Erasmus School of Economics

Bachelor Thesis Final Report
BSc^2 Econometrics and Economics

Shrinking the cross-section in a quantile factor framework

Author
Luc Schumans (479017)

Supervisor
Dr. M Grith

Second assessor
Dr. MD Zaharieva

July 3rd 2021

Abstract

This paper reproduces and verifies the main results of Kozak et al. (2020), who construct a robust stochastic discount factor (SDF) to find the joint explanatory power of many cross-sectional stock return predictors in the US stock market. Their research is extended by applying a novel type model introduced by Chen et al. (2019); Quantile factor analysis. The high dimensionality problem of the data is overcome by setting a novel type Bayesian prior that shrinks the contribution of low variance principle components of the characteristic-based factors. The results indicate that raw characteristics, principal components, and quantile factors cannot adequately describe the cross-section of stock returns in a characteristic-sparse SDF specification when the data has no inherently strong factor structure. Furthermore, the quantile analysis shows that there is a relatively higher explanatory power in the tails of the distribution. The tail quantiles are also able to create beneficial portfolio properties for characteristic-based portfolios.

Contents

1 Introduction

It has been an ongoing practice in asset pricing literature to seek stock characteristics that can explain the cross-sectional variation in stock returns. Many have sought to explain this variation with as little characteristics as possible. This, in essence, means that they have tried to find a characteristic-sparse stochastic discount factor (SDF) representation, that captures the linear relationship between a small number of factors and the cross-sectional variation in stock returns. These so-called factor models have developed tremendously over time. As Kozak et al. (2020) point out, each time a new cross-sectional factor is developed, the models need to be extended and modified to take into account this new evidence. Furthermore, these new models are tested in a universe where only a small number of predictors are considered and it is therefore not clear how they behave in a setting with a large number of possible explanatory characteristics. Statistical inference for this problem is challenging, as we are confronted with high dimensionality, which can lead to overfitting.

In their paper, Kozak et al. (2020) question the economic rationale behind characteristic-sparse SDF models. The main motivation for their research is that there would be extreme redundancy among the well-known anomalies in asset pricing if the cross-sectional variation in asset returns could be explained by a small number of factors. The present-value identity or q-theory, that many models are based on, does predominantly not imply that only a small set of characteristics matters. For example, present-value identities can show that expected profitability plays a major role in explaining expected return. However, expected profitability is not observable and, theoretically, many observable stock characteristics could have an influence on the variation of expected profitability and hence also on expected returns. Therefore, Kozak et al. (2020) set out to investigate whether these characteristic-sparse SDF models can adequately describe the cross-sectional variation in stock returns.

The high-dimensional nature of the data, that is the great number of possible factors, could result in spurious overfitting when conventional cross-sectional regression would be used to estimate the SDF coefficients. Kozak et al. (2020) overcome this problem by introducing a Bayesian approach that uses a novel prior. Their Bayesian posterior shrinks the SDF coefficients to zero in a way that satisfies economic beliefs. That is, more shrinkage is applied to the SDFs corresponding to low eigenvalue (i.e. low variance) principal components (PCs) because they typically explain the variation in the SDF less well than the high eigenvalue PCs. The Bayesian estimator of Kozak et al. (2020) bears some resemblance with the ridge regression procedure, but with a difference. In the least squares objective function, the penalty added in the ridge regression would be the L^2-Norm, whereas in this estimator it is based on the maximum squared Sharpe Ratio as implied by the SDF.

In itself, the aforementioned estimator is not tailored for characteristic-sparse SDF at all because the weight of any candidate characteristic is never set exactly equal to 0. Kozak et al. (2020) also allow for sparsity by employing a second penalty factor in the least squares objective function. This penalty is based on the L^1-Norm of the SDF, which is often used in lasso regressions (R. Tibshirani, 1996), and introduces automatic factor selection. Inherently, this penalty sets some of the weights to 0, which leads to sparse outcomes. In their combined penalty specification (L^1- and L^2-Norm), the penalization

strength, of both shrinkage and sparsity, is chosen by maximizing the out-of-sample (OOS) R^2. This type of penalization bears resemblance to the elastic net regression by Zou and Hastie (2005), which combines lasso and ridge regression penalties.

In an empirical framework, Kozak et al. (2020) consider dozens of portfolios based on well-known anomalies and financial ratios and thousands of powers and interactions of these characteristics to examine OOS performance. Next to the raw data, also the PCs of these characteristics-portfolios are considered. It is economically sensible that there is greater evidence for sparsity in characteristics when using PCs instead of the original data. All in all, the results of Kozak et al. (2020) show that characteristic-sparse SDF cannot adequately describe the cross-sectional variation in stock returns. This image changes when the PCs are considered, where it is found that a successful sparse SDF can be achieved.

