Erasmus School of Economics

Bachelor Thesis BSc[2] in Econometrics and Economics

Covariance Matrix Uncertainty in Cross-Sectional Shrinkage

Author

Olivier Wollaert (458 704)

Supervisor

Dr. Maria Grith

Second assessor

Dr. Martina Zaharieva

Date Final version: July 4th 2021

Abstract

I investigate incorporating the Analytical Nonlinear Shrinkage (ANS) estimator in the Stochastic Discount Factor (SDF) framework in a high dimensional setting and find no evidence that using the ANS estimator to estimate covariance matrices leads to improved performance. To arrive at this conclusion, interaction portfolios are created based on financial ratios of all stocks in the three largest US stock exchanges. The covariance matrix of the portfolios is estimated using ANS, which is consequently used in the SDF framework to summarize the joint, cross-sectional explanatory power of stock return predictors. I find no substantial difference between using the ANS estimator or the sample covariance estimator in the SDF based on Sharpe Ratios and out-of-sample R^2 values. When splitting up the entire sample into shorter sample periods, there is strong evidence that different characteristics best explain the SDF in different periods. Furthermore, different periods require different degrees of shrinkage. The SDF performs best for smaller periods in which specific industries have abnormal returns.

Contents

1 Introduction

Since the introduction of the Capital Asset Pricing Model (CAPM) by Sharpe (1964) and Lintner (1965), hundreds of alternative stock characteristics have been published in asset pricing literature, claiming to add value in better capturing stock returns (J. Cochrane, 2011). One way to derive the price of an asset is by discounting its future cash flows using stochastic discount factors (SDF). With many proposed stock characteristics claiming to perform well as a factor in the SDF model, a common approach is to introduce sparsity in characteristic based factor models, that is, including a small number of factors. These models tackle the problem of high dimensionality due to too many stock characteristics.

When the sparse factor representation assumption is relaxed, and potentially hundreds of factors are considered, conventional methods to estimate factor loadings become infeasible due to their high dimensionality (Feng et al., 2020). Machine learning techniques like lasso-style estimation, ridge regression, and a combination thereof, an elastic net, have been proposed to overcome this problem. These techniques are frequently used in risk premia estimation and applied directly on cross-sectional regressions.

In contrast to applying these techniques directly on the cross-sectional regressions, Kozak et al. (2020) recently investigated the estimation of risk prices and applied these machine learning techniques on the SDF. In their 2020 paper, Kozak et al. investigate the construction of an SDF that consists of a sparse amount of factors and shrinks the factors with low explanatory power. Moreover, in contrast to the majority of literature, they pose no prior restrictions on the sparsity of the SDF model and hence consider the possibility that it includes many factors.

This paper elaborates upon the work of Kozak et al. (2020) and mainly focuses on a data set constructed of interactions. First, I replicate the research of Kozak et al. (2020), after which I extend their research by shifting the focus to the incorporation of covariance matrix estimation into the SDF framework, in contrast to their assumption of known covariance matrices. Specifically for the interaction data set, which has a high dimension relative to the sample size, sample covariance matrices are well known to provide poor estimation results (Ledoit and Wolf, 2020c). This leads to the motivation of this research, namely that in this high dimensional setting sample, more extensive covariance matrix estimators are likely to provide more accurate results and hence lead to additional, more accurate insights. A second considered area of improvement is the estimation of the SDF in three different, non-overlapping periods. Instead of estimating one set of parameters for the entire sample size, parameters are re-estimated for the three samples of equal length. The interest in doing this is two folded. Firstly, it is of interest to evaluate whether different periods require different levels of shrinkage. A second interesting aspect is to assess if the characteristics that best explain the cross sectional returns differ across periods. The Internet Appendix of Kozak et al. (2019) already gives an indication of a changing impact of certain characteristics when comparing pre-2005 annualized returns to post-2005 annualized returns for different characteristics. The aforementioned areas of improvement lead to the following research question that is answered throughout this paper:

"To what extent does the allowance of covariance matrix uncertainty influence the performance of a characteristic based stochastic discount factor in a high-dimensional framework?"

This research is relevant in two different areas. Firstly, it is of academic relevance as a novel type of SDF model is used for estimating cross sectional stock returns. As this is a relatively new method and results of Kozak et al. (2020) are promising, further researching the performance of these SDF models bears importance. Conducting more in-depth research on the effect of covariance matrix uncertainty robustifies the outcomes of Kozak et al. (2020) and places their method in a more general framework. Secondly, this research is of practical relevance as it gives portfolio managers a new, potentially better technique to construct SDF models and consequently estimate and forecast the cross section of equity returns. It is worthwhile to allow for covariance uncertainty and shift the focus to constructing models that estimate parameters better. In turn, this can provide higher Sharpe Ratios, which leads to better investment decisions when considering the mean-variance portfolios of Markowitz (1952).

This paper considers the Analytical Nonlinear Shrinkage (ANS) estimator of Ledoit and Wolf (2020a) for covariance matrix estimation. Nonlinear shrinkage estimators estimate N free parameters for a portfolio of size N, whereas linear shrinkage estimators estimate only one free parameter. This paper opts for a nonlinear shrinkage estimator as opposed to a linear shrinkage estimator due to its higher flexibility and its dominant performance over linear shrinkage estimators (Ledoit and Wolf, 2017). This state-of-the-art nonlinear shrinkage estimator is the first with an analytical as opposed to a numerical form and is the first that can handle covariance matrix estimation well for dimensions above 1000. In fact, it can handle dimensions up to 10000. This directly leads to the justification of using the ANS estimator: it is the first of its kind to be able to handle the dimensions of the interaction data set well. Furthermore, the ANS estimator is asymptotically optimal and performs well when the dimension of the covariance matrix corresponds in size to the sample size.

In this research, three different data sets are considered. The first data set contains the daily returns of 25 portfolios based on the intersections of five market equity portfolios and five book-to-market ratio portfolios of Fama and French. Secondly, a data set based on anomaly stock characteristics as provided by Kozak (2020) is considered. These two data sets are used for replication purposes in this paper. The last data set is a self-constructed data set based on the interactions of 70 financial ratios retrieved from Wharton Research Data Services (WRDS). This data set is used in the extension and split up into three different subsamples with a sample size of 15 years each, i.e. 3780 days. With the covariance matrix of interactions having a dimension of approximately 2500, the sample size and covariance matrix dimensions correspond in size.

As an initial test, the Fama-French 25 portfolios and the 50 anomaly portfolios are considered. For the raw characteristics of the Fama-French 25 portfolios, two to three characteristics already describe the SDF well. In the PC transformed case, one factor already explains the SDF almost optimally. For the raw characteristics of the 50 anomaly portfolios, sparse models perform poorly out of sample. In contrast, the PC transformed 50 anomaly characteristics do allow for a relatively sparse SDF representation.

Next, an interaction data set is considered, which shows good OOS behaviour. For the interactions set as given, substantial shrinkage is needed to allow for good OOS performance, and sparse models do not perform well. In the case of a PC transformed interactions set, high out-of-sample performance can be attained with a sparse model and little shrinkage, and hence a sparse representation of the SDF exists. When the sample is split up into three periods, the dual-shrinkage estimator performs most along the lines of expectations in the 1989-2004 period. A likely explanation for this is that this period contains abnormal returns due to industry-specific conditions periods. Incorporating the ANS estimator logically leads to different coefficients, but OOS performance is influenced little, also in the three smaller subsamples.

All in all, I find no evidence that implementing the Analytical Nonlinear Shrinkage (ANS) estimator of Ledoit and Wolf (2020a) into a high-dimensional SDF framework leads to better performance. Furthermore, the SDF performs best for periods in which specific industries have abnormal returns.

The remainder of this paper is organised as follows. In Section 2, the existing literature relevant to the scope of this paper is outlined. Thereafter, Section 3 elaborates upon the used methods and techniques, after which the data used in this research is presented in Section 4. Next, results are presented in Section 5. Lastly, concluding remarks are made in Section 6.

2 Literature

The introduction of the CAPM and the test of its significance by Fama and MacBeth (1973) is generally seen as a starting point of asset pricing literature. Since then, extensive literature has been written on this subject, and many models to estimate asset prices in different situations have been proposed.

One of these models is the SDF model. An SDF is a random variable that discounts future payoffs of portfolios or assets and consequently takes the expected value. This estimates the present value of a portfolio or asset by considering its future cash flows. As already touched upon in Section 1, many proposed SDF representations try to tackle the problem of many stock characteristics by introducing sparsity; models with a small number of factors based on stock characteristics that try to explain the present value of an asset as well as possible. Famous examples are the three-factor model proposed by Fama and French (1996) and the five-factor model from the same authors in 2015. The two aforementioned factor models of Fama and French are examples of observed or characteristic based factor models, which use a set of observed parameters as input. Factor models are widely used in various fields and are also used in stock return prediction frameworks.

Discussion on the validity of sparse SDF is still ongoing. Lin and Zhang (2013) try to motivate the use of a characteristic sparse representation of the SDF. Their q-theory model provides a strategy where one forecasts returns based on expected profitability and future investments. Hou et al. (2015) for example pursue this strategy and find better performance than the Fama French three-factor model in most cases. Although the q-theory model provides good results, both expected profitability and future investments are unobservable, as mentioned by Kozak et al. (2020). However, many other characteristics can likely capture these variables and hence can also estimate expected returns. Therefore, Kozak et al. (2020) reason many factors are likely required to approximate the SDF.

More and more stock characteristics that claim to predict stock returns well are being proposed in the literature. Many of these are significant when considered individually but become redundant when compared to already existing factors (Hou et al., 2015, Harvey et al., 2016 and Feng et al., 2020). This ever-increasing literature results in the multidimensional challenge, as described by J. Cochrane

(2011). This challenge arises when one considers constructing a factor model based on a large number of potential cross-section return predictors: using a conventional cross-sectional regression with a large number of predictors results in spurious overfitting of SDF coefficients.

An interesting approach to tackling the high dimensionality problem is using a Bayesian approach that incorporates a prior specification in an asset pricing model. A prior specification is the probability distribution that represents beliefs about a certain variable before evidence or data is taken into account. These prior beliefs are used to construct posterior beliefs, that is, after taking into account evidence. Harvey et al. (2008) show that a Bayesian approach works well for one-period ahead asset allocation. Pastor and Stambaugh (2000) and Pastor (2000) use sample evidence to update the prior and shrink all SDF coefficients towards zero with an equal degree of shrinkage. Kozak et al. (2020) impose a prior similar to the ones in Pastor and Stambaugh (2000), but impose relatively more shrinkage on low-eigenvalue principal components (PCs) compared to high-eigenvalue PCs.

Another relatively new development in asset pricing literature that regards tackling high dimensionality is the use of machine learning techniques. Kelly et al. (2019) provide a dimensionality reduction method, Instrumented Principal Component Analysis, for modelling the cross section of returns. This model, which describes the SDF with only a few PCs, explains the cross section significantly better compared to existing factor models. Kozak et al. (2018), who also use a principal component (PC) based SDF, come to a similar conclusion. Gu et al. (2020) show that machine learning techniques also work well empirically, and large economic gains can be made using machine learning forecasts.

One of the machine learning techniques frequently used to tackle high dimensionality is the lasso-style estimation as introduced by Tibshirani (1996). This method minimizes the residual sum of squares and imposes a constraint on the absolute value of coefficients. This constraint is constructed such that some coefficients are set to zero, which therefore automatically provides sparse solutions. For example, Rapach et al. (2013) apply lasso to select a sparse amount of predictors from a high number of potential predictors to forecast international stock markets. Freyberger and Weber (2020), Feng et al. (2020) and DeMiguel et al. (2020) use a lasso-style estimation to select characteristics-based factors and include penalties to ensure sparsity. Their findings back the views of amongst others Harvey et al. (2016), namely that relatively many cross-sectional stock return predictors are redundant. A drawback of lasso is its poor performance when regressors are correlated. Other machine learning techniques like ridge regression and an elastic net do perform well in this case (Zou and Hastie, 2005). Originally introduced by Hoerl and Kennard (1970), ridge regression shrinks coefficients with varying degrees to zero. Zou and Hastie (2005) propose the use of an elastic net, which combines the penalties constructed in ridge regression and lasso estimation. The elastic net outperforms lasso, while it still has a sparse representation.

As discussed, there is economic support to allow relaxing the assumption of a sparse SDF. Kozak et al. (2020) recently proposed an SDF representation that potentially allows for many factors to be included in an SDF. Their research is at the intersection of Bayesian approaches and machine learning. They construct a Bayesian prior for the factor means, the biggest source of estimation uncertainty. Consequently, they calculate the corresponding posterior SDF coefficients. Their estimator is constructed using two different penalties, one that controls the amount of shrinkage imposed on the coefficients and one that controls the model's sparsity. This dual penalty specification is in a similar fashion to the elastic net of Zou and Hastie (2005). By combining prior beliefs with shrinkage and sparsity penalties, their estimator performs well out of sample while allowing for many SDF coefficients to be potentially included. The estimator of Kozak et al. (2020) is equivalent to an estimator that minimizes the distance of Hansen and Jagannathan (1991) and penalizes the amount of SDF coefficients via the sum of squared coefficients.

Next to estimation uncertainty in the mean, which is the focus of Kozak et al. (2020), another part of finance literature focuses on uncertainty in the covariance, the second biggest contributor to estimation uncertainty. When the dimensionality of analysed assets increases compared to the sample size, sample covariance matrices are known to have poor performance (Engle et al., 2019). One of the many proposed methods to enhance the performance of large covariance matrix estimators in this high dimensional framework is called shrinkage methods. Ledoit and Wolf are at the forefront of academic research in this field. Ledoit and Wolf (2020c) summarize their most important work in this field over the last two decades. A widely applied method in literature is the linear shrinkage estimator to estimate the covariance matrix by Ledoit and Wolf (2004). Building upon their own research, Ledoit and Wolf (2012, 2015) turn to the more complicated but better performing nonlinear shrinkage estimators, providing robust results. In contrast to the linear case where the transformation of sample eigenvalues is optimised in a two-dimensional space, nonlinear estimators put no constraint on the size of this dimension and consider nonlinear transformations of the sample eigenvalues to estimate covariance matrices. Another nonlinear shrinkage method with adequate results is the Nonparametric Eigenvalue-Regularized Covariance Matrix Estimator (NERCOME) proposed by Lam (2016). This estimator builds upon the work of Abadir et al. (2014) and splits the data into two independent sets; one to estimate the eigenvectors and the other to estimate the corresponding eigenvalues. This is repeated for a large number of sample splits and then averaged. The main advantage of the NERCOME is that it is much neater than other estimators. Both the NL-shrinkage estimator of Ledoit and Wolf and the NERCOME are numerical estimators.

In a recent paper, Ledoit and Wolf (2020a) proposed the already mentioned ANS estimator. This is the first analytical nonlinear shrinkage estimator for high-dimension covariance matrices. It has the speed of the linear shrinkage estimator of Ledoit and Wolf (2004) due to its analytical form, performs as well as the NL-shrinkage estimator of Ledoit and Wolf, and has the simplicity of the NERCOME estimator. Furthermore, it estimates covariance matrices of dimension 10000 and beyond, in contrast to other nonlinear shrinkage estimators working for dimensions up to 1000.

