Machine Learning in Asset Pricing: Expanding Multi-factor Models

Copenhagen Business School
by
Johann David Grøndahl (111291) & Nikolaj Bo Andersen (111984)

A master thesis submitted for the degree of MSc in Economics and Business Administration - Finance and Investments

Date of submission: May 17th 2021
Supervisor: Johan Stax Jakobsen
Number of characters: 206,027
Number of pages: 96
Abstract

Machine learning in asset pricing is an extremely popular subject in the current finance literature, with several novel publications applying a variety of machine learning methods in order to challenge conventional asset pricing methods. A pioneering article in the field of machine learning, *Shrinking the cross-section*, challenging the conventional, sparse multi-factor models, were published by Kozak, Santosh & Nagel in 2020. We replicate this newly proposed dual-penalty estimator method for pricing the cross-section of stock returns, and subsequently test it in two high-dimensional data settings consisting of previously proven anomalies and a vast number of firm-specific financial ratios. The resulting factor models are based on the *stochastic discount factor* approach for asset pricing, which is the preferred approach in modern financial literature.

Consistent with the findings of Kozak *et al.* (2020) it is evidently futile to create a sparse characteristics-based factor model sufficiently spanning the *stochastic discount factor* in order to achieve high out of sample performance in explaining cross-sections of stock returns. The best out-of-sample performing model originates from the $L_2$-only shrinkage estimator, summarizing the pricing information contained in a large number of characteristics-based factors. Testing the candidate factor models by construction of mean-variance efficient portfolios shows that the $L_2$-only estimator significantly outperforms the Fama and French (2016) six-factor model. As such, this paper verifies the conclusion of Kozak *et al.* (2020) that the six-factor model of Fama and French (2016) leaves much of the cross-section of stock returns unexplained.

Finally, rotating the factor returns into the space of principal components (PC) enables the possibility of creating PC-sparse factor models sufficiently spanning the *stochastic discount factor*, yielding good out-of-sample performance, as they perform uniformly better than characteristics-sparse models.
Acknowledgements

We would like to express our sincere gratitude to our supervisor, Johan Stax Jakobsen, for his valuable input, guidance, and supervisory contributions throughout the thesis process. In addition, we would like to thank Andreas Pilgaard and Oliver Alexander Press for their helpful comments and inspiring conversations.
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 - Analysis</td>
<td>51</td>
</tr>
<tr>
<td>6.1 - Findings from Kozak et al. (2020)</td>
<td>51</td>
</tr>
<tr>
<td>6.2 - Analysis of dual-penalty estimator method and SDF-coefficients</td>
<td>51</td>
</tr>
<tr>
<td>6.2.1 Data analysis</td>
<td>52</td>
</tr>
<tr>
<td>6.2.2 - 25 Fama &amp; French portfolios</td>
<td>56</td>
</tr>
<tr>
<td>6.2.3 - 50 anomaly portfolios</td>
<td>61</td>
</tr>
<tr>
<td>6.2.4 WRDS Financial ratios (WFR)</td>
<td>65</td>
</tr>
<tr>
<td>6.2.5 Largest SDF-factors</td>
<td>68</td>
</tr>
<tr>
<td>6.2.6 Subsection conclusion</td>
<td>74</td>
</tr>
<tr>
<td>6.3 Asset pricing implication through portfolio optimization</td>
<td>75</td>
</tr>
<tr>
<td>6.3.1 Model setup</td>
<td>75</td>
</tr>
<tr>
<td>6.3.2 Results</td>
<td>78</td>
</tr>
<tr>
<td>6.3.3 Subsection conclusion</td>
<td>82</td>
</tr>
<tr>
<td>7. Discussion</td>
<td>83</td>
</tr>
<tr>
<td>7.1 Deviation compared to the findings in KNS.</td>
<td>83</td>
</tr>
<tr>
<td>7.1.1 Summary of deviations</td>
<td>83</td>
</tr>
<tr>
<td>7.1.2 Potential sources of deviation</td>
<td>84</td>
</tr>
<tr>
<td>7.1.3 Simulation</td>
<td>86</td>
</tr>
<tr>
<td>7.1.4 Subsection summary</td>
<td>90</td>
</tr>
<tr>
<td>7.2 Discussion of methodology</td>
<td>90</td>
</tr>
<tr>
<td>7.2.1 Biased beta estimates</td>
<td>91</td>
</tr>
<tr>
<td>7.2.2 Data-snooping</td>
<td>92</td>
</tr>
<tr>
<td>7.2.3 Alternative selection of basis assets</td>
<td>93</td>
</tr>
<tr>
<td>7.2.4 Subsection summary</td>
<td>94</td>
</tr>
<tr>
<td>7. Conclusion</td>
<td>95</td>
</tr>
<tr>
<td>8. Bibliography</td>
<td>97</td>
</tr>
<tr>
<td>9. Appendix</td>
<td>101</td>
</tr>
</tbody>
</table>
1. Introduction and motivation

In 2019 Forbes Magazine ran an article called *Data is the new oil* ¹, insinuating that data has become the world’s most valuable resource, as a result of the massive technological evolution in the past decades. The concept behind the *Data is the new oil*-statement is that just like oil, raw data is not valuable in and of itself, but, rather, the value is created when it is gathered completely and accurately, and more importantly processed correctly. When properly refined, data and information become valuable tools in almost every imaginable decision-making processes. The need for improved processing of data has motivated technological evolution in the fields of data science. Particularly, the field of machine learning techniques has become essential for the improvement of data processing. Machine learning is an advanced statistical learning method, in which researchers utilize the ever-growing power of computers to handle vast numbers of iterations.

The exponentially growing amount of data, increasing the need for data-processing, has huge implications for many academic fields, including the field of finance. Investors in financial markets are faced with an abundance of potentially value-relevant information from a wide variety of different sources. In such, data-rich high-dimensional environments, techniques from the rapidly advancing field of machine learning are well-equipped for solving prediction problems. Accordingly, machine learning methods are quickly becoming part of the ‘toolkit’ in asset pricing research and quantitative investing. Notably, asset pricing problems are significantly different from the settings for which the machine learning techniques were originally developed, why researchers need to adapt machine learning methods to accommodate the specific conditions in asset pricing applications. With the buzzing empirical activity currently surrounding machine learning in asset pricing, we find this subject highly relevant for further investigation in our thesis. As such, we are motivated to explore asset pricing methods, more specifically factor models, drawing on machine learning techniques. With the application of machine learning methods, we initially sought to improve out of sample performance of factor models by accommodating potentially far more explanatory variables, than the conventional *work-horses* of multi-factor asset pricing, such as the four-factor model of Hou, Xue and Zhang (2015) or the five- and six-factor model of Fama and French (2015, 2016)².

---

¹ The article builds on a publication by the Economist in 2017 called *The world’s most valuable resource is no longer oil, but data.*
² These two factor models represent a response to the failure of the traditional multi-factor models such as the three-factor model of Fama and French (1993) and four-factor model of Carhart (1997)
One of the most recent publications within the field of machine learning in asset pricing is *Shrinking the cross-section* by Serhiy Kozak, Stefan Nagel & Shrihari Santosh (2020) (‘KNS’). They propose and evaluate a dual-estimator method for the estimation of coefficients in stochastic discount factor-based factor models, which successfully handles numerous factors from a high-dimensional setting. Thus, the focus of this publication aligns with our initial interest in application of machine learning methods in asset pricing.

As machine learning in asset pricing is currently a very popular subject, the literature in recent years have shown a variety of possible applications of machine learning methods. The incentive to find new valuation- and asset pricing techniques and strategies are high, both within academia (publications and tenure) and for practitioners in asset management (a marketable story and a larger paycheck). This raises significant data-snooping concerns. Harvey, Liu and Zhu (2015) argue that ‘most claimed research findings in financial economics are likely false’. Although this statement is quite controversial, it motivates the need for replication of empirical results, to evaluate the robustness of given theories or, in this case, asset pricing methods. Replication papers, despite their significant contribution to the literature, are often not published in the most highly rated financial journals, due to their less appealing nature (Alm & Reed, 2015). Alm et al. (2015) highlights the importance of replicating research results, as one of the main principles of the scientific method. Furthermore, in times of increasing retractions and frequent instances of scientific fraud and misconduct, scientific quality assurance mechanisms are proving ever more important (Mueller-Langer et al., 2019). Consequently, formal published replication studies and informal replication are regarded as highly important post-publication quality checks.

As such, this thesis contributes to the novel and highly relevant field of *machine learning in asset pricing* by replicating and investigating the findings and conclusions of the newly proposed dual-penalty estimation method presented by Kozak *et al.* (2020) in *Shrinking the cross-section*. Furthermore, absent the limitations of the relatively strict page-count of a journal-paper, the proposed dual-penalty estimator is outlined in a more descriptive and easier interpretable way, consequently making the method more accessible than in *KNS*.

In the following section we outline the problem statement investigated in this paper.

---

3 For the remainder of this paper, *KNS* refers to the main article *Shrinking the cross-section* (Kozak *et al.*, 2020).

4 For an in-depth explanation of the phenomenon please see section 7.2.3.
Problem statement

Motivated by the increasing use of machine learning for asset pricing applications in the financial literature, we seek to define an SDF-based factor model that successfully handles a high-dimensional problem with many, potentially relevant, pricing factors, based on the dual-penalty estimator proposed by Kozak et al. (2020). Subsequently, we seek to investigate and validate the findings of Kozak et al. (2020), by testing and comparing the out of sample pricing performance of this proposed method.

2. Scope and assumptions

In order to define the scope of the analysis conducted in this paper, some delimitations and fundamental assumptions need to be specified. As the scope of the paper is to replicate the findings in KNS we limit ourselves to characterizing the cross-section of stock returns, rather than identifying asset specific investment strategies. As such, this paper does not propose a directly applicable investment strategy for a practical setting. Furthermore, transaction costs are not accounted for in the empirical assessment of SDF-implied MVE portfolios in section 6.3.

As this paper considers the stochastic discount factor approach for asset pricing, only a limited number of strict assumptions need to be imposed. Thus, we assume that the law of on price is not violated and that investors have no arbitrage opportunities.5

3. Machine learning in Finance

As previously highlighted, the growing technological evolution has increased the attention on machine learning in asset pricing among practitioners and researchers within the field of finance. However, machine learning in general was formerly regarded as a “black-box” framework surrounded with much uncertainty, why many initially refrained from using it (Dixon, Halperin and Bilokon, 2020). Prado (2019) perceives this as a misconception, and argues that the field of machine learning is, in fact, applicable in verifying a broad variety of scientific theories. He considers machine learning as highly advantageous in determination of significant explanatory or predictive variables, as well as in visualization of complex high-dimensional data. Dixon et al. (2020) further amplifies

---

5 The implications of the SDF-approach for asset pricing and the stated assumptions are explained in detail in section 4.3.1

6 In science and computing a black box refers to a system that can be viewed in terms of the inputs and outputs, absent any knowledge of its internal workings.
this argument, stating that advances in the machine learning techniques mitigate the shortcomings of classic econometric models through superior outlier detection, feature selection, regression etc. in complex and high-dimensional data settings. All of which, are widely applied in financial asset pricing. Particularly, many classic multi-factor regression models and feature selection techniques have been applied historically in asset pricing, but with increasing amounts of available data, the complexity of the data has increased correspondingly. As such, the demand for machine learning techniques is exponentially progressing.

In essence, machine learning can be divided into supervised learning, unsupervised learning and dimensional reduction techniques (Dixon et al., 2020).

Supervised learning covers the task of predicting the relationship between features vectors, $\mathbf{x}_1 \ldots \mathbf{x}_n$, and their corresponding responses $y_1 \ldots y_n$. As such, supervised learning resembles classic regression techniques.

In unsupervised learning one seeks to characterize the structure of an unlabeled data set by grouping similar observations or defining patterns. Unsupervised learning includes cluster analysis, in which one seeks to group similar observations in clusters, minimizing the ‘within-cluster’-variation.

Dimensional reduction techniques attempt to present high-dimensional data in a lower, more interpretable version, projecting the structure of data into fewer dimensions. This technique is used for the purpose of visualizing high-dimensional data.

In this paper, we apply a specific regression type supervised learning technique and, additionally, a dimensional reduction technique, why we limit the following literature review to supervised learning and dimensional reduction techniques applied in asset pricing.

Li and Tam (2018) consider several supervised learning techniques in predicting returns on the stock market of mainland China. Particularly, they apply decision trees, support vector machines, and two types of neural networks theories. Decision Trees are supervised learning techniques used for predicting the value of a response variable through learning decision rules, derived from a set of features. Support vector machines are models that apply a certain algorithm to define hyperplanes, classifying the underlying data for classification of regression purposes. Li et al. (2018) find

---

7 ‘Features’ and ‘Responses’ corresponds to, what is commonly known as ‘independent variables’ and ‘dependent variables’, respectively.  
8 Hyperplanes are planes separating multidimensional observations.
hyperplanes with a ‘kernel’ function which projects the inputs into a high-dimensional feature space, where the support vector machine can efficiently solve non-linear problems. Neural networks models are composed of an input layer, several hidden layers and an output layer. The input layer receives a signal upon which the output layer makes a prediction. Li et al. (2018) find that neural networks models work well in the fluctuating market conditions, while all the analyzed machine learning models are effective in bull markets, and finally in order to avoid great loss in a market crash, support vector machines are most effective.

However, decision trees are found incredibly useful in asset pricing by other financial empiricists. In the publication, Forest Through the Trees, Bryzgalova, Pelger and Zhu (2019) apply decision trees in the construction of an asset pricing stochastic discount factor. They apply decision trees in defining basis assets that captures complex information from a set of stock characteristics. Bryzgalova et al. (2019) provides a new perspective on conventional mean-variance optimization by applying an elastic net penalty in solving the Markowitz optimization problem. For now, we highlight that elastic net is a penalized regression method in which coefficients are penalized through two penalty terms. Additional research has also been performed, using penalized regressors, to investigate the impact of characteristics on returns. Freyberger, Neuhierl, and Weber (2019) estimate conditional expected returns as a function of a high-dimensional feature space of characteristics. They use a reduced form elastic net regression, called the LASSO, where only one penalty term is included.

As such, the method proposed by KNS maps into the literature on machine learning in finance, by applying a dual-penalty regression method, similar to that of Bryzgalova et al. (2019) and extending the application by implementing the method on dimensionally reduced variables in the form of principal components. We highlight that none of the above-mentioned publications dates back further than 2018, attesting to the 'state of the art'-nature of the research within this field.

In the following section we outline the theoretical framework underlying this thesis.

---

9 The mathematics behind support vector machines is highly complex, why further explanation lies beyond the scope of this paper.
10 We outline the Markowitz portfolio optimization process thoroughly in section 4.4.
11 We outline the theory behind penalized regressors in section 5.1.2.
12 A further description of principal components can be located in section 4.3.6.
4. Theoretical framework

Throughout the history of financial literature and particularly, literature on asset allocation, a highly engaged subject has been optimal portfolio construction. Portfolio optimization and asset allocation are deeply entwined with the theory on asset pricing. Within asset pricing literature, the conventional asset pricing methods can be divided into two overall paradigms relying on different underlying assumptions; the equilibrium state models, and the no arbitrage models. Among the most popular asset pricing theories, drawing from the no-arbitrage assumption, are State Preference Theory from Arrow and Debreu (1954), Option Pricing presented by Black and Scholes (1973) and Heath, Jarrow and Mortons’ (1992) Term Structure Model. Within equilibrium state price models, the most well-known model is the Capital Asset Pricing Model (CAPM) by Sharpe (1964), Lintner (1965) and Mossin (1966), the Intertemporal CAPM (ICAPM) presented by Merton (1973) and Breeden’s (1979) Consumption CAPM (CCAPM). Furthermore, there are two well-known and diligently applied theories that are neither no-arbitrage models nor equilibrium state models. These are optimal portfolio theory by Markowitz (1952) and Arbitrage Pricing theory by Ross (1976). Common for all the above-mentioned theories on asset pricing is that they can be retraced to a single asset pricing approach, *The stochastic discount factor approach* (SDF), that originates from the fundamental theorem of asset pricing (Campbell, 2018), as illustrated in figure 4.1.

![Figure 4.1: An overview of the most popular asset pricing models in the literature, all originating from the SDF approach. The No-Arbitrage-models are depicted in the blue circles, the “Equilibrium state price models in red, and in green, the theories outside these two paradigms of asset pricing are illustrated. The illustration shows that all the conventional and widely used asset pricing models can be deduced from the stochastic discount factor approach.](image-url)
Despite the widespread use of the SDF approach in modern empirical studies in the field of finance (Campbell, 2018), the academic world still focuses on the traditional methodologies, such as CAPM primarily, when introducing new aspiring students to the field finance.

Putting a pin in the stochastic discount factor approach for now, the following section briefly outlines the fundamentals of the traditional Beta Pricing Model, and its implied beta-representation used for asset pricing. From there we move into multi-factor models building on this beta-representation, then, subsequently, outline the SDF-approach and its application in factor-based asset pricing models and finally argue why applying the stochastic discount factor approach is convenient in this thesis.

4.1 Conventional beta representation

Initially, we present the beta-representation in its simplest form. Let $R^e_t$ be the excess return of $N$ risky assets at time $t$. Traditional asset pricing methodologies, as the ones mentioned above, begin by proposing a return-generating process for excess returns, typically one that provides good explanatory power on the excess returns. For example, one may propose that excess returns are generated by a one-factor model,

$$R^e_t = \alpha + \beta f_t + \epsilon_t$$  \hspace{1cm} (1)

where $f_t$ is the realized value of a systematic risk factor at time $t$, $\epsilon_t$ is the idiosyncratic risk of the assets with $E_t[\epsilon_t|f_t] = 0_N$ and, where $0_N$ is an $N$-vector of zeros, $\beta = \frac{\text{cov}(r_{t|f}, f_t)}{\text{var}(f_t)}$ is the factor loadings of the returns with respect to the common factor and $N$ is the number of factors, in this instance $N = 1$. $\alpha$ is the additional excess return, not captured by the explanatory power of the risk factor.

A highly popular one-factor model, and a classical approach to asset pricing, the CAPM, uses an equilibrium state argument to identify the mean-variance efficient portfolio of risky assets. This approach has materialized as the backbone of asset pricing, portfolio construction and various other problems in the financial literature. The classic derivation of CAPM requires several assumptions to ensure that all investors perceive the same investment opportunity set, and hence worth select a mean-variance efficient portfolio from that set.

Investors who evaluate portfolios using the means and variances of single period returns, have identical beliefs about the means, variances and covariances of returns. They believe that there are no
non-traded assets, no taxes and no transaction costs and, in the basic model, that investors can borrow or lend at a given risk free rate (Sharpe, 1964) and (Lintner, 1965).

Given these assumptions all investors hold a mean-variance efficient portfolio and further agree on which portfolios are mean-variance efficient, i.e., all investors are assumed to be rational. With the addition of a riskless asset, the mutual fund theorem implies that all mean-variance portfolios combine the riskless asset with the tangency portfolio. Consequently, all investors hold risky assets in the same proportions to one another (Campbell, 2018).

The Sharpe-Lintner CAPM (Sharpe, 1964) yields certain asset pricing implications. The mean-variance-efficiency of a portfolio implies that mean returns on individual assets must be related in a certain way to the covariances of individual asset returns with that portfolio. An increase in the weight, \( w_i \) of asset \( i \) in portfolio \( p \), financed by a decrease in the weight on the riskless asset affects the mean and variance of the return on portfolio \( p \) as follows:

\[
\frac{d\bar{R}_p}{dw_i} = \bar{R}_i - R_f 
\]

\[
\frac{d\text{Var}(R_p)}{dw_i} = 2\text{Cov}(R_i, R_p) 
\]

From the formulas above it can be shown that:

\[
\bar{R}_i - R_f = \frac{\text{Cov}(R_i, R_p)}{\text{Var}(R_p)}(\bar{R}_p - R_f) = \beta_{ip} \cdot (\bar{R}_p - R_f)
\]

Where \( \beta_{ip} \equiv \frac{\text{Cov}(R_i, R_p)}{\text{Var}(R_p)} \) is the regression coefficient of asset \( i \)'s return on portfolio \( p \)'s return.


Equation (4) is true, by construction for a mean-variance efficient portfolio \( p \). In order to get a testable and practical model with economic content, we impose the restriction of the CAPM that the market portfolio \( m \) is mean-variance efficient. Given this assumption, equation (4) can be written as

\[
\bar{R}_i - R_f = \beta_{im} \cdot (\bar{R}_m - R_f)
\]
Where $\beta_{im} \equiv \frac{\text{Cov}(R_i, R_m)}{\text{Var}(R_m)}$ is the regression coefficient of asset $i$’s return on the market portfolio $m$’s return.

Equation (5) implies that the excess return on an asset, $i$, can be described by the assets’ covariance with the market, times the market risk premium. Hence, if we consider the regression of excess returns on the market excess return, we obtain the following formula, which highly resembles the general one-factor model presented in equation (1).

$$ R_{it} - R_{ft} = \alpha_i + \beta_{im} \cdot (\bar{R}_m - R_f) + \epsilon_{it} \quad (6) $$

The CAPM, as such, is an example of the beta-representation in its simplest form.

The intercept $\alpha_i = R_{it} - R_{ft} - \beta_{im} \cdot (\bar{R}_m - R_f)$ should by default be equal to zero for all assets. $\alpha_i$ is also known as Jensen’s alpha and is considered a proxy for identifying assets that are mispriced relative to the CAPM.

### 4.2 Multi-factor models

Seeking alpha, commonly known as abnormal excess return, has been the motivating factor of many research topics in the literature for decades. To this day, practitioners and finance researchers still strive to identify new methods to better predict the cross-section of stock returns. A bundle of these methods, that have been highly researched, can be categorized as multi-factor asset pricing, where the literature has identified potential asset specific factors with a predictive function on the cross-sections of stock returns. These multi-factor models originate from the problem that no single-factor models can adequately fit the cross-section of stock returns and simultaneously deliver uncorrelated residual returns (Campbell, 2018).

In the general form, a multi-factor model with $K$ ‘factor portfolios’ capturing the common influence of $K$, underlying sources of risk, and with the possibility of riskless borrowing and lending, $R_f$, follows,

$$ R_{it}^e = \alpha_i + \sum_{k=1}^{K} \beta_{ik} R_{kt}^e + \epsilon_{it} \quad (7) $$

where $\beta_{ik}$ is the factor specific coefficient and $R_{kt}^e$ is the excess return on the $K$’th factor at time $t$. 
We assume that the residual is uncorrelated across stocks. The prediction of this model is that $\alpha = 0$ for almost all stocks. If we do not have factor portfolios but instead consider the factors directly as mean-zero shocks then we have (Campbell, 2018),

$$R_{it} = \mu_i + \sum_{k=1}^{K} \beta_{ik} f_{kt} + \epsilon_{it} \quad (8)$$

and the prediction for the model is that (8) holds for alphas defined by

$$\alpha_t = \mu_t - \lambda_0 - \sum_{k=1}^{K} \beta_{ik} \lambda_{kt} \quad (9)$$

where $\lambda_0$ is the zero-beta rate\footnote{Zero-beta rate is the return for zero-beta assets}, and $\lambda_k$ is the price of risk of the $k$’th factor. We revisit the factor model notation in section 4.3, when elaborating on the SDF-approach.

4.2.1 Multi-factor model literature

This section briefly summarizes a vast empirical literature on the cross-section of stock returns, with focus on characteristics of stocks that appear to predict returns in a way that cannot be explained by the CAPM - so called anomalies. It is worth noting that the below-mentioned anomalies do not necessarily exist today, as literature has shown that some predictors of returns in a given period of time do not have the same abilities in different time periods (Marquering, 2006). Furthermore, we acknowledge that a quite significant number of anomalies have been identified and sought exploited by scholars and practitioners in the last 50 years, why presenting them all would not be expedient in this thesis. Hence, we focus on the most popular anomalies:

**Beta.** Black, Jensen and Scholes (1972), followed up by Fama and Macbeth (1973), presented, in the earliest empirical studies of the CAPM, a positive and close to linear relationship between the subsequent betas of portfolios, (created by grouping stocks based on their historical betas), and their average return. At the time, this was considered a success for the CAPM, however later empirical work found a rather weak relationship between beta and return since the early 1960’s. Naturally, this motivated the search for new factors to predict cross-sections of returns. Banz (1981) discovered that firms with low market capitalization had a tendency of having higher average returns than their betas justify. This was the first introduction of the size-factor. Another notable factor is value. Value investors seek out stocks with a low market capitalization relative to the book value of the company.
There has been a variety of ways to measure or indicate value-stocks, but since the work of Fama and French (1992, 1993) the book-to-market ratio has been the preferred standard. Fama and French (1992, 1993) developed a 3-factor asset pricing model, based on the market beta, size and value and showed that it was superior to the CAPM in predicting cross-sections of returns. In the last 50-year period value-stocks have had lower market betas than growth-stocks, but higher average returns (Campbell, 2018), which, following theory of CAPM, is counterintuitive. The momentum factor builds on the relatively simple approach of using previous history of stock returns to predict future returns of the same stock. Autocorrelation of portfolios of stocks grouped by past returns have proven to exist in some periods of time. In momentum investing the investor buys previous winners and sell previous losers. DeBondt and Thaler (1985) documented a long-term negative autocorrelation, but the anomaly quickly lost popularity as it appears to be closely related to the value anomaly. However, Jegadeesh and Titman (1993) presented medium term positive autocorrelation, or momentum, which appears to be a solid and robust phenomenon (Campbell. 2018). Since then, there have been countless other factors presented in the literature, such as turnover, idiosyncratic volatility, different measures of earnings quality and profitability to mention a few.

Common for most models, based on these anomalies, is that they are what Kozak et al. (2020) describe as characteristics-sparse. Recent financial literature such as Kozak (2018) has questioned and challenged sparsity in characteristics-based factor models. With the growing applicability of artificial intelligence, such as machine learning, in finance, approaches to mitigate sparsity and search for a vast number of significant factors to predict the cross-sections of returns are surfacing.

### 4.3 Stochastic Discount Factor

In order to perform the required statistical modifications, necessary to evaluate which asset characteristics best predict the cross-section of returns, this thesis applies an alternative paradigm for asset pricing, more specifically we introduce the stochastic discount factor (SDF) that were teased at the beginning of the theoretical framework section. Modern literature on the stochastic discount factor reformulates arbitrage and equilibrium-based asset pricing, building on the work of Ross (1978), Harrison and Kreps (1979), Hansen and Richard (1987), Cochrane (1996) and Hansen and Jagannathan (1991, 1997). The stochastic discount factor approach is, as mentioned, the dominant approach to asset pricing in modern academic research (Campbell, 2018).

To understand the properties of the stochastic discount factor, we initially consider the fundamental theorem of asset pricing, stating that the price of any asset is its expected payoffs...
discounted by a proper discount factor. Ideally, the payoff is discounted by a factor, which depends on parameters present in the market, and it should be unique, in the sense that financial derivatives should be able to be priced using the same discount factor. In theory, risk neutral valuation implies the existence of a positive random variable, which is called the *stochastic discount factor* (SDF) and is used to discount the payoffs of any asset.

The general and fundamental equation of asset pricing can be written as

\[ P_{i,t} = E_t[M_{t+1} \cdot X_{t+1}] \]  \hspace{1cm} (10)

Where \( P_{i,t} \) is the price of asset \( i \) at time \( t \), \( X_{t+1} \) is the random payoff on asset \( i \) at time \( t+1 \) and \( M_{t+1} \) is the stochastic discount factor at time \( t+1 \). The SDF is a random variable.

Absent uncertainty the SDF is a constant that converts into the present value of the expected payoffs. In such a case the asset pricing formula is given as

\[ P_{i,t} = \frac{1}{R_f} X_{t+1} \]  \hspace{1cm} (11)

Where \( R_f \) is the gross risk-free rate. We note that \( M = \frac{1}{R_f} \) is the discount factor in this case. Intuitively, riskier assets have lower prices than equivalent risk-free assets, why they are valued using the risk adjusted formula:

\[ P_{i,t} = \frac{1}{R_i} \cdot E[X_{i,t+1}] \]  \hspace{1cm} (12)

In this case the risky asset, \( i \), is discounted by an asset specific risk-adjusted discount factor \( M = \frac{1}{R_i} \) (Cochrane, 2005). Equation (10) brings a significant implication for asset pricing; One can incorporate all risk corrections by defining a single stochastic discount factor, importantly the same for each asset, and putting it inside the expectation. \( M_{t+1} \) is stochastic because it is not known with certainty at time \( t \). Since the discount factor is the same for all assets it is the correlation between this common stochastic discount factor, \( M \), and the asset specific payoff \( X_i \) that generate the asset-specific risk corrections (Cochrane, 2005).

An advantage of the SDF-approach in relation to beta-representation or other alternative asset pricing models is the relatively limited number of conditions needed for the existence of the stochastic discount factor. The SDF is only conditioned upon two important theorems, drawing from the law of
one price and the absence of arbitrage. The law of one price states that if two assets or asset pools have the same payoffs in every economic state, they must have the same price. A violation of this law creates an arbitrage opportunity, as an investor could buy the cheaper version and sell the expensive version simultaneously and by default gain a riskless profit (Cochrane, 2005). Furthermore, the absence of arbitrage implies that if a payoff $X$ is equal to or greater than a payoff $Y$, and sometimes $X$ is greater, the price of $X$ must be higher than the price of $Y$. As such, the definition of absence of arbitrage is: positive payoff implies positive price (Cochrane, 2005).

**Theorem 1:** Given free portfolio formation, the law of one price implies the existence of a discount factor $M_{t+1} > 0$, $P_{i,t} = E_t[M_{t+1}X_{i,t+1}]$ for every payoff $X_i$.

**Theorem 2:** There is a strictly positive discount factor, $M > 0$, such that $P_{i,t} = E_t[M_{t+1}X_{t+1}]$ if and only if there are no arbitrage opportunities and the law of one price holds.  

The first theorem ensures the existence of an SDF for every payoff $X_i$ and the second theorem ensures that there exist a strictly positive, $M > 0$, however, it does not state that $M$ is unique.

In order to obtain a unique SDF one must assume complete markets. A complete market is a market in which the complete set of possible bets on future outcomes can be constructed with existing assets absent friction and transaction costs (Campbell, 2018).

However, as Cochrane (2005) highlights, since strictly positive discount factors can be seen as possible contingent claims, they need not be unique, why we can use the SDF-approach for asset pricing without assuming complete markets, as opposed to classic factor models such as the CAPM. The only conditions that must be met are, as such, no arbitrage opportunities and law of one price. This is a highly advantageous feature of the SDF since the complete market assumption is a quite strong condition.

---

14 For a detailed proof of both theorems, please see ‘Asset Pricing’, chapter 4, by John Cochrane (2005).
4.3.1 Properties of the SDF

To move in the direction of empirical research in finance, this subsection focuses on the properties of the SDF. The formula stated in equation (10) can be rewritten as:

\[ P_{i,t} = E_t[M_{t+1}, X_{t+1}] \Rightarrow P_{i,t} = E_t[M_{t+1}]E[X_{t+1}] + Cov(M_{t+1}, X_{t+1}) \]  

Since the conditional mean of the SDF is the reciprocal of the gross riskless interest rate from equation (11) the price of any asset is its expected payoff, discounted at the riskless interest rate, plus a correction for the conditional covariance of the payoff with the SDF (Campbell, 2018) as we see in equation (13).

Initiating from equation (13), we can, for assets with positive prices, divide both sides of the equation by \( P_{i,t} \):

\[ \frac{P_{i,t}}{P_{i,t}} = \frac{E[M_{t+1}, X_{t+1}]}{P_{i,t}} \Rightarrow \frac{P_{i,t}}{P_{i,t}} = E\left[\frac{M_{t+1}}{P_{i,t}} X_{t+1}\right] \]

Since \( R_{i,t+1} = \frac{X_{i,t+1}}{P_{i,t}} \), and given that \( P_{i,t} \neq 0 \), this yields the asset pricing equation in return form:

\[ \Rightarrow 1 = E[M_{t+1} R_{i,t+1}] \]

Similar to equation (13) this can be reformulated as:

\[ 1 = E_t[M_{t+1}]E_t[R_{i,t+1}] + Cov(M_{t+1}, R_{i,t+1}) \]  

Where \( R_{i,t+1} = \frac{X_{i,t+1}}{P_{i,t}} \) is the gross return of asset \( i \) at time \( t + 1 \).

Rearranging equation (15) above, and using the relation between the riskless rate and the conditional mean of the SDF, yields:

\[ E_t[1 + R_{i,t+1}] = (1 + R_{f,t+1}) \left(1 - Cov(M_{t+1}, R_{i,t+1}) \right) \]

This indicates that the expected return on any given asset is the riskless return times an adjustment for the covariance of the return with the SDF. To reach the excess return we need only subtract the gross riskless interest rate, which reveals that the risk premium is the gross riskless interest rate times the covariance of the excess return with the SDF (Campbell, 2018)

\[ E_t[R_{i,t+1} - R_{f,t+1}] = -\left(1 + R_{f,t+1}\right) \left(1 - Cov(M_{t+1}, R_{i,t+1} - R_{f,t+1}) \right) \]
Equation (17) can subsequently be rewritten as

\[ E_t[R_{i,t+1} - R_{f,t+1}] = \beta_{it} \lambda_{it} \iff E_t[R_{i,t+1}^e] = \beta_{it} \lambda_{it} \]  \hspace{1cm} (18)

Where the factor-beta, the \(H\)-dimensional vector, given by, \(\beta_{it} \equiv \frac{\text{Cov}_t(M_{t+1}, R_{i,t+1})}{\text{Var}_t(M_{t+1})}\) is the regression coefficient on asset return \(i\) on the SDF and the vector \(\lambda_{it} \equiv -(1 + R_{f,t+1}) \cdot \text{Var}_t(M_{t+1})\) is the factor risk premium of the SDF. Hence, returns always follow a linear factor pricing model with the SDF as the single factor. \(\lambda_{it}\), the factor risk premium of the SDF, is equal to the conditional variance of the SDF multiplied by the gross risk-free interest rate.