A recent innovation in factor model estimation for dimension reduction comes from Chen et al. (2019). They consider a quantile factor model (QFM), that represents a new class of factor models for high-dimensional panel data. This model makes use of quantile factor analysis (QFA), which can capture factors that shift other parts of the underlying distribution than just the location. In comparison, PCA only captures factors that allow for a shift in mean. The research of Chen et al. (2019) shows promising results in the fields of macroeconomic forecasting, climate research, and, most importantly for this research, financial factor models. Since the results of Kozak et al. (2020) show that PCA can result in a well-performing sparse SDF, it seems viable that QFA models can achieve the same.

This research aims to reproduce, verify, and critically evaluate a part of the results found by Kozak et al. (2020). This results in the following research question:

To what extent can a characteristic-sparse stochastic discount factor representation successfully summarize the cross-sectional variation in stock returns?

Next to that, this paper extends the research of Kozak et al. (2020) by considering a novel type of factor model introduced by Chen et al. (2019). The QFA model is used to answer the following research question:

In what manner can quantile factor analysis be applied to characteristic-based factors to summarize the cross-sectional variation in stock returns?

The relevance of this research has two sides. Firstly, the paper is academically relevant, because it verifies an important finding in the field of factor model research and extends this finding by including a promising, new, and novel model. To the best of my knowledge, I am the first one who investigates whether QFA can successfully create a meaningful sparse SDF representation. Secondly, the paper is socially relevant, because it extends the knowledge of factor models which are, in one form or another, used by practitioners throughout the world. Insights into the value of sparse factor models are useful in terms of risk management and investment choices.

The theoretical models used consist of 11 different factor specifications: raw characteristics, principal components, and nine decile specifications of quantile factors. The estimation of the quantile factors follows an equivalent strategy as Chen et al. (2019), who introduce an estimation algorithm called the

Iterative Quantile Regression. The estimation of the SDF for all models follows the original work of Kozak et al. (2020), which makes use of the aforementioned novel Bayesian specification that results in a dual penalty specification. The estimation is conducted in a three-fold cross-validation setting to maximise the out-of-sample R^2 values of the dual penalty specification.

For further insight, the quantile factors are compared as characteristic-based portfolios. For every individual factor as well as the complete MVE portfolios for each decile a Sharpe Ratio (SR) can be calculated. The maximum SR is statistically compared for equality with all other SRs within each decile and across the MVE portfolios by making use of prewhitened HAC inference by Ledoit and Wolf (2008). Furthermore, within its own factor universe, each characteristic-based portfolio is utilized to plot an efficient frontier combined with a scatter plot of the portfolios.

Two sets of data are considered in this research. These consist of the daily data of the 25 ME-BM-sorted Fama and French portfolios from July 1926 to December 2017, provided in the French (2021) online data library, and daily data of 50 different stock characteristics that underlie certain anomalies in the period November 1973 to December 2017, supplied by Kozak et al. (2020).

All in all, the 25 Fama and French portfolios have such an inherently strong factor structure that a characteristic-sparse specification can adequately capture the variation in the SDF. For data sets with no inherently strong factor structure, i.e. the 50 anomaly portfolios, a characteristic-sparse SDF specification cannot successfully be created. The QFA models are able to attain an overall better out-of-sample performance. Hence, there is a great explanatory power in moments higher than the mean for the variation in the SDF. Furthermore, the estimation results show that the relative outliers (tails) have an important role in this. This result is in accordance with the original quantile factor model research of Chen et al. (2019). Additionally, the factors across different deciles show extremely high levels of canonical correlation, indicating that only a few factors account for the difference between deciles. Lastly, it is also shown that quantile factors create beneficial properties for characteristic-based portfolios.

The remainder of this paper starts with a description of all relevant literature related to this research in Section 2. Secondly, Section 3 outlines the theoretical models. After that, all estimation and statistical comparison techniques are explained in Section 4. Consequently, the data is discussed in Section 5, after which the results are presented in Section 6. Finally, Section 7 provides a discussion and conclusion of the research and an examination of limitations and suggestions for further work.

2 Literature

The search for explaining variation in stock returns starts with the disruptive work of H. Markowitz (1952). In this early work, H. Markowitz (1952) shows via his mean-variance model that higher returns can only be expected when an investor takes more risk. This principle is widely known today as the risk-return trade-off. H. M. Markowitz (1959) already proposes the idea that returns can be explained by a single-factor model, often referred to today as an index model. The first 'real' factor model is presented by Sharpe (1963), namely the capital asset pricing model (CAPM). In this model, a single type of risk, known as the market risk, has influence on the expected asset returns. This research distinguishes between systemic and idiosyncratic risk factors and presents that only an increase in systemic risk should

lead to an increase in return because idiosyncratic risk can be diversified away. A major step forward in asset pricing literature comes from Roll and Ross (1980), who present the arbitrage pricing theory (APT). The APT also has a focus on systemic risk, yet it proposes that not one, but multiple risk factors have an influence on security returns. This innovation is combined with the fact that the factors used are not readily observable variables as is the case for the CAPM.