3 Methodology

This section presents the methods and techniques used in this paper. First, Section 3.1 discusses the theoretical background of the SDF. Thereafter, Section 3.2 elaborates upon the SDF estimation. Lastly, the methodology of the ANS estimator and its incorporation into the SDF framework are discussed in Section 3.3 and Section 3.4, respectively.

3.1 Theoretical model

Following J. H. Cochrane (2001), one can express the basic asset pricing formula as

$$P_{t-1} = \mathbb{E}_{t-1}[M_t x_t], \tag{1}$$

where P_{t-1} denotes an $N \times 1$ vector of the prices of N assets at time $t-1$ and x_t is defined as the payoff of an $N \times 1$ asset vector at future time t. The SDF M_t is defined as a random variable that discounts the future payoffs of the N portfolios. This formula holds for both individual assets and for portfolios, as proven in Back (2017). Using that returns R_t equal $\frac{x_t}{P_{t-1}}$ and assuming all prices are positive, Equation (1) simplifies to

$$\mathbb{E}_{t-1}[M_t R_t] = 1, \tag{2}$$

as in accordance with Back (2017). Consequently defining excess returns for the for $N \times 1$ vector of stocks R_t^* as $R_t - R_{m,t}$, the conditional pricing equation can be defined as

$$\mathbb{E}_{t-1}[M_t R_t^*] = 0, \tag{3}$$

where R_t^* is defined as the excess returns for N stocks. Throughout the remainder of this paper, excess returns are used. Following Kozak et al. (2020), an SDF is created in the linear span of excess returns in the spirit of Hansen and Jagannathan (1991), which is defined as

$$M_t = 1 - b_{t-1}'(R_t^* - \mathbb{E}R_t^*), \tag{4}$$

where b_{t-1} is defined as the $N \times 1$ vector of time-varying SDF loadings that satisfies the conditional pricing equation in Equation (3). As in accordance with Kozak et al. (2020), it is assumed that factors are observable and based on $N \times H$ characteristic matrix Z_t, which contains elements $z_{s,t}^i$, defined as the value corresponding to characteristics i for stock s at time t. This leads to the definition of an SDF linear in the span of factor returns as

$$F_t = Z_{t-1}' R_t^*. \tag{5}$$

The interpretation of factors F_t is two folded here. The factors represent the assets of which the stock returns are tried to be predicted as well as the potential factors that can be included in the SDF. Furthermore, slope coefficients b_t of the SDF are assumed to depend on characteristics Z_t, such that for a characteristic based asset pricing model, the SDF loadings are defined as

$$b_t = Z_t b, \tag{6}$$

where b is an $H \times 1$ time invariant coefficient vector and H is the amount of characteristic based factor returns. Combining Equation (4) and Equation (6) and deriving from Equation (5) that R_t^* is equal to $(Z_{t-1}' Z_{t-1})^{-1} Z_{t-1}' F_t$, leads to the following representation of the SDF:

$$M_t = 1 - b'(F_t - \mathbb{E}F_t). \tag{7}$$

When the first and second population moments are known, the SDF of Equation (7) can be solved for SDF coefficients b defined as

$$b = \Sigma^{-1}\mathbb{E}(F_t), \tag{8}$$

where $\Sigma \equiv \mathbb{E}[(F_t - \mathbb{E}F_t)(F_t - \mathbb{E}F_t)']$. A complete derivation of Equation (8) is included in Appendix A.1. However, population moments are unknown, and sample moments give poor, overfitted estimates of b since H is not extremely small relative to T. Therefore, alternative estimation methods are needed, as stated by Kozak et al. (2020). [1]

3.2 Estimation methodology

To avoid the problem of overfitted estimates when calculating $\hat{b}$ using sample moments, Kozak et al. (2020) suggest introducing economic prior beliefs on the expected returns of the factors. A prior is partially shaped by an investor's beliefs on the accuracy of a pricing model. Next to the prior beliefs, one can form posterior beliefs that are based upon the actual data. These posterior beliefs eventually form the posterior SDF coefficients. Choosing informative priors helps reduce the uncertainty about the posterior SDF coefficients. To ensure a closed form of the posterior distribution, that is, to ensure the prior and the posterior follow the same distribution, I consider the multivariate normal family of priors. This is defined identically to Kozak et al. (2020) as

$$\mu \sim \mathcal{N}\Big(0, \frac{\kappa^2}{\tau}\Sigma^{\eta}\Big), \tag{9}$$

where κ is a constant controlling the scale of the prior μ, τ is the trace of Σ and η a constant controlling the shape of the prior.

In the spirit of Kozak et al. (2020), the posterior mean of μ is calculated using the standard formula for a conjugate multivariate normal prior with known covariance matrix.[2] This boils down to the formula $\hat{\mu} = (\Sigma_0^{-1} + T\Sigma^{-1})^{-1}(\Sigma_0^{-1}\mu_0 + T\Sigma^{-1}\bar{\mu})^{-1}$. Here, the prior parameters equal $\mu_0 = 0$ and $\Sigma_0 = \frac{k^2}{\tau}\Sigma^{\eta}$. Plugging in these prior parameters gives $\hat{\mu} = (I + \gamma\Sigma^{1-\eta})^{-1}\bar{\mu}$. For the prior specification of Kozak et al. (2020), $\eta = 2$, and using $\hat{b} = \Sigma^{-1}\hat{\mu}$, this simplifies to

$$\hat{b} = (\Sigma + \gamma I)^{-1}\bar{\mu}, \tag{10}$$

where $\gamma = \frac{\tau}{\kappa^2 T}$ and I is the identity matrix of dimension N. In accordance with Kozak et al. (2020), the posterior variance is given by

$$\mathrm{var}(b) = \frac{1}{T}(\Sigma + \gamma I)^{-1}. \tag{11}$$

This posterior specification with $\eta = 2$ imposes stronger shrinkage on lower-variance coefficients and less shrinkage on high-variance coefficients. In contrast, the more commonly used $\eta = 1$ by Pastor and Stambaugh (2000) applies constant shrinkage across all coefficients.

The Bayesian estimator of Equation (10) is identical to a lasso-style based estimator, which is frequently used in machine learning. It is also equivalent to imposing shrinkage L^2 via penalty $\gamma b'b$ and minimizing the HJ distance of Hansen and Jagannathan (1991) of the model. This leads to the equation written as

$$\hat{b} = \arg\min_b (\bar{\mu} - \Sigma b)'\Sigma^{-1}(\bar{\mu} - \Sigma b) + \gamma b'b. \tag{12}$$

[1] The estimation methods in Section 3.2 make use of the code published by Kozak (2020). The read_me file attached to this paper contains a more in-depth explanation of the entire code and the utilization thereof.

[2] A prior that gives closed forms is called a conjugate prior

Next to imposing shrinkage on the model to account for high dimensionality, one also wants to consider the possibility that some factors do not contribute in explaining the SDF. Therefore, the focus is now shifted to a model where also sparsity can be imposed with an additional L_1 penalty via $\gamma_1 \sum_{i=1}^{H} |b_i|$. When combining both L_1 and L_2 penalties, this results in the following estimator:

$$\hat{b} = \arg\min_b (\bar{\mu} - \Sigma b)' \Sigma^{-1} (\bar{\mu} - \Sigma b) + \gamma_2 b' b + \gamma_1 \sum_{i=1}^{H} |b_i|, \tag{13}$$

where b denotes the vector of SDF coefficients, $\hat{b}$ the estimator of b, Σ the covariance matrix and γ_i the L_i penalty specification.[3] Note that one can attain the sole L_1 or L_2 penalty by setting respectively γ_2 or γ_1 to 0. For the prior specification $\eta = 2$ and penalty parameter $\gamma = \frac{\tau}{\kappa^2 T}$, the expected squared Sharpe Ratio can be defined as

$$\mathbb{E}[\mu \Sigma^{-1} \mu]^{1/2} = \kappa, \tag{14}$$

which implies that γ indirectly represents the expected squared Sharpe Ratio via an inverse relation. When the expected maximum Sharpe Ratio is high, the corresponding k is also high, which implies γ is low and hence little shrinkage is imposed on the model and vice versa. γ_1 and γ_2 are chosen such that they maximize the k-fold cross-validated out of sample R^2, which is given by

$$R^2_{OOS} = \frac{(\bar{\mu}_2 - \bar{\Sigma}_2 \hat{b})' (\bar{\mu}_2 - \bar{\Sigma}_2 \hat{b})}{\bar{\mu}_2' \bar{\mu}_2}, \tag{15}$$

where $\bar{b}$ is computed by applying Equation (10) on the $K - 1$ training sets. The subscript 2 indicates an OOS sample moment, in accordance with Kozak et al. (2020).

3.3 Extension: covariance matrix uncertainty

This paper builds upon Kozak et al. (2020) by including covariance matrix uncertainty in their framework. As mentioned in their Internet Appendix B, Kozak et al. (2020) already conducted some initial research on using the linear shrinkage methods in Ledoit and Wolf (2004) to shrink the sample covariance. This paper considers an alternative, newer method to estimate the covariance matrix. In particular, the Analytical Nonlinear Shrinkage model of Ledoit and Wolf (2020a) is considered. [4] This method provides a static estimator of the covariance matrix in a similar fashion to Ledoit and Wolf (2004) but has the potential to capture the true variance much better. The ANS covariance matrix estimator is as fast as the linear shrinkage method of Ledoit and Wolf (2004), has almost identical performance to the QUEST function and has a clear interpretation in line with the NERCOME. Therefore, it combines the optimal properties of these three estimators.

This method performs well for dimensions up to 10000, whereas other nonlinear shrinkage estimators only work well for dimensions up to 1000. This consequently leads to the justification of using this method. It has the potential to perform well in this framework, where a data set consisting of 2553 interactions is considered. The ANS estimator of Ledoit and Wolf (2020a) uses a kernel distribution and

[3]This estimator is solved using the LARS-EN algorithm of Zou and Hastie (2005). See read_me for more information

[4]This paper makes use of the code for the ANS method published by Ledoit and Wolf (2020b). The read_me file attached to this paper contains a more in-depth explanation of this code and the utilization thereof.

its corresponding Hilbert transform. A Hilbert transform of a distribution $k(x)$ can be interpreted as its convolution with $\frac{dt}{\pi t}$, also known as the Cauchy kernel (Ledoit and Wolf, 2020a).

To arrive at an analytical solution to the non-linear shrinkage estimator, the kernel distribution and Hilbert transform must satisfy two requirements. Let $k(x)$ be a continuous, symmetric kernel that has a non-negative probability density function and has bounded support $[-R,\ R]$. Furthermore, $k(x)$ is defined to have a zero mean and variance one. The kernel $k(x)$ and Hilbert transform $\mathcal{H}_k$ are assumed to satisfy the following requirements:

1. The kernel's Hilbert transform $\mathcal{H}_k$ exists and is continuous on $\mathbb{R}$.
2. Both the kernel and its Hilbert transform are functions whose total variation is bounded, i.e. finite.

The sample covariance matrix of observed factors $F = (f_1, \ldots, f_T)$ with sample size T is computed as $\bar{\Sigma} = F'F/T$, which makes use the fact that the observations in F are all demeaned. Along the lines of Ledoit and Wolf (2020a), the covariance matrix is factorised into its eigenvalues and its corresponding eigenvectors as a first step. This eigendecomposition of the sample covariance matrix is computed as

$$\bar{\Sigma} =: U\Lambda U', \tag{16}$$

where Λ is a diagonal matrix with as elements the eigenvalues $(\lambda_1, \ldots \lambda_n)$ sorted in nondecreasing order and U is an orthogonal matrix with columns $[u_1, \ldots u_n]$ representing the n eigenvectors corresponding to the eigenvalues of $\bar{\Sigma}$.

Before defining the kernel estimator, the bandwidth used in the kernel estimator is defined. The bandwidth is a free parameter, that is, one which is not pre-defined by the model, that strongly influences the estimator. The bandwidth used is variable and proportional to the sample eigenvalue, that is

$$h_j = \lambda_j h \quad \forall j = 1, \ldots, T, \tag{17}$$

where h_j is the locally adaptive bandwidth, λ_j the j^{th} sample eigenvalue and h is defined as the global bandwidth which takes on the form $h = T^z$. The global bandwidth h is chosen in correspondence with Jing et al. (2010) and Ledoit and Wolf (2020a). For the kernel estimator to nonlinearly shrink the sample eigenvalues, the following two conditions must be satisfied:

$$\lim_{T\to\infty} Th^{5/2} = \infty \quad \text{and} \quad \lim_{T\to\infty} h = 0. \tag{18}$$

To satisfy the first condition, the absolute value of the exponent z in $h = T^z$ must be strictly smaller than $\frac{2}{5}$. The second condition is satisfied when h is a negative exponent of T. From all possible values for the exponent, $-\frac{2}{5} < z < 0$, $z = -\frac{1}{3}$ is chosen as it is the first simple fraction in this bounded region above $-\frac{2}{5}$. This leads to the global bandwidth being defined as:

$$h = T^{-1/3}. \tag{19}$$

The kernel estimator is now defined as follows. Following the reasoning of Ledoit and Wolf (2020a), the kernel of Epanechnikov (1967) is chosen for the density estimation. This Epanechnikov kernel satisfies the two kernel requirements mentioned earlier. From this, the eigenvalue density is estimated with the Epanechnikov kernel as

$$\hat{f}\left(\lambda_i\right)=\frac{1}{n}\sum_{j=1}^{n}\frac{3}{4\sqrt{5}h_j}\left[1-\frac{1}{5}\left(\frac{\lambda_i-\lambda_j}{h_j}\right)^2\right]^+, \tag{20}$$

where $\hat{f}\left(\lambda_i\right)$ is the estimated eigenvalue density of the eigenvalue λ_i, h_j is the locally adaptive bandwidth and $[x]^+$ is the positive part of a real number, that is $[x]^+ = \max(x, 0)$. Consequently, the kernel distribution of its Hilbert transform is derived as by Ledoit and Wolf (2020a) is as follows:

$$\mathcal{H}_{\hat{f}}\left(\lambda_i\right)=\frac{1}{n}\sum_{j=1}^{n}\left\{-\frac{3\left(\lambda_i-\lambda_j\right)}{10\pi h_j^2}+\frac{3}{4\sqrt{5}\pi h_j}\left[1-\frac{1}{5}\left(\frac{\lambda_i-\lambda_j}{h_j}\right)^2\right]\times\log\left|\frac{\sqrt{5}h_j-\lambda_i+\lambda_j}{\sqrt{5}h_j+\lambda_i-\lambda_j}\right|\right\}. \tag{21}$$

A more in-depth derivation of the Kernel estimator for the eigenvalue density and its corresponding Hilbert transform are provided in Appendix A.2. The eigenvalue density and Hilbert transform computed using Equation (20) and (21), respectively, are now used to estimate the shrunk eigenvalues. The eigenvalues are shrunk according to the oracle asymptotically optimal nonlinear shrinkage formula as per Ledoit and Wolf (2020a):

$$\hat{d}_i=\frac{\lambda_i}{\left[\pi\frac{n}{T}\lambda_i\hat{f}\left(\lambda_i\right)\right]^2+\left[1-\frac{n}{T}-\pi\frac{n}{T}\lambda_i\mathcal{H}_{\hat{f}}\left(\lambda_i\right)\right]^2}\qquad\forall i=1,\ldots,n. \tag{22}$$

The diagonal of matrix $\widehat{D}_T$ is filled with the elements $\hat{d}_{T,i}$ for $i = 1, \ldots, n$ computed in Equation (22). As a last step, the covariance matrix estimator is recomposed in accordance with Equation (16) as follows:

$$\widehat{\Sigma}=U\Lambda U'=\sum_{i=1}^{n}\widehat{d}_i\cdot u_i u_i', \tag{23}$$

where $\hat{d}_i$ is the i^{th} shrunk eigenvalue, and u_i represents the i^{th} eigenvector corresponding to the i^{th} eigenvalue of $\bar{\Sigma}$ for $i = 1, \ldots, n$.

