645	0.1667	0.1662	0.1624	0.161	0.0887	0.1522	0.0887	0.1196	0.1489	0.1585	0.1589
4	0.1687	0.1656	0.1688	0.1697	0.17	0.1693	0.1626	0.0863	0.1612	0.0872	0.1254	0.1501	0.1643	0.1573
	Predicted R^2 individual returns ($\boldsymbol{r}_t$)							Predicted R^2 individual returns ($\boldsymbol{r}_t$)						
k	Basic	P-value	λ=0.01	λ=0.1	λ=0.2	λ=0.3	λ=0.7	Basic	P-value	λ=0.01	λ=0.1	λ=0.2	λ=0.3	λ=0.7
1	0.0418	0.0367	0.0418	0.0417	0.0408	0.0398	0.0386	0.0494	0.0428	0.0494	0.0497	0.0493	0.0484	0.0444
2	0.0483	0.0433	0.0483	0.049	0.0476	0.0455	0.0436	0.0494	0.0443	0.0495	0.0498	0.0497	0.0483	0.0453
3	0.0487	0.0454	0.0487	0.0492	0.0488	0.0484	0.0472	0.0494	0.0459	0.0494	0.05	0.0497	0.0493	0.0483
4	0.0486	0.0492	0.0489	0.0493	0.0492	0.0478	0.0458	0.0492	0.0501	0.0494	0.0499	0.0496	0.0478	0.046

Note: The total R^2 and predictive R^2 are calculated based on Equation (40) and (41) that are slightly adjusted as described in Section 2.8. Three different methods are utilized for comparison purpose. Basic represents the normal IPCA model based on all characteristics, the p-value represents the model only including characteristics that have shown to be statistically significant at the 10% level, and the last models depicted by λ are based on the lasso method. Moreover, the $t = x$ indicates that x is the start of the out-of-sample sample. For ease of comparison a colour scheme is used where a high R^2 is indicated in green and a low R^2 in red.

From this table we can conclude that (1) a restricted model that is more parsimonious based on lasso does often increase out-of-sample fit ($\lambda \in \{0.1, 0.2\}$, but excluding too many characteristics often yields lower performance ($\lambda = 0.7$). Moreover, (2) the unrestricted model shows that including all characteristics is really affecting the performance negatively, a reduced format based on both lasso as well as p-value technique is increasing the fit substantially.

Table 25: Out-of-sample R^2 for managed portfolio return ($\boldsymbol{x}_t$) comparison.

t=90	Restricted							Unrestricted						
	Total R^2 managed portfolio return ($\boldsymbol{x}_t$)							Total R^2 managed portfolio return ($\boldsymbol{x}_t$)						
k	Basic	P-value	$\lambda = 0.01$	$\lambda = 0.1$	$\lambda = 0.2$	$\lambda = 0.3$	$\lambda = 0.7$	Basic	P-value	$\lambda = 0.01$	$\lambda = 0.1$	$\lambda = 0.2$	$\lambda = 0.3$	$\lambda = 0.7$
1	0.8517	0.9698	0.8524	0.9076	0.9342	0.9463	0.9776	0.5279	0.9822	0.5241	0.7442	0.7033	0.8934	0.9878
2	0.8926	0.9682	0.8926	0.9267	0.9566	0.9708	0.991	0.4715	0.9698	0.4605	0.7578	0.3916	0.6419	0.7946
3	0.9009	0.9643	0.9009	0.9297	0.9604	0.9702	0.9834	0.4277	0.9392	0.4277	0.6973	0.8916	0.965	0.9791
4	0.9062	0.9543	0.9078	0.9339	0.9578	0.9851	0.9913	0.4208	0.9441	0.4183	0.682	0.8891	0.9822	0.9818
	Predicted R^2 managed portfolio return ($\boldsymbol{x}_t$)							Predicted R^2 managed portfolio return ($\boldsymbol{x}_t$)						
k	Basic	P-value	$\lambda = 0.01$	$\lambda = 0.1$	$\lambda = 0.2$	$\lambda = 0.3$	$\lambda = 0.7$	Basic	P-value	$\lambda = 0.01$	$\lambda = 0.1$	$\lambda = 0.2$	$\lambda = 0.3$	$\lambda = 0.7$
1	0.4531	0.441	0.4517	0.452	0.4661	0.4717	0.4562	0.4923	0.4716	0.4924	0.491	0.5049	0.5105	0.479
2	0.4829	0.4772	0.4818	0.4937	0.5036	0.5067	0.4869	0.49	0.479	0.4895	0.5004	0.5093	0.5154	0.4941
3	0.4851	0.4842	0.4851	0.4959	0.5035	0.5012	0.4943	0.4899	0.4782	0.4899	0.4999	0.507	0.5035	0.4962
4	0.4881	0.4803	0.4883	0.4962	0.4992	0.4921	0.4835	0.4897	0.4847	0.4898	0.4978	0.5005	0.4933	0.4846