Following these groundbreaking papers, many researchers have proposed different factor model specifications to explain the variation in the cross-section of stock returns. An important example to discuss in the framework of this research is the work of Fama and French (1993). They extend the CAPM model of Sharpe (1963) with two systemic risk factors based on size (SMB) and book-to-market (B/M) ratio (HML). With these two factors, Fama and French (1993) take into account that, in general, companies with a small market cap outperform larger companies and companies with a high B/M ratio perform better than companies with a low B/M ratio in terms of returns. In essence, the SMB and HML factors capture a part of systemic risk that can explain the cross-sectional variation in returns.

The above-mentioned factor specifications are all examples of observed factor models, i.e. the explanatory variables are known and the loadings need to be estimated. A key part of this research lies in another class of factor models, namely latent factor models. In these types of models both the factors and their corresponding loadings are unknown and need to be estimated. The most common technique to estimate latent factor models is principal component analysis (PCA), which was originally set out by Pearson (1901) and Hotelling (1933). PCA plays a major role in modern finance literature because of its ease of use and interpretability. Stock and Watson (2002), for example, use PCA to forecast macroeconomic variables. Kozak et al. (2020) explain that a characteristic-sparse model containing high-variance principal components (PCs) can explain the variation in cross-sectional asset return quite well. Kozak et al. (2018) present an economic reason for this. They state that, in the absence of near-arbitrage opportunities, a factor with high explanatory power should be a major source of co-movement itself. Therefore, when portfolio returns have a strong factor structure that can be explained by a few high-variance PCs, a successful characteristic-sparse SDF can be attained with only a few high-variance PCs. Furthermore, Kozak et al. (2018) state that this is true for models with either rational investor behaviour or biased investor beliefs.

The major contribution of this paper is considering another novel type of latent factor model, namely quantile factor analysis (QFA) by Chen et al. (2019). The class of quantile factor models (QFM) is a generalization of other latent factor models. This implies that QFA relates to PCA as a quantile regression (QR), introduced by Koenker and Bassett (1978), relates to least squares (LS) (Chen et al., 2019)). The motivation of Koenker and Bassett (1978) to consider QR, lies in the fact that LS does not work well if error terms are non-normally distributed, which is a well-known stylized fact of stock returns. In further literature, many advantages of QR are presented. Chiang and Li (2012) find that, when you move from low to high return quantiles, the risk-return relation turns from negative to positive for the US stock market. Barnes and Hughes (2002) find that many of the known problems of CAPM, such as heteroskedasticity due to omitted variables, strong sensitivity to outliers, and non-normally distributed error terms become less critical when applying QR to the CAPM.

Chen et al. (2019) proof that they can produce asymptotically normal and consistent estimators for factors and their loadings in a QFA framework. They also present that the rate of average convergence for QFA equals the one of PCA by Bai and Ng (2002). Nevertheless, the most important innovation of QFA is that the factors can capture the shifting of higher moments in the underlying distribution, whereas PCA can only capture shifts in location. This is useful because asset returns often have a heavy-tailed underlying distribution. Chen et al. (2019) show promising empirical results in the fields of macroeconomic forecasting, climate research, and, most importantly, financial factor models.

The common ground of all aforementioned papers is that they focus on the estimation of risk premia. This implies that they investigate to what extent a factor corresponding to a certain variable can explain the variation in stock returns. This research follows Kozak et al. (2020) and focuses on the estimation of risk prices instead of risk premia. Risk prices indicate to what extent a factor contributes to the variation of the SDF and hence the price of an asset. The fundamental difference is that a factor can earn a high risk premium because it is correlated with the pricing factors in the SDF, while it is not one of the pricing factors itself. The SDF model, introduced by Cochrane (1996), proposes that the price of an asset can be computed by taking the expectation of future cash flows discounted by some stochastic factor. My research, along the lines of Kozak et al. (2020), focuses on characteristic-based SDF models. These models try to estimate the risk prices of certain characteristic-based factors on the SDF, whereas the original SDF model uses the asset returns. In this way, the factors can be seen as characteristic-based portfolios for which we want to explain the variation in returns, as well as the candidate explanatory variables to explain this variation. The estimated coefficients of any SDF model correspond to the weights of a mean-variance efficient portfolio, as mentioned by H. Markowitz (1952). One of the key contributions of the paper of Kozak et al. (2020) is to adapt the standard ridge and lasso regression estimators' objective functions to be suitable for a characteristic-based SDF and to be consistent with their choice of prior.