3.4 Incorporation of the ANS estimator in the SDF framework

The focus is now shifted back to the paper of Kozak et al. (2020). In a similar fashion to Kozak's Internet Appendix (Kozak et al., 2019), the estimated covariance matrix from Section 3.3 is treated in the same way as the known covariance matrix in the estimation of $\hat{b}$. This means the estimated covariance matrix from Equation (23) is used as a plugin estimator for the dual penalty case in Equation (13). The analytical nonlinear shrinkage estimator performs well when the sample size has a dimension corresponding to the dimension of the covariance matrix. With the covariance matrix of interactions having a dimension of approximately 2500, three subsamples are chosen with a sample size of 15 years each, which is approximately 3780 days. Here, it is assumed a year contains 252 trading days, as is common in asset pricing literature. The dimensions of the sample size and the covariance matrix coincide in that case, which justifies using the ANS estimator.

4 Data

This section elaborates upon the data used in this paper. Three different data sets are considered, two of which are identical to the ones used in Kozak et al. (2020) to allow for a clear replication and comparison. The third data set is used for the extension and deviates from Kozak et al. (2020).

4.1 Data replication

The first data set contains the daily returns of 25 portfolios based on the intersections of five market equity portfolios and five book-to-market ratio portfolios of Fama and French, as provided by French (2020). The portfolios include all stocks from the three major U.S. exchanges, NASDAQ, NYSE, and AMEX, for which the required data is available. The data coincides with Kozak et al. (2020) for replication purposes and ranges from 07/1926 to 12/2017.

Secondly, a data set based on anomaly stock characteristics as provided by Kozak (2020) is considered. In the replication part of this paper, a daily data set of 50 portfolios is used, which ranges from 02/07/1963 until 29/12/2017. These portfolios are based on U.S. firms in the CRSP universe, which includes stocks from the NASDAQ, NYSE, and AMEX. Due to some missing values, the sample used ranges from 11/01/1973 until 12/29/2017 in accordance with Kozak et al. (2020). This data set is useful for the scope of this paper as its 50 characteristics are known to be related to average returns, as described in existing literature. Table C1 in Appendix C contains a list of the 50 anomaly stock characteristics and their mean annualized returns.

4.2 Data extension

The last data set is a constructed data set based on the interactions of the 70 WRDS financial ratios. This set is based on monthly financial ratios on a firm level for U.S. firms in the CRSP universe, which includes stocks from the NASDAQ, NYSE, and AMEX. The final set consists of 69 monthly updated, managed portfolios linked to the daily returns of corresponding stocks. The ratios and their mean annualized returns are provided in Table D1 Appendix D.

The data set used deviates in two ways from the WRDS data set considered by Kozak et al. (2020). The first difference lies within the ratios used. Whereas Kozak et al. (2020) consider 70 financial ratios on an industry level, this paper considers the same 70 financial ratios on a firm level. This allows for a more precise representation of specific firms, as characteristics are now tailored to a specific company instead of aggregated over an entire industry. Secondly, the data used ranges from 31/01/1970 to 31/12/2020 instead of ranging from 09/1964 until 12/2017. The starting date is later since the minimum allowed date for the 70 financial ratios is 31/01/1970 in the WRDS database. Using data up until 2020 allows for more recent observations to be included.

The WRDS 70 financial ratios data set needs to be altered before it is used in the SDF model. First, stocks with non-existing or non-logical market capitalizations are removed from the investment universe. Secondly, a bound on minimum market capitalization is set in accordance with Kozak et al. (2020), by excluding all stocks with a market capitalization smaller than 0.01% of the total market capitalization at each point in time. For the stocks left in the universe, the firm-specific financial ratios are linked to the corresponding stock data. The date of the ratios assigned to a stock precedes the stock return date. [5]

The raw data characteristics are transformed to ensure the focus of this research stays on the cross-sectional aspect of return prediction, to provide results that are insensitive to outliers and to keep leverage

[5]For example, the daily return of a stock on 10 May 1975 is linked to the financial ratios of that stock of April 1975

fixed across all portfolios (Kozak et al., 2020). Along the lines of Kozak et al. (2020) and Freyberger and Weber (2020), the characteristics are first ranked as follows:

$$rc_{s,t}^{i} = \frac{\text{rank}(c_{s,t}^{i})}{n_t + 1}, \tag{24}$$

where $c_{s,t}^{i}$ is defined as characteristic i for stock s at time t, and n_t is the number of stocks at time t for which characteristic i is available. Furthermore, the rank is calculated as $\text{rank}(\min_{i=1,\dots,N_t}(c_{s,t}^{i})) = 1$ and $\text{rank}(\max_{i=1,\dots,N_t}(c_{s,t}^{i})) = N_t$. Two characteristics $c_{s,t}^{i}$ and $c_{s,t}^{j}$ with identical values are both assigned their average rank. Consequently, the procedure of Kozak et al. (2020) is followed to normalize the rank-transformed characteristics $rc_{s,t}^{i}$. That is, first the mean rank is deducted from all ranks after which they are divided by the sum of absolute differences with respect to the mean:

$$z_{s,t}^{i} = \frac{rc_{s,t}^{i} - \bar{rc}_{s,t}^{i}}{\sum_{s=1}^{n_t} |rc_{s,t}^{i} - \bar{rc}_{s,t}^{i}|}, \tag{25}$$

where the mean rank is defined as $\bar{rc}_{t}^{i} = \frac{1}{n_t}\sum_{s=1}^{n_t} rc_{s,t}^{i}$. As a last step, the matrix of transformed characteristics $z_{s,t}^{i}$ is multiplied with returns to get one factor for each characteristic at time t, that is

$$F_t = Z_{t-1}' R_t. \tag{26}$$

In accordance with Kozak et al. (2020), the WRDS data set is extended by considering the linear first-order interactions of characteristics. That is, for two characteristics i and j of a stock s at time t, also the normalized product of both characteristics is considered. This leads to an exponential increase in the number of characteristics to potentially explain the SDF. When considering first-order interactions and the second power of characteristics next to the characteristics themselves, the total number of candidate factors equals $\sum_{i=1}^{n} i + n$, where $\sum_{i=1}^{n} i$ is the number of first-order interactions and second powers, and n is the number of original characteristics. $\sum_{i=1}^{n} i$ can be rewritten to $\frac{1}{2}n(n+1)$, leading to the total candidate factors being equal to $\frac{1}{2}n^2 + 2\frac{1}{2}n$.

There are three reasons to use this interaction-based data set. Firstly, it is statistically more challenging to use this data set due to its high dimension. Secondly, the linearity assumption of the factor portfolio weights of the characteristics can be relaxed, which is an assumption without a solid foundation (Kozak et al., 2020). Thirdly, it justifies using the ANS estimator for covariance matrix estimation, as this is specifically designed for high dimensional covariance matrices. Using the sample covariance matrix is known to work relatively well for covariance matrices with a low dimension relative to the sample size. However, in the case of using interactions, the sample covariance matrix is more likely to perform relatively poor.

The interactions are constructed in line with Kozak et al. (2020) as follows. For two rank characteristics $z_{s,t}^{i}$ and $z_{s,t}^{j}$ that are transformed and normalized according to Equation (**??**) and (25), the interaction characteristic $z_{s,t}^{ij}$ is defined as the product of both individual characteristics that is normalized in accordance with Equation (25). That is:

$$z_{s,t}^{ij} = \frac{\left(z_{s,t}^{i} z_{s,t}^{j} - \frac{1}{n_t}\sum_{s=1}^{n_t} z_{s,t}^{i} z_{s,t}^{j}\right)}{\sum_{s=1}^{n_t} |z_{s,t}^{i} z_{s,t}^{j} - \frac{1}{n_t}\sum_{s=1}^{n_t} z_{s,t}^{i} z_{s,t}^{j}|}. \tag{27}$$

Again, as a last step, the matrix Z_{t-1} is multiplied with returns to get one factor for each characteristic at each point in time, as in Equation (26).

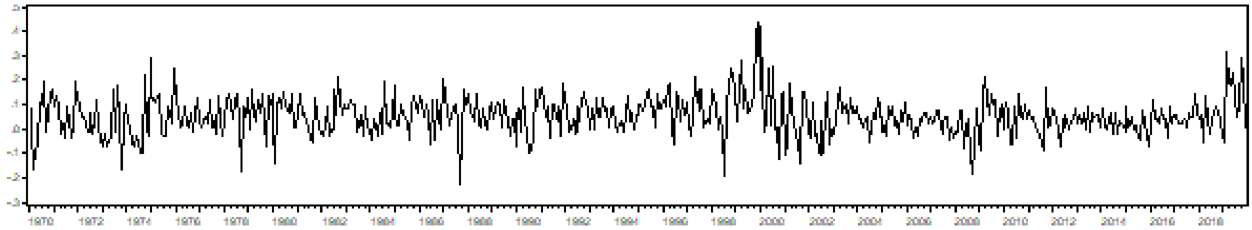


Figure 1: De-marketed monthly return series of the WRDS data set for the period 31/01/1970 to 31/12/2020

I now turn to the justification of the three different periods for the WRDS data set. As already mentioned in Section 1, three non-overlapping periods of equal length are chosen to evaluate whether different periods require different levels of shrinkage and whether different characteristics best explain the cross sectional return in different periods. In Figure 1, the de-marketed return series of the WRDS data set is shown. Based on this figure and on macroeconomic conditions, the following three different periods are chosen:

1. 31/01/1970 to 31/01/1986: Figure 1 shows a relatively volatile return series for this period. Furthermore, this period has no significant financial crises for U.S. based firms.
2. 01/01/1990 to 31/12/2004: Figure 1 shows a period with large growth at first and a drop in growth thereafter. This coincides with the boom and bust of the dot-com bubble.
3. 01/01/2005 to 31/12/2020: Figure 1 shows a relatively low-volatile period. Two volatile periods in the graph correspond to two big crises, the great financial crisis of 2007 and the COVID-19 crisis.

These three periods are chosen as they have different macroeconomic conditions and their return series behave differently. Furthermore, no large economic events overlap the samples and are present in more periods. For these reasons, these periods are appropriate for the evaluation of the model in different periods and allow for a clear comparison.

5 Results

This section presents the research results. Firstly, the replication of the models considered in Kozak et al. (2020) is shown in Section 5.1. Secondly, the extension is elaborated upon in Section 5.2.

5.1 Replication

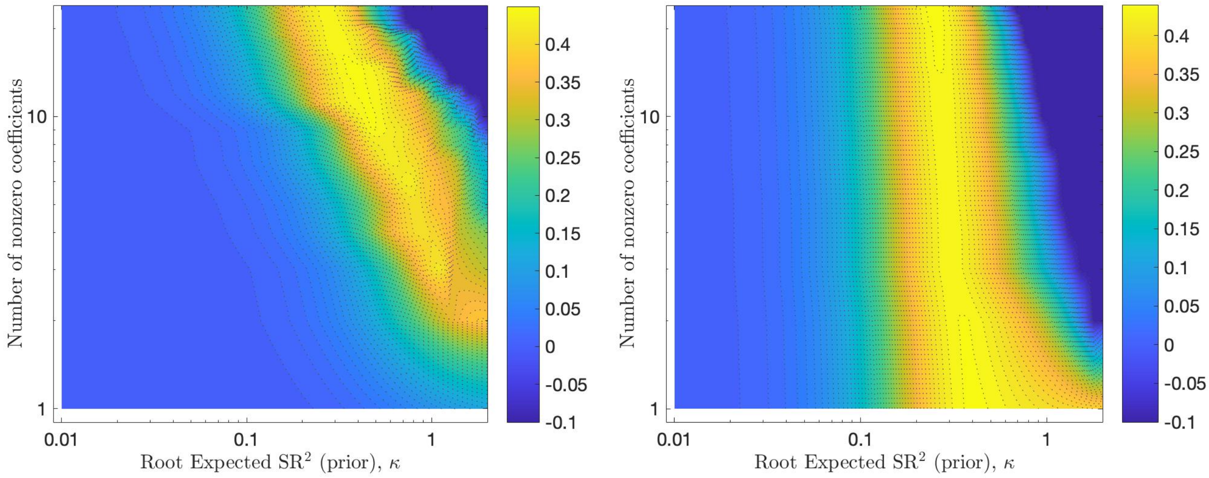
5.1.1 Fama-French portfolio

As an initial test, the Fama and French 25 portfolios, as described in Section 4, are considered. This data set is well-documented in literature, making it easy to compare the results of this research to existing literature and verify whether the proposed methods perform as desired.

Figure 2 contains results of the estimator based on both the L^1 sparsity penalty and the L^2 shrinkage penalty. The left-hand side of the figure contains the results of the SDF based directly on the factor returns. The right-hand side considers the SDF in the space of PCs of the factor returns. Along the x-axis of a contour plot, the root of the expected squared Sharpe Ratio of the prior is indicated. As stated in Equation (14), this equals κ for $\eta = 2$. Due to the inverse relationship between κ and the L^2 shrinkage penalty γ_2, a higher γ_2 and hence stronger shrinkage is obtained when moving to the left along the x-axis. The y-axis indicates the amount of sparsity imposed by the estimator, where moving up allows for more nonzero coefficients and hence represents less sparsity. This is regulated through the strength of the L^1 penalty. A point on the contour map represents the OOS R^2, calculated using Equation (15) for given penalties γ_1 and γ_2. A higher OOS R^2 value is represented by a more yellow colour, as indicated by the colour bar on the right of the contour map.

Figure 2a shows OOS results for an SDF based on the raw Fama-French 25 portfolios. The diagonal area from the mid top to bottom right shows combinations of penalties for which the estimator provides the highest OOS R^2 values. A sparse model including two to three factors performs well, which coincides with Lewellen et al. (2010) who find that the 25 portfolios are approximately linear combinations of the Small-Minus-Big (SMB) and High-Minus-Low (HML) factors of the Fama-French three factor model. Because of this linearity and the fact that the Fama-French three factor model performs well OOS, two to three out of 25 portfolios should span the SDF and perform well. The empirical findings in Figure 2a back up this theory. Furthermore, from the diagonal slope of the high OOS R^2, it can be concluded that shrinkage and sparsity are interchangeable to ensure good OOS performance. That is, when less sparsity is imposed but at the same time stronger shrinkage is enforced or vice versa, OOS R^2 values stay optimal. For non-shrunk, non-sparse models, shown in the top right, OOS performance is extremely poor. This is due to the overfitting of noise in the in-sample (IS) data when almost all factors are included and the absence of shrinkage to offset overfitting. Models with both strong L^1 and L^2 shrinkage perform poorly.