t=75	Restricted							Unrestricted						
	Total R^2 managed portfolio return ($\boldsymbol{x}_t$)							managed portfolio return ($\boldsymbol{x}_t$) total R^2						
k	Basic	P-value	$\lambda = 0.01$	$\lambda = 0.1$	$\lambda = 0.2$	$\lambda = 0.3$	$\lambda = 0.7$	Basic	P-value	$\lambda = 0.01$	$\lambda = 0.1$	$\lambda = 0.2$	$\lambda = 0.3$	$\lambda = 0.7$
1	0.8686	0.9782	0.8698	0.9219	0.9412	0.9539	0.98	0.5977	0.7023	0.5945	0.7702	0.7364	0.907	0.9859
2	0.9034	0.9712	0.9034	0.9341	0.9596	0.9696	0.9807	0.5683	0.8384	0.5585	0.7958	0.5519	0.7294	0.8294
3	0.921	0.974	0.921	0.9462	0.9706	0.9792	0.988	0.5508	0.9482	0.5508	0.7335	0.9068	0.9676	0.9847
4	0.9251	0.9688	0.9258	0.9494	0.9693	0.9879	0.994	0.5286	0.9608	0.5307	0.7521	0.8923	0.9747	0.9851
	Predicted R^2 managed portfolio return ($\boldsymbol{x}_t$)							Predicted R^2 managed portfolio return ($\boldsymbol{x}_t$)						
k	Basic	P-value	$\lambda = 0.01$	$\lambda = 0.1$	$\lambda = 0.2$	$\lambda = 0.3$	$\lambda = 0.7$	Basic	P-value	$\lambda = 0.01$	$\lambda = 0.1$	$\lambda = 0.2$	$\lambda = 0.3$	$\lambda = 0.7$
1	0.3598	0.3379	0.3579	0.3504	0.3617	0.3647	0.3473	0.3884	0.354	0.387	0.3755	0.3862	0.3888	0.3613
2	0.3813	0.3592	0.3796	0.3821	0.3857	0.386	0.371	0.387	0.3632	0.3854	0.3872	0.3906	0.3926	0.3751
3	0.3844	0.366	0.3844	0.3845	0.3858	0.3838	0.3743	0.3872	0.3678	0.3872	0.3869	0.3873	0.3859	0.3761
4	0.3853	0.3685	0.3858	0.3842	0.3828	0.371	0.3651	0.3868	0.371	0.3867	0.3867	0.3839	0.3713	0.3655

Note: The total R^2 and predictive R^2 are calculated based on Equation (40) and (41) that are slightly adjusted as described in Section 2.8. Three different methods are utilized for comparison purpose. Basic represents the normal IPCA model based on all characteristics, the p-value represents the model only including characteristics that have shown to be statistically significant at the 10% level, and the last models depicted by λ are based on the lasso method. Moreover, the $t = x$ indicates that x is the start of the out-of-sample sample. For ease of comparison a colour scheme is used where a high R^2 is indicated in green and a low R^2 in red.
From this table we can conclude that (1) a model that is more parsimonious based on either lasso or p-values does increase out-of-sample fit for returns. This is partly due to portfolios being constructed based on the most important variables that can be explained best. Moreover, (2) expected returns often yield better fit when a lasso technique with λ parameters between 0.1 and 0.3 is considered.

Table 26: Sharpe ratio of basic and reduced (based on 2-step procedure) restricted models.

Sharpe ratio with start date t=90							
k	Basic	p-value	λ=0.01	λ=0.1	λ=0.2	λ=0.3	λ=0.7
1	1.130	1.0584	1.1248	1.1129	1.1042	1.0983	1.1068
2	1.76	1.5509	1.7638	1.7647	1.7401	1.7298	1.6025
3	1.74	1.6004	1.7389	1.8194	1.7449	1.7111	1.6516
4	1.940	1.7826	1.932	1.8316	1.7649	2.0341	1.9888

Sharpe ratio with start date t=75							
k	Basic	p-value	λ=0.01	λ=0.1	λ=0.2	λ= 0.3	λ= 0.7
1	0.8225	0.7822	0.8238	0.8243	0.8201	0.8135	0.8085
2	1.4798	1.339	1.4779	1.4864	1.4215	1.3724	1.3864
3	1.477	1.3883	1.477	1.4796	1.3673	1.4849	1.472
4	1.613	1.4657	1.5887	1.4599	1.4408	1.6889	1.5503