4.3.2 SDF in multi-factor models

The SDF paradigm offers a different way to understand and apply the multi-factor models, briefly introduced in equation (8). We assume that the SDF is a linear combination of \(K\) factors \(f_{k,t+1}, k = 1 \ldots K\), and that the factors have conditional mean zero and are orthogonal to one another, i.e., uncorrelated factors. If we define the SDF, \(M_{t+1}\), as

\[ M_{t+1} = a_t - \sum_{k=1}^{K} b_{kt} f_{kt+1} \]  \hspace{1cm} (19)

Where \(a_t\) is a constant, \(b_{kt}\) is the SDFs loading on a certain factor \(f_{kt+1}\). Then

\[ E_t[R_{i,t+1} - R_{f,t+1}] = -\left(\frac{1}{a_t}\right) \text{Cov}_t(M_{t+1}, R_{i,t+1} - R_{f,t+1}) = \left(\frac{1}{a_t}\right) \sum_{k=1}^{K} b_{kt} \sigma_{ikt} \]

\[ = \left(\frac{1}{a_t}\right) \sum_{k=1}^{K} (b_{kt} \sigma_{ikt}^2) \left(\frac{\sigma_{ikt}}{\sigma_{kt}^2}\right) \iff \]

\[ E_t[R_{i,t+1}^e] = \sum_{k=1}^{K} \lambda_{kt} \beta_{ikt} \]  \hspace{1cm} (20)

In the equation above \(\sigma_{ikt}\) is the conditional covariance of asset return, \(i\), with the \(k\)’th factor, \(\sigma_{kt}^2\) is the conditional variance of the \(k\)’th factor, \(\lambda_{kt} \equiv \frac{b_{kt} \sigma_{kt}^2}{a_t}\) is the price of risk of the \(k\)’th factor and \(\beta_{ikt} \equiv \frac{\sigma_{ikt}}{\sigma_{kt}^2}\) is the regression coefficient of asset return \(i\) on that factor. Looking at this equation it shows an equivalence between a linear factor model written in return-beta form and a linear factor model of the SDF with identical factors, \(K\) (Campbell, 2018).
4.3.3 Benefits of the SDF-approach

Originating from the generic asset pricing formula, the stochastic discount factor approach is very general, which is especially useful for considerations of theoretical valuations questions. To apply the approach for practical valuation questions one can either estimate a discount factor directly from data, as described by Cochrane (2005), or employ a specialization of the approach that is reasonable for whatever particular problem one seeks to investigate.

We previously mentioned the two theorems, dependent on the assumptions of law of one price and absence of arbitrage. The first theorem states that there is a discount factor that prices all the payoffs by \( p = E(mx) \) if and only if the law of one price holds. The second theorem states that there is a positive discount factor that prices all payoffs by \( p = E(mx) \) if and only if there are no arbitrage opportunities. These theories so usefully shows that we can use stochastic discount factors without implicitly assuming anything about utility functions, aggregation, structure of investors, complete markets, and so on (Cochrane, 2005). Essentially, all we must know about investors in order to represent prices and payoffs via a stochastic discount factor is that they will not impose law of one price violations or exploit arbitrage opportunities. This increases the applicability of the approach compared with other, more conventional, asset pricing theories as researchers need not impose as strict and often practically implausible assumptions.

4.3.4 SDF in characteristics-based factor models

In this paper, we draw on the stochastic discount factor approach as we, inspired by Kozak et al. (2020), will explore asset pricing with characteristics-based factors. Building on the SDF we start by rewriting the basic asset pricing equation to fit the framework that underlies characteristics-based factor models.

For any point in time \( t \), let \( R^e_t \) denote an \( N \times 1 \) vector of excess returns for \( N \) stocks. As previously shown, typical reduced form factor models express the SDF as a linear function of excess returns on stock portfolios. The conditional pricing equation, in return form, which we introduced earlier in equation (14), as the basis of the stochastic discount factor approach, can be rewritten by applying the fact that the expected SDF multiplied by a riskless return equals 1, i.e., \( 1 = E_t[M_{t+1}R_{f,t+1}] \):

\[
1 = E_t[M_{t+1}R_{i,t+1}] \Leftrightarrow E_t[M_{t+1}R_{f,t+1}] = E_t[M_{t+1}R_{i,t+1}] \Leftrightarrow
\]
\[
0 = E_t[M_{t+1}R_{t+1}] - E_t[M_{t+1}R_{f,t+1}] = E[M_t(R_{t+1} - R_{f,t+1})] \Leftrightarrow \\
E_t[M_{t+1}R_{e,t+1}] = 0 \tag{21}
\]

As such equation (21) shows that the expected excess return, \( R_{t+1}^e \), discounted with the stochastic discount factor is equal to zero.

Inspired by Hansen and Jagannathan (1991), Kozak et al. (2020) solves for the \( N \times 1 \) vector of SDF loadings \( \beta_{t-1} \) that satisfies the conditional pricing equation above (equation (21)), and defines an SDF in the linear span of excess returns as:

\[
M_{t+1} = 1 - \beta_t'(R_{t+1}^e - E[R_{t+1}^e]) \tag{22}
\]

Hence, the stochastic discount factor is dependent on the factor loadings or factor coefficients (we use these two terms interchangeably), \( \beta_t \), and the difference between the excess return of the stocks and the expected excess return of the same stocks. Where the expected excess return is the mean excess return on the stocks. To proceed with characteristics-based asset pricing, we transform the SDF loadings by parametrizing them as

\[
\beta_t = Z_t \beta \tag{23}
\]

Where \( Z_t \) is a \( N \times H \) matrix of asset characteristics (on \( N \) stocks, with \( H \) characteristics) and \( b \) is a \( H \times 1 \) vector of time-invariant coefficients. Through the Z-score we standardize the characteristics with respect to their cross-sectional means\(^ {15} \). In order to better assess the significance of each factor on the cross-section of returns, we define \( F_t \) as the linear span of the \( H \) characteristics-based factor returns, which are created based on different stock characteristics, then:

\[
F_{t+1} = Z_t' R_{t+1}^e \tag{24}
\]

Plugging equation (23) into (22) delivers an SDF in the space of factor returns such that,

\[
M_{t+1} = 1 - \beta'(F_{t+1} - E[F_{t+1}]) \tag{25}
\]

As such, the transformed SDF is a function of the factor coefficients \( \beta \), and the difference between the excess factor returns \( F_{t+1} \) and the mean excess factor returns \( E[F_{t+1}] \).

\(^ {15} \) Please see section 5.2 ‘Building the estimator’ for further reasoning behind this step.
So far, we have been applying the conditional pricing equation (equation (21)) to determine a stochastic discount factor in the space of factor returns, equation (25).

Pricing equations should be constructed, so that a relation like the ones above, for example in equation (20) or (21), holds for every period, that is, between two adjacent time points. However, the expectation, covariance and variance in such an expression can either be conditional or unconditional, why the literature distinguishes between conditional and unconditional pricing equations (Munk, 2013). \(^{16}\)

In equation (21), \(E_t[M_{t+1}R_{t+1}^e] = 0\), the \(E_t\) depicts expectation conditional on the investors time-\(t\) information. If payoffs and discount factors were independent and identically distributed (i.i.d) over time, then conditional expectations and unconditional expectations would be the same and we need not worry about the distinction between the two concepts. However, stock prices, dividend ratios, bond prices and option prices to name a few, all changes over time, which must reflect changing conditional moments on the right-hand side (Cochrane, 2005).

Testing factor models on actual data requires an unconditional pricing model as we, as an example, need to replace expected returns by average returns. Similarly, when we seek to estimate factor betas and the specific factor from a time series of observations of returns through certain regression methods, the unconditional factor model is required (Munk, 2013). As such, we transform the conditional pricing equation to an unconditional pricing equation, in order to apply the model on actual data. This is in line with most of the empirical work on characteristics-based factor modelling (Kozak et al., 2020).

Starting from the conditional fundamental pricing equation in the space of factor returns.

\[
E_t[M_{t+1}F_{t+1}] = 0
\]

The notation \(E_t[\cdot]\) denotes, as mentioned, the conditional expectation, given a market-wide information set \(\Omega_t\). Empiricists do not get to observe \(\Omega_t\), why it is convenient to consider expectations conditional on an observable subset of these expectations, \(Z_t\). \(^{17}\)

---


\(^{17}\) The notation for the subset of information, \(Z_t\), is not to be confused with the \(N \times H\) matrix of asset characteristics, \(Z_t\) applied to transform the SDF loadings in equation (23). This is simply a notational matter - for the remainder of this paper \(Z_t\) denotes the Z-score of transformed characteristics.
These expectations are denoted $E[\cdot | Z_t]$. When $Z_t$ is the null information set, we have the unconditional expectation, denoted as $E[\cdot]$ (Ferson, 2005).

As such, an unconditional pricing equation implies that:

$$E[M_{t+1}F_{t+1}] = 0$$

(26)

Where $F_t$ serve as both the assets whose returns we try to explain as well as candidate factors that potentially could occur as priced factors in the SDF.

With knowledge of population moments, solving equation (24) and (25) for the SDF-coefficients, $\beta$, yields

$$\beta = \Sigma^{-1}E[F_{t+1}]$$

(27)

Where the factor covariance matrix $\Sigma$, is

$$\Sigma = E[(F_t - E[F_t])(F_t - E[F_t])']$$

(28)

We can rewrite the expression as

$$\beta = (\Sigma \Sigma)^{-1}\Sigma (F_t)$$

(29)

to show that the SDF-coefficients, $\beta$, can be interpreted as the coefficients in a cross-sectional regression of the expected asset returns to be explained by the SDF, which, in this case, are the $H$ elements of $E[F_t]$, on the $H$ columns of covariances of each factor with the other factors and with itself (Kozak et al., 2020).

In practice, without knowledge of population moments, estimating the SDF-coefficients by running such a cross-sectional regression in sample would result in overfitting of noise, with the consequence of poor out-of-sample performance, unless $H$ is small (Kozak et al., 2020).

4.3.5 Sparsity in characteristics-based factor returns

Kozak et al. (2020) acknowledges that sparsity in many of the existing characteristics-based factor models, such as Fama and French limiting the explanation of cross-sectional returns to 3 factors (1992) or subsequently limiting the model at six explanatory factors (2016) is driven by the need for simplicity, i.e., avoiding the high-dimensionality problem. Kozak et al. (2020) argue that a characteristics-sparse model is potentially limited in describing the stochastic discount factor in a cross-section with many stock characteristics (Kozak et al., 2020). This motivates the pursuit of an
approach that does not impose that the SDF is necessarily characteristics-sparse. Furthermore, it addresses the need for a method that can accommodate an SDF that includes a potentially vast number of characteristics-based factors, while simultaneously ensuring good out-of-sample (OOS) performance and robustness against in-sample overfitting. However, the method should ideally also have the capacity to handle cases in which some of the candidate factors are not contributing to the SDF at all. This could occur in cases where the analysis includes characteristics not previously covered by the literature or if some cross-sectional return predictors, previously documented in the literature, proves to be redundant (Kozak et al., 2020). Thus, to accommodate such cases, the applied approach, ideally, allows the possibility of sparsity but without necessarily relying on sparsity to perform well out-of-sample.

4.3.6 Principal components

Motivated by previous analysis of Kozak et al. (2018) a sparse SDF-representation in the space of Principal Components (PCs) of characteristics-based factor returns can be considered to accommodate the functionality and requirements of the approach outlined in the section above. This section briefly explains the underlying theory of PCs before transforming equation (24) to arrive at an SDF-representation in the space of PCs of characteristics-based factors.

Principal component analysis is, simply put, a dimensionality-reduction method, transforming large data sets into smaller ones, while preserving most of the information from the large data set. Naturally, reducing the number of variables will increase simplicity, but it comes at a cost, i.e., decrease in accuracy. This is the tradeoff we must accept in dimensionality reduction. However, when simplifying the data, we obviously seek to preserve as much information as possible.

As such, principal components are new variables that are constructed as linear combinations of the initial variables. The PCs are uncorrelated and most of the information within the initial variables are squeezed or compressed into the first components. Thus, $H$-dimensional data gives you $H$ principal components, where the transformation applies maximum possible information in the first component, then maximum remaining information in the second and so on. The amount of information applied is measured by the eigenvalue (indicator of variance) of the PCs. In geometrical terms principal components represent the directions of the data that explain a maximum amount of variance, i.e., the lines that capture the most information. The following section thoroughly describes the computation of principal components, as well as their relevance for our problem.
The first step in dimensionality reduction through PCs is to standardize the range of continuous initial variables, in our case the $H$ characteristics of stocks, so that each characteristic contributes equally to the analysis. A broad variety of such transformation methods exists, and the exact choice of transformation is of less importance. For the purpose of our analysis, we perform a rank transformation and standardize the characteristics with respect to their cross-sectional means, by computing $Z$-scores. This transformation is outlined in detail in section 5.2.1.

The next step is to decompose the factor covariance matrix between the transformed characteristics. For this purpose, we reintroduce the factor covariance matrix from Equation (28):

$$\Sigma = E[(F_t - E[F_t])(F_t - E[F_t])'] .$$

From the covariance matrix, eigenvectors and eigenvalues are computed based on the following eigen decomposition:

$$\Sigma = QDQ'$$

with $D = diag(d_1, d_2, ..., d_H)$

Here $Q$ is the matrix of eigenvectors of $\Sigma$, and $D$ is the diagonal matrix of eigenvalues sorted in decreasing magnitude. Eigenvectors and eigenvalues are the linear algebra concepts needed to determine the PCs of factor characteristics. The number of eigenvectors and eigenvalues are equal to the number of dimensions. In our case we have $H$ characteristics, consequently there are $H$ eigenvectors with $H$ corresponding eigenvalues. The eigenvectors of the covariance matrix are the directions of the axes containing the most variance (i.e., the most information). Eigenvalues are simply the coefficients attached to each eigenvector, indicating the amount of variance carried in each eigenvector (Hastie, Tibshirani & Friedman, 2009). The principal components are then constructed by multiplying the vector of transformed factor returns with the corresponding eigenvectors from the factor covariance matrix:

$$PC_t = Q'F_t$$

Ranking the eigenvectors of the factor covariance matrix in order of their eigenvalues, yields the PCs in order of importance. An important feature of PCs is that the level of the specific PC values is of less interpretable value.

As mentioned above, PCs are constructed as linear combinations of the initial variables and thereby, serves as a simple representation of the structure of a complex data set. For the sake of intuition, we consider a data set of 25 portfolios, sorted on Book to market and Market Capitalization.
provided by Kenneth French\textsuperscript{18}. These 25 portfolios should, by construction, have a significant factor structure, as they are all related to the classic SMB and HML factors introduced by Fama and French (1992).

Thus, principal components should be effective in characterizing the structure of data, by capturing the majority of variance in just a few principal components. Computing the PCs using the method described above yields the 25 PCs summarized in table 4.1.

Table 4.1: Table 4.1 illustrates the variance contained in each of the 25 PCs constructed based on the Fama-French 25 ME/BM sorted portfolios withdrawn from Kenneth French’s website. PCs are sorted in descending order based on the variance contained, with the highest variance PC denoted as PC1. Furthermore, the table highlights the proportion of variance explained and the cumulative variance explained for the 25 principal components. Sample data is daily from 1926 – 2017.

<table>
<thead>
<tr>
<th>Principal components</th>
<th>PC1</th>
<th>PC2</th>
<th>PC3</th>
<th>PC4</th>
<th>PC5</th>
<th>PC6</th>
<th>PC7</th>
<th>PC8</th>
<th>PC9</th>
<th>PC10</th>
<th>PC11</th>
<th>PC12</th>
<th>PC13</th>
<th>PC14</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard deviation</td>
<td>4.39</td>
<td>1.01</td>
<td>0.87</td>
<td>0.82</td>
<td>0.70</td>
<td>0.57</td>
<td>0.54</td>
<td>0.47</td>
<td>0.46</td>
<td>0.43</td>
<td>0.41</td>
<td>0.40</td>
<td>0.37</td>
<td>0.36</td>
</tr>
<tr>
<td>Proportion of variance</td>
<td>76.9%</td>
<td>4.1%</td>
<td>3.0%</td>
<td>2.7%</td>
<td>2.0%</td>
<td>1.3%</td>
<td>1.1%</td>
<td>0.9%</td>
<td>0.8%</td>
<td>0.7%</td>
<td>0.6%</td>
<td>0.5%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cumulative proportion</td>
<td>76.9%</td>
<td>81.0%</td>
<td>84.1%</td>
<td>86.8%</td>
<td>88.7%</td>
<td>90.0%</td>
<td>90.9%</td>
<td>91.2%</td>
<td>91.9%</td>
<td>92.9%</td>
<td>93.7%</td>
<td>94.3%</td>
<td>95.0%</td>
<td>95.5%</td>
</tr>
</tbody>
</table>

(Continued)

<table>
<thead>
<tr>
<th>Principal components</th>
<th>PC15</th>
<th>PC16</th>
<th>PC17</th>
<th>PC18</th>
<th>PC19</th>
<th>PC20</th>
<th>PC21</th>
<th>PC22</th>
<th>PC23</th>
<th>PC24</th>
<th>PC25</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard deviation</td>
<td>0.35</td>
<td>0.34</td>
<td>0.32</td>
<td>0.31</td>
<td>0.31</td>
<td>0.30</td>
<td>0.29</td>
<td>0.28</td>
<td>0.27</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>Proportion of variance</td>
<td>0.5%</td>
<td>0.5%</td>
<td>0.4%</td>
<td>0.4%</td>
<td>0.4%</td>
<td>0.3%</td>
<td>0.3%</td>
<td>0.3%</td>
<td>0.3%</td>
<td>0.2%</td>
<td></td>
</tr>
<tr>
<td>Cumulative proportion</td>
<td>96.6%</td>
<td>97.0%</td>
<td>97.4%</td>
<td>97.8%</td>
<td>98.2%</td>
<td>98.6%</td>
<td>98.9%</td>
<td>99.2%</td>
<td>99.5%</td>
<td>99.8%</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

In table 4.1 the proportion of variance explained by each PC is shown as well as the cumulative proportion explained. We note that the proportions of variance sum to one, as expected. As we intuitively expected, the standard deviation table for each PC shows that the first principal components carry the highest standard deviation, and that the standard deviation depreciates by descending order of principal components. Thereby, the first principal component contributes the highest proportion of total variance explained, accounting for 76.9\%.

This is additionally evident from figure 4.2 illustrating the proportion of total variance explained by each PC, where the first PC is significantly larger than the following PCs.

\textsuperscript{18} The 25 portfolios are provided in the data library of Kenneth French’s web-page: \url{http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html}
Figure 4.2: The Scree-Plot illustrates the percentage of total variance contained in the first 10 principal components based on the Fama-French 25 ME/BM sorted portfolios withdrawn from Kenneth French’s website. The horizontal axis depicts PCs from 1 – 10 and the vertical axis represents the variance contained in each PC. The grey line illustrates the cumulative variance explained. The sample data is daily from 1926 to 2017.

Beyond the highest variance PC each subsequent PC contributes less than 5%, illustrating that the total variance in the 25 portfolios can be represented relatively well through the first, or at most three, PCs. Hence, plotting the first PCs against each other yields a quite accurate representation of the structure in the data. Figure 4.3 below depicts the highest variance PCs plotted against each other, with figure 4.3 (a) illustrating PC1 plotted against PC2 and figure 4.3 (b) further including PC3 in a 3-dimensional plot.

Figure 4.3: The plot in figure 4.3 is a visualization of the strong factor structure in the 25 BE/ME sorted portfolios in the space of PCs. Figure 4.3 (a), in the left panel, illustrates PC1, on the horizontal axis, plotted against PC2, on the vertical axis. In figure 4.3 (b) PC3 is included in a 3-dimensional plot. The sample data is daily from 1926 to 2017.

Figure 4.3 (a) illustrates that the data points primarily move around the first axis (representing PC1) indicating that most of the variance is captured in the first PC. Moving along the 2nd axis (representing
PC2) alters the data point slightly. In figure 4.3 (b) we note that there is not much variance explained in the third PC (PC3). We note, that even PC2 and PC3 have relatively low significance attesting to the strong factor structure of the chosen 25 ME/BM portfolios.

We highlight that principal components analysis usually is a preferred method in unsupervised learning as it is useful in identifying clusters in the data set. However, Kozak et al. (2020) applies PCs in the construction of a sparse SDF-model. As such, the following section outlines how PCs maps into the SDF-approach.

4.3.7 Sparsity in principal components of characteristics-based factor returns

With the fundamentals of principal components in mind, we, in line with Kozak et al. (2020), consider sparse SDF-representations in the space of PCs of characteristics-based factor returns.

Using all PCs, one can express the SDF in the space of principal components as

\[ M_t = 1 - \beta'_p(P_t - E[P_t]) \quad \text{with} \quad \beta'_p = D^{-1}E[P_t] \]  

(32)

As such, the stochastic discount factor is a function of the factor coefficients in the space of principal components and the difference between PC-factor returns and the mean of PC-factor returns.

According to Kozak et al. (2018), absence of near-arbitrage (extremely high Sharpe ratios) implies that factors earning substantial risk premia must be major drivers of co-movements. This conclusion can be deemed valid under very mild assumptions and applies equally to both ‘rational’ and ‘behavioral’ models.

Furthermore, for typical sets of test assets, returns have a strong factor structure dominated by a small number of PCs contributing the highest variance, as shown in the previous example for the 25 portfolios (Kozak, 2020). Hence, under these two conditions, an SDF with a small number of these high-eigenvalue PCs should be sufficient in describing the cross-section of expected returns. Similar to Kozak et al. (2020), we seek to compare the pricing performance of a PC-sparse SDF with a characteristics-sparse SDF on raw factor returns.

In order to assess the asset pricing implications of the multi-factor models, based on the proposed dual-penalty method, we later seek to compare the out-of-sample performance of mean-variance efficient (MVE) portfolios. As such, in section 6.3, we construct portfolios based on the MVE-method initially introduced by Harry Markowitz (1952). The following section is a brief introduction to the Mean-Variance optimization methods for portfolio construction.
4.4 Markowitz mean-variance efficient portfolios

The most renown model for a quantitative strategic asset allocation approach was presented by Harry Markowitz (1952). Markowitz’s portfolio theory laid the foundation to what is known as mean-variance analysis and modern portfolio theory. The framework of optimal portfolio construction seeks to address the problem of the trade-off between risk and return, as well as, defining and accounting for the concept of diversification.

Markowitz (1952) emphasizes the importance of considering the covariance of assets. He argues that minimization of the variance of returns is achieved through strategic allocation based on the covariance of assets. Markowitz’ (1952) technical and formal definition of investment optimization in a two-dimensional space, return and variance, states that a portfolio, A, is another Portfolio, B, superior when $\mu_A > \mu_B$ and $\sigma_A \leq \sigma_B$ (Basile & Ferrari 2016).

Before presenting the Markowitz-approach to optimal portfolio allocation, the matrix notation of the parameters of return, risk and portfolio weights is provided below. The parameters can be presented as follows:

The composition of the portfolio is given by the vector $\mathbf{w}$ of size $H \times 1$, where $w_i$ is the percentage of asset $i$ in the portfolio:

$$
\mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_i \\ \vdots \\ w_H \end{bmatrix}
$$

The sum of the individual weights must equal 1, so that $\sum_{i=1}^{N} w_i = 1$

The return and risk characteristics of the individual assets are also $H \times 1$ vectors:

$$
\mathbf{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_i \\ \vdots \\ \mu_H \end{bmatrix} \quad \text{and} \quad \mathbf{\sigma} = \begin{bmatrix} \sigma_1 \\ \sigma_2 \\ \vdots \\ \sigma_i \\ \vdots \\ \sigma_H \end{bmatrix}
$$

The dimension of the Correlation, $\mathbf{C}$, and the Covariance, $\mathbf{\Sigma}$, matrices are dependent on $H$ as they are $H \times H$ matrices and show the relationship between all pairs of assets:

The expected return of the portfolio $U_p$, is calculated using matrix notation $U_p = \mathbf{w}' \mathbf{\mu}$. This notation simply states that the expected returns for $H$ assets are multiplied by their corresponding portfolio weights. The expected portfolio variance is calculated as $\sigma_p^2 = \mathbf{w}' \mathbf{\Sigma} \mathbf{w}$. 


According to Markowitz (1952), for every level of expected return, a rational investor would choose to invest in the portfolio with the minimum-variance, thus, optimizing the risk-return relationship. Every portfolio satisfying the before-mentioned risk-return relationship are, by definition, mean-variance efficient portfolios. Together these portfolios represent the efficient frontier. Portfolios placed below the efficient frontier are, as such, inefficient but feasible, while portfolios above the efficient frontier are non-feasible.

Consequently, investors have two optimization solutions, either with respect to risk or return. An investor can either optimize the portfolio with the target of minimising the variance (risk) at a given target return or maximising return for a given risk-level. In this analysis we apply a certain relation between the mean-variance efficient portfolios and the stochastic discount factor to construct MVE portfolios\(^{19}\). This relation is rooted in the minimization problem, why we refrain from outlining the maximization problem below\(^{20}\).

1. *Minimization of expected variance given a target level of expected return:*  
\[
\min_w \, w' \Sigma w  
\text{subject to}  
\quad w' \mu = \mu_P  
\quad w'e = 1
\]

Furthermore, if one were to solve for the global-minimum-variance portfolio, the minimization optimization algorithm can be used, without a target level of expected return as a constraint.

**4.5 Section conclusion**

In this section, we outlined the theoretical framework underlying the analysis. First, we briefly touched upon the most popular and widely used asset pricing models in the last few decades. Second, we described the simplest single factor asset pricing model and its associated beta-representation before moving into multi-factor models for asset pricing and the proven anomalies underlying them. We then introduced an alternative paradigm, centered around the fundamental asset pricing equation, the stochastic discount factor (SDF), and its properties. With the introduction of an alternative paradigm, we then described characteristics-based factor models in the space of stochastic discount

\(^{19}\) The relation between MVE portfolios and the SDF is further outlined in section 6.3

\(^{20}\) For mathematical notations regarding the maximization problem we refer to Basile et al. (2016).
factors and detailed how Kozak et al. (2020) applies this approach to show that traditional characteristics-sparse factor models perform poorly. We explained the basis framework of principal components and subsequently transform the characteristics-based factor model to consider sparse SDF-representations in the space of PCs of characteristics-based factor returns. Finally, we explain mean-variance portfolio optimization by Markowitz (1952) as this method is applied to test the asset pricing abilities of the stochastic discount factor derived from the dual-penalty estimator outlined in section 5.2.8.

5. Methodology

Olsen and Pedersen (2018) describe the methodology as the intent of the examination. It is the link between the motivation behind the examination and the actual act of examining the issue at hand. The motivation of this thesis is to explore a more nuanced multi-factor model with higher success in explaining the cross-sections of stock returns than its preceding, traditional multi-factor models, by replicating the method and findings of KNS. This will be examined using two high-dimensional data sets and a variety of statistical methods described below.

The first high-dimensional data set consists of 69 financial ratios, representing asset specific characteristics, provided by Wharton Research Data Services in their data set “The Financial Ratios: Firm Level”21 (WFR). The 69 financial ratios are drawn on the companies currently in the S&P500, which totals 505 stocks (as some companies have issued multiple classes of shares) with daily data samples from January 2000 to December 2017. This corresponds to the financial ratios used in the analysis in KNS, however with two key differences. KNS uses industry-specific financial ratios whereas this thesis applies firm-specific financial ratios. Furthermore, the data sample in KNS is daily from September 1964 to December 2017.

Additionally, in the second high-dimensional data setting we examine 50 previously proposed anomalies in the literature provided by Kozak et al. (2020). Similar to KNS the data sample is daily from November 1973 to December 2017. Descriptions of the constituent anomalies and financial ratios in the two high-dimensional data sets are provided in Appendix 2 and 3. Using daily returns for each individual stock enables us to capture the fluctuations of the data much more precisely

---

compared to monthly data. This allows us to focus on the uncertainty in the mean returns rather than in the variance of the data, since the model estimates the second moment (variance) more precisely.

We further characterize the two high-dimensional data sets in the data analysis section 6.2.1.

According to Olsen and Pedersen (2018) research questions that are examined using quantitative methods are categorized as descriptive, as they strive to show how a certain phenomenon exists and co-exists with other phenomena. Olsen and Pedersen (2018) introduce three quality criteria to determine if a research examination is trustworthy: validity, reliability and adequacy. An examination is deemed valid if it sufficiently answers the research problem in question. It is reliable if the results are not random, and if other researchers are able to reproduce the same results. To obtain adequacy the sub questions and methodology, applied in the examination, must provide comprehensive answers to the research question at hand.

Quantitative methods, like those applied in this thesis, measures quantities and attempt to describe the existence of a phenomenon, or back testing of theories (Olsen & Petersen, 2018). Hence, quantitative models generally have relatively high reliability, as there is limited room for the researcher to affect the results of a quantitative measurement or statistical test, even though data selection offers the possibility of subjectivity through choice of data period, asset universe etc. In an ideal world, another researcher should, with access to an identical data set, be able to easily reproduce the same results. Nevertheless, they generally have poorer validity than qualitative models, as quantitative models in comparison struggle to really get to the core of the research question, as a result of being limited by the underlying data (Olsen & Petersen, 2018). The adequacy of an examination will have to be determined on a case-to-case basis.

These preceding quality criteria have all been considered in the formulation process of the sub question and gathering as well as selection of necessary data, to ensure that this examination on characteristics describing the cross-sections of asset returns, and as such, replication of the KNS article, is trustworthy. In the remainder of this section, we outline the statistical and econometric methods applied in this thesis, as well as the underlying assumptions. Furthermore, we discuss some of the shortcomings of certain econometric methods and subsequently introduce alternative methods presented in the literature to accommodate these shortcomings.
5.1 Regression methods

The following section focuses on multiple regression analysis, as this is the cornerstone in explaining a dependent variable, as a function of multiple independent variables. Multiple regression analysis has been an important tool for testing economic theories, as well as solving many other statistical issues in a broad variety of settings. Since multiple regression can accommodate many explanatory variables that may be correlated, one strives to infer causality in cases where simple regression analysis would be misleading (Woolridge, 2012). Intuitively, the addition of useful factors to a model, adds explanatory power. The multiple regression model is still the most widely used vehicle for empirical analysis in economics and other social sciences, however the literature has provided many different approaches for estimating the parameters in the multiple regression model.

The ordinary least squares method (OLS) is one of the most popular methods for estimating such parameters.

5.1.1 OLS

We briefly summarize some computational and algebraic features of the OLS-method, before presenting the most crucial shortcomings of the OLS and subsequently provide alternative approaches to mitigate these issues.

We consider an estimation model with $k$ independent variables. The estimated OLS equation is written in the following form,

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_k x_k$$

(33)

where $\hat{\beta}_0 = \text{the estimate of } \beta_0, \hat{\beta}_1 = \text{the estimate of } \beta_1 \ldots \hat{\beta}_k = \text{the estimate of } \beta_k$.

Basically, the ordinary least squares (OLS) method chooses $\beta$-coefficients in such a way that the sum of squared errors (residuals) between the model estimated response, $\hat{y}$, and the actual observed value of the response variable, $y$, is minimized. That is, given $n$ observations on $y, x_1, x_2 \ldots x_k, \{x_{i1}, x_{i2}, \ldots, x_{ik}, y_i; i = 1, 2, \ldots, n\}$, the estimates $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k$ are chosen simultaneously such that

$$\sum_{i=1}^{n} (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_k x_k))^2 = \sum_{i=1}^{n} \left( y_i - \sum_{j=1}^{p} x_{ij} \hat{\beta}_j \right)^2$$

(34)

is minimized (Woolridge, 2012).

---

22 In this thesis the dependent variable is the expected mean cross-sectional returns
Generally, OLS plays a large role in empirical finance and economics, because when the Gauss-Markov assumptions are satisfied, it is the best linear unbiased estimator (Woolridge, 2012). However, some of these assumptions can be quite restrictive, especially if one assumes normally distributed errors, which is necessary to do inference. As a result of these somewhat ambitious assumptions, often, one or more of the assumptions are violated in practice, resulting in poor OOS performance. What follows is a brief description of the strongest assumptions needed for the consistency of the OLS estimator (Woolridge, 2010).

**Linearity** - The first assumption is that the population model is linear in parameters. Hence, in the population, the model can be described as

\[ y = x\beta + \epsilon \]

Where \( y \) is the dependent variable observations, \( x \) is a 1 x K vector of regressor (independent variables) \((x_1, x_2, ..., x_K)\), \( \beta \) is a K x 1 vector of K coefficients and \( \epsilon \) is an error term. Plotting the dependent variable against the independent variables might reveal whether there is a linear relationship or not. If the linearity assumption is violated, OLS or any other linear regression model is not viable, and non-linear estimation models must be employed instead. As such, this assumption is key.