Figure 2b shows OOS results for an SDF based on the PCs of the Fama-French 25 portfolios. A straight area from mid top to mid bottom gives maximal OOS R^2 values. Even one PC already gives almost optimal OOS results (OOS R^2 of 0.4427 for a model based on one PC compared to the absolute maximum of 0.4462 for a model based on 24 PCs). Since the area performing best is almost straight, models with a varying number of variables all require a similar level of shrinkage. This is due to the fact that the SDF coefficients of low-variance PCs are already set close to zero by the estimator $\hat{b}$. Therefore, the impact of setting them to exactly zero by imposing additional sparsity is minimal (Kozak et al., 2020). As in Figure 2a, non-shrunk, non-sparse models in the top right of the graph give poor OOS performance. Models with both strong L^1 and L^2 shrinkage also perform poorly. Comparing the results from Figure 2 to the results from Kozak et al. (2020), the output is identical, which in turn also leads to similar conclusions.



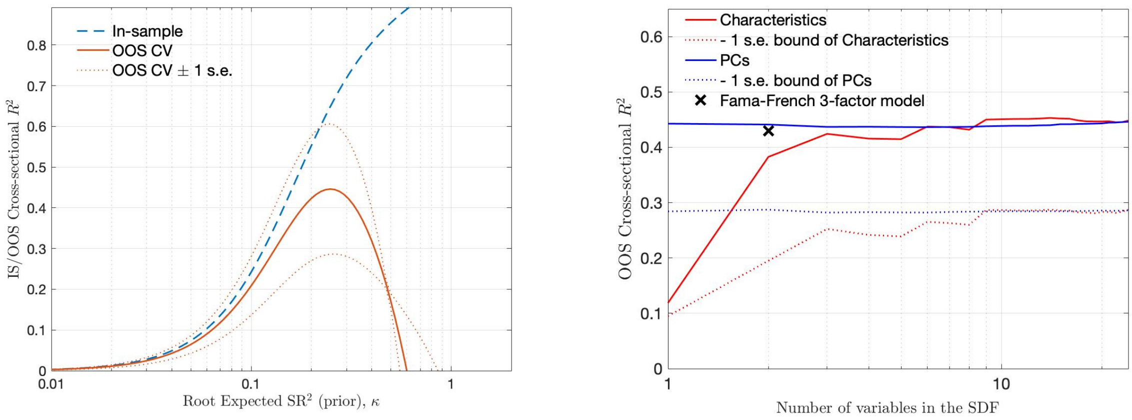
(a) Raw Fama-French 25 portfolios

(b) Principal Components of Fama-French 25 portfolios

Figure 2: OOS R^2 for models with different levels of L^1 and L^2 shrinkage for the Fama-French 25 portfolios. The x-axis shows the square root of the expected SR^2. Due to the inverse relationship between κ and L^2, the x-axis indirectly represents levels of L^2 penalties, where more shrinkage is obtained when moving to the left along the x-axis. The y-axis indicates the number of nonzero coefficients in the model, which is regulated through penalty L^1. More nonzero coefficients imply the model is less sparse and the L^1 penalty is less strong. The colour bar on the right indicates the value of the OOS R^2, where more yellow colours indicate higher values. A minimum bound for OOS R^2 in the graph is set at -0.1. Some values can be even lower, but since one can interpret all values lower than -0.1 as poor performing, and to allow for a clear comparison, a minimum bound is set at -0.1. The data used ranges from July 1926 to December 2017 and is daily.

Figure 3 zooms in on more specific cuts from Figure 2. Figure 3a corresponds with a cut through the top of Figure 2a. That is, no sparsity is included in the model, and the maximum R^2 for different levels of shrinkage is reported. In accordance with the results from Figure 2, one can see that the R^2 attains its maximum of 0.44 for $k \approx 0.25$. From the graph, one can clearly see the distorted view one gets when solely considering the IS R^2 when evaluating a model. For IS performance, R^2 values are much higher than for OOS R^2, and less shrinkage coincides with better performance, while this is not true when the OOS R^2 of a test set is computed. The red dotted lines indicate the ± 1 standard error (s.e.) upper and lower bounds for the OOS R^2. Note that the lower and upper bounds at first glance seem to intersect at $\kappa \approx 0.49$. Although the values come very close together around $\kappa = 0.49$, they do not actually intersect, which makes sense mathematically.

Figure 3b contains a plot of the maximum OOS R^2 for different numbers of variables included in the model. This is obtained by considering different levels of L^2 shrinkage for an SDF with a certain number of variables and choosing the maximal corresponding OOS R^2. The characteristics based model represented by the solid red line shows good OOS R^2 performance from two to three variables in the SDF onward, which corresponds with the results in Figure 2a. A PC based model already performs close to optimal with one variable in the SDF. The black cross indicates the performance of the Fama-French three factor model, which gives an OOS R^2 of 0.42. This seems to be slightly higher than the performance noted by Kozak et al. (2020), although the difference seems to be negligible.[6] In line with Kozak et al. (2020), both the Characteristic and the PC based model outperform the Fama-French three factor benchmark model. A PC based model outperforms the Fama-French three factor model even when one variable is included. Comparing the other results from Figure 3 to Kozak et al. (2020), identical model performance is found, which in turn also leads to similar conclusions.



(a) IS/OOS R^2 for different levels of L^2 penalties

(b) OOS R^2 for different levels of sparsity

Figure 3: Maximal cross sectional R^2 values for different levels of L^2 shrinkage and L^1 sparsity for the 25 Fama-French portfolios. Panel (a) shows the IS R^2 and cross-validated OOS R^2 for different root expected SR^2 values with the blue dashed line and solid red line, respectively. It poses no restrictions on sparsity. Due to the inverse relationship between κ and L^2, the x-axis indirectly represents the strength of the L^2 penalty, where more shrinkage is obtained when moving to the left along the x-axis. The red dotted lines indicate the 1 s.e. lower and upper bounds of the cross-validated OOS cross sectional R^2. Panel (b) shifts the focus from L^2 shrinkage in panel (a) towards L^1 shrinkage. That is, for all number of variables in the SDF denoted on the x-axis, the maximum OOS R^2 is projected by considering all levels of L^2 shrinkage on the y-axis. The solid red line shows the OOS R^2 values of a characteristic based model with corresponding -1 s.e. lower bound denoted by the dotted red line. The solid and dotted blue lines show these values for a PC based model. The black cross indicates the OOS performance of the Fama-French three factor model with the SMB and HML factor next to the market factor. The data used ranges from July 1926 to December 2017 and is daily.

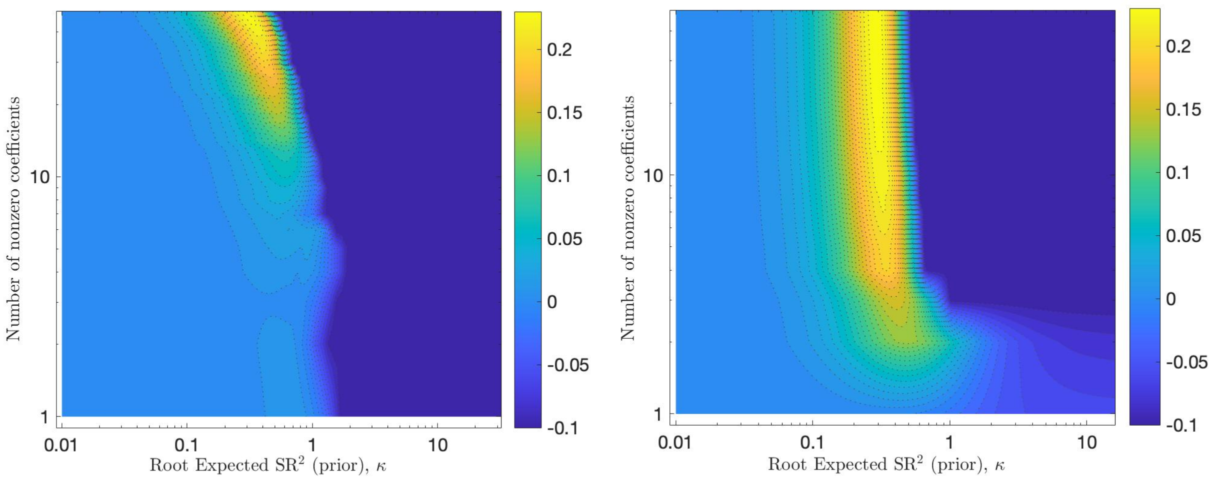
5.1.2 Anomaly portfolio

The focus is now shifted to the second data set considered, the 50 anomaly portfolios. Table 1 shows the coefficients and t-statistics of the most significant SDF coefficients. The left side of the table shows the coefficients and t-statistics of the 50 anomaly factors. The right-hand side contains coefficients and t-statistics of their PCs. The results of Table 1 are identical to the results from Kozak et al. (2020). The characteristics with the highest significance for the raw results are found in literature to be the most robust in explaining the SDF (Kozak et al., 2020). This implies that the SDF reveals the best-explaining characteristics naturally. The PC results show that low-index PCs, and hence high-variance PCs, have the highest absolute values. The first four PCs are significantly different from zero based on a 5% significance level.

[6] Kozak et al. (2020) do not mention the actual value in their paper, and this is also not directly retrievable from their code. From Figure 2b in Kozak et al. (2020), their found value seems to be around 0.415, although it is difficult to pinpoint the exact value based on the graph.

Table 1: Largest factors in the SDF for the 50 anomaly portfolios. The left hand side of the table indicates the biggest coefficients and their corresponding t-statistics for estimation based on the raw portfolios. The right hand side of the table shows the biggest coefficients and their corresponding t-statistics of the portfolios that are first transformed into their PCs.

Raw 50 anomaly portfolios			PCs of 50 anomaly portfolios		
	b	t-statistic		b	t-statistic
Industry rel. rev. (L.V.)	-0.88	3.53	PC 4	1.01	4.25
Ind. mom-reversals	0.48	1.94	PC 1	-0.54	3.08
Ind. rel. reversals	-0.43	1.70	PC 2	-0.56	2.65
Seasonality	0.32	1.29	PC 9	-0.63	2.51
Earnings surprises	0.32	1.29	PC 15	0.32	1.27
Value-profitability	0.30	1.18	PC 17	-0.30	1.18
Return on market equity	0.30	1.18	PC 6	-0.30	1.18
Investments/Assets	-0.24	0.95	PC 11	-0.19	0.74
Return on equity	0.24	0.95	PC 13	-0.17	0.65
Composite issuance	-0.24	0.95	PC 23	0.15	0.56
Momentum (12 months)	0.23	0.91	PC 7	0.14	0.56



(a) Raw 50 anomaly portfolios

(b) Principal Components of 50 anomaly portfolios

Figure 4: OOS R^2 for models with different levels of L^1 and L^2 penalties for the 50 anomaly portfolios. The x-axis shows the square root of the expected SR^2. Due to the inverse relationship between κ and L^2, the x-axis indirectly represents levels of L^2 penalties, where more shrinkage is obtained when moving to the left along the x-axis. The y-axis indicates the number of nonzero coefficients in the model, which is regulated through penalty L^1. More nonzero coefficients imply the model is less sparse and the L^1 penalty is less strong. The colour bar on the right indicates the value of the OOS R^2, where more yellow colours indicate higher values. A minimum bound for OOS R^2 in the graph is set at -0.1. Some values can be even lower, but since one can interpret all values lower than -0.1 as poor performing, and to allow for a clear comparison, a minimum bound is set at -0.1. The data used ranges from November 1973 to December 2017 and is daily.

Figure 4 contains results of the estimator based on both penalties for the 50 anomaly portfolios constructed by Kozak et al. (2020). The left-hand side of the figure contains the results of the SDF based directly on the factor returns. The right-hand side considers the SDF in the space of PCs of the factor returns. As for Figure

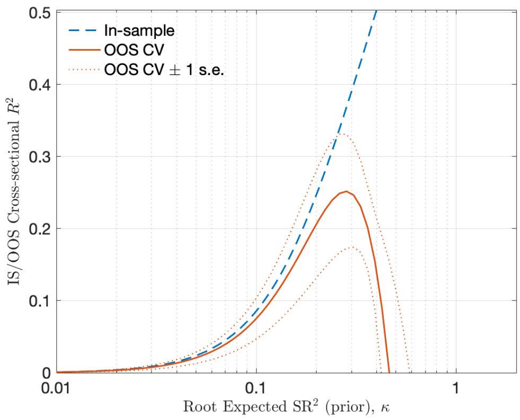
2, the x-axis indicates the root of the expected squared Sharpe Ratio of the prior, where stronger shrinkage is obtained when moving to the left. The y-axis indicates the amount of sparsity imposed by the estimator, where moving up represents less sparsity. A point on the contour plot represents the OOS R^2, calculated using Equation (15) for given penalties γ_1 and γ_2. A higher OOS R^2 value is represented by a more yellow colour, as indicated by the colour bar on the right.

Figure 4a shows OOS results for an SDF based on the raw 50 anomaly portfolios. A relatively small area at the mid top contains OOS R^2 values that are high. Models that impose no shrinkage perform extremely poor for all levels of sparsity. Non-sparse models with $\kappa \approx 0.33$ perform best, indicated by the yellow area. When sparsity is imposed for this level of shrinkage, performance decreases drastically. Compared to the 25 Fama-French portfolios in Figure 2a, similar results are found on the poor performance of completely unrestricted models. These non-shrunk, non-sparse models in the top right of the graph have poor OOS performance. This is again due to the overfitting of noise in the IS data when almost all factors are included and the absence of shrinkage to offset this overfitting. Furthermore, very restricted models perform equally poor as well. A clear difference is the absence of the interchangeability of shrinkage and sparsity to ensure good OOS performance in the 50 anomaly portfolio case. Another difference is that while Fama-French portfolios allowed for significant shrinkage and still gave good results, this is not the case for the anomaly portfolios.

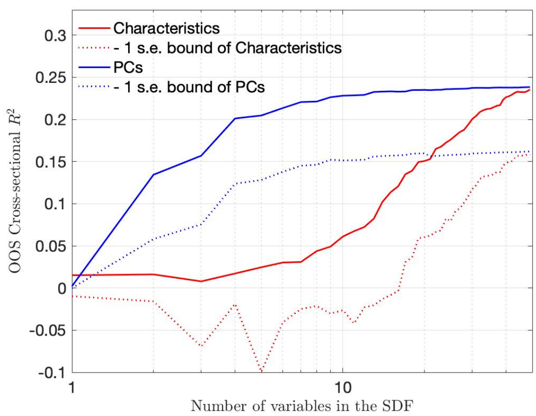
Figure 4b shows OOS results for an SDF based on the PCs of the 50 anomaly portfolios. Similar to the PC based model for the 25 Fama-French portfolios in Figure 2b, a straight area from mid top to mid bottom gives maximal OOS R^2 values. A model containing around eight PCs already gives reasonable OOS results, while this is not the case for the raw 50 anomaly portfolios. Since the yellow area is almost straight, models with a different number of variables all require a similar level of shrinkage. The same reasoning holds here as for the Fama-French portfolios, namely that the SDF coefficients of low-variance PCs are already set close to zero by the estimator $\hat{b}$. Similar to the Fama-French portfolios, non-shrunk models show extremely poor performance. However, shrinkage must be stronger compared to the Fama-French portfolios before good OOS R^2 results are obtained. Note that the colour bar in Figure 4 has a different scale than in Figure 2, where maximal OOS R^2 values are approximately twice as high. The results of Figure 4 are identical to Kozak et al. (2020), leading to similar conclusions drawn.