The regular estimation for an SDF model would be to take a simple cross-sectional regression. However, spurious overfitting could occur because this research employs a large number of possible explanatory variables. To overcome this problem, this research takes a Bayesian estimation approach with a novel prior, as by Kozak et al. (2020). In the literature, the considered prior bears resemblance to the prior described in Pástor and Stambaugh (2000) and Pástor (2000). A key difference with these two papers is that they consider all shrinkage of any asset to be equal, whereas this is not the case with the prior of Kozak et al. (2020). Kozak et al. (2018) provide an economic reason for this difference. They show that a substantial risk price is attributable to high-variance principal components of explanatory factors. The posterior that follows from the prior of Kozak et al. (2020) takes this into account by shrinking the SDF coefficients of the low variance principal components more significantly.

The Bayesian estimator that follows has similarities to a standard ridge regression, but imposes a penalty based on the maximum squared Sharpe Ratio as implied by the SDF. This is, in turn, equivalent to a minimization of the Hansen and Jagannathan (1991) distance with an L^2-Norm penalty on the sum of squared SDF coefficients. This estimator maps into the estimator of the MVE portfolio weights constrained by an L^2-Norm as by Brandt et al. (2009) and DeMiguel et al. (2009). Moreover, DeMiguel et al. (2009) show that these weights also bear resemblance to a portfolio optimization where the covariance

matrix is shrunk towards an identity matrix, as in e.g. Ledoit and Wolf (2004). However, there is an important difference with covariance shrinkage literature. The amount of shrinkage in covariance-based models depends on the uncertainty in covariance estimation. This paper, as in Kozak et al. (2020), assumes a known covariance matrix and lets the amount of shrinkage depend on the uncertainty in the estimation of the mean. Kozak et al. (2020) show that the uncertainty in the covariance matrix is negligible when the uncertainty in the mean estimation is accounted for.

Since this estimator is still not tailored for sparsity, another penalty term based on the L^1-Norm of the SDF coefficients is considered. This term is based on lasso type regressions as in R. Tibshirani (1996). When combining the L^1- and L^2-norm penalties, the resulting estimator is similar to the elastic net technique as by Zou and Hastie (2005), which is often used in machine learning. An example of another research that uses machine learning techniques to overcome the high dimensionality challenge in asset pricing comes from Rapach et al. (2013). They set out to forecast global stock market returns by applying a lasso technique to select a few explanatory variables from a large set of candidate variables. In other literature, DeMiguel et al. (2020), Freyberger et al. (2020), and Feng et al. (2017) apply lasso estimation techniques on characteristic-based factors. All three papers find that there is a large redundancy among explanatory characteristics. However, as the result of Kozak et al. (2020) and other statistical literature suggests, an estimation where only the L^1-Norm is considered is inferior to an L^2-Norm or elastic net estimation when the candidate factors are correlated (R. Tibshirani, 1996 & Zou and Hastie, 2005).

As is explained by Kozak et al. (2020), this research also relates to literature that concerns itself with the problem of data-mining return predictors. This type of literature seeks statistical significance of individual factors, when researchers may have tried many possible factors. For example, Harvey et al. (2016) and Green et al. (2017) make an adjustment to the significance range to account for data-mining. This research does not directly adjust for data-mining. However, as stated by Kozak et al. (2020), there is no basis for first and second moments for data-mined factors to have a relation. By tying them together in the Bayesian prior, the effect of data-mined factors is downweighed. Next to that, the period considered in this research has a minimal overlap to the periods that the anomalies, on which the characteristic-based factors are based, were revealed (McLean and Pontiff, 2016).

3 Theoretical model

This section presents the theoretical SDF model and the PCA & QFA specifications. Firstly, the model for the replication is discussed. After that, the model specification for the extension is shown.

3.1 Replication

The factor returns of the stocks are defined as

$$F_t = Z'_{t-1} R_t, \tag{1}$$

where Z_{t-1} is an $N \times H$ matrix of stock characteristics at time $t-1$ and R_t corresponds to the $N \times H$ matrix of asset returns. Section 5 elaborates on the specification and computation of these matrices. The SDF in a linear span of excess returns, in the spirit of Hansen and Jagannathan (1991), can be

transformed into a characteristic-based SDF in the linear span of H stock characteristic returns, when we solve

$$M_t = 1 - b'(F_t - \mathbb{E}F_t), \tag{2}$$

for the $N \times 1$ vector of time-invariant SDF loadings b. In line with the literature, this paper focuses on the unconditional asset pricing equation constraint when solving the above equation. That is,

$$\mathbb{E}[M_t F_t] = 0, \tag{3}$$

where the factors F_t serve simultaneously as the assets for which we want to explain the returns, as well as possible explanatory factors that are priced in the SDF. For interpretability, each column of Z is demeaned such that each F_t can be seen as a long-short zero-investment portfolio. Also, each factor F_t is orthogonalized with respect to the market factor. Appendix A shows the derivation from the original model of Hansen and Jagannathan (1991) in more detail.