**No endogeneity** – The second assumption is no endogeneity of regressors. This is mathematically expressed as covariance of the error and the dependent variable is zero, for any combination of \( \epsilon \) and \( x \).

\[ \sigma_{x\epsilon} = 0 \forall x \]

**Orthogonality** – The third assumption is that the error term \( \epsilon \) has zero mean and is uncorrelated with every independent variable (Woolridge, 2010). The error term accounts for the variation in the dependent variable, that the independent variables fail to explain. As such the average value of the error terms must equal zero for the model to be unbiased. Furthermore, the error term is assumed to be an unpredictable and random error, as it is completely uncorrelated with the independent variable.

**No autocorrelation** – The covariance between any two error terms, \( \epsilon_i \), is zero.

\[ \sigma_{\epsilon_i\epsilon_j} = 0 \forall i \neq j \]

This implies that there is no trend in the error terms meaning that one cannot predict the errors based on historical observations, i.e., the error term can be assumed to be random.
**Full rank / No Multicollinearity** – No independent variable can be a perfect linear function of other independent variables. i.e., two independent variables cannot be perfectly correlated, as the OLS would not be capable of distinguishing the one variable from the other. As such, the model must drop one of the perfectly correlated variables in order to function. Whereas, perfect correlation is impossible for the OLS, statistical software can however fit OLS regression models with imperfect but strong relationships between independent variables. If these correlations are high enough, it will however cause problems, as it reduces the precision of the estimates in OLS linear regression (Woolridge, 2010). This condition is known as multicollinearity and is one of the major reasons why the OLS regression is not expected to be the optimal option in this thesis. As we eventually seek to create an estimation method, analysing 69 different asset specific features/characteristics one would expect some or even many of these to be highly correlated with one another. As such, the OLS method would yield estimates that are highly uncertain if correlated independent variables are included in the model.

To illustrate the problem of multicollinearity and highlight potential problems when dealing with multiple explanatory variables, we provide a brief example based on the following simulation. In the simulation study we control the true correlation between explanatory variables and, in particular, we determine the true value of the coefficients in the model.

Consider the following model,

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \epsilon_i$$ \hspace{1cm} (35)

where $x_1$ and $x_2$ are normally distributed with mean 0 and variance 5, $N(0,5)$, and the error term is a white noise process following $\epsilon_i \sim N(0,1)$. We determine the coefficients $\beta_0 = 0$ and $\beta_1 = \beta_2 = 1$. This represents a standard OLS regression with two explanatory variables, $x_1$ and $x_2$, and an intercept $\beta_0$. We wish to explore the effect of multicollinearity in the explanatory variables. As such, we simulate random variables $x_1$ and $x_2$ with distribution $N(0,5)$, while simultaneously controlling the correlation between the variables, $\rho_{1,2}$, with the following formula:

$$N_3 = \rho N_1 + \sqrt{1 - \rho^2} N_2$$ \hspace{1cm} (36)

Where $N_1$, $N_2$ and $N_3$ are distributed variables following $N(\mu, \sigma^2)$. The formula ensures that $N_3$ is following the same distribution as $N_1$ and $N_2$, such that;

$$E[N_3] = \mu \wedge Var[N_3] = \sigma^2$$ \hspace{1cm} (37)
Further, it ensures that the correlation between $N_1$ and $N_3$ is always $\rho$.

We simulate 50 observations of $x_1 = N_1$ and $x_2 = N_3$ 10,000 times for two different cases, where $\rho = 0.95$ and $\rho = 0$. We plot the coefficients $\beta_1$ and $\beta_2$ against each other yielding the following plots illustrated in figure 5.1.

![Figure 5.1](image.png)

**Figure 5.1**: Figure 5.1 illustrates the coefficients of the simulated data, derived as described in the section above, for two different settings. In figure 5.1 (a) the simulation is executed with no correlation between dependent variables, $\rho = 0$. In figure 5.1 (b) the simulation is executed with high correlation between dependent variables, $\rho = 0.95$, imitating high multicollinearity. For both settings, the plots are based on 50 observations simulated 10,000 times for each of the two different cases.

From figure 5.1 it is evident that, for uncorrelated explanatory variables, $(\rho_{x_1x_2} = 0)$, the coefficients are close to the true values of 1, as the observations are clustered around coordinates (1,1) in figure 5.1 (a). When high correlation is imposed, between the explanatory variables $(\rho_{x_1x_2} = 0.95)$ the problem of multicollinearity becomes evident as the uncertainty in the estimation becomes large, illustrated by the linear relationship in figure 5.1 (b). Since the variables in both the WFR and 50 anomalies data set are likely internally correlated, the OLS is highly insufficient in handling these multidimensional data settings\(^{23}\). One or more of these assumptions are often violated in practice, as illustrated above for the multicollinearity assumption. As such, in many cases the OLS regression method is not the optimal choice for estimating the $\beta$-coefficients.

Scientists and researchers often seek a way to define the criteria for evaluating and effectively choosing an optimal model. Following Zou and Hastie (2005) the criteria for evaluating regression models are typically based on the following two aspects:

1. **Accuracy of prediction on future data** – In the words of Zou and Hastie (2005) ‘It is difficult to defend a model that predicts poorly out of sample’, which intuitively makes perfect sense.

\(^{23}\) Multicollinearity between the variables in the two high-dimensional data sets is illustrated in figure 6.2 in the data analysis section 6.2.1.
Interpretation and simplicity of the model – scientists or empirical researchers prefer a simpler model because it sheds more light on the relationship between the response variable and the covariates.

Consequently, a model that achieves solid explanatory power from a minimal number of relevant variables is desired. This is defined as a parsimonious model, and it is a highly desirable and important feature to consider when dealing with high-dimensional data sets (Zou and Hastie, 2005).

According to Zou and Hastie (2005), it is well known that OLS often does poorly in both prediction and interpretation. Reasons for the usually subpar prediction performance of the OLS, could, for example, be attributed to the spurious overfitting one would expect the OLS-estimate to yield in cases with a large number of factors, such as ours. The OLS-estimator would include every single factor in the model, yielding high ‘in sample’ accuracy, but possibly poor out of sample accuracy as a consequence of a high degree of overfitting. Furthermore, the estimates of the OLS regression are highly sensitive to outliers (Zou and Hastie, 2004) and internally correlated predictors.

5.1.2 Penalization

As highlighted in the previous section, the classic OLS estimation method experiences issues dealing with high-dimensional data settings. To mitigate these issues, the literature has proposed useful methods imposing different sorts of penalization on the beta coefficients in the regression. Various penalization techniques have been proposed in the literature to improve OLS. Hoerl and Kennard (1970) introduced ridge regression, that minimizes the residual sum of squares subject to a bound on the $L_2$-norm of the coefficients, to mitigate the above-mentioned overfitting issue. In the formula below the highlighted part represents the $L_2$-regularization term.

$$
\hat{\beta} = \arg\min \sum_{i=1}^{n} \left( y_i - \sum_{j=1}^{p} x_{ij} \beta_j \right)^2 + \lambda \sum_{j=1}^{p} \beta_j^2
$$

Intuitively, if $\lambda = 0$, it is easy to see that we simply end up at the regular OLS, and if the value of $\lambda$ is too big, the consequence will be underfitting, since it would result in relevant explanatory variables being shrinked too much and thereby decrease the explanatory power of the model. As such, an appropriate $\lambda$ must be chosen, to attain optimal shrinkage. Kozak et al. (2020) proposes a method for optimally choosing the level of $L_2$-shrinkage through a data-driven approach, which will be outlined in section 5.2.8.
As a continuous shrinkage method, ridge regression achieves better prediction performance through a bias-variance trade-off. Thus, the optimal level of shrinkage lies between imposing a high degree of \( L_2 \)-penalty, which will bias the estimates, and imposing no \( L_2 \)-penalty resulting in overfitting of variables. However, parsimony cannot be achieved through ridge regression as all predictors are kept in the model.

To accommodate this issue Tibshirani (1996) has introduced a technique called LASSO (Least Absolute Shrinkage and Selection Operator). The LASSO is a penalized least squares method imposing an \( L_1 \)-penalty on the regression coefficients.

\[
\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{n} \left( y_i - \sum_{j=1}^{p} x_{ij} \beta_j \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \tag{39}
\]

Whereas the ridge regression technique includes an \( L_2 \)-penalty term that shrinks the coefficients towards zero, but not exactly zero, the \( L_1 \)-penalty term, introduced in the LASSO regression, shrinks some coefficients to exactly zero. This is due to the \( L_2 \)-term penalizing the squared coefficients, while the \( L_1 \)-term penalizes their absolute values. Thereby, the first derivative of the \( L_2 \)-term, for very small coefficient values, is approximately zero, meaning that changing the small coefficient values to exactly zero will have little benefit in the minimization problem. As such, the \( L_2 \)-term retains explanatory variables that do not contribute much/any explanatory power in the model, but assigns them a small non-zero coefficient. Contrary, the derivative of the \( L_1 \)-term for very small coefficients is significantly larger than zero. Thus, setting the coefficients equal to exactly zero will have substantial benefits in the minimization problem, since it comes at a large ‘cost’ to include new factors, even with small coefficients. In the formula above, the \( L_1 \)-penalization term is highlighted. As such, LASSO does both continuous shrinkage and automatic variable selection simultaneously (Zou and Hastie, 2005). We highlight that LASSO and the ridge regressions yield basic OLS regressions, when \( \lambda = 0 \).

The different penalization models have shown to perform equally in terms of prediction performance Tibshirani (1996) and Fu (1998), however, as feature selection becomes increasingly
important, one could argue that the LASSO is more appealing, due to its sparse representation. Although showing success in many occasions even the LASSO has its limitations.\textsuperscript{24}

5.1.3 Elastic Net

To accommodate various limitations Zou and Hastie (2005) propose a new regularization technique, the elastic net, which combines the Ridge and LASSO regression techniques. The elastic net is a penalized least squares method using an elastic net penalty combining the $L_1$- and $L_2$-penalty terms. This technique is almost identical to the dual-penalty regression that we later impose in order to estimate the multi-factor model, following Kozak et al. (2020). We therefore provide a brief description of the conventional elastic net estimator proposed by Zou and Hastie (2005).

Suppose a data set with $n$ observations and $k$ predictors. Let $\mathbf{y} = (y_1, ..., y_n)^T$ be the response variable and $\mathbf{X} = (x_1 | ... | x_k)$ be the model matrix, where $\mathbf{x}_j = (x_{1j}, ..., x_{nj})^T$, $j = 1, ..., k$ are the predictors. Assuming the response is centred, and predictors are standardized, then

$$\sum_{i=1}^{n} y_i = 0, \quad \sum_{i=1}^{n} x_{ij} = 0, \quad \sum_{i=1}^{n} x_{ij}^2 = 1, \quad f or \ j = 1,2, ..., k$$

for any fixed combination of non-negative $\lambda_1$ and $\lambda_2$, the elastic net criterion is defined as (Zou et al., 2005)

$$L(\lambda_1, \lambda_2, \beta) = |\mathbf{y} - \mathbf{X}\beta|^2 + \lambda_2|\beta|^2 + \lambda_1|\beta|_1$$

(40)

where

$$|\beta|^2 = \sum_{j=1}^{k} \beta_j^2$$

$$|\beta|_1 = \sum_{j=1}^{p} |\beta_j|$$

\textsuperscript{24} For an overview of the different limitations please see Regularization and Variable Selection via the Elastic Net (Zou and Hastie, 2005).
The elastic net estimator $\hat{\beta}$ is the solution to the minimization problem defined in equation (40).

$$\hat{\beta} = \arg\min_\beta \{L(\lambda_1, \lambda_2, \beta)\}$$

We note that this elastic net function resembles the dual-penalty estimator in equation (56) which we, following Kozak et al. (2020)\(^{25}\), apply in the analysis to estimate the loadings on the stochastic discount factor of our multi-factor models.

5.1.4 Subsection conclusion

In summary, we initially introduced Olsen and Pedersens (2018) methodology criteria, validity, reliability and adequacy, that must be accommodated to ensure the quality of the findings in empirical papers. Subsequently we explained how this paper comply with the listed criteria. Then we introduced one of the most popularly used multiple regression methods, ordinary least squares (OLS) and highlighted a variety of issues with this method, especially connected to high-dimensional data sets. Different proposed solutions in the literature, such as Ridge- and LASSO-regression were outlined before finally explaining the elastic net estimator, which is relevant as it resembles the main estimator proposed in KNS. In the following section, 5.2, we present a thorough examination of the method for building the dual-penalty estimator originating from KNS.

5.2 Building the estimator

As mentioned in the theoretical framework in section 4, this thesis seeks to determine a stochastic discount factor that can price assets based on a broader variety of asset specific characteristics, as opposed to traditional sparse factor models.

Before initiating the actual analysis, we define a characteristics-based strategy on how to set up a factor portfolio for each characteristic. Kozak et al. (2020) presents the methodology in a compressed form, due to the limited extend of conventional publications in financial journals. As such, the following section provides a more interpretable description of the model building process, outlined in KNS.

5.2.1 Computation of factor returns

We apply the method, following KNS, in the definition of each characteristics-based strategy, in which z-scores are used as portfolio weights, such that focus is put exclusively on the cross-sectional

\(^{25}\) This equation can be located in KNS p. 279, equation (28).
aspect of return predictability, outliers are removed, and leverage is kept constant across all portfolios. The transformation is performed for each characteristic at each point in time, $t$.

First, let $c_{S,t}^i$ denote each characteristic, $i$, of a stock, $S$, at any given point in time $t$. We then rank all $c_{S,t}^i$ cross-sectionally across all $S$ from 1 to $n_t$, where $n_t$ represents the number of stocks with available data for a given characteristic at time $t$. The initial rank transformation is then completed by normalizing the rank by a factor $\frac{1}{n_t+1}$:

$$rc_{S,t}^i = \frac{rank(c_{S,t}^i)}{n_t + 1}$$ (41)

Each rank-transformed characteristic $rc_{S,t}^i$ is further normalized by centering cross-sectionally and scaling, dividing by the sum of absolute deviations from the mean of all stocks:

$$z_{S,t}^i = \frac{rc_{S,t}^i - \bar{rc}_{S,t}^i}{\sum_{S=1}^{n_t} |rc_{S,t}^i - \bar{rc}_{S,t}^i|}$$ (42)

Where the mean rank-transform is $\bar{rc}_{S,t}^i = \frac{1}{n} \sum_{S=1}^{n_t} rc_{S,t}^i$.

These resulting Z-scores $z_{S,t}^i$ then serve as the weights of each stock, $S$, in the factor portfolios. By this rank-transformation, normalization and scaling we obtain explanatory variables that are insensitive to outliers, which allows us to keep the factor portfolio for each characteristics-based strategy fixed at a net zero investment. Finally, to compute the factor returns, $F_t$, of all characteristics-based strategies at each point in time, we multiply the $Z_{t-1}$-matrix of z-scores for each stock $S$ and characteristic $i$ at a given point in time by the vector of stock returns $R_t$ similar to equation (24):

$$F_t = Z_{t-1}^t R_t$$

5.2.2 Computation of principal components

In addition to the raw factor returns, we expand our analysis to investigate how principal components work as explanatory variables in the construction of an optimal SDF. In particular, this thesis investigates how rotating factor portfolios into the space of principal components might yield sparse representations of the SDF.

We represent the SDF by principal components following equation (32), reintroduced below:

$$M_t = 1 - \beta_p(P_t - [E]P_t)$$
We construct PCs based on the method described in *KNS* as outlined in section 4.3.6. As such, we set up an eigen decomposition of the factor covariance matrix, in order to compute the PC-factors $P_t$ at each point in time $t$ by equation (31), $(P_t = Q'F_t)$.

With the construction of the inputs to the SDF, by computing the characteristics-based factor portfolios and PCs, we now turn the attention to the task of estimating the factor coefficients in the SDF by deriving the final estimator.

### 5.2.3 Method for coefficient estimator

In this section, we thoroughly outline the intuition behind the estimator of the SDF-coefficients. Define,

$$
\bar{\mu} = \frac{1}{T} \sum_{t=1}^{T} F_t \quad \& \quad \bar{\Sigma} = \frac{1}{T} \sum_{t=1}^{T} (F_t - \bar{\mu})(F_t - \bar{\mu})'
$$

(43)

where $\bar{\mu}$ is mean return of the factor portfolios and $\Sigma$ is the factor covariance matrix as introduced in section 4.3.4.

According to Kozak *et al.* (2020), a natural, but naïve, estimator of the time invariant $\beta$-coefficients in the SDF-model could be constructed based on the sample moment conditions:

$$
\mu - \frac{1}{T} \sum_{t=1}^{T} F_t = 0
$$

(44)

$$
\frac{1}{T} \sum_{t=1}^{T} M_tF_t = 0
$$

(45)

From these conditions one can show, by regressing factor means on the covariances of factors with each other, that the resulting estimator is:

$$
\hat{\beta} = \bar{\Sigma}^{-1} \bar{\mu}
$$

(46)

However, factor mean estimates, $\bar{\mu}$, are fairly imprecise (Kozak *et al.*, 2020). Intuitively, it makes sense, as many of the financial ratios in the data are highly volatile. In our high-dimensional setting, with potentially 69 predictors, $H = 69$, the cross-sectional regression has a vast number of explanatory variables. Consequently, the regression might overfit the factors by fitting the model to noise in factor means, which will lead to highly imprecise $\hat{\beta}$ estimates and poor out-of-sample performance.
Additionally, estimation errors in the covariance matrix can further inflate the problem. One could implement methods to accommodate this issue, such as shrinking the covariance matrix towards the CAPM implied covariance matrix, as proposed by Ledoit and Wolf (2003), however as the main source of fragility originates from the mean factor returns, this thesis is limited to just regularizing the covariance matrix and focusing on mean factor returns. To further mitigate the potential issues of uncertainty in covariance estimates we follow the regularization method by Kozak et al. (2020), where the covariance matrix is regularized by

$$\Sigma^{\text{Regualized}} = a \cdot \frac{\text{tr}(\Sigma^{\text{Unregularized}})}{H} \cdot I + (1 - a) \cdot \Sigma^{\text{Unregularized}}$$  \hspace{1cm} (47)$$

where \( \text{tr}(\Sigma^{\text{Unregularized}}) \) is the trace function summing up the values of the main diagonal (variances) in the covariance matrix, \( a = \frac{H}{H + \tau} \). What this regularization basically achieves is, shrinkage of the covariances towards zero, and shrinkage of the factor specific variances towards the cross-sectional mean variance.

To reduce the uncertainty in the factor means we follow Kozak et al. (2020), introducing economically motivated prior beliefs about factors’ expected returns. Including prior information in the estimation model sufficiently regularizes the estimation, yielding more robust estimates when tested out of sample (OOS) (Kozak et al., 2020). Bringing in existing prior beliefs and updating these using data to yield posterior beliefs is referred to as a Bayesian method (Woolridge, 2012). For the sake of intuition, we proceed under the assumption that \( \Sigma \) is known. Since the mean of the factors in the SDF-model, as mentioned, is the main source of uncertainty, we now focus on the estimation of factor means, \( \mu \).

Earlier asset pricing studies have encountered the family of priors given by:

$$\mu \sim N \left( 0, \frac{\gamma^2}{\tau} \Sigma^{\eta} \right)$$  \hspace{1cm} (48)$$

Where \( \tau = tr(\Sigma) \), and \( \gamma \) is a constant controlling the scale of \( \mu \) dependent on \( \tau \) and \( H \). This general family of priors resembles many asset pricing theories relying on the assumption that first moments of factor returns are connected to their second moments. In this framework, the parameter \( \eta \) controls the shape of the prior, in the sense that it dictates the relationship between first and second moments of factor returns (Kozak et al., 2020).
For the sake of intuition, we now focus on PCs in order to understand the implications of varying the parameter $\eta$. Expressing the family of priors in terms of PCs yields

$$\mu_p \sim N\left(0, \frac{\kappa^2}{\tau} D^\eta\right)$$

(49)

where the covariance matrix, $\Sigma$, is replaced with the diagonal matrix of eigenvalues, $D$.

For the distribution of Sharpe ratios of the PCs one can show that:

$$D^{-\frac{1}{2}} \mu_p \sim N\left(0, \frac{\kappa^2}{\tau} D^{\eta-1}\right)$$

(50)

From equation (50), we investigate assumptions about $\eta$ by considering the Sharpe ratios implied by small-eigenvalue PCs, since the distribution of eigenvalues is typically highly skewed for asset returns. Hence, a few principal components typically account for most of the variance in returns, consequently resulting in the majority of principal components having extremely low eigenvalues.\(^{26}\)

The following is a brief discussion of unrealistic and realistic assumptions regarding the constant $\eta$.

$\eta = 0$

We initially consider the situation of $\eta = 0$. From the assumption about skewed eigenvalues, $\eta = 0$, is evidently economically implausible, since this would imply that the expected Sharpe ratio in equation (50) would be infinitely high for very small eigenvalue PCs. This implies existence of near-arbitrage opportunities, which is implausible in rational expectation models (Kozak et al., 2018).

$\eta = 1$

Considering $\eta = 1$, would imply that the level of eigenvalues of the PCs are of no significance to the Sharpe ratio.\(^{27}\) Meaning that low eigenvalue PC Sharpe ratios are expectedly of the same magnitude as those of high eigenvalue PC. One can argue that this is not an economically plausible assumption. Drawing from rational expectations models, where cross-sectional differences in expected returns arise from exposure to macroeconomic risk factors, risk premia is often focused in one or a few common factors, consequently, implying that low eigenvalue PCs should have lower Sharpe ratios than high-eigenvalue PCs, that are major drivers of risk premia (Kozak et al., 2020).

\(^{26}\) This is also evident from the Scree plot in figure 4.2.

\(^{27}\) The mathematical intuition behind this can be seen when plugging $\eta = 1$ into equation (50) resulting in $D^{1-1} = D^0 = 1$.\[44\]
η = 2

Kozak et al. (2018) argue that an unconstrained rational investor in equilibrium holds a portfolio with finite portfolio weights. Since the optimal portfolio weights of a rational investor and SDF-coefficients are equivalent, the prior should ensure that \( \mathbf{\beta}' \mathbf{\beta} \) remains bounded. Therefore, the expectation \( E[\mathbf{\beta}' \mathbf{\beta}] \) should be bounded as a minimal requirement. Based on the naïve estimator in equation (27), the decomposition of the covariance matrix in equation (28) and the prior, Kozak et al. (2020) shows that:

\[
E[\mathbf{\beta}' \mathbf{\beta}] = \frac{\kappa^2}{\tau} \sum_{i=1}^{H} d_i^{\eta-2}
\]

Where \( d_i \) is the eigenvalue of asset \( i \) in the diagonal matrix \( \mathbf{D} \). According to Kozak et al. (2020) the lowest eigenvalue in the decomposition of a covariance matrix on asset return data is typically extremely close to zero, meaning that \( d_i^{\eta-2} \) is extremely big when \( \eta < 2 \). This implies that an investor would be willing to place huge bets on low eigenvalue PCs, which is counterintuitive to the postulate that a few high-eigenvalue-PCs explains most of the variance in factor returns. By setting \( \eta \geq 2 \) Kozak et al. (2020) avoids such counterintuitive portfolio weights.

In this thesis we determine \( \eta = 2 \) following the framework of Kozak et al. (2020). As such, the prior on the SDF-coefficients follows \( \mathbf{\beta} \sim N \left( 0, \frac{\kappa^2}{\tau} \mathbf{I} \right) \) where the posterior mean coefficient is given by:

\[
\hat{\mathbf{\beta}} = (\mathbf{\Sigma} + \gamma \mathbf{I})^{-1} \bar{\mathbf{\mu}}
\]

Where \( \gamma = \frac{\tau}{\kappa^2 \tau} \) and \( \mathbf{I} \) is an identity matrix of size \( H \times H \) with the main diagonal containing ones, and all other entries containing zeros.

5.2.4 Interpreting the estimator

Expressing the new estimator in equation (52) and the naïve estimator in equation (46) in terms of PC-returns, \( \mathbf{P}_t = \mathbf{Q}' \mathbf{F}_t \) we see an interesting difference. For the naïve GMM estimator, the corresponding PC coefficient estimator is, following Kozak et al. (2020):

\[
\hat{\mathbf{\beta}}_{ols}^{P,j} = \frac{\hat{\mathbf{\mu}}_{P,j}}{d_j}
\]

And for the new estimator in (52) it is:
\[ \hat{\beta}_{p,j} = \left( \frac{d_j}{d_j + \gamma} \right) \frac{\bar{\mu}_{p,j}}{d_j} \]  

(54)

From the two equations it is evident that compared to the classic GMM estimator, the estimator in equation (53) shrinks the SDF-coefficients toward zero by a shrinkage factor \( \frac{d_j}{d_j + \gamma} < 1 \). From this estimator it is important to notice that small eigenvalue PC-coefficients are shrunk particularly more than high eigenvalue PC-coefficients, as the numerator in the shrinkage estimator, \( d_j \), represents the eigenvalue. This feature of the estimator draws from the assumption that PCs with low eigenvalues contribute an insignificantly amount to overall volatility of the SDF. Hence, shrinking these small eigenvalue PCs towards zero increases the parsimony of the model while only insignificantly affecting the prediction performance.

5.2.5 Adding penalization

So far, in the introduction of the estimator we have counted on classic Bayesian statistics. Subsequently, we outline how Kozak et al. (2020) transform this estimator into a penalized estimator similar to common penalized estimators, in the machine learning literature, such as ridge regression and LASSO. In order to maximizing the cross-sectional \( R^2 \) subject to a penalty term, \( \gamma \beta' \Sigma \beta \), we minimize:

\[ \hat{\beta} = \arg\min_{\beta} \{ (\bar{\mu} - \Sigma \beta)' \Sigma^{-1} (\bar{\mu} - \Sigma \beta) + \gamma \beta' \beta \} \]  

(55)

Equation (55) estimates coefficients by minimizing the HJ-distance subject to an \( L_2 \)-penalty term (Hansen & Jagannathan, 1997). Looking at this estimator in terms of factor returns, transformed into their principal components, it is intuitively shown, how the \( L_2 \)-penalty term induces higher shrinkage on low-eigenvalue PCs as compared to high-eigenvalue PCs. For instance, if the estimator shrinks the coefficient \( \beta_{p,j} \) for low-eigenvalue PCs toward zero, this would lead to a decrease in the penalty term \( \gamma \beta' \beta \), which is preferable in the minimization problem. Furthermore, it would come at low ‘cost’ since a low eigenvalue PC has an insignificant contribution to the total variance of the SDF. Thus, it has little effect on the HJ-distance, which can be thought of as a measure of the pricing error of the SDF.

Alternatively, shrinking high-eigenvalue PCs would bring similar penalty benefits but at a much larger cost, since it would remove a major source of variance in the SDF and thereby have a significant
impact on the HJ-distance, which lowers the explanatory power of the SDF. Consequently, this estimator automatically tilts towards shrinking low eigen-value PCs, which is preferable.

The estimator presented in equation (55) is very similar to that of a ridge regression, described in section 5.1.2, which is a common technique in machine learning (Hastie et al., 2009). However, a standard ridge regression would include a penalty term on the $L_2$-coefficients $\beta' \beta$, or weight the pricing errors with the identity matrix as opposed to the estimator in equation (55), which weighs the pricing errors by $\Sigma^{-1}$. Kozak et al. (2020) argues that the standard ridge regression corresponds to choosing $\eta = 3$ in their estimator, which would lead to further shrinkage of the low eigenvalue PCs.

5.2.6 Sparsity

So far, the estimator obtained in equation (55) accommodates the high-dimensionality challenge by shrinking SDF-coefficients of low eigenvalue PCs toward zero. However, none of the SDF-coefficients are set exactly equal to zero, which can have some implications, since the model will tend to overfit the data by including insignificant factors. Hence, the next objective is to impose some degree of sparsity on the estimator making certain factors redundant in terms of their contribution to the explanatory power of the SDF.

To obtain true sparsity, a proposed solution introduces an additional $L_1$-penalty, $\gamma_1 \sum_{j=1}^{n} |\beta_j|$, in the estimator (Kozak et al., 2020). This approach combines LASSO and ridge regression and resembles Zou and Hasties (2005) elastic net approach previously described in section 5.1.3. Introducing this term, forces some SDF-coefficients equal to zero, resulting in a sparse model with automatic factor selection. The degree to which the model imposes sparsity is controlled by the strength of the $\gamma_1$-parameter. With the introduction of the $L_1$-term, the estimator now penalizes the absolute value of SDF-weights, therefore this new term shrinks big weights substantially more than small weights.

With only the first penalty term, $L_2$, applied, the model potentially obtains very small weights, since the first derivative will be approximately zero, why changing their weight to exactly zero has almost no effect on the SDF. However, with the new penalty term $L_1$ the first derivative of small coefficients is significantly different from zero, why changing the coefficients to exactly zero has a larger impact on the SDF. Hence, factors that do not contribute much explanatory power to the SDF are excluded from the model (given a zero $\beta$-value) instead of having a very small but non-zero weight.
Hence the new estimator can be defined as:

\[
\hat{\beta} = \arg\min_{\beta} \left\{ (\bar{\mu} - \Sigma \beta)' \Sigma^{-1} (\bar{\mu} - \Sigma \beta) + \gamma_2 \beta' \beta + \gamma_1 \sum_{j=1}^{H} |\beta_j| \right\}
\] (56)

This is the main estimator, which is applied throughout the analysis. An important feature of the estimator is that by setting \( \gamma_2 \) or \( \gamma_1 \) equal to zero the estimator resembles LASSO- or ridge-regression, respectively.

With the definition of the main estimator-function outlined, the subsequent problem is to solve it efficiently. Since the estimator does not imply a closed form solution, we need a computer-based solver in order to solve the problem. As previously mentioned, the analysis bases upon 69 and 50 anomalies respectively, meaning that the optimal solution of the estimator potentially involves the estimation of 69 and 50 SDF-coefficients respectively. This implies that a computer-based solver would have to loop over all possible values of every coefficient for each possible value of one coefficient. Hence, assuming for the WFR data, that we limit our solver to solutions of \( \beta \in [0.01, 0.02, 0.03, \ldots, 1] \) our solver would have a number of iterations equal to \( 69^{100} = 7.671 \cdot 10^{183} \). Even though some, high-capacity, computers might have the power to handle such an extensive number of iterations, we mitigate this problem by applying the LARS-EN algorithm initially proposed by Zou and Hastie (2005).

### 5.2.7 The LARS-EN algorithm

The LARS-EN algorithm is based on the subsequent algorithm LARS designed by Efron, Hastie, Johnstone and Tibshirani (2004). They proposed a new algorithm, LARS, to solve the entire LASSO solution path efficiently by using the same relatively simple order of computations as a single OLS fit. As this thesis will apply the LARS-EN algorithm through the statistical tool, R, we briefly outline the basic intuition of the LARS algorithm below following Efron et al. (2004).

Least Angle Regression (LARS) is based on the intuition behind the classic model-selection method, Forward selection, described in Weisberg (1980). Given a collection of possible predictors, one selects the predictor having the highest correlation with the response \( y \), say \( x_{j1} \), and performs simple linear regression of \( y \) on \( x_{j1} \). This leaves a residual vector orthogonal to \( x_{j1} \), which is now

---

considered the response. Then the other predictors are projected orthogonally to \( x_{ij} \) and the selection process is repeated. After \( k \) steps this method results in a set of \( k \) predictors \( x_{j1}, x_{j2}, \ldots, x_{jk} \) that are used to construct a \( k \)-parameter linear model. Forward selection has been criticized for being an aggressive fitting technique that often is overly ‘greedy’, resulting in elimination of useful predictors that happens to be correlated with \( y \) (Efron et al., 2004). To accommodate this issue, the LARS procedure was introduced, working roughly as follows: Similar to Forward Selection, we start with all coefficients equal to zero, and identify the predictor most correlated with the response variable, say \( x_{j1} \). The model takes the largest possible step in the direction of this predictor, until some other predictor, \( x_{j2} \) is equally as correlated with the current residual. Different from forward selection, instead of continuing along \( x_{j1} \), LARS proceeds in a direction equiangular\(^{29}\) with the two predictors, until a third predictor \( x_{j3} \) makes its way into the “most correlated” set. The LARS proceeds equiangularly between \( x_{j1}, x_{j2} \) and \( x_{j3} \) until a fourth variable, \( x_{j4} \), enters the set and so forth. As such, LARS builds up the solution in successive steps, with each step adding an additional covariate to the model, so that after \( k \) steps, we have just \( k \) estimates of, \( \hat{\beta}_j \) that are non-zero coefficients (Efron et al., 2004). The intuition and the procedure of the LARS algorithm is relatively comprehensible, however it is a highly advanced statistical tool, why the mathematics behind it, is overly complex. As such, an in-depth decomposition of the algorithm and the underlying algebra is outside the scope of this thesis\(^{30}\).

The LARS-EN algorithm proposed by Zou et al. (2005) solves the elastic net problem explained in section 5.1.3, by applying the LARS algorithm. As the elastic net problem is similar to the minimization problem of the main estimator, we, aligned with Kozak et al. (2020), apply the LARS-EN algorithm when estimating the SDF-loadings.

5.2.8 Choosing the parameters

In order to implement the estimator in (56) we need to identify the optimal values of the parameters \( \gamma_1 \) and \( \gamma_2 \). When setting the parameter \( \gamma_2 = \frac{\tau}{\kappa^2 T} \) it has a meaningful economic interpretation. By setting \( \eta = 2 \), the root expected maximum squared Sharpe ratio under the prior can be defined as

\(^{29}\) Equiangular, means that they have an equal angle to each other, for example, an equiangular triangle, is a triangle, where all angles are exactly equal = 60°

\(^{30}\) For an in-depth decomposition of the algebra underlying the LARS-algorithm we refer to Least Angle Regression (Efron et al., 2004)
\[ E[\mu \Sigma^{-1} \mu]^\frac{1}{2} = \kappa \] (57)

Intuitively, this implies that \( y_2 \) by construction represents some views on the expected squared Sharpe ratio. If one expects the Sharpe ratio to be very high, \( \kappa \) would equally be high, which would impose low degree of shrinkage in the estimation since \( y_2 \) would consequently be small (Kozak et al., 2020).