Figure 5 contains specific cuts from Figure 4. Figure 5a corresponds with a cut through the top of Figure 4a. That is, no sparsity is included in the model, and the R^2 for different levels of shrinkage is reported. The blue dashed line indicates the IS R^2. The OOS R^2 is indicated by the red line, and the red dotted line shows its corresponding lower and upper bounds. The maximum OOS R^2 is found around $k \approx 0.28$. In contrast to Kozak et al. (2020), the level shrinkage using $\eta = 1$ by Pastor and Stambaugh (2000), is not included in this graph. Through a rather arbitrary trial and error procedure, the same output is eventually found. Using posterior specification $\hat{b} = (\gamma\Sigma)^{-1}\bar{\mu}$ gives values that have similarities to the values from Kozak et al. (2020). However, this is a different posterior specification than the theoretically correct one using $\eta = 1$, as shown in Equation (B.1). When mirroring the output and manipulating the x-axis, eventually the same output as by Kozak et al. (2020) is obtained. However, the choices made here seem rather arbitrary and the results are obtained through a trial and error procedure. Therefore, no other conclusion can be drawn than that this part of Kozak et al. (2020) is not well replicable. For the sake of completeness, the graph and a more detailed explanation of its arbitrarity are included in Appendix B.

Figure 5b contains a plot of the maximum OOS R^2 for different numbers of variables included in the model. This is obtained by considering different levels of L^2 shrinkage for an SDF with a certain number of variables and choosing the maximal corresponding OOS R^2. The graph shows that for many included variables, a characteristics based model (solid red) and PC based model (solid blue) attain a similar OOS R^2. For sparse models, the PC model already performs well, while the characteristics model gives poor results. Compared to Kozak et al. (2020), similar results are found. However, the way they are displayed differs slightly. Whereas Kozak et al. (2020) pose a minimum bound of the y-axis at zero and thereby exclude all values smaller than 0, this graph shows all attained values. A characteristic based model with few variables in the SDF shows negative values for the -1 s.e. bound. This means it is well within the margin of error that OOS R^2 values of a sparse characteristic based model can be negative, which indicates these models potentially perform very poor. It is important to keep this in mind when considering this model, and it further indicates the relatively better performance of a PC based model.



(a) IS/OOS R^2 for different levels of L^2 penalties



(b) OOS R^2 for different levels of sparsity

Figure 5: Maximal cross sectional R^2 values for different levels of L^2 shrinkage and L^1 sparsity for the 50 anomaly portfolios. Panel (a) shows the IS R^2 and cross-validated OOS R^2 for different root expected SR^2 values with the blue dashed line and solid red line, respectively. It poses no restrictions on sparsity. Due to the inverse relationship between κ and L^2, the x-axis indirectly represents levels of L^2 penalties, where more shrinkage is obtained when moving to the left along the x-axis. The red dotted lines indicate the 1 s.e. lower and upper bounds of the cross-validated OOS R^2. Panel (b) shifts the focus from L^2 shrinkage in panel (a) towards L^1 shrinkage. That is, for all number of variables in the SDF denoted on the x-axis, the maximum OOS R^2 is projected by considering all levels of L^2 shrinkage on the y-axis. The solid red line shows the OOS R^2 values of a characteristic based model with corresponding -1 s.e. lower bound denoted by the dotted red line. The solid and dotted blue lines show these values for a PC based model. The data used ranges from November 1973 to December 2017 and is daily.

5.2 Extension

The focus is now shifted to the last data set, the interactions of the WRDS portfolios. This data set is evaluated in two different scenarios. The first case considers the interaction portfolios without covariance matrix uncertainty, which is elaborated upon in Section 5.2.1. The second scenario considers the interactions portfolios with covariance matrix estimation using the ANS estimator. These results are presented in Section 5.2.2.

5.2.1 Interactions

Table 2 shows the coefficients and t-statistics of the most significant SDF coefficients for the raw and PC-transformed interactions data set in the left part and right part of the graph, respectively. The raw interaction coefficients show that out of the 69 most significant factors, 64 are the characteristics themselves, and only five are the factors based on the interactions. This provides some foundation that the assumption of linearity in characteristics made in Section 5.1 is not a completely arbitrary one. The best performing PCs are different from the interaction data set Kozak et al. (2020). The PC results show that overall low-index PCs, and hence high-variance PCs, have the highest t-statistics. Compared to Kozak et al. (2020), a difference exists in coefficient b, although t-statistics are similar. These differences may be due to the different sample period used, due to the absence of lagged returns in the data set of this paper or due to the use of financial ratios on a firm level instead of on an industry level.

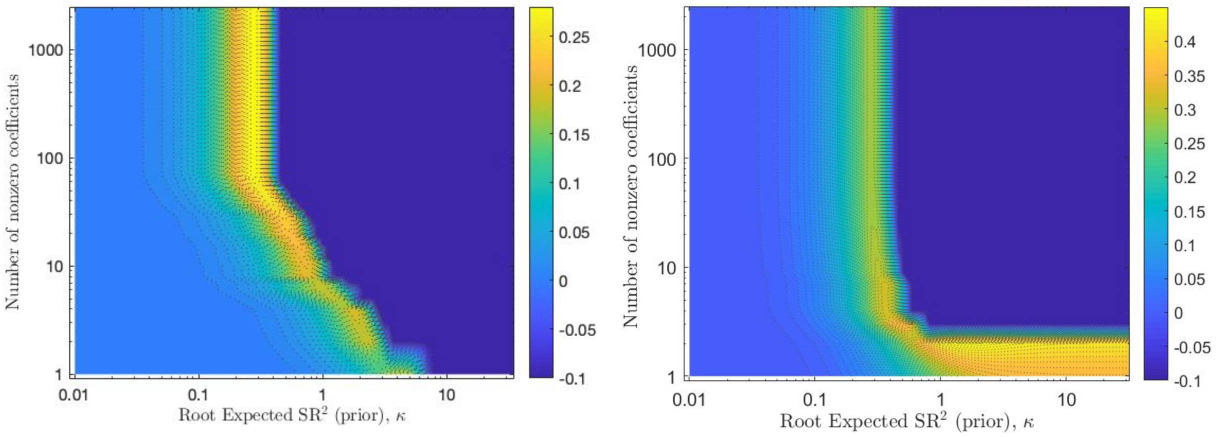
Table 2: Largest factors in the SDF for the model with interactions based on the 70 WRDS financial ratios. The left hand side of the table indicates the biggest coefficients and their corresponding t-statistics for estimation based on the raw interaction portfolios. The right hand side of the table shows the biggest coefficients and their corresponding t-statistics of the portfolios that are first transformed into their PCs.

Raw interactions			PCs of interactions		
	b	t-statistic		b	t-statistic
Price/Sales	-1.48	2.72	PC 2	-4.30	10.40
Cash flow margin	-1.46	2.66	PC 3	3.00	5.88
Profit before depreciation / current liabilities	-1.31	2.39	PC 1	1.43	4.12
Net profit margin	-1.27	2.31	PC 6	-0.98	1.83
Enterprise value multiple	-1.22	2.21	PC 4	0.93	1.77
Book/Market	1.19	2.18	PC 14	-0.84	1.52
Pre-tax profit margin	-1.17	2.12	PC 19	0.81	1.46
Operating profit margin before depreciation	-1.15	2.09	PC 8	0.78	1.44
Labor expenses/ Sales	-1.13	2.06	PC 17	-0.64	1.16
Operating cash flow/Current liabilities	-1.11	2.03	PC 24	0.60	1.09

Figure 6 contains results of the estimator based on both penalties for the interactions of the WRDS data set. As for Figure 2, the x-axis indicates the root of the expected squared Sharpe Ratio of the prior, where stronger shrinkage is obtained when moving to the left. The y-axis indicates the amount of sparsity imposed by the estimator, where moving up represents less sparsity. A point on the contour plot represents the OOS R^2, calculated using Equation (15) for given penalties γ_1 and γ_2. A higher OOS R^2 value is represented by a more yellow colour, as indicated by the colour bar on the right.

Figure 6a shows OOS results for an SDF based on the raw interactions of the WRDS data set. A small area from the mid top to the middle and consequently from the middle to the right bottom contains OOS R^2 values that are high. Non-sparse models with strong shrinkage of $\kappa \approx 0.29$ perform best. A model with around 90 variables gives high OOS R^2 values when the corresponding L^2 penalty is chosen optimally, but more sparse models perform poorly OOS. For all levels of sparsity, substantial shrinkage is needed to allow for good OOS performance. This is to offset the overfitting of noise when estimating parameters using in-sample data.

Figure 6b shows OOS results for an SDF based on the PCs of interactions of the WRDS data set. An L-shaped area from the mid top to the bottom right contains maximum OOS R^2 values. It is notable that the bottom right of the figure, that is sparse models with little shrinkage, performs best OOS. A possible explanation for this is that this data set is based on a variety of financial ratios, and not only on financial ratios that are known to explain the SDF well. This means the overfitting of noise in the in-sample estimation, which is present in the anomaly portfolios, is not as strong in this case. Therefore, less shrinkage may be needed to offset the overvalued significance of factors. The findings from Figure 6b are in line with those for the WRDS portfolios of Kozak et al. (2020), although not for the interactions of the WRDS data set. Non-sparse, non-shrunk models have poor performance, as is in line with Kozak et al. (2020). Models with both strong shrinkage and sparsity also perform relatively poor.



(a) Raw interactions of 70 WRDS portfolios

(b) Principal Components of interactions of 70 WRDS portfolios

Figure 6: OOS R^2 for models with different levels of L^1 and L^2 shrinkage for the interactions of the WRDS data set. The x-axis shows the square root of the expected SR^2. Due to the inverse relationship between κ and L^2, the x-axis indirectly represents levels of L^2 penalties, where more shrinkage is obtained when moving to the left along the x-axis. The y-axis indicates the number of nonzero coefficients in the model, which is regulated through penalty L^1. More nonzero coefficients imply the model is less sparse and the L^1 penalty is less strong. The colour bar on the right indicates the value of the OOS R^2, where more yellow colours indicate higher values. A minimum bound for OOS R^2 in the graph is set at -0.1. Some values can be even lower, but since one can interpret all values lower than -0.1 as poor performing, and to allow for a clear comparison, a minimum bound is set at -0.1. The data used ranges from February 1970 to December 2020 and is daily.

All in all, it can be concluded that this interaction data set behaves well, but different with respect to the interaction portfolio created by Kozak et al. (2020). There are several possible explanations for this. These differences may, for example, be due to a different sample period used, due to using financial ratios on a firm level instead of on an industry level. Another explanation is the absence of interactions of lagged returns and third powers in this data set of this paper. Those have been incorporated in an initial setup but led to uninterpretable results such that they have been left out. The differences between these interactions constructed for this paper and Kozak et al. (2020) could for example be to due undocumented data constructing steps that they performed.

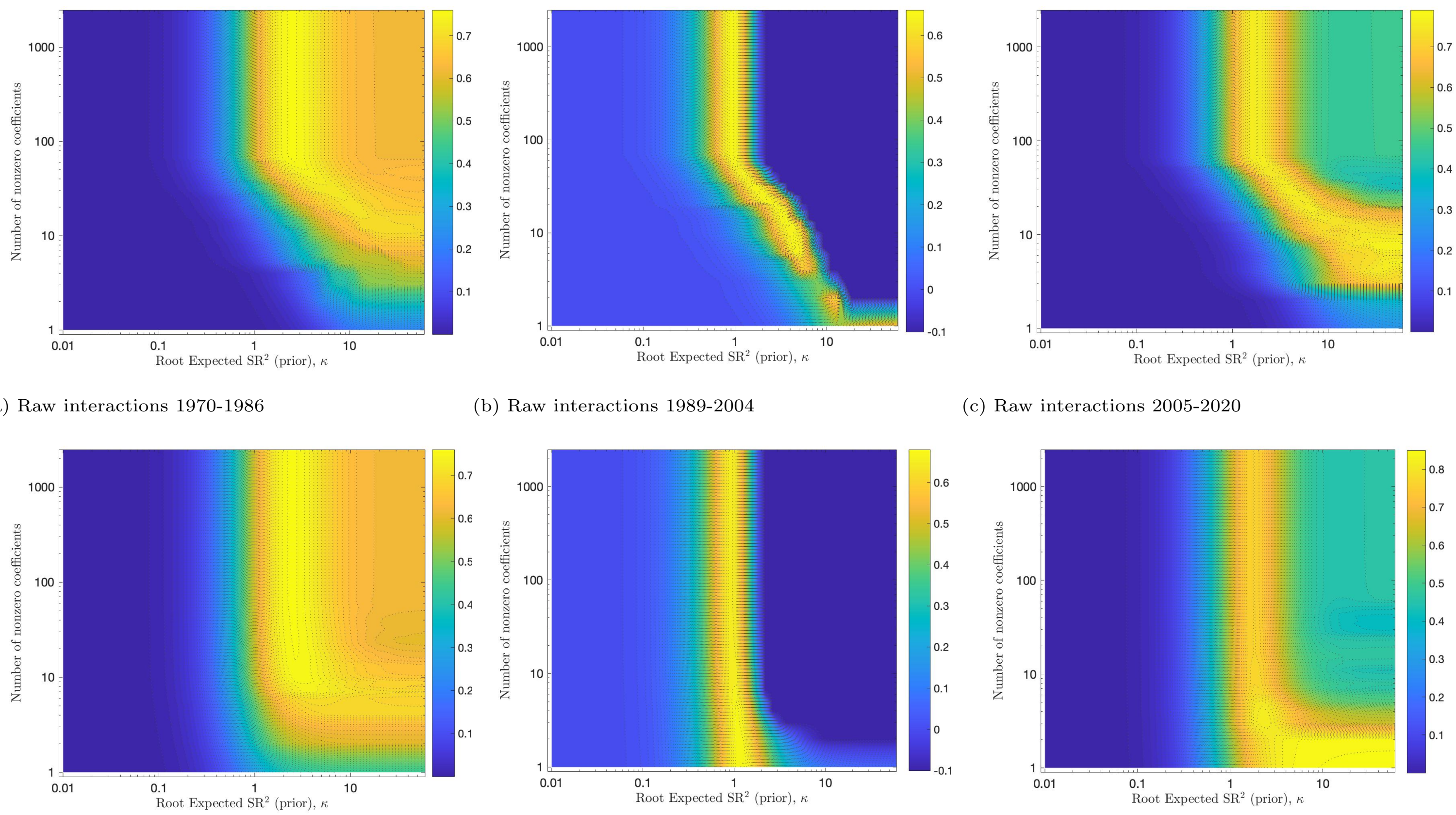
5.2.2 Interactions with covariance matrix uncertainty

Table 3 contains the coefficients and t-statistics of the most significant SDF coefficients for the interactions data set with ANS estimator to estimate the covariance matrix. The raw and PC-transformed interactions data set in the left part and right part of the graph, respectively. Compared to Table 2, all coefficients and t-statistics are higher. Higher coefficients make sense since $\hat{b}$ is now estimated using a shrunk covariance matrix. Due to the inverse relation between $\hat{b}$ and $\hat{\Sigma}$ this leads to higher $\hat{b}$ values. The higher t-statistics indicate the characteristics have a higher significance in the ANS estimator case. For the raw characteristics, eight out of ten variables are the same in Table 2 and 3. For the PC transformed variables, all PCs are the same. This indicates that the model with an ANS estimator picks the same characteristics as best explaining variables as the model without an ANS estimator.