Note: The Sharpe ratios for a start date $t = 90$ and $t = 75$ are displayed for the basic restricted and their reduced models based on the two-step method. The basic column represents the Sharpe ratios based on the restricted model, the p-value column based on only characteristics that were significant at the 10% level, and $\lambda = x$ columns are based on characteristics that are not set to 0 when penalty parameter of x is used. For ease of comparison a colour scheme is used where a high Sharpe ratio is indicated in green and a low Sharpe ratio in red.
The table shows that for 1 or 2 risk factors the reduced model based on a lasso procedure can beat the basic model for $\lambda = 0.01$ or $\lambda = 0.1$ whereas for more risk factors the reduced model sometimes outperforms the basic for varying lambda but with a focus on relatively higher lambdas. Note that the model that is reduced based on p-values does not perform well, probably because too few characteristics are included in the model.

C.3 Robust IPCA models

C.3.1 Out-of-sample analysis

Table 27: Out-of-sample R^2 comparison between basic and robust methods.

Restricted (t=90) Total R^2 individual returns ($\boldsymbol{r}_t$)			
k	Basic	Huber	WLS
1	0.1338	0.1306	0.1286
2	0.1466	0.1359	0.1456
3	0.1536	0.1507	0.1514
4	0.1567	0.1544	0.1544

Restricted (t=90) Total R^2 managed portfolio return ($\boldsymbol{x}_t$)			
k	Basic	Huber	WLS
1	0.8517	0.8324	0.8455
2	0.8926	0.8447	0.8861
3	0.9009	0.8987	0.9028
4	0.9062	0.8994	0.9051

Restricted (t=90) Predicted R^2 individual returns ($\boldsymbol{r}_t$)			
k	Basic	Huber	WLS
1	0.075	0.0631	0.0729
2	0.0823	0.0643	0.0799
3	0.0827	0.0701	0.0805
4	0.0832	0.0711	0.0813

Restricted (t=90) Predicted R^2 managed portfolio return ($\boldsymbol{x}_t$)			
k	Basic	Huber	WLS
1	0.4531	0.3853	0.4678
2	0.4829	0.3909	0.4751
3	0.4851	0.4181	0.4717
4	0.4881	0.4236	0.4731

Restricted (t=75) Raw returns			
k	Basic	Huber	WLS
1	0.1405	0.1411	0.1256
2	0.1545	0.1459	0.1536
3	0.1645	0.1592	0.1629
4	0.1687	0.164	0.1696

Restricted (t=75) Portfolio returns			
k	Basic	Huber	WLS
1	0.8686	0.8649	0.8387
2	0.9034	0.8754	0.9029
3	0.921	0.9167	0.9215
4	0.9251	0.919	0.9309

Restricted (t=75) Raw predicted returns			
k	Basic	Huber	WLS
1	0.0418	0.0285	0.0401
2	0.0483	0.0298	0.0442
3	0.0487	0.0348	0.0464
4	0.0486	0.0358	0.0472

Restricted (t=75) Portfolio predicted returns			
k	Basic	Huber	WLS
1	0.3598	0.2895	0.3576
2	0.3813	0.2941	0.3738
3	0.3844	0.3152	0.3756
4	0.3853	0.3197	0.3762

Note: The total R^2 and predictive R^2 are calculated based on Equation (40) and (41) that are slightly adjusted as described in Section 2.8. Three different methods are utilized for comparison purpose. Basic represents the normal IPCA model based on all characteristics, the two other methods, Huber and Ridge, are the robust counterpart. Moreover, the $t = x$ indicates that x is the start of the out-of-sample sample. For ease of comparison a colour scheme is used where a high R^2 is indicated in green and a low R^2 in red.
From this table we can conclude that (1) robustness based on Huber does not improve fit out-of-sample for neither returns nor expected returns. Moreover (2) the WLS does outperform the basic model very infrequently.

Table 28: Sharpe ratio comparison robust methods for restricted model.

Restricted model t=90				Restricted model t=75			
k	basic	WLS	Huber	k	basic	WLS	Huber
1	1.13	1.1461	1.0955	1	0.8225	0.938	0.8085
2	1.76	1.5167	1.4298	2	1.4798	1.2944	1.1495
3	1.74	1.6782	1.7327	3	1.477	1.2758	1.4325
4	1.94	1.5268	1.5764	4	1.613	1.4161	1.2943

Note: The Sharpe ratio displayed are based on the tangency portfolios of the restricted OLS, Huber and WLS IPCA model with k systematic factors as defined in Equations (11) and Algorithms 2 and 3. For the calculation of the tangency return factors Algorithm 4 is used. Moreover, the $t = x$ indicates that x is the start of the out-of-sample sample. For ease of comparison a colour scheme is used where a high Sharpe ratio is indicated in green and a low Sharpe ratio in red.