To identify the optimal values of \( y_1 \) and \( y_2 \), Kozak et al. (2020) propose a data-driven approach as their preferred choice. By \( k \)-fold cross-validation, which is a common machine learning technique, the historic data is divided into \( k \) contiguous equal samples, and for each sample the optimal choice of parameters \( y_1 \) and \( y_2 \) is chosen based on the best out of sample (OOS) \( R^2 \) given the formula:

\[
R_{OOS}^2 = 1 - \frac{(\bar{\mu}_2 - \bar{\Sigma}_2 \bar{\beta})'(\bar{\mu}_2 - \bar{\Sigma}_2 \bar{\beta})}{\bar{\mu}'_2 \bar{\mu}_2} \] (58)

The procedure is repeated \( k \) times, each time treating different folds.

In the following analysis, the first part will focus on identifying the optimal values of \( y_1 \) and \( y_2 \) as described above, yielding the beta coefficients of the stochastic discount factor, both in the space of factor returns and PCs. Subsequently we intend to test the true OOS asset pricing implications of the models that the estimator in equation (56) proposes, through Mean-Variance Efficient (MVE) portfolio construction.

### 5.2.9 Subsection conclusion

In this section we inferred an estimator in equation (56) for the estimation of coefficients spanning the SDF. This estimator combines ridge regression and LASSO regression to an elastic net equivalent function, which we seek to minimize in order to deduct the SDF loadings. We introduced the LARS-EN algorithm which we intend to apply when solving this minimization problem. We then showed how we intend to apply a data-driven approach in choosing the optimal parameters controlling the level of sparsity and shrinkage through the two penalty terms \( L_1 \) and \( L_2 \).
6 - Analysis

Before initiating the analysis, the main findings of Kozak et al. (2020) are briefly highlighted, as we refer to these when assessing whether we successfully manage to replicate the paper.

6.1 - Findings from Kozak et al. (2020)

Kozak et al. (2020) finds that sparse characteristics models (like the conventional 3, 4 and 5-factor models) return poor OOS performance: “There simply is not enough redundancy among the large number of cross-sectional return predictors that have appeared in the literature for such a characteristics sparse model to adequately price the cross-section.” – Kozak et al. (2020) p. 291. Thus, to perform well, Kozak et al. (2020) finds that the SDF needs to load on a large number of characteristics-based factors. In order to achieve a model handling a vast number of characteristics spanning the SDF, Kozak et al. (2020) shows that an adequate degree of $L_1$- and $L_2$-shrinkage must be imposed. Subsequently, the results in KNS suggests that a sparse SDF, in the space of principal components, suffices to achieve good out-of-sample performance. Finally, the asset pricing implication of the proposed method is tested in a true OOS setting, by assessing the performance of MVE portfolios. Here, KNS finds that the portfolios based on an $L_2$-only estimator outperforms both traditional factor models and the characteristics-sparse models based on the dual-estimator.

6.2 - Analysis of dual-penalty estimator method and SDF-coefficients

In the following analysis we seek to empirically test the performance of the proposed estimator in equation (56). As previously mentioned, a ‘good’ estimation model performs well in terms of predictive precision and parsimony (Zou et al., 2005). As such, we evaluate the estimation method based on these two criteria. In this thesis, following Kozak et al. (2020) the analysis is divided into two subsections. Initially, we evaluate the out-of-sample performance based on $OOS \ R^2$ through k-fold cross-validation. However, since the optimal values of the shrinkage parameters, $\gamma_1$ and $\gamma_2$ are identified through a data-driven approach, using the entire data period (through 3-fold cross-validation), one could argue that the resulting $OOS \ R^2$, for this particular analysis, is not a true out of sample measure. As such, we subsequently create SDF-implied MVE portfolios to evaluate the asset pricing implications of the optimal SDF in a true out of sample setting.

Before moving into the first part of the analysis we briefly describe the practical approach of building the estimator and the underlying data processing in the following section. Furthermore, we carefully describe the two main data sets underlying our analysis.
6.2.1 Data analysis

Following Kozak et al. (2020), the set of characteristics, in the 50 anomalies data set, relies on features underlying common anomalies, that have been proposed in the literature and sought exploited by asset managers in the past. This set involves 50 different characteristics, representing anomalies. Here, we follow the anomaly definitions of Kozak et al. (2020), who have constructed the 50 anomalies portfolio based on previous work by, Novy-Marks and Velikov (2016), McLean and Pontiff (2016), Kogan and Tian (2015) and Hou, Xue and Zhang (2015). We consider data in the period January 1973 to December 2017.

The second set of characteristics are based on 69 asset specific financial ratios from WRDS. The data set is defined by WRDS: “WRDS Industry Financial Ratios (WFR) is a collection of most commonly used financial ratios by academic researchers (often for purposes other than return prediction). There are in total over 70 financial ratios grouped into the following seven categories: capitalization, efficiency, financial soundness/solvency, liquidity, profitability, valuation, and others.”31 We have excluded one of these 70 parameters, as the data available for this ratio, in the time-period selected for the empirical test, was not sufficient. We therefore define the WFR data set as these 69 different characteristics. Definitions of all variables in both data sets are provided in Appendix 2 and 3.

As outlined in section 5, we require data on firm specific returns in order to compute the WFR factor returns. Since the financial ratios in the WFR belong to firms in the CRSP asset universe, we withdraw returns for 505 companies in the S&P 500 through the DataStream portal at the CBS Library Data lab. We consider data in the period January 2000 to December 2020. The reason for the relatively short time period, is computational considerations. Particularly, in the calculation of the WFR factor returns, multiplying returns for 505 companies, with 69 financial ratios on daily data for a 20-year period requires millions of iterations, which already challenged the capacity of accessible computers. As such, extending the time frame would require infeasible computational capacity32. This is potentially a limitation of the analysis, however, we discuss the effects of this limited time frame in section 7.1.2 of the discussion.

---

31 The data as well as the description cited can be located at: https://wrds-www.wharton.upenn.edu/pages/grid-items/financial-ratios-firm-level/ last visited 06.05.2021

32 As we are limited to the use of private computers, due to the COVID-19 pandemic, we were not able to handle more data for this step. We therefore assessed, that 20 years of data was sufficient.
For each data set, we perform certain alterations to the raw data. In particular, following Kozak et al. (2020), we orthogonalize all factor returns with respect to the market return. We apply the CRSP value weighted index (including distributions) as the ‘market return’\textsuperscript{33}. This market index is calculated based on the following groups of securities: NYSE, AMEX, and NASDAQ. The orthogonalization is executed, for each individual factor return, running a time series regression, where the factor return is the dependent variable, and the market return is the explanatory variable. The beta coefficient of this simple OLS-regression is then applied in the following way to construct orthogonalized factor returns:

\[ F_{i,t} = \bar{F}_{i,t} - \beta_{i} R_{m,t} \] (59)

Next, in order to make the factor returns even more comparable internally, we adjust the variance to follow the market return, by multiplying each factor return \( F_{i,t} \) by the factor \( \frac{\sigma_{m}}{\sigma_{F_{i,t}}} \), since,

\[ Var\left( \frac{F_{i,t}}{\sigma_{F_{i,t}}} \cdot \frac{\sigma_{M}}{\sigma_{F_{i,t}}} \right) = \frac{\sigma_{M}^{2}}{\sigma_{F_{i,t}}^{2}} Var\left( F_{i,t} \right) = \frac{\sigma_{M}^{2}}{\sigma_{F_{i,t}}^{2}} \sigma_{F_{i,t}}^{2} = \sigma_{M}^{2} \] (60)

Consequently, we end up with comparable factor returns, with equal variance and all orthogonalized with respect to the market return.

Before inserting the factor returns in the dual-penalty estimator, we conduct an initial characterization of the factor returns in order to obtain an understanding of the underlying factor structure, on which we apply the dual-penalty estimator. In Appendix 6, we provide summary statistics of all factor returns for the 50 anomalies and WFR data sets. The summary statistics show that the magnitudes of the factor returns are relatively low. In order to further describe the magnitude of the factor returns, figure 6.1 illustrates the mean factor returns for the portfolios in each data set.

\textsuperscript{33} We withdraw the CRSP value weighted index from the Wharton Research Data Services
From the histograms we note that most factor portfolio mean returns are close to 0, with only small deviations. This was expected as we are considering mean returns. We refrain from commenting on the return distribution of factor returns, as we need not consider the distribution of raw returns, since the SDF-approach does not assume any underlying distribution. As such, we include the figure above, in order to illustrate the magnitude of the factor returns.

We subsequently consider the collinearity of the factor portfolios. As outlined in the theoretical framework, the dual-penalty estimator is capable of handling multicollinearity in the explanatory variables. This is one of the main advantages of the method, as opposed to a standard OLS regression, which becomes inaccurate for highly correlated explanatory variables. In order to characterize the structure of the data, in terms of correlation between variables, we provide a heatmap of correlations between factor portfolios (the independent variables) for both data sets in figure 6.2. For a more detailed version of the correlation heatmaps in figure 6.2 (a) and (b), we refer to Appendix 4 and 5, where the name of each anomaly or financial ratio, underlying the factors, are displayed.

---

34 The effect of multicollinearity of independent variables on OLS regression is illustrated in figure 5.1
Figure 6.2: The heat map illustrates intercorrelation between dependent variables for the two high-dimensional data sets, 50 anomalies in figure (a) and WFR in figure (b). The factor returns of the 50 anomalies are not sorted in any specific order as opposed to the WFR data set, where the dependent variables are sorted based on the type of financial ratio underlying the factor, to illustrate the high intercorrelation between variables within the same categories. As such, the WFR financial ratios are sorted in the following categories (position on the axes in (b)); Capitalization (1–4), Efficiency (5–11), Financial Soundness (12–27), Liquidity (28–31), Other (32–35), Profitability (36–50), Solvency (51–56) and Valuation (57–69). The sample data is daily from 1973 to 2017 (a) and 2000 to 2020 (b).
From figure 6.2 (a) and (b) it is evident that there is a significant degree of multicollinearity in both data sets. In the 50 anomalies data there are several dark areas indicating highly negatively correlated portfolios. We do note smaller clusters of significant correlation (yellow colors), however only a few of these are highly correlated (indicated by bright yellow colors). We highlight that the diagonal in both heat maps obviously equals one. For the WFR data, the plot clearly shows several bright yellow areas implying highly positively correlated variables, as expected, since the WFR data consists of several financial ratios that are based on the same variables. For instance, share price is included in Price/Earnings, Diluted Price Earnings, Price/Operating Earnings etc. why a high level of correlation is expected in the data. The WFR data set in figure 6.2 (b) is sorted based on seven financial ratio-categories: Capitalization (1-4), Efficiency (5-11), Financial Soundness (12-27), Liquidity (28-31), Other (32 – 35), Profitability (36-50), Solvency (51 – 56) and Valuation (57-69). The parenthesis depicts the placement of the ratios along both axes. As such the heatmap in figure 6.2 (b) shows high intercorrelation among financial ratios categorized as Profitability and Valuation, respectively. We note that the financial ratios are correlated within the two categories, and additionally, across the two categories. As such, the correlation plots in figure 6.2 clearly motivate the need for a model capable of handling strong collinearity in the explanatory variables.

The choice of estimator, as outlined in the methodology section, form the SDF from a combination of factors rather than choosing factors that are individually significant. In addition, it shrinks contribution of some factors to exactly zero. Hence, factors explaining the same, i.e., do not contribute additional explanatory power, should automatically be left out of the model. This feature is highly desirable for a strongly correlated factor structure, as the ones displayed above.

With the careful description of data underlying the analysis in place, the following section evaluates the performance of the dual-penalty estimator, following the methodology outlined in section 5.2.

**6.2.2 - 25 Fama & French portfolios**

Initially, we conduct a preliminary analysis, applying the proposed estimator in equation (56) on returns of the 25 Fama-French ME/BE-sorted portfolios (FF25) in the time period from July 1926 to December 2017. These 25 portfolios are sorted on the SMB and HML factors introduced by Fama & French (1992). We treat the 25 portfolio membership indicators as stock characteristics and estimate the SDF loadings (regression coefficients) on these 25 portfolios35. One could argue that these

---

35 The 25 portfolios are based on the initial three factor model presented by Fama & French (1992) that includes Size and value anomalies in describing cross-section of returns. (continued in the footnotes of the following page)
portfolios are not the challenging high-dimensional problem that the estimator was designed for, however, for verification purposes, we first apply the estimator on these relatively comprehensible portfolios, before moving into more complex and statistically challenging sets of asset returns. Intuitively, the FF25 factor returns have such a strong factor structure that the 25 portfolio returns, orthogonalized with respect to the market index, are almost linear combinations of the SMB and HML factors (Kozak et al., 2020). Consequently, any selection of a number of portfolios with somewhat alternating loadings on the SMB an HML should be sufficient to span the SDF. Hence, treating the portfolio membership indicators as characteristics, the estimator should present optimal solutions with significant sparsity.

Kozak et al. (2018) has shown that the SMB and HML factors match the 1st and 2nd PCs of the FF25 portfolio returns, i.e., the factors with the most predictive power on the cross-sections of returns. As such, we would expect even more sparsity, when rotating the SDF into the space of principal components. To show the impact of rotating returns into principal components we run the analysis using the PCs of the FF25 portfolio returns as the basis assets.

Figure 6.3 below presents results from applying the dual-penalty estimator in equation (56). The results from the raw FF25 factor returns are shown in figure 6.3 (a) and the SDF in the space of principal components are pictured in figure 6.3 (b).

The portfolio membership indicators refer to the sorting of portfolios based on the assets’ rank in Size (Market Cap) and Value (Book-to-market Ratio). The portfolio returns are available on the website of Kenneth French: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html#Developed
Figure 6.3: The contour plot shows the OOS $R^2$ from the dual-penalty specification, in equation (56), on the Fama-French 25 ME/BM sorted portfolios. OOS cross-sectional $R^2$ for families of models that employ both $L_1$ and $L_2$ shrinkage simultaneously using 25 Fama-French ME/BM-sorted portfolios (panel a) and 25 PCs based on the returns of the Fama-French portfolios (panel b). We illustrate the degree of $L_2$-shrinkage by the prior root expected squared Sharpe Ratio ($\bar{S}^2$) ($\kappa$) on the horizontal axis. We show the number of included variables in the SDF on the vertical axis, which quantifies the degree of $L_1$-penalty imposed. Warmer (yellow) colors depict higher levels of OOS $R^2$. Both axes are plotted on logarithmic scale to enhance interpretability of the contour plot. The sample data is daily from 1926 to 2017.
Every point on the plane, in the plots above, represent a particular combination of the penalty coefficients $\gamma_1$ and $\gamma_2$ that control the level of sparsity and shrinkage, respectively. The horizontal axis represents the level of $L_2$-shrinkage, going from no shrinkage at the right border to extreme shrinkage at the left border of the plot. The degree of $L_2$-shrinkage is, in this case, depicted as $\kappa$ for the sake of interpretation. Kozak et al. (2020) argues that in the $L_2$-only penalty case, $\kappa$ can be economically interpreted as the square root of the expected maximum squared Sharpe Ratio under the prior in equation (48). $\kappa$ is inversely related to the shrinkage coefficient $\gamma_2$, as $\gamma_2 = \frac{x}{k^{2+\tau}}$. Note that smaller levels of $\kappa$ implies a high degree of shrinkage, as $\gamma_2$, consequently, will be large. The right-hand side of the plot, indicating that no $L_2$-shrinkage is imposed can intuitively be interpreted as a pure LASSO-regression with only the $L_1$-penalty term.

Variation along the vertical axis represents the degree of sparsity imposed on the model. The degree of sparsity is expressed in terms of how many factors remain in the SDF with non-zero coefficients. As such, the bottom end of the range, with only one non-zero coefficient implies the maximum level of sparsity, and the top of the plot implies no sparsity at all, with all potential factors included in the SDF. As mentioned in section 5.1.3, the number of non-zero coefficients is directly related to the number of steps in the LARS-EN algorithm, as each step adds another non-zero coefficient to the SDF.

To make the plot more interpretable both axes are depicted on logarithmic scale. The contour map shows the OOS $R^2$ calculated as in equation (58) for each combination of $\gamma_1$ and $\gamma_2$. The data-driven penalty method chooses the optimal combination of penalty coefficients that maximizes OOS $R^2$, however, for the sake of intuition we show OOS $R^2$ for a wide range of penalties to illustrate how $L_2$-shrinkage and sparsity ($L_1$-penalty) influence the OOS $R^2$. Warmer (yellow) colors indicate higher OOS $R^2$.

Focusing in on figure 6.3 (a), depicting OOS $R^2$ for different combinations of $\gamma_1$ and $\gamma_2$ for the raw FF25 factor returns, we see that sparsity and shrinkage are inversely correlated/substituted, in terms of creating good OOS performance. We see that the ridge (yellow area of high OOS performance) draws from the top part of the contour map, indicating no sparsity ($L_1$) and substantial $L_2$-shrinkage, to the right-hand side of the map, indicating $L_1$-sparsity, but no shrinkage. Since the factor structure is strong in the FF25 we would expect a selection of only a handful of these portfolios to be sufficient to span the SDF, which is also illustrated by the analogous contour plot in KNS. As expected, the ridge of high-OOS $R^2$ extends as far down as three non-zero coefficients. Adding more
non-zero coefficients to the SDF hurts the OOS performance, unless more $L_2$-shrinkage is imposed to mitigate overfitting. Furthermore, we highlight, that a completely unregularized model, illustrated in the top right corner, has very poor out-of-sample performance.

Figure 6.3 (b), representing the principal components of the FF25 portfolio returns, shows, what we would also expect from our preliminary analysis of PCs in section 4.3.6, that a small number of PCs sufficiently span the SDF. Imposing sparsity, such that the model only contains two non-zero coefficients is sufficient to span the SDF and achieve maximum OOS $R^2$ given some substantial level of $L_2$-shrinkage. Furthermore, given the vertical direction of the ridge of close to maximum OOS $R^2$, we note that relaxing the $L_1$-penalty term, (resulting in less sparsity) while holding a constant level of $L_2$-shrinkage does not significantly hurt the OOS performance in the PC space. This can be explained by the $L_2$-penalty in the estimator from equation (55), when applied in terms of principal components, as shown in equation 54 (highlighted again below). For principal components, the $L_2$-penalty already downweighs low variance PCs by forcing their SDF-coefficients close to zero.

$$\hat{\beta}_{p,j} = \left( \frac{d_j}{d_j + \gamma} \right) \frac{\hat{\beta}_{p,j}}{d_j}$$

Applying the beta estimate of principal components restated above, in equation (55), the $L_2$-penalty term already downweighs low variance PCs by forcing their SDF-coefficients close to zero. Consequently, it makes little difference whether these coefficients are left close to zero, without $L_1$-penalization, as in the top of the map, or completely forced to zero by imposing sparsity. Figure 6.3 (a) and (b) shows, that the estimator in a setting of PCs of factor returns successfully manages to impose more sparsity at an optimal level of $L_2$-shrinkage, without hurting the OOS prediction performance.

So far, we have shown that the estimator successfully imposes shrinkage and sparsity to improve the prediction capabilities, illustrated by higher OOS $R^2$, as well as the parsimony of the model. Furthermore, the initial analysis shows that rotating the SDF into a space of PCs of factor returns allows us to reduce the number of non-zero coefficients, while keeping a constant level of $L_2$-shrinkage around $\kappa = 0.6$. Credited to the strong factor structure of the FF25 portfolios, we were, similar to KNS, able to obtain a characteristics-sparse SDF in the space of raw factor returns for this data set. However, the next step is to test the estimator on more complex and statistically challenging settings, which is the intention of the dual-penalty estimator (Kozak et al., 2020). As the true value of the dual-penalty estimator is to accommodate a multi-dimensional setting with a vast abundance
of characteristics, where classic estimators fail, we apply two different sets of characteristics with a large number of features to validate our findings across different data settings.

### 6.2.3 - 50 anomaly portfolios

We start by focusing on the fifty portfolios based on the anomaly characteristics described in section 6.2.1. Identical to figure 6.3, figure 6.4 presents the OOS $R^2$ as a function of the degree of $L_2$- and $L_1$-penalty imposed, depicted as $\kappa$ and the number of non-zero coefficients, respectively. We highlight that both axes are depicted in logarithmic scale to improve interpretability. Looking at figure 6.4 (a) there are some similarities, but also significant differences from the corresponding plot for the FF25 factor returns in figure 6.3 (a).

**Figure 6.4:** The contour plot shows the OOS $R^2$ from the dual-penalty specification, on the 50 anomaly portfolios. OOS cross-sectional $R^2$ for families of models that employ both $L_1$- and $L_2$-shrinkage simultaneously on 50 anomaly portfolios (panel a) and 50 PCs based on the returns of the anomaly portfolios (panel b). We illustrate the degree of $L_2$-shrinkage by the prior root expected squared Sharpe Ratio ($SR^2 = (\kappa)$) on the horizontal axis. We show the number of included variables in the SDF on the vertical axis, which quantifies the degree of $L_1$-penalty imposed. Warmer (yellow) colors depict higher levels of OOS $R^2$. Both axes are plotted on logarithmic scale to enhance interpretability of the contour plot. The sample data is daily from 1973 to 2017.
Figure 6.4 (a) clearly illustrates that an unregularized, i.e., completely unshrunked model with no sparsity imposed, has an extremely poor OOS performance with OOS $R^2$ below zero, due to spurious overfitting.

We once again see a slightly negatively correlated relationship between sparsity and $L_2$-shrinkage, as the ‘ridge’ of high OOS $R^2$ starts in the top of the plot, at a substantial level of shrinkage and no sparsity, and moves diagonally downwards in the direction of the right-hand side of the plot as more sparsity is imposed. However, unlike the FF25 portfolios, this relation is rather weak. The biggest difference is the limited number of sparsity one can impose without hurting OOS $R^2$. The highest OOS $R^2$ values, depicted with deep yellow colors, are only reached at a high number of non-zero coefficients. This implies that a large number of non-zero coefficients is required to capture the variance sufficient to span the SDF. Imposing sparsity results in a major deterioration of the OOS $R^2$, as almost every characteristic has significant predictive power, indicating almost no redundancy among the 50 anomalies. Or put differently; to adequately capture the pricing information of the 50 anomalies one needs to roughly include all, or close to all, of the 50 factors.

As such, we conclude that substantial shrinkage is needed for good OOS $R^2$ for the 50 anomalies data. However, forcing coefficients to zero, in order to obtain a parsimonious model comes at a heavy cost of poor OOS performance. These findings are consistent with the analogous plot presented by
Similar to the findings of KNS, figure 6.4 (a) shows that a characteristics-sparse SDF with good pricing performance does not exist, implying that many of these anomalies actually make significant marginal contributions to OOS explanatory power of the SDF. This demonstrates the urgent need for a method designed to impose sparsity absent significant tradeoff of pricing performance.

Naturally, this leads us to figure 6.4 (b), which illustrates the OOS $R^2$ at given combinations of $\gamma_1$ and $\gamma_2$ representing the degree of $L_1$- and $L_2$-shrinkage, respectively, in the space of principal components of factor returns for the 50 anomaly portfolios. The most notable benefit of rotating factor returns into PCs is the significantly increased sparsity, without deteriorating the OOS pricing performance. Just a few PCs are sufficient to adequately span the SDF. We observe relatively high OOS $R^2$ values at only three non-zero coefficient, and close to the maximum OOS $R^2$ is obtained at four non-zero coefficients. Thus, a PC-sparse SDF seems to price the anomalies fairly well. The ridge of high OOS $R^2$ values can be roughly described as an “L”-Shape. Similar to the PC space of the FF25 we see that $L_2$-shrinkage can be held relatively constant, at some substantial level, while imposing more or less sparsity without hurting the OOS $R^2$ due to the $L_2$-penalty term in equation (55). However, when imposing extreme sparsity, resulting in five non-zero coefficients or less, we must relax the degree of $L_2$-shrinkage to avoid underfitting the high-variance PCs. This is evident as the ridge takes a rather sharp right-turn around the 4th or 5th principal component, creating the described “L”-shape. The findings above are broadly consistent with the conclusions presented in KNS, and the levels of OOS $R^2$ are similar, though the contour plots are not completely identical. In a later section of the thesis, we discuss possible reasons for the slight deviations.

To further illustrate the effect of $L_1$-sparsity and $L_2$-shrinkage on the OOS $R^2$ for the data set of 50 anomaly portfolios, figure 6.5 depicts different cuts of the contour plots in figure 6.4. In figure 6.5 (a) we illustrate the importance of $L_2$-shrinkage by taking a cut along the top edge of the contour plot in figure 6.4 (a), including all possible variables as non-zero coefficients, i.e., imposing no sparsity. Along the horizontal axis, we vary $\kappa$, representing the degree of $L_2$-shrinkage, with maximum shrinkage in the left-hand side of the plot and no shrinkage in the right-hand side. The vertical axis represents the resulting OOS $R^2$ value, imposing no sparsity. As no $L_1$-shrinkage is imposed, this plot can be interpreted as the performance of a resulting ridge regression, since only the $L_2$-penalty term

---

36 The analogous plot to figure 6.4 is illustrated in figure 3 in * Shrinking the cross-section* (Kozak et al., 2020) which is located on page 284.
is in play. The blue line in figure 6.5 (a) illustrates the OOS $R^2$ from the cut along the top edge of the contour plot, and the green line illustrates the in-sample cross-sectional $R^2$.

![L2 model selection - 50 Anomalies](image)

![Sparsity - 50 Anomalies](image)

Figure 6.5: $L_2$ model selection and sparsity (50 anomaly portfolios).

Panel (a) plots the in sample cross-sectional $R^2$ (green) and OOS cross-sectional $R^2$ based on 3-fold cross-validation (blue), with no sparsity imposed. Panel (b) shows the maximum OOS cross-sectional $R^2$ attained by a model with $n$ factors (on the horizontal axis) across all possible values of $L_2$ shrinkage, for models based on original characteristics portfolios (blue) and PCs (green). The sample data is daily from 1973 to 2017.

Note that both high levels of sparsity, and no sparsity enforces either under-fitting or over-fitting of coefficients respectively, resulting in poor OOS performance. As such, there is an optimal level of shrinkage. As figure 6.5 (a) shows, the OOS $R^2$ is maximized around $\kappa \approx 0.6$. We have illustrated the in-sample performance, to show that a regression model without shrinkage (as $k$ increases), such as the OLS, is inclined to significantly overfit the estimator. With the significant difference between the blue and the green line, we highlight that in-sample $R^2$ can be a crucially misleading metric for OOS performance, especially in a high-dimensional data setting.

Figure 6.5 (b) illustrates the OOS $R^2$ for various degrees of sparsity, choosing the optimal level of $L_2$-shrinkage for each number of non-zero coefficients. As such the horizontal axis represents the number of non-zero coefficients and the vertical axis depicts the belonging value of OOS $R^2$. The green line represents the OOS $R^2$ at given levels of non-zero coefficients in the space of PCs. The blue line depicts the OOS $R^2$ for the raw factor returns at given numbers of non-zero coefficients.

This plot amplifies the illustration in figure 6.4, as we clearly see, in the space of raw factor returns, that a high number of characteristics is needed to span the entire SDF. We highlight that the OOS $R^2$ increases for each added non-zero coefficient, reaching its maximum at the highest available number of coefficients, 50. Imposing sparsity will, as such, only hurt the OOS predictive power. Consistent
with the findings from figure 6.4, transforming the SDF into PCs of factor returns allows the model to impose significant sparsity. A model, including two PC-based factors achieves more than half of the total cross-sectional OOS $R^2$. And a model with only five PC-based factors yields close to the optimal OOS $R^2$ in this case. This is, furthermore, in line with the economic reasoning discussed earlier in this thesis, that the PCs with the biggest absolute coefficients are the PCs with the highest variance. As such, only a limited number of PCs should sufficiently contain the majority of total variance, as illustrated above.

In summary, there is next to no redundancy among the fifty anomalies, why it is impossible to identify a parsimonious SDF with just a few characteristics-based factors that performs well out-of-sample. This highlights the need for $L_2$-shrinkage to accommodate the high-dimensional nature of the estimation problem. Imposing a certain level of $L_2$-shrinkage in the dual-penalty estimator delivers significantly higher OOS $R^2$ than just imposing $L_1$-shrinkage, such as the LASSO method would. Transforming the SDF into PCs of factor returns of the 50 anomalies gives the opportunity of creating a parsimonious model with good OOS $R^2$ performance. This is in line with the arguments and findings presented by KNS.

6.2.4 WRDS Financial ratios (WFR)

In the 50 anomalies data set, each characteristic, i.e., each anomaly has previously been proposed and tested in the literature. Meaning that, consequently, all the potential factors spanning the SDF have proven, to a smaller or larger extend, explanatory power of future mean returns. The dual-penalty estimator, applied in the previous section, successfully tested for potential redundancy among the 50 anomaly-factors, however, it was not challenged in regard to identifying novel pricing factors from a high-dimensional data set. Therefore, we subsequently test the estimator on a WFR data set consisting of 69 financial ratios. Treating each financial ratio as an asset specific characteristic, we seek to estimate a model with these, potentially 69, factor returns spanning the SDF.

It is known from previous studies that some of the financial ratios in the WFR data set are closely related to expected returns, (such as multiple versions of the P/E ratio (Kozak et al. (2020)), while some are not. It is therefore plausible that some of the 69 characteristics have absolutely no pricing power. Hence, it is highly interesting to see if the dual-penalty estimator method proposed by Kozak et al. (2020) is able to successfully; (1.) De-emphasize the factors with no relevant pricing power and avoid overfitting them, (2.) identify potential new pricing factors with significant explanatory power
of cross-sectional expected returns. The contour maps in Figure 6.6 has a striking resemblance with the previous plots for the 50 anomaly portfolios.

Figure 6.6: The contour plot shows the OOS $R^2$ from the dual-penalty specification, on the 69 WRDS financial ratios. OOS cross-sectional $R^2$ for families of models that employ both $L_1$- and $L_2$-shrinkage simultaneously on 69 WRDS financial ratios (panel a) and 69 PCs based on the returns of the 69 WRDS financial ratios (panel b). We illustrate the degree of $L_2$-shrinkage by the prior root expected squared Sharpe Ratio ($SR^2$) ($\kappa$) on the horizontal axis. We show the number of included variables in the SDF on the vertical axis, which quantifies the degree of $L_1$-penalty imposed. Warmer (yellow) colors depict higher levels of OOS $R^2$. Both axes are plotted on logarithmic scale to enhance interpretability of the contour plot. The sample data is daily from 2000 to 2020.
Looking at figure 6.6 (a), unregularized models performs extremely poorly out of sample with highly negative OOS $R^2$-values in the top right corner of the plot. $L_2$-penalty-only models, (top line of the plot), performs well for both data sets. When an $L_1$-penalty term is added, imposing sparsity, the OOS $R^2$ quickly deteriorates, implying once again that there is not much redundancy between the characteristics. Hence, consistent with the findings in KNS, we conclude that it is not possible to obtain a characteristics-sparse SDF with good OOS pricing performance for the factor returns of the WFR data set either. We note that there seems to be some substitute relation between $L_1$- and $L_2$- shrinkage, as the ridge moves diagonally from the top towards the right corner of the plot. As more sparsity is imposed $L_2$-shrinkage must be eased to avoid over shrinking / underfitting factors.

Turning to the SDF in the space of PCs displayed in figure 6.6 (b) we note that the model successfully implements a significantly higher degree of $L_1$-shrinkage without hurting the OOS $R^2$. Rotating into principal components of factor returns results in a good OOS pricing performance for a relatively parsimonious model with optimal OOS $R^2$ reached for as few as nine non-zero coefficients. The ridge of high OOS $R^2$ is wider and located at higher $\kappa$ values than in the previous setting with 50 anomaly portfolios. This implies that, not only does a characteristics-sparse SDF perform well in terms of OOS $R^2$, but it does so with quite limited amount of $L_2$-shrinkage imposed.

Figure 6.7 (a) illustrates this more comprehensibly by taking a cut along the top edge of the contour plot in figure 6.6 (a). The blue plot illustrates the OOS $R^2$ at different degrees of $L_2$-shrinkage, imposing no sparsity.

![Figure 6.7: $L_2$ model selection and sparsity (69 WRDS financial ratios). Panel (a) plots the in sample cross-sectional $R^2$ (green) and OOS cross-sectional $R^2$ based on 3-fold cross-validation (blue), with no sparsity imposed. Panel (b) shows the maximum OOS cross-sectional $R^2$ attained by a model with $n$ factors (on the horizontal axis) across all possible values of $L_2$-shrinkage, for models based on raw factor returns (blue) and PCs (green). The sample data is daily from 2000 to 2017.](image)
The blue plot illustrates the \( OOS R^2 \) at different degrees of \( L_2 \)-shrinkage, imposing no sparsity. Its peak (\( \kappa \approx 0.9 \)), indicating optimal OOS \( R^2 \)-levels, is significantly more to the right than in the comparable figure for the anomalies data in figure 6.5 (a) (\( \kappa \approx 0.6 \)), further illustrating that less \( L_2 \)-shrinkage is sufficient to obtain good OOS performance for the WFR data set. Figure 6.7, once again, clearly illustrates that the in-sample \( R^2 \) is a very poor predictor of OOS \( R^2 \).