Table 3: Largest factors in the SDF for the model with interactions and the Analytical Nonlinear Shrinkage Estimator based on the 70 WRDS financial ratios. The left hand side of the table indicates the biggest coefficients and their corresponding t-statistics for estimation based on the raw interaction portfolios. The right hand side of the table shows the biggest coefficients and their corresponding t-statistics of the portfolios that are first transformed into their PCs.

Raw interactions			PCs of interactions		
	b	t-statistic		b	t-statistic
Cash flow margin	-2.54	4.26	PC 2	-8.34	13.97
Profit before depreciation/Current liabilities	-2.43	4.07	PC 1	4.23	7.09
Dividend payout ratio	-2.39	4.00	PC 3	4.11	6.88
Pre-tax profit margin	-2.28	3.82	PC 6	-1.22	2.05
Operating profit margin before depreciation	-2.16	3.61	PC 4	1.20	2.01
Pre-tax return on total earning assets	-2.15	3.61	PC 14	-0.98	1.64
Operating profit margin after depreciation	-2.13	3.58	PC 8	0.94	1.58
Operating cash flow/Current liabilities	-2.08	3.49	PC 19	0.94	1.58
Price/Sales	2.04	3.42	PC 17	-0.75	1.25
Enterprise value multiple	-1.97	3.31	PC 24	0.70	1.17

Figure 7 contains results of the estimator for three different samples based on both penalties for the interactions of the WRDS data set. Furthermore, the ANS estimator is used to estimate the covariance matrix. Note that only the figures for the ANS estimator are included as the ones without the ANS estimator are almost identical and therefore provide no additional insights. As for Figure 2, the x-axis indicates the root of the expected squared Sharpe Ratio of the prior, where stronger shrinkage is obtained when moving to the left. The y-axis indicates the amount of sparsity imposed by the estimator, where moving up represents less sparsity. A point on the contour plot represents the OOS R^2, calculated using Equation (15) for given penalties γ_1 and γ_2. A higher OOS R^2 value is represented by a more yellow colour, as indicated by the colour bar. For all subsamples, R^2 values are much higher than for the full sample. This makes sense as the model is more specifically tailored to a given period. Figures 7a and 7d show OOS results for an SDF based on the interactions of the WRDS data set from 1970 until 1986. Figure 7a contains the raw interactions and shows an area from the mid top to the middle containing high OOS R^2 values. Non-sparse models with $\kappa = 2.61$ perform best. Models with a sparse amount of factors perform poorly, as is in line with expectations for the raw interaction set. Models with both strong shrinkage and sparsity have poor performance. Figure 7d contains the PCs of interactions and shows an area from the mid top to the mid bottom containing high OOS R^2 values. Compared to Figure 7a, the PC based model performs better with more sparsity imposed, as is expected. Strongly shrunk models again perform poorly. As opposed to the findings in Kozak et al. (2020), non-shrunk, non-sparse models still perform relatively well in this sample period. A possible explanation for this is that the overfitting of noise in the in-sample estimation is not as strong for this sample size. Therefore, less shrinkage may be needed to offset the overvalued significance of factors. Another possible explanation is that there is a relatively large amount of missing values in this early sample. Especially when considering interactions, the number of ill-behaved characteristics also increases exponentially. All in all, the SDF does not perform as expected for the sample of 1970-1986 in the non-shrunk case. However, when comparing the raw and PC based model, performance is as expected.



(a) Raw interactions 1970-1986

(b) Raw interactions 1989-2004

(c) Raw interactions 2005-2020

(d) PCs 1970-1986

(e) PCs 1989-2004

(f) PCs 2005-2020

Figure 7: OOS R^2 for models with different levels of L^1 and L^2 shrinkage for the interactions of the WRDS data set where the covariance matrix is estimated using the ANS Estimator. Three different, non-overlapping periods are used to evaluate the models. The left two figures, Figure 7a and 7d, contain the models evaluated based on the sample from 1970 until 1986. Figures 7b and 7e are based on the sample from 1989 until 2004. Figures 7c and 7f are based on the 2005-2020 sample. The top figures, Figure 7a, 7b and 7c, are based on the raw characteristics of the interactions. Figure 7d, 7e and 7f are the interactions first transformed into their PCs. The x-axis shows the square root of the expected SR^2. Due to the inverse relationship between κ and L^2, the x-axis indirectly represents levels of L^2 penalties, where more shrinkage is obtained when moving to the left along the x-axis. The y-axis indicates the number of nonzero coefficients in the model, which is regulated through penalty L^1. More nonzero coefficients imply the model is less sparse and the L^1 penalty is less strong. The colour bar on the right indicates the value of the OOS R^2, where more yellow colours indicate higher values. Note that the scale for some graphs differs from each other. A minimum bound for OOS R^2 in the graph is set at -0.1. Some values can be even lower, but since one can interpret all values lower than -0.1 as poor performing, and to allow for a clear comparison, a minimum bound is set at -0.1.

Figures 7b and 7e show OOS results for an SDF based on the interactions of the WRDS data set from 1989 until 2004. Figure 7b contains the raw interactions and shows an area from the mid top to the middle and from the middle to the right bottom containing high OOS R^2 values. This model performs similar to the model in Figure 6a. That is, models with more than 100 coefficients must all be shrunk in a similar fashion and perform well. From the diagonal slope of the high OOS R^2 for models with eight to 50 characteristics, it can be concluded that shrinkage and sparsity are interchangeable to ensure good OOS performance. Models with strong shrinkage have poor performance. Figure 7e contains the PCs of interactions and shows an area from the mid top to mid bottom where OOS R^2 are maximized. This model performs very well when sparsity is imposed, as one PC already explains the SDF almost optimally. For all number of variables included, a similar shrinkage is needed corresponding to $\kappa \approx 0.9$. Non-shrunk and non-sparse models perform poorly OOS for both models from 1989-2004. The models for this sub-sample perform very similar to the interaction results from Kozak et al. (2020) and also perform along the lines of expectations. Furthermore, it behaves in accordance with theoretical expectations.[7] Therefore, it can be concluded that the model performs very well for this sample period.

Figures 7c and 7f show OOS results for an SDF based on the interactions of the WRDS data set from 2005 until 2020. Figure 7c contains the raw interactions and shows an area from the mid top to the middle, and from the middle to the right bottom containing high OOS R^2 values. This model performs in a similar fashion to the model in the sample 1989-2004 regarding sparsity and shrinkage combinations that lead to a maximum OOS R^2. For L^1 and L^2 penalty combinations diverging from the optimal ones, values differ. That is, models with no L^1 and L^2 penalties perform relatively well. Figure 7f contains the PCs of interactions and shows an L-shaped area from the mid top to the right bottom where OOS R^2 are high. Especially non-shrunk models with a small number of variables explain the SDF well. For more than ten variables included, similar shrinkage is needed. Strong shrinkage gives poor performance for both models from 2005-2020. All in all, the models for this subsample perform better than the subsample from 1970-1986 but still not completely as expected.

Overall, the OOS performance of the three subsamples is very different. For the first period, non-restricted models already perform relatively well, but shrinkage does improve the OOS performance. In contrast, the second period requires substantial shrinkage before good OOS performance. The last period has similarities to the first period, as non-restricted models perform reasonably. However, the performance of non-shrunk models is better in the first period. For the 1989-2004 period, the dual-shrinkage estimator of Kozak et al. (2020) performs most along the lines of expectations. A possible explanation for this performance is the economic conditions during this period. Internet companies went through a period of massive growth and were highly valued. In 2000, these same companies collapsed. With one group of quickly growing companies, a clear characteristic-based estimation can be made for cross sectional returns in this period. This would mean that this period is best described by any characteristic based SDF, not just the one by Kozak et al. (2020). With the great financial crisis and the COVID-19 crisis, the more recent period might be more difficult to describe, as almost all companies suffered from this in 2007 and 2020 to a strong degree and not just companies with specific characteristics.

Table 4 contains the optimal output parameters for the three subsamples with and without the ANS estimator for covariance matrix uncertainty. Based on these values, using the ANS estimator does not seem to have a large impact on model performance. The kappa values are higher for the first and last period and similar for the second period. This indicates that the model with ANS estimator shrinks parameters less strong in the two periods that are less good behaved based on Figure 7. Sharpe Ratios differ slightly between the three different periods, but no consistent conclusion can be drawn from this. OOS R^2 values are equal for both models. When comparing the three different periods, it can be seen that the 1970-1986 period shrinks parameters least. 2005-2020 also imposes little shrinkage on the SDF. The middle period, from 1989 until 2004, shrinks the parameters the strongest. The difference in shrinkage could be due to the smaller IS bias of parameters in the first and last periods. Sharpe Ratios and R^2 values are much higher in the first and last period than in the first period. Overall, there is no evidence for the better performance of the SDF model with the incorporation of the ANS estimator.

[7] See Figure 7b and 7d of Kozak et al. (2020)

Table 4: Output values for the SDF model with interactions of the 70 WRDS financial ratios. The left part of the graph contains the information for an SDF without ANS estimator. The right hand side is an SDF using the ANS estimator for covariance matrix estimation. Different rows correspond to different sample periods.

		No ANS estimator			ANS estimator		
		Kappa	Sharpe Ratio	R^2	Kappa	Sharpe Ratio	R^2
Sample Period	1970-1986	2.42	7.69	0.76	2.61	7.74	0.76
	1990-2004	0.91	1.98	0.66	0.89	1.98	0.66
	2005-2020	1.70	6.48	0.77	1.83	6.46	0.77

Table 5 contains the coefficients and t-statistics of the most significant SDF coefficients for the raw interactions data set with ANS estimator to estimate the covariance matrix. The 1970-1986, 1989-2004 and 2005-2020 samples are shown in the table's left, middle, and right column, respectively. Compared to the scenario without ANS estimator, the coefficients are much higher. This makes sense as the coefficients are inversely related to the estimated covariance matrix, which is now shrunken with the ANS estimator. In the three subsamples, where the dimension of the covariance matrix is relatively big compared to the sample size, this effect is stronger and coefficients are higher than in the full sample case. Comparing the characteristics between periods, very few characteristics are shown to be in common. The 1970-1986 period has some in common with the 1989-2004 period, and the 1970-1986 period has some in common with the period from 2005 to 2020. The last two periods have none in common. From these differences, it can be concluded that the different subperiods have different characteristics that best describe the cross sectional return. In line with the strongest shrinkage for the 1989-2004 period in Table 4, the coefficients of that period are the smallest.

Table 5: Largest factors in the SDF with ANS estimator for the interactions of 70 WRDS portfolios. The table contains three columns corresponding to the three sub periods. The table indicates the biggest coefficients and their corresponding t-statistics for estimation based on the raw portfolios.

1970-1986			1990-2004			2005-2020		
	b	t-statistic		b	t-statistic		b	t-statistic
Price/Sales	-374.95	33.64	Cash ratio	-7.05	5.50	Pre-tax return on total earning assets	-91.94	23.58
Book/Market	363.72	32.64	Sales/Working capital	6.81	5.31	Profit before depreciation/current liabilities	-90.35	23.17
Cash flow margin	-357.07	32.04	Cash balance/Total liabilities	-6.64	5.18	Operating cash flow/Current liabilities	-88.04	22.58
Net profit margin	-350.54	31.45	Quick ratio (acid test)	-6.53	5.09	Operating profit margin after depreciation	-76.27	19.56
Pre-tax profit margin	-334.15	29.98	Current ratio	-5.94	4.63	Return on assets	-74.85	19.19
Shillers cyclically adjusted P/E ratio	-329.85	29.60	Price/Sales	-5.64	4.40	Operating profit margin before depreciation	-74.46	19.09
P/E (diluted, inc. extr. items)	-328.62	29.49	Total debt / EBITDA	5.59	4.36	Return on equity	-74.17	19.02
Price / Book	-325.80	29.23	Research and development/Sales	-5.53	4.31	After-tax return on invested capital	-74.02	18.98
After-tax return on total stockholders equity	-325.03	29.16	Total debt/Total assets	5.40	4.22	Pre-tax profit margin	-73.69	18.90
After-tax return on average common equity	-320.93	28.80	Total debt/Equity	5.18	4.04	Pre-tax return on net operating assets	-73.32	18.80

Table 6 contains the coefficients and t-statistics of the most significant SDF coefficients for the raw interactions data set without ANS estimator. The 1970-1986, 1989-2004 and 2005-2020 samples are shown in the table's left, middle, and right column, respectively. The characteristics "Enterprise value multiple" and "Quick ratio (acid test)" reappear in different periods. The other best-explaining characteristics are all different.

Table 6: Largest factors in the SDF without ANS estimator for the interactions of 70 WRDS portfolios. The table contains three columns corresponding to the three sub periods. The table indicates the biggest coefficients and their corresponding t-statistics for estimation based on the raw portfolios.

1970-1986			1990-2004			2005-2020		
	b	t-statistic		b	t-statistic		b	t-statistic
Labor expenses/Sales	-90.72	10.52	Sales/Stockholders equity	1.64	1.28	Operating cash flow/Current liabilities	-16.53	4.88
Free cash flow/Operating cash flow	70.98	8.37	Research and development/Sales	-1.53	1.20	Pre-tax return on total earning assets	-11.72	3.42
Quick ratio (acid test)	54.47	5.96	Cash ratio	-1.50	1.18	Enterprise value multiple	-10.79	3.26
Cash balance/Total liabilities	51.83	5.55	Price/Sales	-1.45	1.15	Pre-tax return on net operating assets	-10.27	3.06
Enterprise value multiple	-45.71	5.41	Sales/Invested capital	1.43	1.11	Cash flow/ Total debt	-10.12	2.99
Trailing P/E to growth (PEG) ratio	-43.95	5.27	Asset turnover	1.42	1.10	Profit before depreciation/current liabilities	-10.03	2.93
P/E (diluted, including extraordinary items)	-44.78	5.09	Sales/Working capital	1.27	1.00	Total debt / EBITDA	-9.11	2.69
Effective tax rate	45.39	5.06	Quick ratio (acid test)	-1.22	0.96	Current liabilities/Total liabilities	8.40	2.49
Cash flow margin	-48.04	5.05	Total debt/Total assets	1.14	0.89	Operating profit margin before depreciation	-8.52	2.48
Accruals/Average assets	40.72	4.55	Operating profit margin after depreciation	1.14	0.89	Total debt/Equity	-8.41	2.47

Comparing Table 5 to Table 6, the first period has only two best explaining characteristics that overlap between the models with and without ANS estimator. The second and third period have five overlapping characteristics. From this, it can be concluded that a substantial difference exists in best explaining characteristics for the case with and without ANS estimator. Appendix E contains the tables of the PC based models with and without ANS estimator.