The general model specification can be extended to adhere to a PCA structure. The PC factors are based on the eigenvalue decomposition of the covariance matrix of the factors

$$\Sigma = VDV', \quad \text{with} \quad D = diag[d_1, \ldots, d_H], \tag{4}$$

where each column of V equals an eigenvector of Σ and each element of the diagonal matrix D equals an eigenvalue in decreasing order. Then the PC factors are created as follows:

$$P_t = V'F_t. \tag{5}$$

Normally, one would try to select a certain optimal number of PC factors to add as explanatory variable based on e.g. Information Criteria. For the sake of this research, the number of PCs is set equal to the number of raw characteristics, which equals the maximum number of PCs possible. When taking into account all PCs, the SDF is constructed in the spirit of Kozak et al. (2020) as

$$M_t = 1 - b_p'(P_t - \mathbb{E}P_t), \quad \text{with} \quad \mathbb{E}[M_t P_t] = 0. \tag{6}$$

3.2 Extension

The QFA approach is not as straightforward as PCA and does not follow immediately from an eigenvalue decomposition. The quantile factor model structure in the spirit of Chen et al. (2019) is as follows:

$$F_t(\tau) = \lambda'(\tau)q_t(\tau) + u_t(\tau), \tag{7}$$

where the factor returns of quantile τ ($\in [0, 1]$) are dependent on the $N \times 1$ vector of unobserved random variables $q_t(\tau)$, a $N \times 1$ vector of non-random factor loadings $\lambda(\tau)$, plus a $N \times 1$ vector of error terms $u_t(\tau)$. These error terms follow the following property:

$$P[u_{it}(\tau) \leq 0 | q_t(\tau)] = \tau, \tag{8}$$

where i indicates the error term belonging to characteristic i. Following Chen et al. (2019), this means that the probability that the fitted model overestimates the factor return is equal to τ.

The vector of quantile factors $q_t(\tau)$ can now be treated equivalent as the factors in the original SDF and the PC specification of Equation (2) and Equation (6), respectively, that is

$$M_t = 1 - b_p'(q_t(\tau) - \mathbb{E}q_t(\tau)), \quad \text{with} \quad \mathbb{E}[M_t q_t(\tau)] = 0. \tag{9}$$

This notation still seeks an SDF in a linear span of the returns as the quantile factors are constructed linearly with respect to the quantiles of the returns.

This research focuses on 8 different QFA specifications at every decile. This implies that

$$\tau \in \zeta = \{0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90\}. \tag{10}$$

4 Estimation Methodology

This section outlines the estimation techniques for the SDF coefficients. Again, this section is split up into two parts that define the methodology needed for the replication and the extension. Firstly, for the replication, the basic estimation for the raw data and PCA is discussed. After that, the Bayesian approach to overcome the high dimensionality problem is explained. This is followed by the penalty definitions of the estimator and the statistical technique to compare different levels of sparsity and shrinkage. Secondly, the extension part starts with the estimation method for a QFM. Lastly, a statistical comparison technique between different decile estimators is elaborated on.

4.1 Replication

This subsection outlines all estimation techniques necessary for the replication of Kozak et al. (2020).

4.1.1 Naive SDF coefficient estimation

When population moments are known, b (Equation (2)) can be solved as

$$b = (\Sigma\Sigma)^{-1}\Sigma\mathbb{E}F_t. \tag{11}$$

Note that, without any knowledge about the population moments, taking this result as the cross-sectional regression of the expected factor returns on the covariance of the factors would result in over-fitting due to the high dimensionality this regression faces.

A similar solution can be given for the PCA framework. When taking into account all PCs and with knowledge about population moments, b_p (Equation (6)) can be solved as

$$b_p = D^{-1}\mathbb{E}P_t, \tag{12}$$

where D equals a diagonal matrix whose elements correspond to the eigenvalues of the factors in decreasing order and P_t refers to the matrix of PC factors as in Equation (5).