Figure 6.7 (b) illustrates the OOS \( R^2 \), varying the number of non-zero coefficients at the optimal level of \( L_2 \)-shrinkage for each level of sparsity. The blue line, illustrating the SDF in space of PCs of factor returns, shows that as few as nine PCs out of the 69 potential PCs sufficiently obtain close to maximum OOS \( R^2 \). The maximum OOS \( R^2 \) is obtained for a model including 13 PCs. The OOS \( R^2 \) for the raw factor returns, green line, moves in similar fashion to the analogous figure for the 50 anomalies data in figure 6.5 (b). The OOS \( R^2 \) slowly increases as more non-zero coefficients are added, further solidifying the fact that a characteristics-sparse SDF with good pricing performance does not exist. Fig. 6.7 (b) additionally illustrate that the optimal OOS \( R^2 \) in the space of raw factor returns, with almost no sparsity, is higher than the optimal OOS \( R^2 \) for the sparse SDF in the space of principal components. This implies, that we cannot avoid some tradeoff of pricing performance when constructing a parsimonious model on the WFR data set, regardless of applying raw factor returns or rotating the SDF into a space of PCs. As such, the \( L_2 \)-only-penalty model performs best in terms of OOS pricing performance.

In summary, the analysis of the WFR data set shows that the proposed model of KNS, applying the dual-penalty estimator from equation (56), successfully handles a data set that mixes relevant and irrelevant pricing factors. If sparsity is desired a relatively moderate degree of \( L_1 \)-shrinkage can be imposed to exclude the most pricing-irrelevant factors, however an \( L_2 \)-penalty-only model returns equally as high OOS \( R^2 \) values. Rotating the SDF into the space of principal components of factor returns of the 69 characteristics allows us to construct a relatively sparse model, obtaining good OOS \( R^2 \) performance with approximately 10 non-zero coefficients spanning the SDF. The main conclusion from the WFR-analysis is that the \( L_2 \)-only method achieves the highest OOS \( R^2 \), implying that sparsity, regardless of the form of the factors spanning the SDF, will hurt the OOS prediction performance to a larger or smaller extent.

**6.2.5 Largest SDF-factors**

Thus far, we have presented results broadly consistent with the findings of KNS, leading to the same conclusions regarding the attributes and advantages of the dual-penalty estimator method and
the effects of rotating returns into PCs. Subsequently, for a more detailed comparison of results, we investigate if the estimator identifies the same relevant pricing factors as KNS. In the following section we illustrate the largest SDF factors for the 50 anomaly portfolios and 69 financial ratios, in table 6.1 and 6.2, respectively. The main goal of this section is to compare the largest SDF factor loadings unearthed by our model with the identified pricing relevant factors in KNS. In a later section, we discuss any differences in relevant pricing factors unearthed from our estimator and the analogous table presented by Kozak et al. (2020).

Table 6.1 (a) lists the anomaly factors with the largest absolute t-statistics, where standard errors are calculated based on the following equation (Kozak et al., 2020):

\[
\text{var}(\beta) = \frac{1}{T}(\Sigma + \gamma I)^{-1}
\]

It should be noted that the identification of the largest SDF factors is based on the \(L_2\)-penalty-only model, why no sparsity is imposed.

### Table 6.1
Largest SDF Factors (50 anomaly portfolios)

<table>
<thead>
<tr>
<th>(a) Raw 50 anomaly portfolios</th>
<th>(b) PCs of 50 anomaly portfolios</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beta</td>
<td>t-stat</td>
</tr>
<tr>
<td>Return on Market Equity</td>
<td>-0.24</td>
</tr>
<tr>
<td>Industry relative reversals</td>
<td>-0.13</td>
</tr>
<tr>
<td>Dividend Yield</td>
<td>-0.09</td>
</tr>
<tr>
<td>Beta Arbitrage</td>
<td>-0.14</td>
</tr>
<tr>
<td>Value-Momentum</td>
<td>0.11</td>
</tr>
<tr>
<td>Seasonality</td>
<td>0.03</td>
</tr>
<tr>
<td>Long Term reversals</td>
<td>-0.04</td>
</tr>
<tr>
<td>Accruals</td>
<td>0.08</td>
</tr>
<tr>
<td>Investment</td>
<td>-0.05</td>
</tr>
<tr>
<td>Value (monthly)</td>
<td>0.06</td>
</tr>
</tbody>
</table>

The t-statistics are calculated dividing the beta coefficients by their standard deviations estimated with equation (61). We note that the estimated t-statistics are quite low, individually, for the factors of the SDF, with only the biggest factor, associated with the anomaly ‘Return on market equity’, being significant according to conventional significance levels. However, as Kozak et al. (2020) also argues, it is quite important to highlight, that what essentially matters for the SDF is the joint significance of linear combinations of the factors.

Table 6.1 (b), illustrating the largest SDF factors in the space of PCs, is exactly just that - linear combinations of the factors in table 6.1 (a). Intuitively, this is given in the definition of PCs (outlined in section 4.3.6). In accordance with our previous expectations of high ordered PCs carrying most of the variance, and consequently predictive power, table 6.1 (b) shows that the largest factors in the SDF are PC1, PC2, PC5, PC3 and PC8. Our previous analysis in figure 6.5 (b) and 6.6 (b) additionally shows that a PC-sparse SDF with only these five PCs yields the optimal OOS $R^2$. However, quite surprisingly, regarding the analogous results illustrated in KNS, only the largest factor in the SDF, PC1, is significantly different from zero, ($t_{stat_{PC1}} = 3.93 > 1.96$). In KNS, we see that all of the five largest factors in PC-space are significant based on conventional significance levels.

---

37 With conventional significance levels we refer to a significance level, $\alpha = 0.05$, where a coefficient is deemed significantly different from zero if it has a t-statistic above 1.96 (two standard deviations).

38 We refer to Table 1, p. 283 in Shrinking the cross-section (Kozak et al., 2020).
Table 6.2
Largest SDF Factors (69 WRDS Financial Ratios)

Coefficient estimates (Beta) and absolute t-statistics (t-stat) at the optimal value of the prior root expected $SR^2$, $\kappa$, (based on cross-validation). Panel (a) focuses on the coefficients of the SDF from the raw 69 WRDS factor returns and panel (b) rotates returns into the space of PCs and illustrates coefficient estimates based on these PCs. Coefficients are sorted in descending order based on their absolute t-statistic values. The sample data is daily from 2000 to 2015.

Table 6.2 lists the largest SDF factors based on the WFR data set, imposing no shrinkage. Or put differently, the estimated coefficients at the optimal level of $L_2$-only shrinkage. Coefficients are ranked in descending order based on their absolute t-statistic value. Table 6.2 (a), focusing on the raw returns of the WFR data set, shows that the proposed method is inclined to estimate factors with high weight on financial ratios that have previously been associated with expected returns. Most of the largest factors are associated with financial ratios categorized as valuation ratios such as ‘Trailing P/E to growth (PEG) ratio’, ‘Shillers Cyclically Adjusted P/E Ratio’, ‘P/E (Diluted incl. Extraordinary Items)’, ‘Price/Cash Flow’, ‘Dividend Yield’, ‘P/E (Diluted, excl. EI)’ and ‘Price / Operating Earnings (Diluted excl. EI)’. The remaining two factors, which, interestingly, are the two largest, ‘Interest / Average Long-term Debt’ and ‘Effective Tax Rate’ are categorized as a Financial Soundness ratio and a Profitability ratio, respectively.39

In terms of conventional significant levels, we see that all the largest identified factor coefficients are significantly different from zero at a 5%-significance level. However, as Kozak et al. (2020) also emphasize, the individual OOS pricing power of each characteristic is not important, because none of these variables, on their own, are likely to be optimal measures of the ‘true’ underlying signal (factor). The brilliance of the proposed method is that it combines information from many of such imperfect measures (averaging them by the means of the $L_2$-penalty) and delivers a sound SDF with

39 For an in-depth description of each of the 69 Financial Ratios we refer to Appendix 2 provided by WRDS.
robust OOS performance. As such, combining several measures of each category (for example valuation ratios) performs significantly better out of sample, than using any individual ratio. Since the model relies on the combination of different measures, adding or removing potential financial ratios to the data set, can potentially alter the dynamic of the SDF and consequently result in a different identification of the largest factors spanning the SDF. KNS adds 12 characteristics, based on lagged monthly returns, (t-1 to t-12), to capture the momentum effect. This could potentially explain why our estimator identifies coefficients associated with different value financial ratios than the list presented in table 2, p. 283 in KNS. We further emphasize and discuss the deviation of coefficients associated with different ratios in section 7.1.2.

Table 6.2 (b) rotates assets into PC space. Most of the entries into the list belongs to the Top-10 PCs. Table 6.2 (b) shows that the largest coefficients are associated with PC5, PC8, PC6, PC3 and PC1, ranked in that specific order. All these coefficients are significant, based on conventional significance level ($\alpha = 0.05$). We note that PC1 is only the 5th largest coefficient and PC2 the 8th largest, which is counterintuitive, since these two PCs is supposed to carry a majority of the variance of returns. However, we still see a quite clear trend that the first PCs are associated with the largest factors, the ten largest factors are associated with the first 11 PCs. Bear in mind that there are 69 total PCs available. Compared with the anomaly portfolios in table 6.1 (b) there are a few more of the lower variance PCs on this second list as well. Imposing some level of $L_1$-sparsity in the dual-penalty specification shows that the low variance PCs drop out of the mix. Estimating a sparse SDF with five non-zero coefficients include PC1, PC2, PC3, PC4 and PC5. This is roughly in accordance with the previous economic intuition and arguments that the most important pricing factors are associated with high-variance PCs.

In summary, table 6.1 and 6.2 illustrates the largest SDF loadings in the space of raw returns and PCs. For the 50 anomalies data set the estimation method automatically unearthed anomaly factors, that have previously proven robust in the literature. Individually, the SDF factors are not significantly different from zero, however it is the joint significance of the factors that is important for the performance of the SDF. For the setting of the WFR data set, the largest SDF factors were primarily associated with financial ratios categorized as ‘Value ratios’, which is broadly equivalent to the findings of KNS. Rotating assets into PC space, for both data sets, broadly substantiated the argument that the most important pricing factors are associated with the highest variance PCs. This relation was, however, stronger for the 50 anomalies data set, than the financial ratio setting, where the first
PCs (PC1, PC2, and PC3) only claim the ‘largest coefficient podium’ when a significant level of $L_1$-shrinkage is imposed. With the identification of the most relevant pricing factors, it is interesting to see if we manage to identify the same pricing factors as Kozak et al. (2020). As such, table 6.3 below illustrates the relevant pricing factors presented in KNS.

Table 6.3

Largest SDF Factors from Shrinking The Cross-Section (Kozak et al., 2020), p. 283.

Panel (a) illustrates the Largest SDF Factors on the 50 anomaly data set and panel (b) illustrates the Largest SDF Factors for the WFR data set (Kozak et al. (2020) adds 12 momentum financial ratios, resulting in 81 factors potentially spanning the SDF).

<table>
<thead>
<tr>
<th>(a) 50 anomaly portfolios</th>
<th>(b) 81 WRDS Financial ratios</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beta</td>
<td>Beta t-stat</td>
</tr>
<tr>
<td>Industry rel. reversals (LV)</td>
<td>-0.88 3.53</td>
</tr>
<tr>
<td>Ind. Momentum-reversals</td>
<td>0.48 1.94</td>
</tr>
<tr>
<td>Industry rel. reversals</td>
<td>-0.43 1.70</td>
</tr>
<tr>
<td>Seasonality</td>
<td>0.32 1.29</td>
</tr>
<tr>
<td>Earnings surprises</td>
<td>0.32 1.29</td>
</tr>
<tr>
<td>Value profitability</td>
<td>0.30 1.18</td>
</tr>
<tr>
<td>Return on market equity</td>
<td>0.30 1.18</td>
</tr>
<tr>
<td>Investment</td>
<td>-0.24 0.95</td>
</tr>
<tr>
<td>Return on equity</td>
<td>0.24 0.95</td>
</tr>
<tr>
<td>Composite issuance</td>
<td>-0.24 0.95</td>
</tr>
</tbody>
</table>

50 anomaly portfolios

Comparing the Top-10 largest SDF factors in table 6.1 (a) with the findings from KNS above, it is easily noted that the two estimators do not return identical results. There are deviations, both regarding which factors are included, and the position of the coinciding factors. We note that both estimators identify factors based on the anomalies; ‘Industry relative reversals’, ‘Seasonality’, ‘Return on market equity’ and ‘Investment’. These have all proven to be robust in the literature (Kozak, et al., 2020). We see that the model in KNS returns none of the traditional anomalies from the Fama-French 3-factor model, whereas our estimator has the value anomaly as the 10th largest contributor.

WFR Data set

Similarly, for the SDF based multi-factor model on the WFR data set we observe deviations between the largest SDF factor given in KNS and our results in table 6.2 (a). The following ‘financial ratio’-based factors are identified by both analyses: ‘P/E (Diluted, incl. EI)’, ‘Trailing P/E to growth (PEG) ratio’ and ‘P/E (Diluted, incl. EI)’. Although, this seems like quite few coinciding SDF factors, there is a clear trend in both findings, that the majority of the largest SDF factors are associated with ‘Value-financial ratios’. Furthermore, in both Top-10s for the WFR data we see that ‘financial
soundness’-ratios also has relevant pricing power. It should be noted that KNS includes 12 momentum ratios in their data set, of which three of these (‘Month t-9’, ‘Month t-11’ and ‘Month t-1’) are among the ten largest SDF factors. Naturally, as we do not include these, and they seem to have significant pricing power this affects the dynamic of the largest SDF factors in the two empirical studies.

In section 7.1.2 we further discuss potential sources responsible for the above-mentioned deviations in the 50 anomalies and WFR data settings.

6.2.6 Subsection conclusion

In the preliminary analysis we tested the proposed dual-penalty estimator from equation (56) on a simple data set containing the FF25 BM/ME portfolios, where previous studies have showed a strong factor structure between the anomalies size and value. The main results were illustrated in figure 6.2, depicting the OOS $R^2$ for varying degrees of $L_1$- and $L_2$-shrinkage. The analysis showed that the estimator successfully imposed $L_1$- and $L_2$-penalty terms, improving the prediction performance, illustrated by higher OOS $R^2$. Furthermore, the initial analysis showed that rotating the SDF into a space of PCs of factor returns enabled the use of a parsimonious model, as the degree of sparsity increased without negative impact on predictive power. Subsequently, we continued examining the proposed method in high-dimensional settings resulting in two parallel analyses, on the 50 anomalies data set and WFR setting, respectively (illustrated in figure 6.3, 6.4, 6.5 and 6.6). The analysis showed that completely unregularized models, as well as excessively shrinked models perform very poorly out of sample for both data sets. Hence, there is some optimal level of $L_1$- and $L_2$-shrinkage needed to obtain optimal OOS $R^2$.

For both data sets we concluded that it is impossible to create a factor-sparse SDF in the space of raw returns, that performs well out of sample. Rotating the SDF into the space of PCs enabled the possibility of a PC-sparse model sufficiently spanning the SDF with a limited number of factors. With only five factors (50 anomalies) and nine factors (WFR) the model obtained close to the maximum OOS $R^2$. Furthermore, the analysis of the WFR data set validated that the dual-penalty estimator is capable of successfully processing a high-dimensional data set, that mixes relevant and irrelevant pricing factors. Finally, we evaluated the largest SDF factors, resulting from the dual-penalty estimator, on the two data sets. From the 50 anomalies data set we saw that the method unearthed factors that have been known to be robust in the literature. Furthermore, rotating assets into PCs enhanced the argument that important pricing factors are associated with the PCs of high variance. In
the WFR data set, the largest coefficients are primarily associated with *value-ratios*, which is broadly consistent with the findings in *KNS*.

So far, we have shown that the dual-estimator successfully imposes $L_1$- and $L_2$-shrinkage improving the prediction performance for both low-complexity and high-complexity data sets. Furthermore, for each data set, rotating the SDF into the space of principal components enables the method to estimate a parsimonious and well out-of-sample performing model for all included data sets.

However, in this out-of-sample test of the SDF performance, using cross-validated OOS $R^2$, we evaluated the model on the part of the sample withheld from the estimation of the SDF-coefficients. Nevertheless, this OOS metric is not purely *out of sample*, as the data-driven choice of penalty parameters were still based on the OOS performance within the sample. As such, the withheld sample in one fold is simultaneously part of the training sample in another fold. Hence, the cross-validated penalty parameters might be optimal in our sample, but not generalize well to other samples. Therefore, to evaluate this potential shortcoming of the method, we subsequently conduct a pure out of sample test in the following section.

### 6.3 Asset pricing implication through portfolio optimization

In this section, we seek to evaluate the true out of sample performance of the SDFs in a portfolio optimization context. As such, we test the asset pricing implication of the dual-penalty estimator on a data sample that is completely separated from the estimation period. The following analysis is based on the two, previously described, high-dimensional data sets ‘50 anomaly portfolios’ and ‘WFR’.

#### 6.3.1 Model setup

In order to obtain a true OOS evaluation of the method, we consider the two data sets 50 anomalies and WFR in the period January 2000 - December 2014 as our training data. We completely leave out data in the period January 2015 - December 2017 from our estimation in order to evaluate the model in a pure OOS setting.\(^{40}\)

Initially, we repeat the entire k-fold cross-validation procedure, dividing the training data sample into three contiguous parts ($k=3$), estimating the three OOS $R^2$ values for each combination of penalty parameters $\gamma_1$ and $\gamma_2$ and subsequently choosing the optimal level of $\gamma_1$ and $\gamma_2$. We apply this optimal

---

\(^{40}\)The entire data sample for this exercise is daily data from January 2000 – December 2017. The length of the data set is bounded as the intersect between the data period of the 50 anomaly portfolios (July 1973 – December 2017) and the WRDS Financial Ratios (January 2000 – December 2020).
strength of the two penalty terms on the entire, undivided, training data (January 2000 - December 2014) to solve the dual-penalty problem yielding the optimal SDF-coefficients for the whole training data.

We subsequently consider the test data in the period January 2015 - December 2017. We seek to determine, how well the optimal SDF-coefficients describes the cross-section of stock returns in a true out of sample setting. In order to adequately evaluate this property of the estimated SDF, we, analogous to KNS, consider a mean-variance efficient portfolio setting. In this setting we initially orthogonalize the portfolio returns for each factor strategy with respect to the market, using betas estimated in the training data (similar to the procedure executed in section (6.2)). Then, we seek to determine optimal portfolio weights. Following the framework of Markowitz as outlined in section 4.4 the mean-variance efficient frontier of optimal portfolios can be constructed by minimizing variance for a given target, i.e., solving the following problem:

\[
\begin{align*}
\min_w w' \Sigma w \\
subject \ to \\
w' \mu &= \mu_p \\
w' 1 &= 1 \\
[w] &\geq 0
\end{align*}
\]

Bryzgalova (2019) shows that this conventional Markowitz problem is equivalent to the main estimator in equation (56), but with the main estimator imposing shrinkage of the covariance matrix by the two penalty terms \( \gamma_2 \beta' \beta \) and \( \gamma_1 \sum_{j=1}^{H} |\beta_j| \).

\[
\hat{\beta} = \arg\min_{\beta} \left\{ (\bar{\mu} - \Sigma \beta)' \Sigma^{-1} (\bar{\mu} - \Sigma \beta) + \gamma_2 \beta' \beta + \gamma_1 \sum_{j=1}^{H} |\beta_j| \right\}
\]  (62)

Hence, they argue, that by minimizing this function, we solve a Markowitz optimization problem with imposed shrinkage of the covariance matrix of the factor returns and select a set of weights that spans the tangency portfolio. The MVE portfolio weights are therefore set equal to the SDF-coefficients, \( \hat{\beta} \). As such, we subsequently multiply the SDF-coefficients, \( \hat{\beta} \), estimated in the training data by the OOS returns of the factor portfolios, \( F_t \), yielding the OOS MVE portfolio returns, \( P_t^{MVE} \):
\[ \hat{p}_t^{MVE} = \hat{\beta}'F_t \]  

Applying this method, we construct three types of portfolios for each high-dimensional data set, yielding six portfolios in total. The first portfolio is based on the \( L_2 \)-shrinkage-penalty only, imposing no sparsity. As such, every possible factor (50 for the anomalies and 69 for the WFR, respectively) is applied to span the SDF. The second portfolio is a characteristics-sparse SDF model built based on the dual-penalty estimator, imposing both \( L_1 \)- and \( L_2 \)-shrinkage, where the \( L_1 \)-shrinkage level is adequately modified to fit five non-zero coefficients in the model. Hence, we have characteristics-sparse portfolios with only five factors spanning the SDF. The final portfolio for each data set is designed identically to the second set of portfolios, but in the space of PCs, resulting in portfolios based on a sparse PC model with only five PCs spanning the SDF.

In order to evaluate the performance of the six constructed portfolios we consider two well-known factor-based models as benchmarks: The classic one-factor model *capital asset pricing model* (CAPM) and the Fama-French six-factor model (FF6) initially proposed by Fama and French (2016)\(^{41} \). In order to make the constructed portfolios adequately comparable to these benchmarks, we construct the implied MVE portfolio for each of the two benchmarks. Since we orthogonalized all returns with respect to the market excess return, which in our analysis is equal to the market factor of the CAPM, the MVE mean portfolio return associated with the CAPM benchmark, in our analysis, is simply equal to zero. For the FF6 portfolio we estimate the unregularized MVE portfolio weights as:

\[ \hat{\mathbf{w}}_{FF6} = \hat{\Sigma}_{FF6}^{-1} \hat{\mu}_{FF6} \]  

Where \( \hat{\Sigma}_{FF6} \) and \( \hat{\mu}_{FF6} \) are the covariance matrix and mean returns, respectively, of the returns in the FF6 model estimated in the period January 2000 to December 2014\(^{42} \). We then apply these weights in the OOS period (January 2015 - December 2017), to construct a single benchmark portfolio return for the FF6 model:

\[ p_{FF6} = \hat{\mathbf{w}}_{FF6}'F_{t,FF6} \]  

\(^{41}\)Fama and French (2016) proposed the addition of a human capital component to the Fama-French 5-factor model (Fama & French 2015) that, besides the original size- and value-factors, also includes profitability- and investment-factors.

\(^{42}\)The FF6 portfolio data set is withdrawn from Kenneth French’s data library, which can be accessed through: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html
Finally, for all MVE portfolios we scale returns, so that standard deviations are equal to the standard deviation of the market, in order to facilitate interpretation of the magnitude of portfolio returns.

\[ P_{t,\text{ScaledMVE}} = P_{t,MVE} \frac{\sigma_{\text{mkt}}}{\sigma_{\text{MVE}}} \]  

(66)

For the remainder of the paper, whenever the notation \( P_{t,MVE} \) is used, we refer to the scaled portfolio returns.

6.3.2 Results

In order to assess the individual pricing performance of the different portfolios described above we calculate the orthogonalized portfolio alphas with respect to the CAPM benchmark, illustrated in table 6.4 below\(^{43}\). All alphas are scaled to the standard deviation of the market.

<table>
<thead>
<tr>
<th>Table 6.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Portfolio Alphas with respect to the CAPM Benchmark.</td>
</tr>
<tr>
<td>The intercepts (( \alpha )) of the time series regressions of the constructed portfolios, including FF6, on the benchmark CAPM portfolio returns. The seven MVE portfolios displayed, from left to right are based on; ‘Raw factor returns of 50 Anomalies (( L_2 )-shrinkage only)’, ‘Raw factor returns of 50 Anomalies – Sparsity imposed’, ‘PCs of 50 Anomalies – Sparsity imposed’, ‘Raw factor returns of WFR (( L_2 )-shrinkage only)’, ‘Raw factor returns of WFR – Sparsity imposed’, ‘PCs of WFR – Sparsity imposed’ and ‘FF6’. The table illustrates both daily and annualized values of ( \alpha ).</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>50 Anomalies</th>
<th>WFR (69 Financial ratios)</th>
<th>Benchmark</th>
</tr>
</thead>
<tbody>
<tr>
<td>( L_2 )-only shrinkage</td>
<td>Characteristic Sparse</td>
<td>PC Sparse</td>
</tr>
<tr>
<td>Average daily alpha to CAPM</td>
<td>0.04%</td>
<td>0.03%</td>
</tr>
<tr>
<td>Average annual alpha to CAPM</td>
<td>10.72%</td>
<td>6.34%</td>
</tr>
</tbody>
</table>

As illustrated in table 6.4, all constructed MVE portfolios achieve abnormal returns with respect to the CAPM model. Particularly, MVE returns based on the \( L_2 \)-only estimation method outperforms the characteristics-sparse model for both the 50 anomalies data set and the WFR data set, implying that imposing sparsity on raw factor returns hurts the pricing performance of the model. This further amplifies the conclusion drawn in the previous analysis section, that it is infeasible to construct a characteristics-sparse SDF without hurting OOS performance. For the WFR data set the \( L_2 \)-only portfolio also outperforms the PC-sparse model. Interestingly, the PC-sparse MVE portfolio for the 50 anomalies data set achieves a slightly abnormal alpha relative to the \( L_2 \)-shrinkage-only portfolio.

\(^{43}\) The \( \alpha \) is the intercept determined through a time series regression of the respective portfolio return against the CAPM portfolio.
This is inconsistent with the findings of KNS, that shows the outperformance of the $L_2$-shrinkage-only model for all benchmarks, however, just slightly with regards to PC-sparse models. Regardless, PC-sparse SDFs, consistent with the findings in the previous analysis, outperforms the sparsity imposed SDFs based on raw factor returns, for both data sets.

For the purpose of comparability, we illustrate the asset pricing performance in similar fashion to Kozak et al. (2020). We focus on the $L_2$-shrinkage-only MVE portfolios for both data sets and benchmark them against portfolios based on the following four-factor models; CAPM, FF6, optimal characteristics-sparse model with five factors and optimal PC-sparse model, with five factors. As mentioned previously, the first two benchmarks are classic factor models in the form of CAPM and FF6. Additionally, we benchmark the $L_2$-only MVE portfolios against their characteristics-sparse and PC-sparse counterparts, to test asset pricing implications of sparse models. The analysis is conducted by regressing the CAPM-orthogonalized returns on each benchmark. Table 6.5 illustrates annualized intercepts ($\alpha$) for time series regression of the $L_2$-only MVE OOS portfolio returns on each benchmark and corresponding standard deviations in parenthesis.

Table 6.5

MVE portfolio’s annualized OOS $\alpha$ in the withheld sample (January 2015 - December 2017)

The table shows annualized alphas computed from the time series regression of the SDF-implied OOS-MVE portfolio’s return (based on $L_2$-shrinkage only) relative to four restricted benchmarks: CAPM, Fama-French six-factor, optimal characteristics-sparse model with five factors and optimal PC-sparse model, with five factors. MVE portfolio returns are normalized to have the same standard deviation as the aggregate market. The 50 anomalies $L_2$-only shrinkage portfolio is regressed against Characteristics-sparse and PC-sparse portfolios based on the same data set (50 anomalies) and similarly for WFR, the regressed portfolios are based on the same data. For both 50 anomaly data set and WFR data set, the sample is daily data from January 2000 – December 2017.

<table>
<thead>
<tr>
<th>Annual Alpha to different benchmarks</th>
<th>CAPM</th>
<th>FF6</th>
<th>Characteristic Sparse</th>
<th>PC Sparse</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L_2$-only shrinkage (50 Anom)</td>
<td>10.72%</td>
<td>7.99%</td>
<td>4.90%</td>
<td>-0.54%</td>
</tr>
<tr>
<td></td>
<td>(7.28%)</td>
<td>(6.04%)</td>
<td>(2.91%)</td>
<td>(0.77%)</td>
</tr>
<tr>
<td>$L_2$-only shrinkage (WFR)</td>
<td>17.13%</td>
<td>17.25%</td>
<td>9.52%</td>
<td>2.84%</td>
</tr>
<tr>
<td></td>
<td>(7.28%)</td>
<td>(7.29%)</td>
<td>(7.29%)</td>
<td>(0.61%)</td>
</tr>
</tbody>
</table>

44 We refer to Table 4, page 290 in Shrinking the cross-section (Kozak et al., 2020).
Table 6.5 further solidifies the findings illustrated in table 6.4 that the \( L_2 \)-only MVE portfolio for the WFR data set achieves abnormal OOS \( \alpha \) with respect to all four restricted benchmarks. Particularly, it achieves \( \alpha \) close to 17\% with respects to the two classic factor models CAPM and FF6. Furthermore, we note that the \( L_2 \)-only model for the WFR data has a slightly positive \( \alpha \) with respect to its PC-sparse counterpart (\( \alpha \approx 2.8\% \)). This once again illustrate the recurring theme throughout the analysis, that imposing sparsity, significantly hurts the pricing performance of the model, with exception of PC-sparse models, that seems to perform almost as good as the \( L_2 \)-only MVE portfolios.

Furthermore, table 6.5 shows that all annualized alphas of the \( L_2 \)-only WFR MVE OOS portfolio returns with respects to each benchmark are more than one standard deviation away from zero.

Focusing on the 50 anomalies data set it is evident that the \( L_2 \)-only MVE OOS portfolio outperforms the classic factor models, CAPM and FF6, by a substantial amount of 10.7\% and 8.0\% respectively. Consistent with the findings from table 6.4 the PC-sparse MVE portfolio achieves a slightly abnormal \( \alpha \) of 0.5\% relative to the \( L_2 \)-only portfolio. In addition, we note that the PC-sparse MVE portfolio outperforms the characteristics-sparse MVE portfolio significantly by multiple standard deviations.

To further illustrate the MVE portfolio OOS performance, we index the initial alphas from the orthogonalization with respect to the CAPM, illustrated in table 6.4, and plot the resulting indexed return time-series for the full test period, January 2015 to December 2017 in figure 6.8.
Figure 6.8 illustrates indexed alphas in the out of sample time period, January 2015 – December 2017, originating from the regression of the following MVE portfolios on the CAPM based MVE portfolio: ‘L2-only shrinkage (on raw factor returns)’, ‘Characteristics-sparse portfolio’, ‘PC-sparse portfolio’ and ‘FF6’. Figure 6.8 (a) and (b) shows indexed $\alpha$ for the 50 anomalies data set and WFR data set, respectively. $\alpha$ values are indexed at 1.00 on the 1st of January 2015.
The plot in figure 6.8 further highlights the findings from table 6.4 and 6.5. The WFR $L_2$-only portfolio and the WFR PC-sparse portfolio achieves the highest OOS $\alpha$ over the full three-year testing period. The FF6 factor portfolio ends up at a lower index compared to all the constructed portfolios. However, we highlight that the indexed $\alpha$ of the two characteristics-sparse models are only slightly above the FF6 portfolio. We note that the evolution of the indexed $\alpha$’s for different portfolios roughly move in the same directions throughout the OOS period. As such, none of the portfolios are inversely correlated with the rest.

6.3.3 Subsection conclusion

As expressed in the theoretical framework the aim of the proposed model is to accommodate an SDF that includes a potentially vast number of characteristics-based factors, while simultaneously ensuring good OOS performance and robustness against ‘in sample overfitting’. From our results of the OOS portfolio optimization analysis, we conclude that the method achieves abnormal $\alpha$ compared to the FF6 and CAPM model for all the SDF-based proposed portfolios. This implies that the method achieves good OOS performance and that the model is robust against ‘in sample overfitting’.

The results further show that sparse models do not outperform full characteristics-based models in an OOS portfolio optimization setting. This indicates that the $L_2$-only method does well in shrinking candidate factors that are not contributing substantially to the SDF. Consequently, the true out of sample analysis substantiate the findings, that creating a characteristics-sparse model with good OOS performance is infeasible. However, for both the WFR and the 50 anomalies data set the PC-sparse model, based on the dual-penalty estimator, experience abnormal $\alpha$ compared to the characteristics-sparse MVE OOS portfolio. This emphasizes our earlier claim, that PC-sparse models do well in predicting the cross-section of stock returns compared to characteristics-sparse models. It motivates the use of PCs in high-dimensional factor models, since a few PCs perform better in describing the variation of the high-dimensional characteristics data, compared to a limited number of raw factors.
7. Discussion

As emphasized in each subsection-summary of the analysis, our empirical study is broadly consistent with the findings of Kozak et al. (2020), as we draw the same overall conclusion regarding attributes and OOS pricing performance of the proposed dual-penalty estimator for both high-dimensional data sets. However, as the results are not completely identical it is relevant to further elaborate on the potential sources of deviations.

7.1 Deviation compared to the findings in KNS.

In the following we briefly summarize the dissimilarities between the results found in this thesis and the results presented in the KNS.

7.1.1 Summary of deviations

We find three major points of deviation, that are worth highlighting; (1) The contour plots on the cross-validated empirical OOS test, (2) the loadings of the largest SDF-factors and the (3) relative pricing performance of the MVE constructed portfolios in a true OOS setting.

Contour plots on the cross-validated empirical OOS test

Looking at the contour plots in figure 6.4 and 6.6 the ridge (yellow ridge) of optimal OOS $R^2$ slightly deviates from the analogous plots in KNS.\(^{45}\) The ridge of optimal OOS $R^2$ in PC-space of 50 anomaly portfolios has an ‘L-Shape’ as described in section 6.2.3 implying that, at a substantial level of sparsity ($L_1$-shrinkage), the degree of $L_2$-shrinkage can be relaxed without hurting OOS $R^2$ performance. In the plots presented in KNS, this ridge of optimal OOS $R^2$ values is directly vertical. However, the two ridges are located very similarly horizontally (equal degree of $L_2$-shrinkage necessary around $\kappa \approx 0.6$) and have the same vertical length (indicating an equal amount of necessary sparsity).