6 Conclusion and discussion

This paper's goal is to identify how the stochastic discount factor (SDF) framework of Kozak et al. (2020) performs when allowing for covariance matrix uncertainty. The motivation for this research comes from the relatively poor estimation results of sample covariance matrices when considering high dimensional data sets (Ledoit and Wolf, 2020c). This motivation leads to the following research question:

"To what extent does the allowance of covariance matrix uncertainty influence the performance of a characteristic based stochastic discount factor in a high-dimensional framework?"

As an initial test, the Fama-French 25 portfolios and the 50 anomaly portfolios are considered. For the raw characteristics of the Fama-French 25 portfolios, two to three characteristics already describe the SDF well. From the diagonal slope of the high OOS R^2, it can be concluded that shrinkage and sparsity are interchangeable to ensure good OOS performance. In the PC transformed case, one factor already explains the SDF almost optimally. For the raw characteristics of the 50 anomaly portfolios, substantial shrinkage is needed for all sparsity levels, and sparse models perform poorly out of sample. In contrast, the PC transformed 50 anomaly characteristics do allow for a relatively sparse SDF representation.

Next, an interaction data set is considered, which ensures a high dimensional covariance matrix. For the entire sample period, that is, from 1970 until 2020, the interaction data set behaves well, although different from the interaction portfolio created by Kozak et al. (2020). For the interactions set as given, substantial shrinkage is needed to allow for good OOS performance, and sparse models do not perform well. This is in line with the 50 anomaly portfolio. In the case of a PC transformed interactions set, high out-of-sample performance can be attained with a sparse model and little shrinkage, and hence a sparse representation of the SDF exists. Implementing the Analytical Nonlinear Shrinkage (ANS) estimator of Ledoit and Wolf (2020a) into the SDF framework, no evidence for better performance of the SDF is found.

Consequently, the sample is split up into three non-overlapping periods of equal length, such that the sample corresponds in size to the dimensionality of the covariance matrix. Comparing the OOS performance of these three subsamples, very different results are found. Similar results for the 1970-1986 and 2005-2020 periods are found, namely that non-restricted models already perform relatively well, but shrinkage does improve the OOS performance. In contrast, the 1989-2004 period requires substantial shrinkage before good OOS performance. For the 1989-2004 period, the dual-shrinkage estimator of Kozak et al. (2020) performs most along the lines of expectations. A likely explanation for this is that this period contains abnormal returns due to industry-specific boom or bust periods. Implementing the ANS estimator in the SDF framework logically leads to different coefficients, but OOS performance is influenced little, also in the three smaller subsamples.

All in all, it must therefore be concluded that there is no evidence that the ANS estimator affects the performance of the SDF in a high-dimensional framework. Furthermore, the SDF performs best for periods in which specific industries have abnormal returns.

The main limitation of this research lies within the comparability of the constructed interaction data set. For example, lagged returns and third powers are excluded from this research, as the inclusion of these interactions led to uninterpretable results. This might be due to a difference in the construction of these interactions between this paper and Kozak et al. (2020). Furthermore, the current data range allowed by WRDS differs from the data range considered by Kozak et al. (2020). For these reasons, a direct comparison between the interaction data set in this paper and Kozak et al. (2020) stays difficult. Therefore, a first suggestion for future research is to replicate this paper with the exact interaction data set as constructed by Kozak et al. (2020). That way, it becomes easier to compare the papers and pinpoint more specific areas of improvement for this research within the data construction or within the methodology.

Another limitation of this research is the sole use of static covariance matrix estimation. Shifting the focus to dynamic covariance matrix estimators instead of static estimators could better capture the time-varying component of volatility. For example, the Dynamic Conditional Correlation Non-Linear (DCC-NL) method of Engle et al. (2019) can be considered as a suggestion for future research. It has to be noted that when daily data is used to forecast the covariance matrices but these forecasts are used one month ahead, a mismatch exists for this dynamic model, as explained by De Nard et al. (2019). Therefore, the three-step procedure of De Nard et al. (2019) should be followed when one considers implementing the DCC-NL method.

A last suggestion is to research whether other values for the global bandwidth of the ANS estimator provide different results. The assumptions made regarding the global bandwidth in Ledoit and Wolf (2020a) are rather arbitrary and are chosen only as they coincide with existing literature. Choosing other values could provide insightful results.

References

Abadir, K., Distaso, W., & Žikeš, F. (2014). Design-free estimation of variance matrices. *Journal of Econometrics*, *181*, 165–180.

Back, K. E. (2017). *Asset pricing and portfolio choice theory.* Oxford University Press.

Cochrane, J. (2011). Presidential address: Discount rates. *The Journal of Finance*, *66*(4), 1047–1108.

Cochrane, J. H. (2001). *Asset pricing.* Princeton University Press.

De Nard, G., Ledoit, O., & Wolf, M. (2019). Factor models for portfolio selection in large dimensions: The good, the better and the ugly. *Journal of Financial Econometrics*, 1–22.

DeMiguel, V., Martín-Utrera, A., Nogales, F. J., & Uppal, R. (2020). A transaction-cost perspective on the multitude of firm characteristics. *The Review of Financial Studies*, *33*(5), 2180–2222.

Engle, R. F., Ledoit, O., & Wolf, M. (2019). Large dynamic covariance matrices. *Journal of Business & Economic Statistics*, *37*(2), 363–375.

Epanechnikov, V. (1967). Non-parametric estimation of a multivariate probability density. *Theory of Probability & Its Applications*, *14*(1), 153–158.

Fama, E., & French, K. (1996). Multifactor explanations of asset pricing anomalies. *The Journal of Finance*, *51*(1), 55–84.

Fama, E., & French, K. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, *116*(1), 1–22.

Fama, E., & MacBeth, J. (1973). Risk, return, and equilibrium: Empirical tests. *The Journal of Political Economy*, *81*(3), 607–636.

Feng, G., Giglio, S., & Xiu, D. (2020). Taming the factor zoo: A test of new factors. *The Journal Of Finance*, *75*(3), 1327–1370.

French, K. R. (2020). https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html#Research

Freyberger, J., & Weber, A. N. M. (2020). Dissecting characteristics nonparametrically. *The Review of Financial Studies*, *33*(5), 2326–2377.

Grith, M. (2021). *Notes on shrinking the cross-section by Kozak, S., Nagel, S., Santosh, S. (2020).*

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, *33*(5), 2223–2273.

Hansen, L. P., & Jagannathan, R. (1991). Implications of security market data for models of dynamic economies. *Journal of Political Economy*, *99*(2), 225–262.

Harvey, C., Liechty, J., & Liechty, M. (2008). Bayes vs. resampling: A rematch. *Journal of Investment Management*, *6*(1), 1–17.

Harvey, C., Liu, Y., & Zhu, H. (2016). ... and the cross-section of expected returns. *The Review of Financial Studies*, *29*(1), 5–68.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55–67.

Hou, K., Xue, C., & Zhang, L. (2015). Digesting anomalies: An investment approach. *The Review of Financial Studies*, *28*(3), 650–705.

Jing, B.-Y., Pan, G., Shao, Q.-M., & Zhou, W. (2010). Nonparametric estimate of spectral density functions of sample covariance matrices: A first step. *The Annals of Statistics*, *38*(6), 3724–3750.

Kelly, B., Pruitt, S., & Su, Y. (2019). Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics*, *134*(3), 501–524.

Kozak, S. (2020). Data and code supplementary to shrinking the cross-section. https://www.serhiykozak.com/data

Kozak, S., Nagel, S., & Santosh, S. (2018). Interpreting factor models. *The Journal of Finance*, *73*(3), 1183–1233.

Kozak, S., Nagel, S., & Santosh, S. (2019). Internet appendix to shrinking the cross-section.

Kozak, S., Nagel, S., & Santosh, S. (2020). Shrinking the cross-section. *Journal of Financial Economics*, *135*(2), 271–292.

Lam, C. (2016). Nonparametric eigenvalue-regularized precision or covariance matrix estimator. *The Annals of Statistics*, *44*(3), 928–953.

Ledoit, O., & Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, *88*(2), 365–411.

Ledoit, O., & Wolf, M. (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. *The Annals of Statistics*, *40*(2), 1024–1060.

Ledoit, O., & Wolf, M. (2015). Spectrum estimation: A unified framework for covariance matrix estimation and pca in large dimensions. *Journal of Multivariate Analysis*, *139*, 360–384.

Ledoit, O., & Wolf, M. (2017). Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets goldilocks. *Working Paper No. 137*.

Ledoit, O., & Wolf, M. (2020a). Analytical nonlinear shrinkage of large-dimensional covariance matrices. *The Annals of Statistics*, *48*(5), 3043–3065.

Ledoit, O., & Wolf, M. (2020b). Code of the analytical nonlinear shrinkage estimator. https://www.econ.uzh.ch/en/people/faculty/wolf/publications.html#Programming_Code

Ledoit, O., & Wolf, M. (2020c). The power of (non-)linear shrinking: A review and guide to covariance matrix estimation. *Journal of Financial Econometrics*, 1–32.

Lewellen, J., Nagel, S., & Shanken, J. (2010). A skeptical appraisal of asset pricing tests. *Journal of Financial Economics*, *96*(2), 175–194.

Lin, X., & Zhang, L. (2013). The investment manifesto. *Journal of Monetary Economics*, *60*(3), 351–366.

Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *The Review of Economics and Statistics*, *47*(1), 13–37.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, *7*(1), 77–91.

Pastor, L. (2000). Portfolio selection and asset pricing models. *The Journal of Finance*, *55*(1), 179–223.

Pastor, L., & Stambaugh, R. (2000). Comparing asset pricing models: An investment perspective. *Journal of Financial Economics*, *56*(3), 335–381.

Rapach, D. E., Strauss, J. K., & Zhou, G. (2013). International stock return predictability: What is the role of the united states? *The Journal of Finance*, *68*(4), 1633–1662.

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

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

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *67*(2), 301–320.

A Proofs and derivations

This appendix includes proofs and derivations that are used in the paper.

A.1 SDF coefficients with population moments

The derivation of $b = \Sigma^{-1}\mathbb{E}(F_t)$ makes use of the notes graciously provided by Grith (2021). Combining the fact that $\mathbb{E}_{t-1}[M_t R_t^*] = 0$, as stated in Equation (3) with $M_t = 1 - b'(F_t - \mathbb{E}F_t)$ in Equation (7), and rewriting Equation (5) to $R_t = (Z'_{t-1}Z_{t-1})^{-1}Z'_{t-1}F_t$, it holds that

$$\mathbb{E}_{t-1}\Big([1 - b'(F_t - \mathbb{E}F_t)]F_t\Big) = 0 \tag{A.1}$$

This equation can be rewritten as follows:

$$\begin{aligned}
&\mathbb{E}_{t-1}\Big(1 - b'(F_t - \mathbb{E}F_t)(F_t - \mathbb{E}F_t)\Big) + \mathbb{E}_{t-1}\Big(M_t\mathbb{E}_{t-1}(F_t)\Big) = 0\\
&\mathbb{E}_{t-1}\Big(1 - b'(F_t - \mathbb{E}F_t)(F_t - \mathbb{E}F_t)\Big) + \mathbb{E}_{t-1}(M_t)\mathbb{E}_{t-1}(F_t) = 0\\
&\mathbb{E}_{t-1}\Big((F_t - \mathbb{E}F_t) - b'(F_t - \mathbb{E}F_t)(F_t - \mathbb{E}F_t)\Big) + \mathbb{E}_{t-1}(F_t) = 0\\
&\mathbb{E}_{t-1}\Big(F_t - b'(F_t - \mathbb{E}F_t)(F_t - \mathbb{E}F_t)\Big) = 0\\
&\mathbb{E}_{t-1}(F_t) = b'(F_t - \mathbb{E}F_t)(F_t - \mathbb{E}F_t)\\
&\mu = b'Cov_t(F_tF_t)\\
&\mu = \Sigma_F b
\end{aligned}$$

This eventually leads to the formula as it is written as in the paper:

$$b = \Sigma^{-1}\mathbb{E}(F_t), \tag{A.2}$$

A.2 Derivation of the Kernel density and Hilbert transform

This section provides a more in-depth derivation of the Kernel estimator for the eigenvalue density and its corresponding Hilbert transform as by Ledoit and Wolf (2020a). The Hilbert transform is defined as a convolution with the Cauchy kernel $\frac{dt}{\pi t}$, and can for a real function g be defined as

$$\mathcal{H}_g(x) = \frac{1}{\pi}\text{PV}\int_{-\infty}^{+\infty} g(t)\frac{dt}{t-x} \quad \forall x \in \mathbb{R}, \tag{A.3}$$

where PV is defined as the Cauchy principal value. This Cauchy principal value evaluates the improper integral as follows:

$$\text{PV}\int_{-\infty}^{+\infty} g(t)\frac{dt}{t-x} = \lim_{\varepsilon\to 0^+}\left[\int_{-\infty}^{x-\varepsilon} g(t)\frac{dt}{t-x} + \int_{x+\varepsilon}^{+\infty} g(t)\frac{dt}{t-x}\right], \tag{A.4}$$

where ε is defined as the error term corresponding to $g(x)$. Next to the Hilbert transform, the kernel estimator of the sample spectral p.d.f. f is defined as follows:

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_i}k\left(\frac{x-\lambda_i}{h_i}\right) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\lambda_i h_T}k\left(\frac{x-\lambda_i}{\lambda_i h}\right) \quad \forall x \in \mathbb{R}, \tag{A.5}$$

where n is the amount of characteristics included, λ_i is the i^{th} eigenvalue, h_j is the locally adaptive bandwidth, h is the global bandwidth, and $k(x)$ the kernel distribution of x. The kernel estimator of its corresponding Hilbert transform, $\mathcal{H}_f$, is then defined as

$$\mathcal{H}_f(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_i}\mathcal{H}_k\left(\frac{x-\lambda_i}{h_i}\right) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\lambda_i h}\mathcal{H}_k\left(\frac{x-\lambda_i}{\lambda_i h_n}\right) = \frac{1}{\pi}\text{PV}\int\frac{\hat{f}(t)}{x-t}dt, \tag{A.6}$$

where $\mathcal{H}_k$ is defined as in Equation (A.3), PV is defined as in Equation (A.4) and all other variables are defined as in Equation (A.5).