4.1.2 L^2-norm Bayesian and penalization approach

As previously mentioned, a simple sample estimator of b would yield very imprecise results, according to Kozak et al. (2020), because of too much uncertainty about the expected values of the used factors

(i.e. regular, PC or QF). This uncertainty in mean values could lead to spurious overfitting. Kozak et al. (2020) solve this problem by creating a Bayesian framework with a well-motivated prior. Due to the extra information of this prior, the estimation procedure becomes robust to overfitting. Assuming that Σ is known, the family of priors for the mean μ of the factors is considered as

$$\mu \sim \mathcal{N}\Big(0, \frac{\kappa^2}{\theta}\Sigma^{\eta}\Big), \tag{13}$$

where $\theta = tr[\Sigma]$ and κ is 'scaling' constant of μ that can depend on θ and H.

Kozak et al. (2020) set $\eta = 2$, which yields the independent and identically distributed prior on the SDF coefficients as

$$b \sim \mathcal{N}\Big(0, \frac{\kappa^2}{\theta}\mathcal{I}\Big). \tag{14}$$

These prior beliefs together with information about the sample mean of the factors $\bar{\mu}$ form, when assuming a multivariate-normal likelihood, a posterior mean of b as

$$\hat{b} = (\Sigma + \gamma\mathcal{I})^{-1}\bar{\mu}, \tag{15}$$

$$\text{with} \quad \gamma = \frac{\theta}{\kappa^2 T}. \tag{16}$$

which is also a visible result from the algebraic derivation as shown in Equation (B.4) in Appendix B. The posterior variance of b is given as:

$$var(b) = \frac{1}{T}(\Sigma + \gamma\mathcal{I})^{-1}, \tag{17}$$

which is used for the construction of confidence intervals in the empirical results. Note that, the Σ in the aforementioned and the subsequent equations refers to a sample covariance matrix of the factors which is regularized with a flat Wishart prior.

The economic interpretation of γ is more understandable when $\hat{b}$ is written as a penalized estimator. The penalized estimator, in the spirit of Kozak et al. (2020), maximizes the models' R^2, with a penalty on the models' implied maximum Sharpe Ratio (SR) as

$$\hat{b} = argmin_b\{(\bar{\mu} - \Sigma b)(\bar{\mu} - \Sigma b)' + \gamma b'\Sigma b\}, \tag{18}$$

which leads to the same specification as Equation (15). It can also be written as the minimization of a HJ-distance, as by Hansen and Jagannathan (1991), subject to an L^2-Norm penalty $\gamma b'b$ as

$$\hat{b} = argmin_b\{(\bar{\mu} - \Sigma b)\Sigma^{-1}(\bar{\mu} - \Sigma b)' + \gamma b'b\}, \tag{19}$$

which is again an equivalent notation. This estimator introduces shrinkage in the estimation by the choice of κ, which is inversely related to γ. In terms of shrinkage, the lower the choice of κ the more shrinkage there takes place. Under the prior with $\eta = 2$, the maximum root expected squared Sharpe Ratio equals:

$$\mathbb{E}[\mu\Sigma^{-1}\mu] = \kappa. \tag{20}$$

Thus, the expected squared Sharpe Ratio has a direct effect on the level of shrinkage imposed. In other words, a low expected squared Sharpe Ratio implies a low κ, which in turn implies a high γ. So, a low expected squared Sharpe Ratio actually indicates a high degree of shrinkage.

If one considers the PC space of factors, it becomes clear how the shrinkage works. Suppose, the weight of a low-variance (eigenvalue) PC is shrunk towards 0. This creates a certain benefit for the penalty term but has little cost in terms of the explanatory power of the volatility in the SDF. If one would shrink a high-variance PC with the same magnitude, the benefit in terms of the penalty would be equal. However, the cost would be much greater because this PC has more explanatory power over the variance of the SDF. Hence, this estimator is tilted towards shrinking low-variance PCs more heavily.

4.1.3 L^1- and L^2-norm penalty estimator

All in all, the aforementioned model is still not tailored for any sparsity in characteristics as it never sets any of the MVE weights in $\hat{b}$ exactly equal to 0. For this, Kozak et al. (2020) introduce a second penalty term in the penalized estimator based on lasso regression as follows:

$$\hat{b} = argmin_b\{(\bar{\mu} - \Sigma b)\Sigma^{-1}(\bar{\mu} - \Sigma b)' + \gamma_2 b'b + \gamma_1 \sum_{i=1}^{H} |b_i|\}, \tag{21}$$

where γ_2 equals the L^2-Norm penalty (γ in Equation (18)) and γ_1 controls the degree of sparsity of the estimation.