Similarly, we observe slight deviations in the ridge of optimal OOS $R^2$ in the PC-space for the WFR data set. Here, the inverse relation can be found, as the high OOS $R^2$ ridge from figure 5 (b) in KNS has a clear ‘L-shape’ and the ridge of the optimal OOS $R^2$ in our analogous plot in figure 6.6 (b) has a completely straight, vertical line.

\(^{45}\) Please see appendix 7 for an illustration of the contour plots from Shrinking the cross-section (Kozak et al., 2020).
Thus, despite the cosmetic differences, both plots show, in accordance with KNS, that more sparsity can be imposed without hurting OOS $R^2$ significantly, when transforming assets into PCs of factor returns.

**Largest SDF-factors**

Another point of deviation, worth further elaboration, is the difference in identification of the most relevant SDF-factors in explaining cross-sectional returns. We highlight, in section 6.2.5, quite significant deviations in the list of the largest SDF Factors, both in the empirical analysis of the 50-anomaly data set and the WFR data set. In the following section, we present potential sources for these deviations, however we do wish to emphasize that the scope of this paper is to investigate whether machine learning techniques can be used in asset allocation settings and not specifically on identifying the individual anomalies / financial ratios, that carry the most pricing power. Furthermore, as Kozak *et al.* (2020) additionally argues, the individual pricing indicator is not significant - it is the combination of relevant factors spanning the SDF that matters for good OOS performance.

**True OOS pricing performance of the MVE constructed portfolios.**

Finally, the most significant deviation, altering a sub conclusion, compared to KNS, is associated with the pricing performance of MVE portfolios in the true OOS setting. As previously mentioned, Kozak *et al.* (2020) shows that the pure $L_2$-only estimator performs better than any other constructed portfolio out of sample, however only marginally better than a PC-sparse model with five factors spanning the SDF. In our analysis, table 6.5 illustrates that the MVE portfolio based on the PC-sparse model vaguely ($\alpha = -0.5$) outperforms the $L_2$-only portfolio in the 50 anomalies data set. As such, our results, opposite the findings in KNS, show that for the 50 anomalies data set it is possible to create a PC-sparse model, without hurting OOS pricing performance.

**7.1.2 Potential sources of deviation**

In the previous subsections we accounted for the main dissimilarities in our results compared to those of Kozak *et al.* (2020). These dissimilarities could be originating from a variety of different sources, which will be discussed in the following.

**Industry vs Firm specific financial ratios in WFR**

For the WFR data set, Kozak *et al.* (2020) builds their SDF model based on financial ratios on industry level, meaning that the factor returns in their analysis is calculated based on overall industry financial ratios. To expand on their analysis and see how it maps into a firm specific characteristic
universe, we chose to base our analysis on firm specific financial ratios. Thus, some dissimilarities could arise simply from the difference in how we, for a given point in time, aggregate firm specific financial ratios to a factor return, compared to the way of aggregating industry financial ratios to a factor return in KNS. Intuitively, the rank transformation, that we apply, will result in higher ranks, since there are more observations of each financial ratio at each given point in time for firm specific data than industry specific data. Consequently, the factor portfolios based on each financial-ratio-specific strategy will have more extreme weights, which might cause excess variation to the data. Contrarily, including more observations for each characteristic should intuitively contribute more adequacy to the analysis, which is one of the reasons why we initially chose this expansion of the original data set in KNS.

**Deviations in the process of model building**

We acknowledge that deviations from the results in KNS could also potentially be caused by differences in the model setup. We might have missed some steps in the model-building procedure, as the entire estimation process is not explicitly reported in the paper. In fact, during the thorough examination of the published code underlying their model and results\(^4\), we unearthed some steps not reported in the paper. For instance, it is not reported that, nor how, the covariance matrix is regularized in the paper, why we had to unravel this, reviewing the code of Kozak et al. (2020). It proved to be a crucial step in the estimation process since regularizing the factor covariance matrix yields less-extreme values. Furthermore, when rotating returns into the space of principal components, the application of a standard eigen decomposition of the covariance matrix is reported in the paper, however in the code, singular value decomposition (SVD) is used.

To illustrate the difference between the two methods, consider an eigen decomposition, as reported in Kozak et al. (2020), \( \Sigma = QDQ' \) which is roughly equivalent to a singular value decomposition \( \Sigma = UDV' \) (Hastie & Tibshirani, 2017). The key difference between the two methods, and possibly the main reason why Kozak et al. (2020) chooses to implement the SVD method in their model, is that the vectors in the eigen decomposition \( Q \) are not necessarily orthogonal, whereas the \( U \) and \( V \) matrices are by construction orthogonal to vectors and thereby represents rotations. Furthermore, there are several other desirable properties of the SVD, such as the property that there always exists an SVD decomposition for all square matrices. In addition, all values in the diagonal of \( D \) in the SVD method

\(^4\) The entire code can be accessed at: [https://www.serhiykozak.com/data](https://www.serhiykozak.com/data)
are always real and non-negative numbers. We emphasize that both methods yield orthogonal rotations \( Q \) and \( U \) of the original matrix, where \( U \) from the SVD method is more well-behaved (Hastie et al., 2009).

**Difference in data settings**

In general, we have slightly deviating data settings which potentially could alter the identification of the largest SDF-factors in the analyses. The data sample periods are not exactly identical, resulting in varying length of time periods, which could affect the structure of the characterization of important factors. In addition, inclusion of more factors potentially spanning the SDF, such as the 12 lagged return factors, included in the WFR data set in KNS, could have possibly contributed significant explanatory power, consequently altering the dynamics of the largest SDF factors.

To investigate these two points of argument, we present a simulation in which we vary either the number of observations included (length of time period), number of factors included or both in order to assess whether the model consequently identifies alternative pricing relevant SDF-factors. Furthermore, the simulation illustrates how the estimator in equation (56) handles highly correlated factors and how coefficient estimates are affected by inclusion of new factors with high correlation to factors already included in the model. A simulation enables us to investigate these effects on a data set for which we can control its properties.

**7.1.3 Simulation**

Initially, we construct a data set of factor returns. We seek to construct the data, such that it resembles the financial ratio data, where many factors are strongly intercorrelated, as illustrated in figure 6.2. We first define a random variable, ‘Random’, following a normal distribution of \( N(0.2,0.0001) \). Then we construct 4 clusters of factor returns; ‘A’, ‘B’, ‘C’ and ‘D’ with 5 factors within each cluster (i.e. A1, A2, ..., A5). We construct the clusters in such a way, that the factors within each cluster are internally correlated to the first variable in the cluster with \( \rho \in [0, 0.9] \), multiplied by a normally distributed random variable, which is varying for each cluster. Within each cluster the first variable is a either a completely random variable or a random variable correlated with the ‘Random’ variable. In addition, we add two noise indicators, ‘Noise’ and ‘Noise1’ both distributed \( N(0,10) \), which are completely random and uncorrelated with any other factor, in order to investigate

---

\(^{47}\) These names are arbitrary as they are just used for intuition.

\(^{48}\) In order to control the correlation, we implement the formular from section 5.1.1, \( N_3 = \rho N_1 + \sqrt{1-\rho^2} N_2 \), which is then multiplied by a random variable.
how the estimator handles irrelevant variables. Finally, we construct a sixth factor in the B-cluster, ‘B6’ to evaluate whether adding more factors to a cluster affects the dynamics of the data set. Consequently, we construct a total of 24 variables, for which we simulate 200 observations per variable. This procedure results in a simulated data set with a strong factor structure within 4 clusters and, in addition, two random noise variables.

Subsequently, we apply the estimator on the simulated data set, in four different settings. The base case initially includes 150 observations pr. variable and 19 variables. The second setting includes the remaining five simulated variables (D5, B5, B6, ‘Noise’ and ‘Noise 1’) totalling 24 variables, while still only including 150 observations per variable, illustrating potential differences when more factors are added to the data set. The third setting includes the remaining 50 observations, totalling 200 observations per variable and 19 variables, to illustrate potential effects of varying the size of the data sample. The fourth and final data set includes the entire simulation, totalling 200 observations per variable and 24 variables. Table 7.1 below illustrates the 10 largest identified SDF-factors from the dual-penalty estimation for the four outlined simulation settings.

---

49 We purposely leave some observations and some variables out of the base case, such that we can later add these to test the implications of varying either number of potential factors, length of time period or both.
Table 7.1: Table 7.1 illustrates the largest SDF-factors identified by the dual-penalty estimator for the four different settings of the simulated data set described in section 7.1.3. The Base Case includes 19 variables and 150 observation per variable. The Extended Time Period setting adds 50 observations, consequently resulting in data setting with 19 variables and 200 observations per variable. The Added Variables setting, adds 5 new variables to the Base Case, consequently resulting in a data set of 24 variables and 150 observations for each variable. Finally, the Combined setting includes both additions, resulting in a data set of 24 variables and 200 observations per variable.

The results in table 7.1, shows that the estimator in equation (56) is fairly unaffected by changes in time periods, as the only difference from the base case in (a) and the extended data sample in (b) is the substitution of the variable B4 with B2 in the 10th spot. As such, the dynamics do not significantly change, when adding observations / extending the data sample. We note that the C-cluster seems to have the coefficients with the highest pricing relevance as C5 through C1 occupies the top spots on the list. Comparing the base case (a) with the setting with added variables (c) there are quite significant changes to the dynamics of the largest SDF-coefficients. Two of the newly added variables enter the top five as B6 becomes the most pricing relevant factor and D5 is placed 5th, consequently pushing some previously relevant pricing factors out of the Top-10 category. We do however note that the order of relevant pricing factors remains intact, as only the least relevant factors are affected by the addition of the new, highly correlated, factors and not the dynamics, per se. In the

<table>
<thead>
<tr>
<th>Base Case</th>
<th>Extended Time Period</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Variable</strong></td>
<td><strong>Beta</strong></td>
</tr>
<tr>
<td>C5</td>
<td>0.46</td>
</tr>
<tr>
<td>C4</td>
<td>0.38</td>
</tr>
<tr>
<td>C3</td>
<td>0.33</td>
</tr>
<tr>
<td>C2</td>
<td>0.27</td>
</tr>
<tr>
<td>C1</td>
<td>0.24</td>
</tr>
<tr>
<td>Random</td>
<td>0.13</td>
</tr>
<tr>
<td>D2</td>
<td>0.14</td>
</tr>
<tr>
<td>D3</td>
<td>0.14</td>
</tr>
<tr>
<td>D4</td>
<td>0.14</td>
</tr>
<tr>
<td>B4</td>
<td>0.14</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Added Variables</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Variable</strong></td>
<td><strong>Beta</strong></td>
</tr>
<tr>
<td>B6</td>
<td>0.39</td>
</tr>
<tr>
<td>C5</td>
<td>0.51</td>
</tr>
<tr>
<td>C4</td>
<td>0.42</td>
</tr>
<tr>
<td>C3</td>
<td>0.36</td>
</tr>
<tr>
<td>D5</td>
<td>0.32</td>
</tr>
<tr>
<td>C2</td>
<td>0.29</td>
</tr>
<tr>
<td>C1</td>
<td>0.26</td>
</tr>
<tr>
<td>Random</td>
<td>0.15</td>
</tr>
<tr>
<td>D2</td>
<td>0.15</td>
</tr>
<tr>
<td>D3</td>
<td>0.15</td>
</tr>
</tbody>
</table>
combined setting (d), the only difference from (c) is the entry of one of the ‘Noise’-factors in the lower range of the list.

Overall, the results are satisfying, in the sense that it shows the estimators ability to, fairly consistently, identify significant factors. However, between the four analyses we see some slight deviations as described above. As such, the simulation indicates that we should find a majority of intersections in our significant factors as compared to those in Kozak et al. (2020). However, the simulation also showed that some significant factors might drop out of the analysis when other significant factors are included. Hence, as previously discussed, our deviating ‘largest-factor identification’ from the findings in KNS can partly be attributed to the 12 lagged return indicators, that seemingly have high pricing relevance for the WFR data set. Nevertheless, based on the simulation above, one would expect more correspondence in significant factors between our findings and KNS.

Regarding the varying length of the data sample it should be highlighted that the simulation fails to accommodate a quite significant event, which we would expect to see in a practical setting. In the simulated data set, the correlation of variables is constant over time. However, the lag of correspondence might be a result of changing pricing relevancy (changing correlations) of some factors over time. As discussed by Munk (2013), a vast variety of papers have tried to identify factor models where the addition of a factor return to the market portfolio should improve the explanatory power. Munk (2013) discusses that for any given set of historical returns it is always possible to identify factors that work in describing returns. However, there is no economic arguments suggesting that a historically effective factor will be equivalently effective in the future. As such, factors with high pricing relevance, in one period could potentially be irrelevant or even inversely correlated in another period. The deviations from the results in KNS could, as such, potentially be credited to the difference in the location of the time period rather than the length of the data sample.

The dynamic pricing relevance of anomalies could also be a main factor in explaining the outperformance of the PC-sparse SDF-implied MVE portfolio compared to the $L_2$-only portfolio in the true OOS empirical test for the 50 anomalies data set in (section 6.3.2). Here, due to limited data-availability, our data sample is daily from January 2000 – December 2017 (of which January 2015

---

50 In table 5.3 depicting the results in KNS t-9, t-11 and t-1 occupies the 4th, 5th and 10th largest spots respectively.

51 In order to make the performance of the portfolios comparable with the portfolios based on the WRF data set we applied the intersecting data period of the two data sets.
December 2017 is the true OOS period), whereas KNS has sample data from November 1973 to December 2017 (with the OOS period ranging from January 2005 – December 2017). Keeping in mind the changing nature of pricing relevant anomalies, the difference in data period could alter the findings of the analysis quite significantly.

7.1.4 Subsection summary

In the section above, we outlined the main points of deviations in the findings from our analysis and KNS, which can roughly be divided into three areas; Cosmetic differences in OOS R² contour plots, deviating largest SDF-factor identification and performance of MVE portfolios in true out of sample setting. We then discussed potential sources of dissimilarities and found several plausible explanations such as; difference in data settings, time varying factor significance and slight deviations in the process of model building. We highlight that none of these dissimilarities has any significant effect on the overall conclusions and findings in this thesis, as they are consistent with KNS.

7.2 Discussion of methodology

In the following section we evaluate the methodology applied in KNS, and in this paper, by outlining issues and concerns and discussing how these issues are mitigated in the literature. We summarize the methodology in figure 7.1 by providing an overview of the different sections.

![Figure 7.1](image)

Figure 7.1: Figure 7.1 illustrates the stepwise model-building process, summarizing the different components in the methodology. As such the methodology is divided into five chronological steps, beginning with (1) the choosing and refining of raw data inputs in Definition of data input. Data is then further refined in (2) Construction of factor portfolios, (3) Alteration of factor returns, (4) Estimation of SDF-coefficients and (5) Evaluation of OOS performance. The stepwise process shows each of the major interventions performed throughout the model-building process. Furthermore, it is a helpful tool, illustrating where, in the process, other empiricists have proposed the application of machine learning methods, for asset pricing.

We examine different components in the methodology, as summarized in figure 7.1, and discuss shortcomings and possible improvements.
7.2.1 Biased beta estimates

We initially consider step 4 in which we estimate the SDF-coefficients. When running penalized regression models such as ridge regression, LASSO and elastic net, the researcher is faced with the problem of choosing the optimal values of the penalization parameters. We call this problem the bias-variance trade-off. It arises since the choice is a trade-off between a completely unbiased model (like the OLS regression) yielding the actual mean across different simulations/iterations, but with the risk of potentially overfitting the model, consequently performing poorly out of sample, and a biased estimator with penalty terms that shrinks the contribution of each characteristic towards zero, and thereby reducing the variance compared to OLS\(^{52}\). In section 5.1.1 we described the problems of the unbiased OLS regression model. Even though penalized regressors overcome these issues, other potential problems could arise. One of these problems arises if the true values of the coefficients, \( \beta \), are in fact very large, the penalized regressors might shrink contribution of significant coefficients too much. However, it seems plausible to assume that in the high-dimensional setting no single factor in the two data sets (WFR and 50 anomalies) individually contributes large amounts of explanatory power.

Further, one could argue that we mitigate this problem by tuning the penalty parameters, \( \gamma_1 \) and \( \gamma_2 \), through data-driven cross-validation of their out of sample explanatory power, \( R^2 \). As such, if the ‘true’ coefficients are in fact very large, this data-driven approach should automatically choose penalty parameters that decrease strength of the penalty terms sufficiently. As such, since KNS do not explicitly state their level of shrinkage in the selection of significant factor coefficients, our deviation from their significant factor selection might be originating from different levels of bias imposed on the coefficients, and thereby variance in the coefficient estimates, ultimately yielding different T-statistics. The \( \kappa \) yielding the highest OOS \( R^2 \) in our analysis is close to 0.6 (for the 50 anomalies data set) and 0.9 (for the WFR data set) whereas the optimal OOS \( R^2 \) in KNS, is obtained at \( \kappa \) close to 0.3 and 1.1, respectively. Therefore, the chosen \( \kappa \) for the \( L_2 \)-only estimation, which is not explicitly stated in KNS, could also deviate substantially. This could potentially explain why we experience deviations in significant factors for the 50 anomalies data.

The dual-penalty regression estimator is, in fact, imposing a bias on the coefficients, raising the question as to whether conventional methods for estimation of standard deviation and t-statistics are

\[^{52}\text{Pirinen, 2019, Visited 05-05-2021} \]

https://www.mv.helsinki.fi/home/mjxpirin/HDS_course/material/HDS6_penalized_regression.html
sufficient. If conventional methods for estimating standard deviation, and thereby t-statistics, are not sufficient for biased estimators, questions, concerning whether the metric for measuring significance of coefficients is appropriate, can be raised. Harvey, Liu and Zhu (2015) argue that generally, there are two main approaches to handling biased estimators: Out of sample validation and statistical frameworks that allows for multiple testing.

In this thesis we applied out of sample validation, in which we tested the out of sample performance in order to determine the optimal level of bias imposed in our model. However, as previously argued, our cross-validation of out of sample performance, when choosing the optimal level of bias, is not a true out of sample test. A true out of sample test would require future data. To mitigate this problem, following Kozak et al. (2020), we performed a true out of sample MVE portfolio analysis. However, the scope of this analysis was not to determine the optimal level of bias imposed by the penalty terms, but rather to assess the asset pricing implications of the dual-penalty estimator. As such, multiple testing methods in which one applies statistical techniques to obtain more sample data could have been advantageous.

Chatterjee & Lahiri (2011) applies one of such approaches. They suggest that a bootstrapping method for consistent estimation of coefficient standard deviation is more precise for penalized regression techniques. They argue that it is difficult to attach appropriate standard error estimates to coefficients in a LASSO regression. Huang and Harrison (2002) further propose a method for bootstrap selection of penalty parameters. Their simulations show a bootstrap method that decreases bias, mean squared estimation errors and prediction errors for some of their simulations. Hence, for the purpose of estimating unbiased results regarding the significance of coefficients, applying a bootstrapping approach might increase precision of the significance test, and potentially mitigate the problem of bias on our beta estimates.

7.2.2 Data-snooping

Apart from bias in coefficient estimates another big issue in the machine learning literature, and particularly in financial literature, is data-snooping. It is argued that many test results might be due to luck, rather than the actual result being significant. As mentioned in the introduction, there is a general issue in financial literature of reproducing results across time and in different data settings.

---

53 Their analysis is for penalized partial likelihood regression within the field of Biostatistical Science. However, the penalized partial likelihood regression resembles the estimator that we used, and the paper also generalizes its method to nonparametric regression modeling.
This could be accredited to data-snooping. As such, we consider how data-snooping might have affected our results, and thereby how step 5, in figure 7.1, could be improved.

Sullivan, Timmermann and White (2001) argue that when data is used more than once to evaluate the performance of a given model, a data-snooping problem arises. The cross-validation method for evaluating the OOS performance of our model is exposed to this data-snooping problem since the three contiguous data sets, that we form from the sample, are reused across all three folds. Sullivan et al. (2001) argue that when a procedure is subject to this data-snooping problem, “there is always the possibility that any satisfactory results obtained may simply be due to chance rather than to any merit inherent in the method yielding the results” (Sullivan et al., 2001, p.1648). One could argue that we accommodate this problem by replicating an existing analysis, and applying the framework in a new sample setting, thereby cross-validating the results for different samples. Nevertheless, Sullivan et al. (2001) extends the work of Brock, Lakonishok and LeBaron (1992), evaluating results by fitting several models to the same raw sample data and resampling the errors to produce several bootstrap samples. Hence, introducing such bootstrap methods could further validate our findings in step 5 of figure 7.1.

7.2.3 Alternative selection of basis assets

In step 2, in figure 7.1, construction of factor returns, we define each characteristics-specific strategy for computing factor portfolios, by simply sorting portfolios based on the characteristics z-score. Bryzgalova et al. (2019) suggest applying decision trees instead of linear long-short factors, or PCs, in order to generalize the concept of conventional sorting. They argue that decision trees are more interpretable than PCs and that they enable the definition of combinations of factors as the basis assets. In particular, one can build factor portfolios based on different cuts in a decision tree. One factor could be an original node in the tree, for instance the market return factor, and another factor could be a cut in the size dimension resulting in small cap and large cap portfolios (Bryzgalova et al., 2019). Additionally, one could choose a portfolio based on double sorting where, for instance, the basis asset is small cap value stocks. By applying decision trees in this way, one could reduce the asset universe to fewer sorted portfolios and thereby construct a sparse model, in which it is always possible to trace back the portfolio to the original stock characteristics. Thus, the interpretational value, as compared to the PC-sparse models, could be improved, as it includes information from many characteristics while considering only a few factors in the model.
Empirically, Bryzgalova et al. (2019) show that traditionally sorted portfolios, like the factor portfolios we estimate, do not sufficiently account for the complex information contained in the structure of the raw basis assets. Therefore, expanding the initial basis asset universe to portfolios, sorted on decision trees, might improve performance of our estimator, since more information would be included in each basis asset and potentially yield a sparser model.

7.2.4 Subsection summary

In this section we argued that our penalized regressor is, in fact, imposing bias on the SDF-coefficients, resulting in a bias-variance trade-off. As such, the researcher is faced with the issue of choosing highly biased estimates with low variance or high-variance estimates with low bias. We found that several methods exist for estimating variance more adequately, for biased coefficients in penalized regressions, which could have been implemented, in order to improve accuracy of the assessment of significant coefficients (t-statistic).

Furthermore, we discussed how the cross-validation method, following Kozak et al. (2020) might be subject to a data-snooping concern, since data is used more than once to evaluate the performance of a given model. We argued why bootstrap methods could improve validity of our findings. Finally, we discussed how decision trees, instead of conventional sorting, could expand the basis asset universe for the estimator and potentially improve OOS performance.
7. Conclusion

Our results suggest, in similar fashion to Kozak et al. (2020), that it is evidently futile to summarize the cross-section of stock returns with characteristics-sparse factor models. There is simply not enough redundancy among cross-sectional return predictors to adequately price the cross-section with only a handful of factors. For the model to perform well out-of-sample, the SDF necessarily must load on a vast number of characteristics-based factors. This finding significantly challenges the conventional multi-factor models including only a handful of return predictors (e.g., 3, 4, 5 or 6 factor-models have previously been proposed).54

In a high-dimensional data setting, shrinkage of the SDF-coefficients towards zero is key in order to obtain good OOS pricing performance. In this paper, following Kozak et al. (2020) we impose both $L_1$- and $L_2$-penalization to shrink the SDF-coefficients with a dual-penalty regression. Our results show that $L_2$-shrinkage (equivalent to ridge-regression) is the most effective penalizer, and even manages to yield good OOS performance on its own, as it successfully shrinks the non-relevant factors towards zero. The $L_1$-penalty (equivalent to LASSO-regression) can be a convenient tool in removing the most redundant factors, but quickly becomes an obstacle, as further sparsity hurts performance. Furthermore, it performs poorly on its own. This is consistent with the findings in KNS.

A sparse SDF application with good out-of-sample performance is however achievable if we, instead of raw characteristics-based portfolio returns, apply principal components of characteristics-based portfolio returns. Our results suggest that a small number of high-variance PCs sufficiently span the SDF while ensuring good predictive power on cross-sectional returns. This consequently yields a sparse factor model with a few variables spanning the SDF, however one must keep in mind that all characteristics / factor-returns are still used in constructing this optimal SDF.

In the true OOS empirical evaluation, assessing the performance of MVE portfolios, the findings, highlighted above, are further solidified. The mean-variance efficient portfolios implied by our estimated SDF serve as useful test assets to evaluate the pricing implications of the different models for cross-section of returns. The empirical analysis shows that portfolios based on $L_2$-only estimation (i.e., a model summarizing the pricing information contained in a large number of characteristics-based factors) and MVE portfolios implied by a PC-sparse SDF, significantly outperform the benchmark portfolios, CAPM and (more importantly) the Fama-French 6-factor model. As such,

---

54 We refer to the Fama-French multi-factor models who has been expanded from originally including three factors (Fama and French, 1992) describing the cross-section up to as many as six factors (Roy, 2018).
similar to Kozak et al. (2020), we have shown that the six-factor model of Fama and French (2016) leaves much of the cross-section of stock returns unexplained.

As discussed in the previous section, we do note some dissimilarities between the results in this paper and the results presented in KNS. These deviations could potentially originate from a variety of different sources, highlighted in the discussion section of the paper. However, the majority of the deviations are mostly cosmetic and none of them alters the conclusions drawn, why we do not deem these dissimilarities heavily significant.

As such, this paper illustrates that one can successfully replicate the estimation method and findings originally presented by Kozak et al. in Shrinking the cross-section (2020).
8. Bibliography


9. Appendix

The following is a descriptive table of contents of the documents attached in the appendix of this thesis.

Appendix 1 – Code Overview

Appendix 1 presents an overview of the different code files (r-software) which can either be found the attachments to the thesis submission or accessed via the following Dropbox link: https://www.dropbox.com/sh/6foktnmzq0p64q9/AADEcYPHnWaVqjAGj5_qEc4Xa?dl=0

Appendix 2 – WRDS Industry Financial Ratios


Appendix 3 – 50 anomalies manual

Appendix 3 is a description of the 50 previously proven anomalies in the literature created by Kozak et al. (2020), withdrawn from: https://www.serhiykozak.com/

Appendix 4 and 5 – Correlation heatmaps (50 anomalies and WFR)

Appendix 4 and 5 shows a more detailed version of the correlation heatmaps in figure 6.2 (a) and (b), where the name of each anomaly or financial ratio, underlying the factors, are displayed.

Appendix 6 – Summary Statistics for 50 anomaly portfolios and WFR

Appendix 6 shows the expanded summary statistics (from section 6.2.1) of the two high-dimensional data sets applied in the thesis.

Appendix 7 – Contour plots from Shrinking the Cross-Section (Kozak et al., 2020)

Appendix 7 includes the contour plots from the KNS-article exactly as they are displayed. These have been included for the sake of comparison.
Appendix 1 - Code overview

Code for analysis part 1 - Cross Validated Contour plots

1 - Template_LarsEN_Analysis1
This is the template applied for all our dual penalty estimations. We refer to this for a walkthrough of our code since this contains comments on every step of the code. In addition to this code walkthrough, we provide the specific code used to produce the main plots in the first analysis regarding WFR and FM50. These files are called “LarsEN50Anomalies_Final” and “LarsEnWFR_Final”.

2 - LarsEN50Anomalies_Final
This is the specific code producing the plots for the 50 anomaly data. The Underlying data file is “managed_portfolios_anom_d_50”

3 - LarsEnWFR_Final
This is the specific code producing the plots for the WFR data. The Underlying data file is “Factor_returns_WRDS_Done”

Code for analysis part 2 - MVE portfolios

4 - Template_MVE_Returns
This is the template applied to estimate MVE portfolio returns for both the WFR and 50 Anomalies data sets. We refer to this for a walkthrough of our code, since this contains comments on every step of the code. In addition to this code walkthrough, we provide the specific code used to produce the main plots in the second analysis regarding WFR and FM50. These files are called “MVE_OOS_Test_50Anomalies”, “MVE_OOS_Test_WFR” and “FF6Portfolio”.

5 - MVE_OOS_Test_50Anomalies
This is the specific code producing MVE portfolios for the WFR data. Underlying data file: “managed_portfolios_anom_d_50”

6 - MVE_OOS_Test_WFR
This is the specific code producing MVE portfolios for the 50 anomalies data. Underlying data file: “Factor_returns_WRDS_Done”
7 - FF6Portfolio
This is the specific code producing MVE portfolios for the FF6 data. The underlying data files is FF6Factors.txt

8 - MVE_Portfolios & Betas_Final
This code gathers all seven portfolios return files and combines them to one. The portfolios are for WFR and 50 anomalies the Full-, Characteristic-sparse- and PC-sparse model and finally the FF6 benchmark portfolio. We analyze this gathered file in the excel sheet “Portfolio_returns_done”

Other codes

9 - TemplateBetaEstimation
This is the specific code we applied in withdrawing beta coefficients and estimating standard deviations and t-statistics throughout the analysis.

10 - WRDSFactorReturnsSplit
This is the code we applied to compute factor returns from our initial raw financial ratios. It has the suffix “split” since we split the data in 3 parts in order for our computer to process the computations. The Underlying data files for this step are the financial ratios withdrawn from the WRDS data base and the returns from the Eikon Datastream database.

11 - CombineSplitFactors
This code gathers the splitted factor returns from “WRDSFactorReturnsSplit”

12 - FM25PCA
This is the initial PCA analysis for the FM25 data set.

13 - DataAnalysis
This is the code for our initial data analysis of correlation heat maps and mean return histograms. The Underlying data files for this step are “Factor_returns_WRDS_Done” and “managed_portfolios_anom_d_50”

14 - SimulationFinal
This code runs the simulation based on our template “Template_LarsEN_Analysis1”. The underlying data for this step is our simulated data from the excel spreadsheet “Simulation_final”
Excel spreadsheets

1 - FF6Factors
This file performs orthogonalization and volatility scaling of the FF6 factor returns

2 - managed_portfolios_anom_d_50
This file performs orthogonalization and volatility scaling of the 50 anomalies factor returns

3 - Factor_returns_WRDS_Done
This file performs orthogonalization and volatility scaling of the WFR factor returns

4 - Portfolio_returns_done
This data sets scale the volatility to match the market again and then computes alpha plots.

5 - Simulation_Multicolinearity
This sheet simulates the data for our simulation in the theoretical framework.

6 - Simulation_Final
This sheet simulates the data for our simulation in the discussion-section.
Appendix 2 - WRDS Industry Financial Ratio

We access this description from the WRDS database:

August 2016
WRDS Research Team

Overview

WRDS Industry Financial Ratio (WIFR hereafter) is a collection of most commonly used financial ratios by academic researchers. There are in total over 70 financial ratios grouped into the following seven categories: Capitalization, Efficiency, Financial Soundness/Solvency, Liquidity, Profitability, Valuation and Others. Ratios for each individual company as well as at industry aggregated level are included in the output.

Parameter Specification

Universe Selection:

Users can choose between the universe of CRSP Common Stock and S&P 500 Index Constituents. As many of the ratios studied here are void of economic meanings among the finance companies, we have hence excluded these firms from our universe.

Industry Classification:

Two systems of industry classification are accepted in the WIFR: GICS Economic Sector Level Index, and Fama-French Industry Classification. More specifically, the GICS classification includes 10 distinct economic sectors: Energy, Materials, Industrials, Consumer Discretionary, Consumer Staples, Health Care, Financials, Information Technology, Telecom and Utilities. As Fama-French carries more than one unique industry classification system, we allow users to choose the exact number of industries.\(^1\)

Industry Level Aggregation:

Aggregation of financial ratios to industry level is an important consideration, especially when it comes to valuation ratios. As researchers have previously pointed out, P/E ratios (or generally ratios that use denominators that can be negative) should never be averaged.\(^2\) While some industry practitioners advocate simply dropping out all the firms with non-positive ratios before aggregation, we propose keeping all the observations and taking the median, rather than mean, as

\(^1\)Please see Kenneth R. French’s website for detailed industry classification description.
http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

\(^2\)See discussion on p 239 of Welch’s “Intro to Finance”,
the industry-level ratio. Users will still have access to all the firm-level ratio results, and if needed, can choose other aggregation metrics to arrive at the industry-level ratio.

Financial Ratio Definition

Theme Classification

WIFR collects over 70 different financial ratios, categorized based on the economic intuition into the following seven groups:

1. **Capitalization**: measures the debt component of a firm’s total capital structure, e.g.: Capitalization Ratio, Total Debt-to-Invested Capital Ratio;
2. **Efficiency**: captures the effectiveness of firm’s usage of assets and liability, e.g.: Asset Turnover, Inventory Turnover;
3. **Financial Soundness/Solvency**: captures the firm’s ability to meet long-term obligations, e.g.: Total Debt to Equity Ratio, Interest Coverage Ratio;
4. **Liquidity**: measures a firm’s ability to meet its short-term obligations, e.g.: Current Ratio, Quick Ratio;
5. **Profitability**: measures the ability of a firm to generate profit, e.g.: ROA, Gross Profit Margin;
6. **Valuation**: estimates the attractiveness of a firm’s stock (overpriced or underpriced), e.g.: P/E ratio, Shiller’s CAPE ratio;
7. **Others**: Miscellaneous ratios, e.g.: R&D-to-Sales, Labor Expenses-to-Sales.

Please refer to the Appendix section for complete list of financial ratios and corresponding categorization.