In this paper, the kernel as originally introduced by Epanechnikov (1967) is considered. This kernel with unit variance and support $[-\sqrt{5}, \sqrt{5}]$ has the following density function:

$$\kappa^E(x) = \frac{3}{4\sqrt{5}}\left[1 - \frac{x^2}{5}\right]^+ \quad \forall x \in \mathbb{R}, \tag{A.7}$$

where $[x]^+$ is again defined as a real number's positive part. This is the default kernel used in many univariate density estimation frameworks. Its corresponding Hilbert transform is defined by Ledoit and Wolf (2020a) as follows:

$$\mathcal{H}_{\kappa^E}(x) = \begin{cases} -\frac{3x}{10\pi} + \frac{3}{4\sqrt{5}\pi}\left(1 - \frac{x^2}{5}\right)\log\left|\frac{\sqrt{5}-x}{\sqrt{5}+x}\right| & \text{if } |x| \neq \sqrt{5} \\ -\frac{3x}{10\pi} & \text{if } |x| = \sqrt{5} \end{cases} \quad \forall x \in \mathbb{R}. \tag{A.8}$$

From this, the eigenvalue density and its corresponding Hilbert transformation can be deducted for all characteristics $i = 1, \ldots, n$ along the lines of Ledoit and Wolf (2020a). These are stated in Equation (20) and Equation (21) of Section 3.3, respectively.

B Pastor specification

This appendix elaborates upon the calculation of the cross-sectional R^2 when using the prior specification of Pastor and Stambaugh (2000), that is $\eta = 1$. In the spirit of Kozak et al. (2020), the posterior mean of b is calculated using the standard formula for a conjugate multivariate normal prior with known covariance matrix. This boils down to the formula $\hat{\mu} = (\Sigma_0^{-1} + T\Sigma^{-1})^{-1}(\Sigma_0^{-1}\mu_0 + T\Sigma^{-1}\bar{\mu})^{-1}$. Here, the prior parameters equal $\mu_0 = 0$ and $\Sigma_0 = \frac{k^2}{\tau}\Sigma^{\eta}$. Plugging in these prior parameters gives $\hat{\mu} = (I + \gamma\Sigma^{1-\eta})^{-1}\bar{\mu}$. For the prior specification of Pastor and Stambaugh (2000), $\eta = 1$, and using $\hat{b} = \Sigma^{-1}\hat{\mu}$, this simplifies to

$$\hat{b} = (\Sigma + \gamma\Sigma)^{-1}\bar{\mu}, \tag{B.1}$$

where $\gamma = \frac{\tau}{k^2 T}$. Simply changing the estimator $\hat{b}$ to its new specification does not lead to the correct OOS R^2. Through a rather arbitrary trial and error procedure, eventually the same line as by Kozak et al. (2020) is found. $\hat{b} = (\gamma\Sigma)^{-1}\bar{\mu}$ gives values that seemingly have similarities to Kozak et al. (2020). However, this is a different posterior specification than the theoretically correct one in Equation (B.1). When mirroring the output and manipulating the x-axis, eventually the same output as by Kozak et al. (2020) is obtained. The choices made here seem rather arbitrary, and the results are obtained through a trial and error procedure. Therefore, no other conclusion can be drawn than this part of Kozak et al. (2020) being unreplicable.

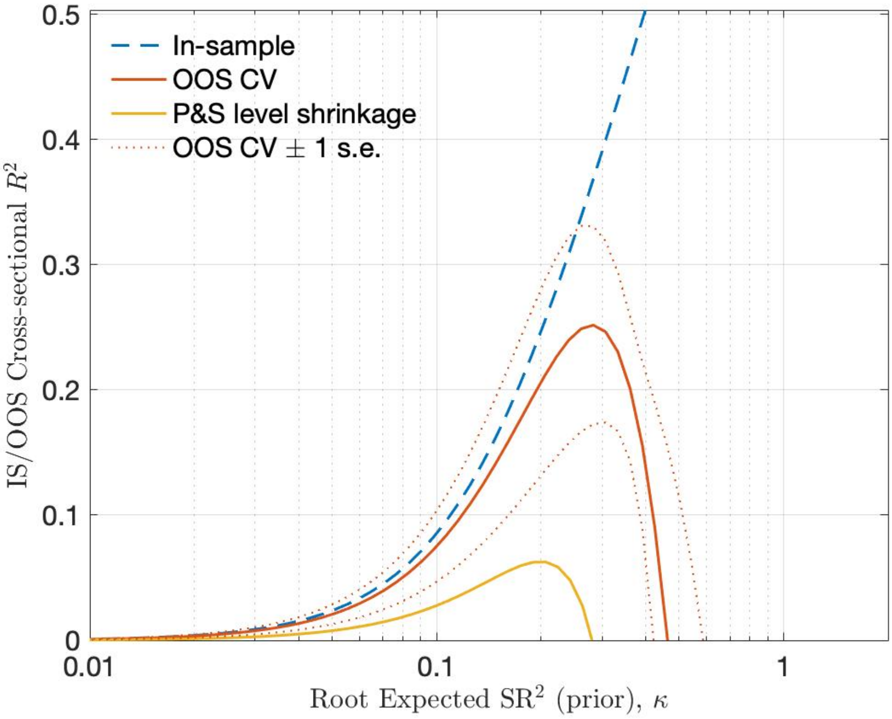


Figure B1: IS and cross-validated OOS R^2 for different root expected SR^2 values with the blue dashed line and solid red line, respectively. It poses no restrictions on sparsity. The red dotted lines indicate the 1 s.e. lower and upper bounds of the cross validation OOS cross-sectional R^2. The yellow line indicates the cross-sectional R^2 when using the prior specification of Pastor and Stambaugh (2000).For the yellow line, the inverse relationship between κ and L^2 does not hold.

Kozak et al. (2020) do mention that due the the different prior specification $\eta = 1$, the relation $\mathbb{E}(SR^2) = \kappa$ no longer holds. This could explain the shift of the graph along the x-axis but does not explain the alternative posterior specification and need for mirroring. For the sake of completeness, Figure B1 contains the specification of Pastor and Stambaugh (2000) in addition to the values already explained in Figure 5a of Section 5.1.

C 50 anomalies variable description overview

This appendix contains the variables included in the data set provided by Kozak (2020) and their corresponding annualized mean returns (in %) for all characteristics in the full sample.

Table C1: Anomaly characteristics used in the replication part of this paper. These characteristics are taken from Kozak et al. (2020). Next to the characteristics, the annualized mean returns (in %) on managed portfolios are shown for all characteristics in the full sample. All returns are based on a monthly-rebalanced buy-and-hold strategy, annualized assuming 252 trading days a year, and are further rescaled to have normalized standard deviations. The data used ranges from November 1973 to December 2017 and is daily.

Characteristic	Mean annualized return (in %)
1. Size	-2,29
2. Value (annual)	6,21
3. Gross profitability	3,64
4. Value-profitablity	13,23
5. Piotroski's F-score	8,15
6. Debt issuance	1,77
7. Share repurchases	6,93
8. Net share issuance (annual)	-9,48
9. Accruals	-5,59
10. Asset growth	-8,57
11. Asset turnover	5,30
12. Gross margins	-1,15
13. Dividend/Price	3,55
14. Earnings/Price	8,28
15. Cash flows/ Market value of equity	7,91
16. Net operating assets	1,90
17. Investment/Assets	-9,96
18. Investment/Capital	-4,11
19. Investment growth	-9,00
20. Sales growth	-5,71
21. Leverage	4,87
22. Return on assets (annual)	2,43
23. Return on book equity (annual)	4,72
24. Sales/Price	9,39
25. Growth in long term net operating assets (LTNOA)	-2,53
26. Momentum (6 months)	2,10
27. Industry momentum	5,63
28. Value-momentum	5,12
29. Value-momentum-profitability	6,55
30. Short interest	0,28
31. Momentum (12 months)	9,00

Characteristic	Mean annualized return (in %)
32. Momentum-reversals	-5,74
33. Long-run reversals	-5,44
34. Value (monthly)	5,48
35. Net issuance (monthly)	-8,73
36. Earnings surprises	12,04
37. Return on equity	10,52
38. Return on market equity	12,24
39. Return on assets	7,07
40. Short-term reversals	-8,04
41. Idiosyncratic volatility	-3,06
42. Beta arbitrage	-0,68
43. Seasonality	11,51
44. Industry relative reversals	-17,85
45. Industry relative reversals (low volatility)	-34,91
46. Industry momentum-reversals	20,06
47. Composite issuance	-8,43
48. Price	-1,11
49. Firm age	3,47
50. Share volume	-1,21

D WRDS variable description overview

This appendix contains the variables included in the data set based on the WRDS 70 financial ratios and their corresponding annualized mean returns (in %) for all characteristics in the full sample.

Table D1: WRDS characteristics used in the extension part of this paper. These characteristics are 70 financial ratios for individual stocks taken from WRDS. Next to the characteristics, the annualized mean returns (in %) on managed portfolios are shown for all characteristics in the full sample. All returns are based on a monthly-rebalanced buy-and-hold strategy, annualized assuming 252 trading days a year, and are further rescaled to have normalized standard deviations. The data used ranges from November 1970 to December 2020 and is daily.

Characteristic	Mean annualized return (in %)
1. Shillers cyclically adjusted P/E Ratio	-10,39
2. Book/Market	-7,60
3. Enterprise value multiple	-7,44
4. Price/Operating Earnings (basic, excluding extraordinary items)	-10,58
5. Price/Operating Earnings (diluted, excluding extraordinary items)	-12,20
6. P/E (diluted, excluding extraordinary items)	-11,49
7. P/E (diluted, including extraordinary items)	-13,22
8. Price/Sales	-8,32
9. Price/Cash flow	-18,33
10. Dividend payout ratio	-15,87
11. Net profit margin	-16,84
12. Operating profit margin before depreciation	-6,63
13. Operating profit margin after depreciation	-17,67
14. Gross profit margin	-17,88
15. Pre-tax profit margin	-12,99
16. Cash flow margin	-14,46
17. Return on assets	-14,28
18. Return on equity	5,01
19. Return on capital employed	-13,40
20. Effective tax rate	-12,18
21. After-tax return on average common equity	-13,25
22. After-tax return on invested capital	-12,52
23. After-tax return on total stockholders equity	-16,33
24. Pre-tax return on net operating assets	-2,53
25. Pre-tax return on total earning assets	0,67
26. Gross Profit / Total Assets	-0,53
27. Common equity/Invested capital	1,30
28. Long-term debt/Invested capital	-1,04
29. Total debt/Invested capital	7,72
30. Capitalization Ratio	7,04
31. Interest/Average long-term debt	4,08

Characteristic	Mean annualized return (in %)
32. Interest/Average total debt	-1,13
33. Cash balance/Total liabilities	-4,60
34. Inventory/Current assets	1,63
35. Receivables/Current assets	-2,68
36. Total debt/Total assets	3,00
37. Total debt / EBITDA	6,66
38. Short-term debt/Total debt	-1,89
39. Current liabilities/Total liabilities	-17,19
40. Long-term debt/Total liabilities	-14,62
41. Profit before depreciation/current liabilities	-10,59
42. Operating cash flow/Current liabilities	2,58
43. Cash flow/Total debt	6,61
44. Free cash flow/Operating cash flow	-0,91
45. Total liabilities/Total tangible assets	1,05
46. Long-term debt/Book equity	1,74
47. Total debt/Total assets	-1,62
48. Total debt/capital	-14,42
49. Total debt/Equity	-13,93
50. After-tax interest coverage	2,34
51. Interest coverage ratio	4,32
52. Cash ratio	5,26
53. Quick ratio (acid test)	4,84
54. Current ratio	-2,96
55. Cash conversion cycle (days)	3,25
56. Inventory turnover	-2,20
57. Asset turnover	0,78
58. Receivables turnover	3,37
59. Payables turnover	5,18
60. Sales/Invested capital	-6,45
61. Sales/Stockholders equity	-0,84
62. Sales/Working capital	-2,67
63. Research and development/Sales	-8,68
64. Advertising expenses/Sales	5,37
65. Labor expenses/Sales	-8,51
66. Accruals/Average assets	-12,58
67. Price/Book	-2,67
68. Trailing P/E to growth (PEG) ratio	-2,60
69. Dividend yield	-0,55

E Additional tables

This appendix contains the tables regarding the estimation of the SDF for the three different subperiods when first the characteristics are transformed into their PCs.

Table E1: Largest factors in the SDF with ANS estimator for the interactions of 70 WRDS portfolios. The table contains three columns corresponding to the three sub periods. The table indicates the biggest coefficients and their corresponding t-statistics for the portfolios that are first transformed into their PCs.

1970-1986			1990-2004			2005-2020		
	b	t-statistic		b	t-statistic		b	t-statistic
PC 1	1314.40	117.94	PC 1	-30.30	23.64	PC 1	392.85	100.73
PC 2	-610.38	54.77	PC 4	-3.22	2.51	PC 4	53.02	13.60
PC 6	-452.07	40.56	PC 3	2.53	1.97	PC 5	-24.55	6.30
PC 4	-291.95	26.20	PC 7	-1.35	1.05	PC 7	-22.26	5.71
PC 5	-254.09	22.80	PC 28	-0.78	0.61	PC 13	11.29	2.90
PC 9	147.19	13.21	PC 6	0.71	0.55	PC 2	-10.82	2.78
PC 19	-83.01	7.45	PC 5	0.67	0.53	PC 23	-9.51	2.44
PC 7	81.06	7.27	PC 25	-0.65	0.51	PC 17	9.15	2.35
PC 25	-69.62	6.25	PC 13	-0.59	0.46	PC 25	8.92	2.29
PC 15	64.19	5.76	PC 22	0.57	0.44	PC 24	-6.82	1.75

Table E2: Largest factors in the SDF without ANS estimator for the interactions of 70 WRDS portfolios. The table contains three columns corresponding to the three sub periods. The table indicates the biggest coefficients and their corresponding t-statistics for the portfolios that are first transformed into their PCs.

1970-1986			1990-2004			2005-2020		
	b	t-statistic		b	t-statistic		b	t-statistic
PC 1	57.99	24.77	PC 1	-5.23	9.96	PC 1	37.10	30.40
PC 6	-134.23	22.10	PC 4	-1.93	1.86	PC 4	16.55	7.59
PC 2	-52.35	16.04	PC 3	1.45	1.42	PC 7	-12.16	4.22
PC 5	-66.67	11.68	PC 7	0.62	0.56	PC 5	-9.48	3.91
PC 4	-45.91	10.39	PC 5	0.59	0.55	PC 13	8.10	2.45
PC 9	63.67	8.69	PC 28	-0.59	0.52	PC 23	-7.59	2.18
PC 19	-55.31	6.08	PC 18	0.54	0.48	PC 25	7.17	2.05
PC 25	-50.31	5.31	PC 11	0.46	0.41	PC 17	6.88	2.03
PC 15	38.81	4.48	PC 22	0.39	0.34	PC 24	-5.46	1.56
PC 7	26.62	4.17	PC 13	-0.36	0.32	PC 8	3.92	1.29

Table E1 contains the coefficients and t-statistics of the most significant SDF coefficients for the PCs of the interactions data set. The ANS estimator is used for covariance matrix estimation. The 1970-1986, 1989-2004 and 2005-2020 samples are shown in the table's left, middle, and right column, respectively. Table E2 shows the coefficients and t-statistics of the most significant SDF coefficients for the PC transformed interactions data set without ANS estimator. The 1970-1986, 1989-2004 and 2005-2020 samples are shown in the table's left, middle, and right column, respectively. Comparing Table E1 to Table E2, it can be noted that PCs with the highest explanatory power are the same in the first period, although some are in a different order. The second period has all but one PC the same. For the last period, some PCs are different.