4.1.4 Model evaluation

To evaluate the estimation of b for different degrees of shrinkage and sparsity, a K-fold cross validation approach is taken. First, the data is split into 3 ($k = 3$) sub-samples. For each possible γ or pair of γ_1 & γ_2 the estimator $\hat{b}$ is computed with the data of 2 of these sub-samples. This estimator is used to calculate an OOS R^2 on the withheld data (whose moments are indicated with subscript 'o') as follows:

$$R^2_{oos} = 1 - \frac{(\bar{\mu}_o - \bar{\Sigma}_o\hat{b})'(\bar{\mu}_o - \bar{\Sigma}_o\hat{b})}{\bar{\mu}_o'\bar{\mu}_o}. \tag{22}$$

This is repeated 3 times such that every single one of the sub-samples is considered as out-of-sample once. Then, the OOS R^2 estimates of the repetitions are averaged, which yields the cross-validated (CV) OOS R^2. To obtain the optimal model, the γ or pair of γ_1 & γ_2 is chosen that maximizes the OOS R^2.

This research follows the choice of $k = 3$ from Kozak et al. (2020). They explain that this choice is a compromise between estimation uncertainty of $\hat{b}$ and estimation uncertainty of $\bar{\Sigma}_o$. When the choice of k would become too high, the estimation sample would become too short to get a well-behaved $\bar{\Sigma}_o$.

It is important to emphasise that all models for both the replication and extension are evaluated with these criteria. Since OOS R^2 is a metric that tests predictive ability for out-of-sample data, it can be compared directly across different models. In this sense, the chosen factor specification is the only differentiating element across the results.

It has to be noted that this practice is not a clean out-of-sample estimation as the whole sample is used to calculate the penalty terms. This implies that the CV OOS R^2 at the optimal penalties is usually upward-biased when compared to a true out-of-sample estimate of the R^2 (R. J. Tibshirani and Tibshirani, 2009 & Varma and Simon, 2006). However, the interest of this paper, in the spirit of Kozak et al. (2020), is the relative performance between different levels of sparsity and shrinkage and not the absolute level of the OOS R^2. Hence, the slight upward bias is not a drawback for this research.

4.2 Extension

This subsection presents the methodology for the QFA model and certain statistical tests to compare the different QFA specifications at every decile.

4.2.1 QFA estimation

As is the case for any statistical factor model, the factors and loadings cannot be estimated separately (Bai and Ng, 2002). In the spirit of Chen et al. (2019) the following normalizations are introduced, without loss of generality:

$$\frac{1}{T}\sum_{t=1}^{T} q_t(\tau)q_t'(\tau) = I_{H-1}, \quad \frac{1}{N}\sum_{i=1}^{N}\lambda_i(\tau)\lambda_i'(\tau) \text{ is diagonal with non-increasing elements,} \tag{23}$$

where the subscript $H-1$ refers to the number of factors. This number cannot equal H because then the estimation algorithm cannot differentiate different quantile specifications.

Now, let $G = (N+T)r$ and $\theta = (\lambda_1'(\tau), \ldots, \lambda_N'(\tau), q_1'(\tau), \ldots, q_T'(\tau))$. Note that, the dependence of θ on G is suppressed, as in Chen et al. (2019). For ease of notation, let $\mathcal{A}, \mathcal{Q} \in \mathbb{R}^r$ and let:

$$\Theta^r = \{\theta \in \mathbb{R}^G : \lambda_i(\tau) \in \mathcal{A}, q_t(\tau) \in \mathcal{Q} \text{ for all } i,t \text{ and } \{\lambda_t(\tau), q_t(\tau)\} \text{ satisfy Equation (23)}\}. \tag{24}$$

With this notation, the objective function can be defined as:

$$\mathbb{M}_{NT}(\theta) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\rho_\tau(f_{i,t} - \lambda_i'(\tau)q_t(\tau)), \tag{25}$$

where the parameters should satisfy the constraints of Equation (23) and $\rho_\tau(u) = (\tau - \mathbf{1}\{u \leq 0\})u$, where $\mathbf{1}$ corresponds to an indicator function that equals 1 when the statement between the accolades is true and 0 otherwise. This check function, as by Koenker and Bassett (1978), is a type of loss function that retrieves the requested quantile τ in the objective function.

The estimator of θ_0 can now be expressed as:

$$\hat{\theta} = (\hat{\lambda}_1'(\tau), \ldots, \hat{\lambda}_N'(\tau), \hat{q}_1'(\tau), \ldots, \hat{q}_T'(\tau)) = argmin_{\theta\in\Theta^r}\mathbb{M}_{NT}(\theta). \tag{26}$$

On a sidetrack, it is now visible how QFA relates to PCA. To visualize this, let $\tau = 0.50$, which corresponds to the median quantile. Following the theory of Koenker and Hallock (2001), in a regular QR, the objective function at the 0.50^{th} quantile is equivalent to the minimization of the sum of absolute errors, or an L^1-Norm. For an LS regression, the objective function is the minimization of the sum of squared residuals, or an L^2-Norm. This framework can be extended into the universe of factor models in exactly the same way. Filling in $\tau = 0.50$ in the check function of Equation (25) yields a minimization of the sum of absolute error terms, whereas a minimization of the sum of squared residuals is required in the PCA framework. Hence, it can be stated that QFA relates to PCA as QR relates to LS. The structure of this objective function also, again, shows that QFA is less sensitive to outliers and can better handle heavy-tailed distributions than PCA.