Individual Ratio

Individual financial ratios are samples of most commonly used metrics by academic researchers or industry practitioners.

Data Source:

All accounting related data are obtained from Compustat Quarterly and Annual file. Pricing related data, such as Market Capitalization and Price, are obtained from both CRSP and Compustat, and we rely on CRSP as the primary data source for pricing data. Earnings related data are from IBES database.

Data Frequency:

The final outputs for both individual firm and industry-level aggregated value are at monthly frequency. In order to populate the data to monthly frequency, we carry forward

---

3 We set default industry aggregation method to be *Median*, and yet provide users with the option of taking simple average across industry classification. Please use the *Mean* option with caution.
the most recent quarterly or annual data item, whichever is most recently available at a
given time stamp, to the subsequent months before the next filing data becomes available.

In addition, in order to make sure that all data is publicly available at the monthly time
stamp, we lag all observations by two months to avoid any look ahead bias.4

Outlier Control:

As ratio metrics often produce unintended extreme outliers, we impose two layers of
outliers control before aggregating at the industry level. First, for all the monthly frequency
firm level individual ratio results, we impose a winsorization at 1% level for extreme values,
and truncate the outliers in the top and bottom percentile to be missing. Secondly, to arrive
at the final ratio output, we enforce a 12 month moving average on the monthly frequency
financial ratios. The second step serves two purpose: to further smooth the final output, and
to fill in the truncated extreme months (from step 1) with firm-specific moving average.

Note that the outlier controls are only applied to the ratios fed to the industry-level
aggregation. Outputs for firm-level financial ratios are raw ratios without any truncation or
smoothing. Hence researchers are advised to censor/smooth the raw ratios to get rid of the
extreme outliers before conducting further analysis.

Ratio Definition/Construction:

We provide definition to each individual ratio in the Appendix section. Please refer to the
previous discussion on Data Frequency and Outlier Control for general guideline on data
alignment and other technical treatment.

For the underlying code used to produce these ratios, please refer to the “Financial Ratio
SAS Code” section listed under “Manuals and Overviews” page.

---

4 Although most recent filings carry only 45 days’ latency in the Compustat database, we set a two-month rule
in order to make sure earlier filings are public information as of the monthly observation date.
## Appendix: List of financial ratios and categorization

<table>
<thead>
<tr>
<th>Financial Ratio</th>
<th>Variable Name</th>
<th>Category</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capitalization Ratio</td>
<td>capital_ratio</td>
<td>Capitalization</td>
<td>Total Long-term Debt as a fraction of the sum of Total Long-term Debt, Common/Ordinary Equity and Preferred Stock</td>
</tr>
<tr>
<td>Common Equity/Invested Capital</td>
<td>equity_invcap</td>
<td>Capitalization</td>
<td>Common Equity as a fraction of Invested Capital</td>
</tr>
<tr>
<td>Long-term Debt/Invested Capital</td>
<td>debt_invcap</td>
<td>Capitalization</td>
<td>Long-term Debt as a fraction of Invested Capital</td>
</tr>
<tr>
<td>Total Debt/Invested Capital</td>
<td>totdebt_invcap</td>
<td>Capitalization</td>
<td>Total Debt (Long-term and Current) as a fraction of Invested Capital</td>
</tr>
<tr>
<td>Asset Turnover</td>
<td>at_turn</td>
<td>Efficiency</td>
<td>Sales as a fraction of the average Total Assets based on the most recent two periods</td>
</tr>
<tr>
<td>Inventory Turnover</td>
<td>inv_turn</td>
<td>Efficiency</td>
<td>COGS as a fraction of the average Inventories based on the most recent two periods</td>
</tr>
<tr>
<td>Payables Turnover</td>
<td>pay_turn</td>
<td>Efficiency</td>
<td>COGS and change in Inventories as a fraction of the average Accounts Payable based on the most recent two periods</td>
</tr>
<tr>
<td>Receivables Turnover</td>
<td>rect_turn</td>
<td>Efficiency</td>
<td>Sales as a fraction of the average of Accounts Receivables based on the most recent two periods</td>
</tr>
<tr>
<td>Sales/Stockholders Equity</td>
<td>sale_equity</td>
<td>Efficiency</td>
<td>Sales per dollar of total Stockholders' Equity</td>
</tr>
<tr>
<td>Sales/Invested Capital</td>
<td>sale_invcap</td>
<td>Efficiency</td>
<td>Sales per dollar of Invested Capital</td>
</tr>
<tr>
<td>Sales/Working Capital</td>
<td>sale_nwc</td>
<td>Efficiency</td>
<td>Sales per dollar of Working Capital, defined as difference between Current Assets and Current Liabilities</td>
</tr>
<tr>
<td>Inventory/Current Assets</td>
<td>invt_act</td>
<td>Financial Soundness</td>
<td>Inventories as a fraction of Current Assets</td>
</tr>
<tr>
<td>Receivables/Current Assets</td>
<td>rect_act</td>
<td>Financial Soundness</td>
<td>Accounts Receivables as a fraction of Current Assets</td>
</tr>
<tr>
<td>Free Cash Flow/Operating Cash Flow</td>
<td>fcf_ocf</td>
<td>Financial Soundness</td>
<td>Free Cash Flow as a fraction of Operating Cash Flow, where Free Cash Flow is defined as the difference between Operating Cash Flow and Capital Expenditures</td>
</tr>
<tr>
<td>Operating CF/Current Liabilities</td>
<td>ocf_lct</td>
<td>Financial Soundness</td>
<td>Operating Cash Flow as a fraction of Current Liabilities</td>
</tr>
<tr>
<td>Cash Flow/Total Debt</td>
<td>cash_debt</td>
<td>Financial Soundness</td>
<td>Operating Cash Flow as a fraction of Total Debt</td>
</tr>
<tr>
<td>Cash Balance/Total Liabilities</td>
<td>cash_lt</td>
<td>Financial Soundness</td>
<td>Cash Balance as a fraction of Total Liabilities</td>
</tr>
<tr>
<td>Cash Flow Margin</td>
<td>cfm</td>
<td>Financial Soundness</td>
<td>Income before Extraordinary Items and Depreciation as a fraction of Sales</td>
</tr>
<tr>
<td>Short-Term Debt/Total Debt</td>
<td>short_debt</td>
<td>Financial Soundness</td>
<td>Short-term Debt as a fraction of Total Debt</td>
</tr>
<tr>
<td>Financial Ratio</td>
<td>Variable Name</td>
<td>Category</td>
<td>Formula</td>
</tr>
<tr>
<td>----------------------------------------------------</td>
<td>-----------------</td>
<td>----------------</td>
<td>-------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Profit Before Depreciation/Current Liabilities</td>
<td>profit_lct</td>
<td>Financial Soundness</td>
<td>Operating Income before D&amp;A as a fraction of Current Liabilities</td>
</tr>
<tr>
<td>Current Liabilities/Total Liabilities</td>
<td>curr_debt</td>
<td>Financial Soundness</td>
<td>Current Liabilities as a fraction of Total Liabilities</td>
</tr>
<tr>
<td>Total Debt/EBITDA</td>
<td>debt_ebitda</td>
<td>Financial Soundness</td>
<td>Gross Debt as a fraction of EBITDA</td>
</tr>
<tr>
<td>Long-term Debt/Book Equity</td>
<td>dltt_be</td>
<td>Financial Soundness</td>
<td>Long-term Debt to Book Equity</td>
</tr>
<tr>
<td>Interest/Average Long-term Debt</td>
<td>int_debt</td>
<td>Financial Soundness</td>
<td>Interest as a fraction of average Long-term debt based on most recent two periods</td>
</tr>
<tr>
<td>Interest/Average Total Debt</td>
<td>int_totdebt</td>
<td>Financial Soundness</td>
<td>Interest as a fraction of average Total Debt based on most recent two periods</td>
</tr>
<tr>
<td>Long-term Debt/Total Liabilities</td>
<td>lt_debt</td>
<td>Financial Soundness</td>
<td>Long-term Debt as a fraction of Total Liabilities</td>
</tr>
<tr>
<td>Total Liabilities/Total Tangible Assets</td>
<td>lt_ppent</td>
<td>Financial Soundness</td>
<td>Total Liabilities to Total Tangible Assets</td>
</tr>
<tr>
<td>Cash Conversion Cycle (Days)</td>
<td>cash_conversion</td>
<td>Liquidity</td>
<td>Inventories per daily COGS plus Account Receivables per daily Sales minus Account Payables per daily COGS</td>
</tr>
<tr>
<td>Cash Ratio</td>
<td>cash_ratio</td>
<td>Liquidity</td>
<td>Cash and Short-term Investments as a fraction of Current Liabilities</td>
</tr>
<tr>
<td>Current Ratio</td>
<td>curr_ratio</td>
<td>Liquidity</td>
<td>Current Assets as a fraction of Current Liabilities</td>
</tr>
<tr>
<td>Quick Ratio (Acid Test)</td>
<td>quick_ratio</td>
<td>Liquidity</td>
<td>Quick Ratio: Current Assets net of Inventories as a fraction of Current Liabilities</td>
</tr>
<tr>
<td>Accruals/Average Assets</td>
<td>Accrual</td>
<td>Other</td>
<td>Accruals as a fraction of average Total Assets based on most recent two periods</td>
</tr>
<tr>
<td>Research and Development/Sales</td>
<td>RD_SALE</td>
<td>Other</td>
<td>R&amp;D expenses as a fraction of Sales</td>
</tr>
<tr>
<td>Advertising Expenses/Sales</td>
<td>adv_sale</td>
<td>Other</td>
<td>Advertising Expenses as a fraction of Sales</td>
</tr>
<tr>
<td>Labor Expenses/Sales</td>
<td>staff_sale</td>
<td>Other</td>
<td>Labor Expenses as a fraction of Sales</td>
</tr>
<tr>
<td>Effective Tax Rate</td>
<td>efftax</td>
<td>Profitability</td>
<td>Income Tax as a fraction of Pretax Income</td>
</tr>
<tr>
<td>Gross Profit/Total Assets</td>
<td>GProf</td>
<td>Profitability</td>
<td>Gross Profitability as a fraction of Total Assets</td>
</tr>
<tr>
<td>After-tax Return on Average Common Equity</td>
<td>aftret_eq</td>
<td>Profitability</td>
<td>Net Income as a fraction of average of Common Equity based on most recent two periods</td>
</tr>
<tr>
<td>After-tax Return on Total Stockholders’ Equity</td>
<td>aftret_equity</td>
<td>Profitability</td>
<td>Net Income as a fraction of average of Total Shareholders’ Equity based on most recent two periods</td>
</tr>
<tr>
<td>Financial Ratio</td>
<td>Variable Name</td>
<td>Category</td>
<td>Formula</td>
</tr>
<tr>
<td>-----------------------------------------------------</td>
<td>--------------------</td>
<td>--------------</td>
<td>-------------------------------------------------------------------------</td>
</tr>
<tr>
<td>After-tax Return on Invested Capital</td>
<td>aftret_invcapx</td>
<td>Profitability</td>
<td>Net Income plus Interest Expenses as a fraction of Invested Capital</td>
</tr>
<tr>
<td>Gross Profit Margin</td>
<td>gpm</td>
<td>Profitability</td>
<td>Gross Profit as a fraction of Sales</td>
</tr>
<tr>
<td>Net Profit Margin</td>
<td>npm</td>
<td>Profitability</td>
<td>Net Income as a fraction of Sales</td>
</tr>
<tr>
<td>Operating Profit Margin After Depreciation</td>
<td>opmad</td>
<td>Profitability</td>
<td>Operating Income After Depreciation as a fraction of Sales</td>
</tr>
<tr>
<td>Operating Profit Margin Before Depreciation</td>
<td>opmbd</td>
<td>Profitability</td>
<td>Operating Income Before Depreciation as a fraction of Sales</td>
</tr>
<tr>
<td>Pre-tax Return on Total Earning Assets</td>
<td>pretret_eamat</td>
<td>Profitability</td>
<td>Operating Income After Depreciation as a fraction of average Total Earnings Assets (TEA) based on most recent two periods, where TEA is defined as the sum of Property Plant and Equipment and Current Assets</td>
</tr>
<tr>
<td>Pre-tax return on Net Operating Assets</td>
<td>pretret_noa</td>
<td>Profitability</td>
<td>Operating Income After Depreciation as a fraction of average Net Operating Assets (NOA) based on most recent two periods, where NOA is defined as the sum of Property Plant and Equipment and Current Assets minus Current Liabilities</td>
</tr>
<tr>
<td>Pre-tax Profit Margin</td>
<td>ptpm</td>
<td>Profitability</td>
<td>Pretax Income as a fraction of Sales</td>
</tr>
<tr>
<td>Return on Assets</td>
<td>roa</td>
<td>Profitability</td>
<td>Operating Income Before Depreciation as a fraction of average Total Assets based on most recent two periods</td>
</tr>
<tr>
<td>Return on Capital Employed</td>
<td>roce</td>
<td>Profitability</td>
<td>Earnings Before Interest and Taxes as a fraction of average Capital Employed based on most recent two periods, where Capital Employed is the sum of Debt in Long-term and Current Liabilities and Common/Ordinary Equity</td>
</tr>
<tr>
<td>Return on Equity</td>
<td>roe</td>
<td>Profitability</td>
<td>Net Income as a fraction of average Book Equity based on most recent two periods, where Book Equity is defined as the sum of Total Parent Stockholders’ Equity and Deferred Taxes and Investment Tax Credit</td>
</tr>
<tr>
<td>Total Debt/Equity</td>
<td>de_ratio</td>
<td>Solvency</td>
<td>Total Liabilities to Shareholders’ Equity (common and preferred)</td>
</tr>
<tr>
<td>Total Debt/Total Assets</td>
<td>debt_assets</td>
<td>Solvency</td>
<td>Total Debt as a fraction of Total Assets</td>
</tr>
<tr>
<td>Total Debt/Total Assets</td>
<td>debt_at</td>
<td>Solvency</td>
<td>Total Liabilities as a fraction of Total Assets</td>
</tr>
<tr>
<td>Total Debt/Capital</td>
<td>debt_capital</td>
<td>Solvency</td>
<td>Total Debt as a fraction of Total Capital, where Total Debt is defined as the sum of Accounts Payable and Total Debt in Current and Long-</td>
</tr>
<tr>
<td>Financial Ratio</td>
<td>Variable Name</td>
<td>Category</td>
<td>Formula</td>
</tr>
<tr>
<td>-------------------------------------</td>
<td>---------------</td>
<td>------------</td>
<td>-------------------------------------------------------------------------</td>
</tr>
<tr>
<td>After-tax Interest Coverage</td>
<td>intcov</td>
<td>Solvency</td>
<td>Multiple of After-tax Income to Interest and Related Expenses</td>
</tr>
<tr>
<td>Interest Coverage Ratio</td>
<td>intcov_ratio</td>
<td>Solvency</td>
<td>Multiple of Earnings Before Interest and Taxes to Interest and Related Expenses</td>
</tr>
<tr>
<td>Dividend Payout Ratio</td>
<td>dpr</td>
<td>Valuation</td>
<td>Dividends as a fraction of Income Before Extra. Items</td>
</tr>
<tr>
<td>Forward P/E to 1-year Growth</td>
<td>PEG_1yrforward</td>
<td>Valuation</td>
<td>Price-to-Earnings, excl. Extraordinary Items (diluted) to 1-Year EPS Growth rate</td>
</tr>
<tr>
<td>Forward P/E to Long-term Growth</td>
<td>PEG_ltgforward</td>
<td>Valuation</td>
<td>Price-to-Earnings, excl. Extraordinary Items (diluted) to Long-term EPS Growth rate</td>
</tr>
<tr>
<td>Trailing P/E to Growth (PEG)</td>
<td>PEG_trailing</td>
<td>Valuation</td>
<td>Price-to-Earnings, excl. Extraordinary Items (diluted) to 3-Year past EPS Growth</td>
</tr>
<tr>
<td>Book/Market</td>
<td>bm</td>
<td>Valuation</td>
<td>Book Value of Equity as a fraction of Market Value of Equity</td>
</tr>
<tr>
<td>Shillers Cyclically Adjusted P/E</td>
<td>capei</td>
<td>Valuation</td>
<td>Multiple of Market Value of Equity to 5-year moving average of Net Income</td>
</tr>
<tr>
<td>Dividend Yield</td>
<td>divyield</td>
<td>Valuation</td>
<td>Indicated Dividend Rate as a fraction of Price</td>
</tr>
<tr>
<td>Enterprise Value Multiple</td>
<td>evm</td>
<td>Valuation</td>
<td>Multiple of Enterprise Value to EBITDA</td>
</tr>
<tr>
<td>Price/Cash flow</td>
<td>pcf</td>
<td>Valuation</td>
<td>Multiple of Market Value of Equity to Net Cash Flow from Operating Activities</td>
</tr>
<tr>
<td>P/E (Diluted, Excl. EI)</td>
<td>pe_exi</td>
<td>Valuation</td>
<td>Price-to-Earnings, excl. Extraordinary Items (diluted)</td>
</tr>
<tr>
<td>P/E (Diluted, Incl. EI)</td>
<td>pe_inc</td>
<td>Valuation</td>
<td>Price-to-Earnings, incl. Extraordinary Items (diluted)</td>
</tr>
<tr>
<td>Price/Operating Earnings (Basic, Excl. EI)</td>
<td>pe_op_basic</td>
<td>Valuation</td>
<td>Price to Operating EPS, excl. Extraordinary Items (Basic)</td>
</tr>
<tr>
<td>Price/Operating Earnings (Diluted, Excl. EI)</td>
<td>pe_op_dil</td>
<td>Valuation</td>
<td>Price to Operating EPS, excl. Extraordinary Items (Diluted)</td>
</tr>
<tr>
<td>Price/Sales</td>
<td>ps</td>
<td>Valuation</td>
<td>Multiple of Market Value of Equity to Sales</td>
</tr>
<tr>
<td>Price/Book</td>
<td>ptb</td>
<td>Valuation</td>
<td>Multiple of Market Value of Equity to Book Value of Equity</td>
</tr>
</tbody>
</table>
Appendix 3 - 50 Anomalies Manual

We access the description from the Internet appendix Kozak et. al (2020): https://www.serhiykozak.com/

Anomaly definitions from Kozak, Nagel, Santosh (2018) “Shrinking the Cross Section.”

Anomaly Characteristics

Anomaly definitions and descriptions are heavily based on the lists of characteristics compiled by Hou et al. (2015); Kogan and Tian (2015); McLean and Pontiff (2016); Novy-Marx and Velikov (2016). All accounting variables are properly lagged. For annual rebalancing, returns from July of year $t$ to June of year $t + 1$ are matched to variables in December of $t − 1$. Returns from January to June of year $t$ are matched to variables in December of year $t − 2$. Financial variables with a subscript “Dec” below are computed using the same timing convention. For monthly rebalancing, returns are matched to the latest quarterly report lagged one month. Additional lagging (if required) is reported for each variable below individually. All subindices below are measured in months. Time subscript $t$ refers to time at which a portfolio is formed.


2. **Value (annual)** ($value$). Follows Fama and French (1993). $value = BE/ME$. At the end of June of each year, we use book equity from the previous fiscal year and market equity from December of the previous year. Rebalanced annually.

3. **Gross Profitability** ($prof$). Follows Novy Marx (2013a). $prof = GP/AT$, where GP is gross profits and AT is total assets. Rebalanced annually.


5. **Piotroski’s F-score** ($F$-score). Follows Piotroski (2000). $F$-score = $1_{IB>0} + 1_{ROA>0} + 1_{CFO>0} + 1_{CFO>IB} + 1_{DTA<0|DLTT=0|DLTT_{−12}=0} + 1_{ATL>0} + 1_{EqIss≤0} + 1_{GM>0} + 1_{ATO>0}$, where IB is income before extraordinary items, ROA is income before extraordinary items scaled by lagged total assets, CFO is cash flow from operations, DTA is total long-term debt scaled by total assets, DLTT is total long-term debt, ATL is total current assets scaled by total current liabilities, EqIss is the difference between sales of of common stock and purchases of common stock recorded on the cash flow statement, GM equals one minus the ratio of cost of goods sold and total revenues, and ATO equals total revenues, scaled by total assets. Rebalanced annually.


7. **Share Repurchases** ($repurch$). Follows Ikenberry et al. (1995). $repurch = 1_{PRSTKC>0}$. Binary variable equal to one if repurchase of common or preferred shares indicated in statement of cash flow. Updated annually.

8. **Share Issuance (annual)** ($nissa$). Follows Pontiff and Woodgate (2008). $nissa = shrout_{Jun}/shrout_{Jun−12}$, where shrout is the number of shares outstanding. Change in real number of shares outstanding from past June to June of the previous year. Excludes changes in shares due to stock dividends and splits, and companies with no changes in shrout.
9. **Accruals** *(accruals)*. Follows Sloan (1996). accruals = \frac{\Delta ACT - \Delta CHE - \Delta LCT + \Delta DLC + \Delta TXP - \Delta DP}{(AT + AT_{-12})/2}, where \Delta ACT is the annual change in total current assets, \Delta CHE is the annual change in total cash and short-term investments, \Delta LCT is the annual change in current liabilities, \Delta DLC is the annual change in debt in current liabilities, \Delta TXP is the annual change in income taxes payable, \Delta DP is the annual change in depreciation and amortization, and \((AT + AT_{-12})/2\) is average total assets over the last two years. Rebalanced annually.


13. **Dividend Yield** *(divp)*. Follows Naranjo et al. (1998). divp = Div/ME_{Dec}. Dividend scaled by price. Both are measured in December of the year \(t - 1\) or \(t - 2\) (for returns in months prior to July). Rebalanced annually.


15. **Cash Flow / Market Value of Equity** *(cfp)*. Follows Lakonishok et al. (1994). cfp = (IB + DP)/ME_{Dec}. Net income plus depreciation and amortization, all scaled by market value of equity measured at the same date. Updated annually.

16. **Net Operating Assets** *(noa)*. Follows Hirshleifer et al. (2004). noa = (AT - CHE) - (AT - DLC - DLTT - MIB - PSTK - CEQ), where AT is total assets, CHE is cash and short-term investments, DLC is debt in current liabilities, DLTT is long term debt, MIB is non-controlling interest, PSTK is preferred capital stock, and CEQ is common equity. Updated annually.

17. **Investment** *(inv)*. Follows Chen et al. (2011); Lyandres et al. (2007). inv = (\Delta PPEGT + \Delta INVT)/AT_{-12}, where \Delta PPEGT is the annual change in gross total property, plant, and equipment, \Delta INVT is the annual change in total inventories, and AT_{-12} is lagged total assets. Rebalanced annually, uses the full period.

18. **Investment-to-Capital** *(invcap)*. Follows Xing (2008). invcap = CAPX/PPENT. Investment to capital is the ratio of capital expenditure (Compustat item CAPX) over property, plant, and equipment (Compustat item PPENT).


20. **Sales Growth** *(sgrowth)*. Follows Lakonishok et al. (1994). sgrowth = SALE/SALE_{-12}. Sales growth is the percent change in net sales over turnover (Compustat item SALE).

21. **Leverage** *(lev)*. Follows Bhandari (1988). lev = AT/ME_{Dec}. Market leverage is the ratio of total assets (Compustat item AT) over the market value of equity. Both are measured in December of the same year.


26. **Momentum (6m)** (mom). Follows Jagadeesh and Titman (1993). mom = \sum_{l=2}^{7} r_{t-l}. Cumulated past performance in the previous 6 months by skipping the most recent month. Rebalanced monthly.

27. **Industry Momentum** (indmom). Follows Moskowitz and Grinblatt (1999). indmom = rank(\sum_{l=1}^{6} r_{t-l}^{ind}). In each month, the Fama and French 49 industries are ranked on their value-weighted past 6-months performance. Rebalanced monthly.


31. **Momentum (1 year)** (mom12). Follows Jagadeesh and Titman (1993). mom12 = \sum_{l=2}^{12} r_{t-l}. Cumulated past performance in the previous year by skipping the most recent month. Rebalanced monthly.

32. **Momentum-Reversal** (momrev). Follows Jagadeesh and Titman (1993). momrev = \sum_{l=14}^{19} r_{t-l}. Buy and hold returns from t − 19 to t − 14. Updated monthly.

33. **Long-term Reversals** (lrrev). Follows DeBondt and Thaler (1985). lrrev = \sum_{l=13}^{60} r_{t-l}. Cumulative returns from t − 60 to t − 13. Updated monthly.

34. **Value (monthly)** (valuem). Follows Asness and Frazzini (2013). valuem = BEQ_{−3}/ME_{−1}. Book-to-market ratio using the most up-to-date prices and book equity (appropriately lagged). Rebalanced monthly.

35. **Share Issuance (monthly)** (nissm). Follows Pontiff and Woodgate (2008). nissm = shrout_{t-13} / shrout_{t-1}, where shrout is the number of shares outstanding. Change in real number of shares outstanding from t − 13 to t − 1. Excludes changes in shares due to stock dividends and splits, and companies with no changes in shrout.
36. **PEAD (SUE)** (*sue*). Follows Foster et al. (1984). $sue = \frac{IBQ - IBQ_{-12}}{\sigma_{IBQ_{-24} - IBQ_{-3}}}$, where IBQ is income before extraordinary items (updated quarterly), and $\sigma_{IBQ_{-24} - IBQ_{-3}}$ is the standard deviation of IBQ in the past two years skipping the most recent quarter. Earnings surprises are measured by Standardized Unexpected Earnings (SUE), which is the change in the most recently announced quarterly earnings per share from its value announced four quarters ago divided by the standard deviation of this change in quarterly earnings over the prior eight quarters. Rebalanced monthly.

37. **Return on Book Equity** (*roe*). Follows Chen et al. (2011). $roe = \frac{IBQ}{BEQ_{-3}}$, where IBQ is income before extraordinary items (updated quarterly), and BEQ is book value of equity. Rebalanced monthly.

38. **Return on Market Equity** (*rome*). Follows Chen et al. (2011). $rome = \frac{IBQ}{ME_{-4}}$, where IBQ is income before extraordinary items (updated quarterly), and ME is market value of equity. Rebalanced monthly.


41. **Idiosyncratic Volatility** (*ivol*). Follows Ang et al. (2006). $ivol = \text{std}(R_{t} - \beta_{t}R_{M,t} - s_{t}\text{SMB}_{t} - h_{t}\text{HML}_{t})$. The standard deviation of the residual from firm-level regression of daily stock returns on the daily innovations of the Fama and French three-factor model using the estimation window of three months. Lagged one month.

42. **Beta Arbitrage** (*beta*). Follows Cooper et al. (2008). $beta = \beta_{t-60} - r_{t-1}$. Beta with respect to the CRSP equal-weighted return index. Estimated over the past 60 months (minimum 36 months) using daily data and lagged one month. Updated monthly.

43. **Seasonality** (*season*). Follows Heston and Sadka (2008). $season = \sum_{l=1}^{5} r_{t-l \times 12}$. Average monthly return in the same calendar month over the last 5 years. As an example, the average return from prior Octobers is used to predict returns this October. The firm needs at least one year of data to be included in the sample. Updated monthly.

44. **Industry Relative Reversals** (*indrrev*). Follows Da et al. (2013). $indrrev = r_{-1} - r_{-1}^{\text{ind}}$, where $r$ is the return on a stock and $r^{\text{ind}}$ is return on its industry. Difference between a stocks’ prior month’s return and the prior month’s return of its industry (based on the Fama and French 49 industries). Updated monthly.

45. **Industry Relative Reversals (Low Volatility)** (*indrrevlv*). Follows Da et al. (2013). $indrrevlv = r_{-1} - r_{-1}^{\text{ind}}$ if vol $\leq$ NYSE median, where $r$ is the return on a stock and $r^{\text{ind}}$ is return on its industry. Difference between a stocks’ prior month’s return and the prior month’s return of its industry (based on the Fama and French 49 industries). Only stocks with idiosyncratic volatility lower than the NYSE median for month are included in the sorts. Updated monthly.

46. **Industry Momentum-Reversal** (*indmomrev*). Follows Moskowitz and Grinblatt (1999). $indmomrev = \text{rank}(\text{industry momentum}) + \text{rank}(\text{industry relative-reversals low-vol})$. Sum of Fama and French 49 industries ranks on industry momentum and industry relative reversals (low vol). Rebalanced monthly.
47. **Composite Issuance** \((ciss)\). Follows Daniel and Titman (2006). \(ciss = \log\left(\frac{\text{ME}_{t-13}}{\text{ME}_{t-60}}\right) - \sum_{t=13}^{60} r_{t-1}\), where \(r\) is the log return on the stock and ME is total market equity. Updated monthly.

48. **Price** \((price)\). Follows Blume and Husic (1973). \(price = \log(\text{ME/shrout})\), where ME is market equity and shrout is the number of shares outstanding. Log of stock price. Updated monthly.

49. **Firm Age** \((age)\). Follows Barry and Brown (1984). \(age = \log(1 + \text{number of months since listing})\). The number of months that a firm has been listed in the CRSP database.

50. **Share Volume** \((shvol)\). Follows Datar et al. (1998). \(shvol = \frac{1}{3} \sum_{i=1}^{3} \frac{\text{volume}_{t-i}}{\text{shrout}_t}\). Average number of shares traded over the previous three months scaled by shares outstanding. Updated monthly.
References


Appendix 5 - Correlation heatmap WFR
### Appendix 6 - Summary statistics for 50 Anomaly portfolios

<table>
<thead>
<tr>
<th>Public_date</th>
<th>rme</th>
<th>re_ew</th>
<th>r_size</th>
<th>r_value</th>
<th>r_prof</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min.</td>
<td>:0.12054</td>
<td>:-0.21262</td>
<td>:0.09962</td>
<td>:0.14359</td>
<td>:-1.0e-01</td>
</tr>
<tr>
<td>1st Qu.</td>
<td>:0.0840</td>
<td>:0.00616</td>
<td>:0.00546</td>
<td>:0.00391</td>
<td>:4.8e-03</td>
</tr>
<tr>
<td>Median</td>
<td>:0.00424</td>
<td>:0.0041</td>
<td>:0.0046</td>
<td>:0.0037</td>
<td>:1.2e-04</td>
</tr>
<tr>
<td>Mean</td>
<td>:0.0034</td>
<td>:0.0053</td>
<td>:0.00015</td>
<td>:0.0004</td>
<td>:9.9e-05</td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>:0.00059</td>
<td>:0.00541</td>
<td>:0.00466</td>
<td>:0.00452</td>
<td>:5.0e-03</td>
</tr>
<tr>
<td>Max.</td>
<td>:0.0925</td>
<td>:0.09247</td>
<td>:0.17417</td>
<td>:0.09312</td>
<td>:1.0e-01</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>r_valph PRESS</th>
<th>r_fscores</th>
<th>r_debtlisa</th>
<th>r_repurch</th>
<th>r_niss</th>
<th>r_accruals</th>
<th>r_growth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min.</td>
<td>:0.11338</td>
<td>:-1.5e-01</td>
<td>:0.12619</td>
<td>:-1.0590</td>
<td>:-0.08484</td>
<td>:-0.07142</td>
</tr>
<tr>
<td>1st Qu.</td>
<td>:0.00414</td>
<td>:4.3e-03</td>
<td>:0.00890</td>
<td>:0.0079</td>
<td>:0.00690</td>
<td>:0.00526</td>
</tr>
<tr>
<td>Median</td>
<td>:0.00026</td>
<td>1.7e-04</td>
<td>:0.00028</td>
<td>:0.00047</td>
<td>:0.00022</td>
<td>:0.00041</td>
</tr>
<tr>
<td>Mean</td>
<td>:0.00047</td>
<td>7.2e-05</td>
<td>:0.00044</td>
<td>:0.00064</td>
<td>:0.00023</td>
<td>:0.00058</td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>:0.00485</td>
<td>3.6e-03</td>
<td>:0.00445</td>
<td>:0.00395</td>
<td>:0.00547</td>
<td>:0.00430</td>
</tr>
<tr>
<td>Max.</td>
<td>:0.15849</td>
<td>:1.2e-01</td>
<td>:0.13126</td>
<td>:0.14172</td>
<td>:0.08292</td>
<td>:0.14573</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>r_aturnover</th>
<th>r_gmargins</th>
<th>r_dvpi</th>
<th>r_r</th>
<th>r_cfp</th>
<th>r_roa</th>
<th>r_inv</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min.</td>
<td>:0.10636</td>
<td>:0.06712</td>
<td>:0.10833</td>
<td>:0.11737</td>
<td>:0.14236</td>
<td>:0.10941</td>
</tr>
<tr>
<td>1st Qu.</td>
<td>:0.00481</td>
<td>:0.00519</td>
<td>:0.00461</td>
<td>:0.00839</td>
<td>:0.00412</td>
<td>:0.00426</td>
</tr>
<tr>
<td>Median</td>
<td>:0.00028</td>
<td>:0.00014</td>
<td>:0.00024</td>
<td>:0.00042</td>
<td>:0.00041</td>
<td>:0.00011</td>
</tr>
<tr>
<td>Mean</td>
<td>:0.00022</td>
<td>:0.00013</td>
<td>:0.00042</td>
<td>:0.00058</td>
<td>:0.00050</td>
<td>:0.00016</td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>:0.00537</td>
<td>:0.00488</td>
<td>:0.00539</td>
<td>:0.00466</td>
<td>:0.00514</td>
<td>:0.00423</td>
</tr>
<tr>
<td>Max.</td>
<td>:0.10885</td>
<td>:0.11086</td>
<td>:0.14534</td>
<td>:0.11405</td>
<td>:0.08608</td>
<td>:0.13202</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>r_invcap</th>
<th>r_igrowth</th>
<th>r_sgrowth</th>
<th>r_leh</th>
<th>r_roae</th>
<th>r_roe</th>
<th>r_sp</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min.</td>
<td>:0.11484</td>
<td>:0.07063</td>
<td>:0.06507</td>
<td>:0.11949</td>
<td>:9.1e-02</td>
<td>:1.1e-01</td>
</tr>
<tr>
<td>1st Qu.</td>
<td>:0.00436</td>
<td>:0.00056</td>
<td>:0.00353</td>
<td>:0.00379</td>
<td>:5.4e-03</td>
<td>:4.9e-03</td>
</tr>
<tr>
<td>Median</td>
<td>:0.00045</td>
<td>:0.00035</td>
<td>:0.00022</td>
<td>:0.00021</td>
<td>:3.4e-05</td>
<td>:8.8e-05</td>
</tr>
<tr>
<td>Mean</td>
<td>:0.00045</td>
<td>:0.00054</td>
<td>:0.00044</td>
<td>:0.00031</td>
<td>:1.7e-04</td>
<td>:2.7e-04</td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>:0.00350</td>
<td>:0.00467</td>
<td>:0.00469</td>
<td>:0.00423</td>
<td>:5.4e-03</td>
<td>:5.2e-03</td>
</tr>
<tr>
<td>Max.</td>
<td>:0.13312</td>
<td>:0.10467</td>
<td>:0.11517</td>
<td>:0.09347</td>
<td>:8.1e-02</td>
<td>:1.4e-01</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>r_glnoa</th>
<th>r_mom</th>
<th>r_indmom</th>
<th>r_valmom</th>
<th>r_valmomprof</th>
<th>r_shortint</th>
<th>r_mom2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min.</td>
<td>:0.12863</td>
<td>:0.13211</td>
<td>:0.14551</td>
<td>:0.17965</td>
<td>:0.16020</td>
<td>:0.14606</td>
</tr>
<tr>
<td>1st Qu.</td>
<td>:0.00598</td>
<td>:0.00349</td>
<td>:0.00320</td>
<td>:0.00347</td>
<td>:0.00363</td>
<td>:0.00584</td>
</tr>
<tr>
<td>Median</td>
<td>:0.00013</td>
<td>:0.00046</td>
<td>:0.00053</td>
<td>:0.00064</td>
<td>:0.00057</td>
<td>:0.00118</td>
</tr>
<tr>
<td>Mean</td>
<td>:0.00011</td>
<td>:0.00010</td>
<td>:0.00029</td>
<td>:0.00032</td>
<td>:0.00025</td>
<td>:0.00055</td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>:0.00566</td>
<td>:0.00421</td>
<td>:0.00421</td>
<td>:0.00466</td>
<td>:0.00539</td>
<td>:0.00447</td>
</tr>
<tr>
<td>Max.</td>
<td>:0.08180</td>
<td>:0.10111</td>
<td>:0.10130</td>
<td>:0.09718</td>
<td>:0.08688</td>
<td>:0.09622</td>
</tr>
<tr>
<td></td>
<td>r glmtno</td>
<td>r mom</td>
<td>r_indnom</td>
<td>r_valmorn</td>
<td>r_valnomprof</td>
<td>r_shortint</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>----------</td>
<td>---------</td>
<td>----------</td>
<td>-----------</td>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>Min.</td>
<td>0.12863</td>
<td>-0.13211</td>
<td>0.14651</td>
<td>-0.17965</td>
<td>-0.16020</td>
<td>-0.14606</td>
</tr>
<tr>
<td>1st Qu.</td>
<td>-0.00598</td>
<td>-0.00349</td>
<td>-0.00320</td>
<td>-0.00347</td>
<td>-0.00363</td>
<td>-0.00584</td>
</tr>
<tr>
<td>Median</td>
<td>-0.00013</td>
<td>0.00046</td>
<td>0.00053</td>
<td>0.00064</td>
<td>0.00057</td>
<td>0.00018</td>
</tr>
<tr>
<td>Mean</td>
<td>0.00011</td>
<td>0.00010</td>
<td>0.00029</td>
<td>0.00032</td>
<td>0.00032</td>
<td>0.00025</td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>0.00566</td>
<td>0.00421</td>
<td>0.00421</td>
<td>0.00466</td>
<td>0.00465</td>
<td>0.00539</td>
</tr>
<tr>
<td>Max.</td>
<td>0.08180</td>
<td>0.10111</td>
<td>0.10130</td>
<td>0.09718</td>
<td>0.08668</td>
<td>0.09622</td>
</tr>
<tr>
<td></td>
<td>r momrev</td>
<td>r lrrev</td>
<td>r valuem</td>
<td>r nissm</td>
<td>r sue</td>
<td>r roe</td>
</tr>
<tr>
<td>Min.</td>
<td>-0.10586</td>
<td>-0.07573</td>
<td>-0.09806</td>
<td>-0.13406</td>
<td>-0.09841</td>
<td>-0.07911</td>
</tr>
<tr>
<td>1st Qu.</td>
<td>-0.00421</td>
<td>-0.00497</td>
<td>-0.00368</td>
<td>-0.00457</td>
<td>-0.00422</td>
<td>-0.00414</td>
</tr>
<tr>
<td>Median</td>
<td>-0.00011</td>
<td>-0.00015</td>
<td>-0.00011</td>
<td>-0.00038</td>
<td>-0.00057</td>
<td>-0.00057</td>
</tr>
<tr>
<td>Mean</td>
<td>-0.00029</td>
<td>-0.00034</td>
<td>-0.00033</td>
<td>-0.00060</td>
<td>-0.00051</td>
<td>-0.00051</td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>0.00390</td>
<td>0.00446</td>
<td>0.00400</td>
<td>0.00869</td>
<td>0.00542</td>
<td>0.00521</td>
</tr>
<tr>
<td>Max.</td>
<td>0.13866</td>
<td>0.14810</td>
<td>0.13268</td>
<td>0.14010</td>
<td>0.10767</td>
<td>0.08830</td>
</tr>
<tr>
<td></td>
<td>r roa</td>
<td>r strev</td>
<td>r ivol</td>
<td>r betaaerb</td>
<td>r season</td>
<td>r indrev</td>
</tr>
<tr>
<td>Min.</td>
<td>-0.07675</td>
<td>-1.4e-01</td>
<td>-0.13733</td>
<td>-0.11981</td>
<td>-0.06974</td>
<td>-0.12875</td>
</tr>
<tr>
<td>1st Qu.</td>
<td>-0.00454</td>
<td>-4.0e-03</td>
<td>-0.00474</td>
<td>-0.00514</td>
<td>-0.00457</td>
<td>-0.00483</td>
</tr>
<tr>
<td>Median</td>
<td>0.00038</td>
<td>2.5e-06</td>
<td>0.00031</td>
<td>0.00072</td>
<td>0.00046</td>
<td>0.00039</td>
</tr>
<tr>
<td>Mean</td>
<td>0.00035</td>
<td>2.3e-04</td>
<td>0.00055</td>
<td>0.00081</td>
<td>0.00037</td>
<td>0.00064</td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>0.09553</td>
<td>3.7e-03</td>
<td>0.00397</td>
<td>0.00838</td>
<td>0.00533</td>
<td>0.00385</td>
</tr>
<tr>
<td>Max.</td>
<td>0.08401</td>
<td>9.5e-02</td>
<td>0.11275</td>
<td>0.13484</td>
<td>0.12153</td>
<td>0.09251</td>
</tr>
<tr>
<td></td>
<td>r indnomrev</td>
<td>r ciss</td>
<td>r price</td>
<td>r age</td>
<td>r shvol</td>
<td></td>
</tr>
<tr>
<td>Min.</td>
<td>-0.10705</td>
<td>-0.10930</td>
<td>-1.5e-01</td>
<td>-0.12786</td>
<td>-0.12764</td>
<td>-0.10764</td>
</tr>
<tr>
<td>1st Qu.</td>
<td>-0.00330</td>
<td>0.00576</td>
<td>3.8e-03</td>
<td>-0.00399</td>
<td>-0.00483</td>
<td>0.10301</td>
</tr>
<tr>
<td>Median</td>
<td>0.00099</td>
<td>0.00061</td>
<td>1.5e-04</td>
<td>0.00005</td>
<td>0.00036</td>
<td>0.00045</td>
</tr>
<tr>
<td>Mean</td>
<td>0.00079</td>
<td>0.00070</td>
<td>4.9e-05</td>
<td>0.00030</td>
<td>0.00045</td>
<td>0.00045</td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>0.00596</td>
<td>0.00452</td>
<td>4.2e-03</td>
<td>0.00415</td>
<td>0.00411</td>
<td>0.00411</td>
</tr>
<tr>
<td>Max.</td>
<td>0.13348</td>
<td>0.14058</td>
<td>9.8e-02</td>
<td>0.14741</td>
<td>0.10679</td>
<td>0.10679</td>
</tr>
</tbody>
</table>
Appendix 6.2 - Summary statistics for WFR

<table>
<thead>
<tr>
<th>Metric</th>
<th>Min.</th>
<th>Q1</th>
<th>Median</th>
<th>Q3</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public Date</td>
<td>2000-01-03</td>
<td>-0.18443</td>
<td>-2.0e-03</td>
<td>0.00651</td>
<td>5.8e-03</td>
</tr>
<tr>
<td>1st Qu.</td>
<td>2005-04-07</td>
<td>0.008465</td>
<td>0.00953</td>
<td>7.5e-05</td>
<td>6.0e-03</td>
</tr>
<tr>
<td>Median</td>
<td>2010-07-07</td>
<td>0.008465</td>
<td>0.00953</td>
<td>7.5e-05</td>
<td>6.0e-03</td>
</tr>
<tr>
<td>Mean</td>
<td>2010-07-07</td>
<td>0.008465</td>
<td>0.00953</td>
<td>7.5e-05</td>
<td>6.0e-03</td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>2015-10-03</td>
<td>0.008465</td>
<td>0.00953</td>
<td>7.5e-05</td>
<td>6.0e-03</td>
</tr>
<tr>
<td>Max.</td>
<td>2020-12-31</td>
<td>0.008465</td>
<td>0.00953</td>
<td>7.5e-05</td>
<td>6.0e-03</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Metric</th>
<th>Min.</th>
<th>Q1</th>
<th>Median</th>
<th>Q3</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>After_tax_Return</td>
<td>0.18495</td>
<td>-0.22200</td>
<td>0.00572</td>
<td>0.00825</td>
<td>0.00515</td>
</tr>
<tr>
<td>1st Qu.</td>
<td>0.00572</td>
<td>0.00572</td>
<td>0.00572</td>
<td>0.00572</td>
<td>0.00572</td>
</tr>
<tr>
<td>Median</td>
<td>0.00572</td>
<td>0.00572</td>
<td>0.00572</td>
<td>0.00572</td>
<td>0.00572</td>
</tr>
<tr>
<td>Mean</td>
<td>0.00572</td>
<td>0.00572</td>
<td>0.00572</td>
<td>0.00572</td>
<td>0.00572</td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>0.00572</td>
<td>0.00572</td>
<td>0.00572</td>
<td>0.00572</td>
<td>0.00572</td>
</tr>
<tr>
<td>Max.</td>
<td>0.00572</td>
<td>0.00572</td>
<td>0.00572</td>
<td>0.00572</td>
<td>0.00572</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Metric</th>
<th>Min.</th>
<th>Q1</th>
<th>Median</th>
<th>Q3</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before_tax_Return</td>
<td>0.18481</td>
<td>-0.13758</td>
<td>-1.1e-01</td>
<td>0.00726</td>
<td>0.11027</td>
</tr>
<tr>
<td>1st Qu.</td>
<td>0.00572</td>
<td>0.00572</td>
<td>0.00572</td>
<td>0.00572</td>
<td>0.00572</td>
</tr>
<tr>
<td>Median</td>
<td>0.00572</td>
<td>0.00572</td>
<td>0.00572</td>
<td>0.00572</td>
<td>0.00572</td>
</tr>
<tr>
<td>Mean</td>
<td>0.00572</td>
<td>0.00572</td>
<td>0.00572</td>
<td>0.00572</td>
<td>0.00572</td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>0.00572</td>
<td>0.00572</td>
<td>0.00572</td>
<td>0.00572</td>
<td>0.00572</td>
</tr>
<tr>
<td>Max.</td>
<td>0.00572</td>
<td>0.00572</td>
<td>0.00572</td>
<td>0.00572</td>
<td>0.00572</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Metric</th>
<th>Min.</th>
<th>Q1</th>
<th>Median</th>
<th>Q3</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cash_Conversion Cycle</td>
<td>0.00007</td>
<td>-0.12036</td>
<td>-0.13185</td>
<td>-0.18481</td>
<td>-0.18481</td>
</tr>
<tr>
<td>1st Qu.</td>
<td>0.00057</td>
<td>0.00057</td>
<td>0.00057</td>
<td>0.00057</td>
<td>0.00057</td>
</tr>
<tr>
<td>Median</td>
<td>0.00057</td>
<td>0.00057</td>
<td>0.00057</td>
<td>0.00057</td>
<td>0.00057</td>
</tr>
<tr>
<td>Mean</td>
<td>0.00057</td>
<td>0.00057</td>
<td>0.00057</td>
<td>0.00057</td>
<td>0.00057</td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>0.00057</td>
<td>0.00057</td>
<td>0.00057</td>
<td>0.00057</td>
<td>0.00057</td>
</tr>
<tr>
<td>Max.</td>
<td>0.00057</td>
<td>0.00057</td>
<td>0.00057</td>
<td>0.00057</td>
<td>0.00057</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Metric</th>
<th>Min.</th>
<th>Q1</th>
<th>Median</th>
<th>Q3</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dividend_Payout</td>
<td>-0.06254</td>
<td>-0.14317</td>
<td>-0.13030</td>
<td>-0.06254</td>
<td>-0.06254</td>
</tr>
<tr>
<td>1st Qu.</td>
<td>-0.06254</td>
<td>-0.06254</td>
<td>-0.06254</td>
<td>-0.06254</td>
<td>-0.06254</td>
</tr>
<tr>
<td>Median</td>
<td>-0.06254</td>
<td>-0.06254</td>
<td>-0.06254</td>
<td>-0.06254</td>
<td>-0.06254</td>
</tr>
<tr>
<td>Mean</td>
<td>-0.06254</td>
<td>-0.06254</td>
<td>-0.06254</td>
<td>-0.06254</td>
<td>-0.06254</td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>-0.06254</td>
<td>-0.06254</td>
<td>-0.06254</td>
<td>-0.06254</td>
<td>-0.06254</td>
</tr>
<tr>
<td>Max.</td>
<td>-0.06254</td>
<td>-0.06254</td>
<td>-0.06254</td>
<td>-0.06254</td>
<td>-0.06254</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Metric</th>
<th>Min.</th>
<th>Q1</th>
<th>Median</th>
<th>Q3</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dividend_Yield</td>
<td>-0.06254</td>
<td>-0.14317</td>
<td>-0.13030</td>
<td>-0.06254</td>
<td>-0.06254</td>
</tr>
<tr>
<td>1st Qu.</td>
<td>-0.06254</td>
<td>-0.06254</td>
<td>-0.06254</td>
<td>-0.06254</td>
<td>-0.06254</td>
</tr>
<tr>
<td>Median</td>
<td>-0.06254</td>
<td>-0.06254</td>
<td>-0.06254</td>
<td>-0.06254</td>
<td>-0.06254</td>
</tr>
<tr>
<td>Mean</td>
<td>-0.06254</td>
<td>-0.06254</td>
<td>-0.06254</td>
<td>-0.06254</td>
<td>-0.06254</td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>-0.06254</td>
<td>-0.06254</td>
<td>-0.06254</td>
<td>-0.06254</td>
<td>-0.06254</td>
</tr>
<tr>
<td>Max.</td>
<td>-0.06254</td>
<td>-0.06254</td>
<td>-0.06254</td>
<td>-0.06254</td>
<td>-0.06254</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Metric</th>
<th>Min.</th>
<th>Q1</th>
<th>Median</th>
<th>Q3</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effective_Tax_Rate</td>
<td>-0.06254</td>
<td>-0.14317</td>
<td>-0.13030</td>
<td>-0.06254</td>
<td>-0.06254</td>
</tr>
<tr>
<td>1st Qu.</td>
<td>-0.06254</td>
<td>-0.06254</td>
<td>-0.06254</td>
<td>-0.06254</td>
<td>-0.06254</td>
</tr>
<tr>
<td>Median</td>
<td>-0.06254</td>
<td>-0.06254</td>
<td>-0.06254</td>
<td>-0.06254</td>
<td>-0.06254</td>
</tr>
<tr>
<td>Mean</td>
<td>-0.06254</td>
<td>-0.06254</td>
<td>-0.06254</td>
<td>-0.06254</td>
<td>-0.06254</td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>-0.06254</td>
<td>-0.06254</td>
<td>-0.06254</td>
<td>-0.06254</td>
<td>-0.06254</td>
</tr>
<tr>
<td>Max.</td>
<td>-0.06254</td>
<td>-0.06254</td>
<td>-0.06254</td>
<td>-0.06254</td>
<td>-0.06254</td>
</tr>
<tr>
<td>Min.</td>
<td>-0.10098</td>
<td>Min.</td>
<td>-1.1e-01</td>
<td>Min.</td>
<td>-0.17499</td>
</tr>
<tr>
<td>-------</td>
<td>---------</td>
<td>-------</td>
<td>-----------</td>
<td>-------</td>
<td>---------</td>
</tr>
<tr>
<td>1st Qu.:</td>
<td>-0.00635</td>
<td>1st Qu.:</td>
<td>-6.5e-03</td>
<td>1st Qu.:</td>
<td>-0.00520</td>
</tr>
<tr>
<td>Median:</td>
<td>-0.00012</td>
<td>Median:</td>
<td>9.6e-05</td>
<td>Median:</td>
<td>0.00068</td>
</tr>
<tr>
<td>Mean:</td>
<td>-0.00016</td>
<td>Mean:</td>
<td>2.4e-04</td>
<td>Mean:</td>
<td>0.00056</td>
</tr>
<tr>
<td>3rd Qu.:</td>
<td>0.00585</td>
<td>3rd Qu.:</td>
<td>7.1e-03</td>
<td>3rd Qu.:</td>
<td>0.00641</td>
</tr>
<tr>
<td>Max.</td>
<td>0.11635</td>
<td>Max.</td>
<td>1.1e-01</td>
<td>Max.</td>
<td>0.08727</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Min.</th>
<th>-9.7e-02</th>
<th>Min.</th>
<th>-0.14367</th>
<th>Min.</th>
<th>-0.16230</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st Qu.:</td>
<td>6.2e-03</td>
<td>1st Qu.:</td>
<td>0.00864</td>
<td>1st Qu.:</td>
<td>0.00623</td>
</tr>
<tr>
<td>Median:</td>
<td>3.4e-04</td>
<td>Median:</td>
<td>0.00044</td>
<td>Median:</td>
<td>0.00046</td>
</tr>
<tr>
<td>Mean:</td>
<td>2.9e-05</td>
<td>Mean:</td>
<td>0.00034</td>
<td>Mean:</td>
<td>0.00034</td>
</tr>
<tr>
<td>3rd Qu.:</td>
<td>6.2e-03</td>
<td>3rd Qu.:</td>
<td>0.00722</td>
<td>3rd Qu.:</td>
<td>0.00700</td>
</tr>
<tr>
<td>Max.</td>
<td>8.1e-02</td>
<td>Max.</td>
<td>0.07987</td>
<td>Max.</td>
<td>0.10889</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Min.</th>
<th>-1.8e-01</th>
<th>Min.</th>
<th>-0.08147</th>
<th>Min.</th>
<th>-8.3e-02</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st Qu.:</td>
<td>6.1e-03</td>
<td>1st Qu.:</td>
<td>0.00652</td>
<td>1st Qu.:</td>
<td>-7.2e-03</td>
</tr>
<tr>
<td>Median:</td>
<td>4.4e-05</td>
<td>Median:</td>
<td>0.00017</td>
<td>Median:</td>
<td>1.7e-04</td>
</tr>
<tr>
<td>Mean:</td>
<td>4.7e-05</td>
<td>Mean:</td>
<td>0.00061</td>
<td>Mean:</td>
<td>4.7e-05</td>
</tr>
<tr>
<td>3rd Qu.:</td>
<td>6.3e-03</td>
<td>3rd Qu.:</td>
<td>0.00596</td>
<td>3rd Qu.:</td>
<td>7.5e-03</td>
</tr>
<tr>
<td>Max.</td>
<td>1.0e-01</td>
<td>Max.</td>
<td>0.12306</td>
<td>Max.</td>
<td>8.1e-02</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Min.</th>
<th>-0.11173</th>
<th>Min.</th>
<th>-8.0e-02</th>
<th>Min.</th>
<th>-7.3e-02</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st Qu.:</td>
<td>6.0e-03</td>
<td>1st Qu.:</td>
<td>-5.8e-03</td>
<td>1st Qu.:</td>
<td>-6.0e-03</td>
</tr>
<tr>
<td>Median:</td>
<td>6.0e-04</td>
<td>Median:</td>
<td>-6.7e-05</td>
<td>Median:</td>
<td>1.4e-04</td>
</tr>
<tr>
<td>Mean:</td>
<td>-1.2e-04</td>
<td>Mean:</td>
<td>0.00023</td>
<td>Mean:</td>
<td>1.5e-05</td>
</tr>
<tr>
<td>3rd Qu.:</td>
<td>8.2e-02</td>
<td>3rd Qu.:</td>
<td>0.00589</td>
<td>3rd Qu.:</td>
<td>7.5e-03</td>
</tr>
<tr>
<td>Max.</td>
<td>8.1e-02</td>
<td>Max.</td>
<td>1.2334</td>
<td>Max.</td>
<td>8.1e-02</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Min.</th>
<th>-0.0865</th>
<th>Min.</th>
<th>-0.15479</th>
<th>Min.</th>
<th>-1.4e-01</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st Qu.:</td>
<td>-0.0016</td>
<td>1st Qu.:</td>
<td>0.00588</td>
<td>1st Qu.:</td>
<td>-6.2e-03</td>
</tr>
<tr>
<td>Median:</td>
<td>0.00069</td>
<td>Median:</td>
<td>-0.00020</td>
<td>Median:</td>
<td>1.3e-04</td>
</tr>
<tr>
<td>Mean:</td>
<td>0.00018</td>
<td>Mean:</td>
<td>0.00017</td>
<td>Mean:</td>
<td>9.8e-05</td>
</tr>
<tr>
<td>3rd Qu.:</td>
<td>0.00658</td>
<td>3rd Qu.:</td>
<td>0.00575</td>
<td>3rd Qu.:</td>
<td>6.6e-03</td>
</tr>
<tr>
<td>Max.</td>
<td>1.0e-01</td>
<td>Max.</td>
<td>8.2e-02</td>
<td>Max.</td>
<td>8.2e-02</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Min.</th>
<th>-0.12236</th>
<th>Min.</th>
<th>-0.09724</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st Qu.:</td>
<td>-0.0059</td>
<td>1st Qu.:</td>
<td>-0.00680</td>
</tr>
<tr>
<td>Median:</td>
<td>0.00027</td>
<td>Median:</td>
<td>-0.00029</td>
</tr>
<tr>
<td>Mean:</td>
<td>0.00025</td>
<td>Mean:</td>
<td>0.00036</td>
</tr>
<tr>
<td>3rd Qu.:</td>
<td>0.00596</td>
<td>3rd Qu.:</td>
<td>0.00602</td>
</tr>
<tr>
<td>Max.</td>
<td>0.2361</td>
<td>Max.</td>
<td>0.09972</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Min.</th>
<th>-0.13748</th>
<th>Min.</th>
<th>-0.13899</th>
<th>Min.</th>
<th>-0.11979</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st Qu.:</td>
<td>-0.00444</td>
<td>1st Qu.:</td>
<td>-0.00455</td>
<td>1st Qu.:</td>
<td>-0.00576</td>
</tr>
<tr>
<td>Median:</td>
<td>0.00089</td>
<td>Median:</td>
<td>0.00089</td>
<td>Median:</td>
<td>0.00061</td>
</tr>
<tr>
<td>Mean:</td>
<td>0.00112</td>
<td>Mean:</td>
<td>0.00111</td>
<td>Mean:</td>
<td>0.00068</td>
</tr>
<tr>
<td>3rd Qu.:</td>
<td>0.00668</td>
<td>3rd Qu.:</td>
<td>0.00668</td>
<td>3rd Qu.:</td>
<td>0.00710</td>
</tr>
<tr>
<td>Max.</td>
<td>0.24452</td>
<td>Max.</td>
<td>0.14167</td>
<td>Max.</td>
<td>0.11590</td>
</tr>
<tr>
<td>Metric</td>
<td>Min.</td>
<td>1st Qu.</td>
<td>Median</td>
<td>Mean</td>
<td>3rd Qu.</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>--------</td>
<td>---------</td>
<td>--------</td>
<td>--------</td>
<td>---------</td>
</tr>
<tr>
<td>Pre-tax Profit Margin Ranked Factor</td>
<td>-0.15105</td>
<td>-0.08612</td>
<td>-0.00018</td>
<td>0.00019</td>
<td>0.00589</td>
</tr>
<tr>
<td>Pre-tax Return on Net Operating Assets Ranked Factor</td>
<td>-0.20907</td>
<td>-0.0411</td>
<td>-0.0010</td>
<td>0.0014</td>
<td>0.00658</td>
</tr>
<tr>
<td>Price Book Ranked Factor</td>
<td>-0.1245</td>
<td>-0.0041</td>
<td>0.0013</td>
<td>0.0014</td>
<td>0.0065</td>
</tr>
<tr>
<td>Price Cash Flow Ranked Factor</td>
<td>-0.1879</td>
<td>-0.0042</td>
<td>0.0013</td>
<td>0.0014</td>
<td>0.0068</td>
</tr>
<tr>
<td>Price Operating Earnings Basic Excl. Earnings Diluted Excl. Earnings Ranked Factor</td>
<td>-0.1104</td>
<td>-0.0044</td>
<td>0.0012</td>
<td>0.0014</td>
<td>0.0069</td>
</tr>
<tr>
<td>Price Sales Ranked Factor</td>
<td>-0.09774</td>
<td>-1.7e-01</td>
<td>8.3e-06</td>
<td>3.8e-03</td>
<td>5.8e-03</td>
</tr>
<tr>
<td>Profit Before Depreciation Current Liabilities Ranked Factor</td>
<td>-1.0e-01</td>
<td>-5.9e-03</td>
<td>3.3e-05</td>
<td>1.0e-04</td>
<td>9.0e-05</td>
</tr>
<tr>
<td>Receivables Current Assets Ranked Factor</td>
<td>-0.1726</td>
<td>-0.00542</td>
<td>0.00061</td>
<td>0.00043</td>
<td>0.00087</td>
</tr>
<tr>
<td>Receivables Turnover Ranked Factor</td>
<td>-0.0634</td>
<td>-0.005</td>
<td>0.00061</td>
<td>0.00043</td>
<td>0.00087</td>
</tr>
<tr>
<td>Research and Development Sales Ranked Factor</td>
<td>-0.0623</td>
<td>-0.00542</td>
<td>0.00061</td>
<td>0.00043</td>
<td>0.00087</td>
</tr>
<tr>
<td>Return on Assets Ranked Factor</td>
<td>-0.2052</td>
<td>-0.00525</td>
<td>0.00033</td>
<td>0.00030</td>
<td>0.00030</td>
</tr>
<tr>
<td>Return on Capital Employed Ranked Factor</td>
<td>-0.71195</td>
<td>-0.00554</td>
<td>0.00061</td>
<td>0.00017</td>
<td>0.00059</td>
</tr>
<tr>
<td>Return on Equity Ranked Factor</td>
<td>-0.0826</td>
<td>-0.00545</td>
<td>0.00069</td>
<td>0.00011</td>
<td>0.00055</td>
</tr>
<tr>
<td>Sales Invested Capital Ranked Factor</td>
<td>-0.08929</td>
<td>-0.00544</td>
<td>0.00061</td>
<td>0.00017</td>
<td>0.00055</td>
</tr>
<tr>
<td>Sales Stockholders Equity Ranked Factor</td>
<td>-0.11734</td>
<td>-0.0286</td>
<td>0.00061</td>
<td>0.00017</td>
<td>0.00055</td>
</tr>
<tr>
<td>Sales Working Capital Ranked Factor</td>
<td>-0.09780</td>
<td>-0.0286</td>
<td>0.00061</td>
<td>0.00017</td>
<td>0.00055</td>
</tr>
<tr>
<td>Statistical Measure</td>
<td>Minimum</td>
<td>1st Quartile</td>
<td>Median</td>
<td>Mean</td>
<td>3rd Quartile</td>
</tr>
<tr>
<td>---------------------</td>
<td>---------</td>
<td>--------------</td>
<td>--------</td>
<td>------</td>
<td>--------------</td>
</tr>
<tr>
<td>Shillers_Cyclically_Adjusted_P.E_RatioRanked_Factor</td>
<td>-0.1369</td>
<td>-0.0041</td>
<td>0.0012</td>
<td>0.0013</td>
<td>0.0065</td>
</tr>
<tr>
<td>Short_Term_Debt_Ranked_Factor</td>
<td>-0.09961</td>
<td>-0.00704</td>
<td>-0.00074</td>
<td>-0.00031</td>
<td>-0.00602</td>
</tr>
<tr>
<td>Total_Debt_Ranked_Factor</td>
<td>-8.4e-02</td>
<td>-5.6e-03</td>
<td>-1.7e-04</td>
<td>-8.4e-05</td>
<td>5.5e-03</td>
</tr>
<tr>
<td>Total_Debt_Capital_Ranked_Factor</td>
<td>0.08945</td>
<td>0.00593</td>
<td>0.00017</td>
<td>0.00019</td>
<td>0.00557</td>
</tr>
<tr>
<td>Total_Debt_EBITDA_Ranked_Factor</td>
<td>-0.08705</td>
<td>-0.00550</td>
<td>-0.00026</td>
<td>-0.00021</td>
<td>0.00509</td>
</tr>
<tr>
<td>Total_Debt_Equity_Ranked_Factor</td>
<td>-8.4e-02</td>
<td>-5.9e-03</td>
<td>-1.0e-05</td>
<td>5.6e-03</td>
<td>9.6e-02</td>
</tr>
<tr>
<td>Total_Debt_Inventory_Ranked_Factor</td>
<td>-8.6e-02</td>
<td>-0.00611</td>
<td>0.00053</td>
<td>0.00011</td>
<td>0.00669</td>
</tr>
<tr>
<td>Total_Debt_Total_Assets_Ranked_Factor</td>
<td>-1.0e-01</td>
<td>-6.5e-03</td>
<td>-1.6e-04</td>
<td>-7.9e-05</td>
<td>6.2e-03</td>
</tr>
<tr>
<td>Total_Liabilities_Total_Tangible_Assets_Ranked_Factor</td>
<td>-1.0e-01</td>
<td>-6.5e-03</td>
<td>-1.6e-04</td>
<td>-7.9e-05</td>
<td>6.2e-03</td>
</tr>
<tr>
<td>Trailing_P.E_to_Growth_PEG_ratio_Ranked_Factor</td>
<td>-0.07659</td>
<td>-0.00573</td>
<td>0.00027</td>
<td>0.00055</td>
<td>0.00673</td>
</tr>
</tbody>
</table>
Appendix 7 - Contour plots from KNS

Copied from Kozak, Nagel, Santosh (2020) “shrinking the cross section.”

Fig. 1. OOS $R^2$ from dual-penalty specification (Fama-French 25 ME/BM portfolios). OOS cross-sectional $R^2$ for families of models that employ both $L_1$ and $L_2$ penalties simultaneously using 25 Fama-French ME/BM-sorted portfolios (Panel a) and 25 PCs based on Fama and French portfolios (Panel b). We quantify the strength of the $L_2$ penalty by prior root expected $SR^2(\kappa)$ on the x-axis. We show the number of retained variables in the SDF, which quantifies the strength of the $L_1$ penalty, on the y-axis. Warmer (yellow) colors depict higher values of OOS $R^2$. Both axes are plotted on logarithmic scale. The sample is daily from July 1926 to December 2017. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

Fig. 3. OOS $R^2$ from dual-penalty specification (50 anomaly portfolios). OOS cross-sectional $R^2$ for families of models that employ both $L_1$ and $L_2$ penalties simultaneously using 50 anomaly portfolios (Panel a) and 50 PCs based on anomaly portfolios (Panel b). We quantify the strength of the $L_2$ penalty by prior root expected $SR^2(\kappa)$ on the x-axis. We show the number of retained variables in the SDF, which quantifies the strength of the $L_1$ penalty, on the y-axis. Warmer (yellow) colors depict higher values of OOS $R^2$. Both axes are plotted on logarithmic scale. The sample is daily from November 1973 to December 2017. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
Fig. 5. OOS $R^2$ from dual-penalty specification (WFR portfolios). OOS cross-sectional $R^2$ for families of models that employ both $L^1$ and $L^2$ penalties simultaneously using 80 WFR portfolios (Panel a) and 80 PCs based on WFR portfolios (Panel b). We quantify the strength of the $L^2$ penalty by prior root expected $SR^2 (\kappa)$ on the $\kappa$-axis. We show the number of retained variables in the SDE, which quantifies the strength of the $L^1$ penalty, on the y-axis. Warmer (yellow) colors depict higher values of OOS $R^2$. Both axes are plotted on logarithmic scale. The sample is daily from September 1964 to December 2017. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)