As Chen et al. (2019) state, the estimator $\hat{\theta}$ of Equation (26) has no closed-form analytical solution, whereas this is the case for PCA. To overcome the estimation problem, Chen et al. (2019) introduce the iterative quantile regression procedure (IQR). In the same spirit as their research, define

$\Lambda = (\lambda_1(\tau), \ldots, \lambda_N(\tau))', Q = (q_1(\tau), \ldots, q_T(\tau))'$ and introduce the averages:

$$\mathbb{M}_{i,T}(\lambda, Q) = \frac{1}{T}\sum_{t=1}^{T} \rho_\tau(f_{i,t} - \lambda'(\tau)q_t(\tau)), \tag{27}$$

$$\mathbb{M}_{t,N}(\Lambda, q) = \frac{1}{N}\sum_{i=1}^{N} \rho_\tau(f_{i,t} - \lambda_i'(\tau)q(\tau)), \tag{28}$$

where $\mathbb{M}_{NT}(\theta) = N^{-1}\sum_{i=1}^{N}\mathbb{M}_{i,T}(\lambda_i, Q) = T^{-1}\sum_{t=1}^{T}\mathbb{M}_{t,N}(\Lambda, q_t)$. The greatest hurdle in the minimization of Equation (25) is that the function is not convex in θ. However, as by Chen et al. (2019), for a given Q, the function $\mathbb{M}_{i,T}(\lambda, Q)$ is convex in λ for every i, and for a given Λ, the function $\mathbb{M}_{t,N}(\Lambda, q)$ is convex in q for every t. Hence, both these functions can efficiently be optimized. This paper follows the proposed method of Chen et al. (2019) to estimate the QFA, which goes as follows:

Iterative quantile regression procedure (IQR):

1. Set random starting parameters $Q^{(0)}$
2. Given a $Q^{(l-1)}$, solve $\lambda_i^{(l-1)}(\tau) = argmin_\lambda \mathbb{M}_{i,T}(\lambda, Q^{(l-1)})$ for $i = 1, \ldots, N$. Then, given $\Lambda^{(l-1)}$, solve $q_t^{(l)}(\tau) = argmin_q \mathbb{M}_{t,N}(\Lambda^{(l-1)}, q)$ for $t = 1, \ldots, T$.
3. Iterate the second step for $l = 1, \ldots L$, where L equals the iteration such that the difference between $\mathbb{M}_{NT}(\theta^{(L)})$ and $\mathbb{M}_{NT}(\theta^{(L-1)})$ is less than or equal 0.0001, i.e. convergence.
   Note that, $\theta^{(l)} = (vech(\Lambda^{(l)})', vech(Q^{(l)})')'$.
4. Normalize $\Lambda^{(L)}$ and $Q^{(L)}$, such that they satisfy Equation (23).

Via this procedure, the quantile factors (QFs) $q_t(\tau)$ can be retrieved for every decile. These factors are then used as described in Section 3.2 in the estimation of the SDF coefficients explained in Section 4.1.

4.2.2 Comparison of QFA factors and SDF coefficients

Just as the raw characteristics and PCs, the QFs can be seen as characteristic-based portfolios. This implies that Sharpe Ratios can be computed for each individual factor and can be compared with each other. First, I consider the Sharpe Ratios for the factors of each decile τ separately, which are expected to be different. These Sharpe Ratios can be tested for (in)equality with the pairwise test inference of Ledoit and Wolf (2008). I use the inference based on HAC standard errors with a prewhitened QS kernel. Ledoit and Wolf (2008) note that this method is often too liberal when sample sizes are moderate to small. However, this is not a problem for this research as sample sizes are relatively large. To avoid a multiple testing problem only the largest Sharpe Ratio is compared with all others per decile. For clearance, a set of 50 factors would require 1225 pairwise tests, which would result in a total of 9800 tests for 8 deciles. This results in the following testing hypothesis:

$$H_0: \qquad max(SR_{\tau,m}) = SR_{\tau,i}, \forall i = 1, \ldots, N, \text{ with } i \neq m \tag{29}$$

where τ corresponds to the decile for which this test in conducted and m to the index of the factor that attains the highest Sharpe Ratio.

It is also of interest to compare performance throughout different deciles. This is not straightforward, as the QFs of different deciles do not necessarily have to have the same underlying interpretation, which makes a direct comparison impossible. However, each model returns an optimal SDF specification for which the highest OOS R^2 is obtained. With these coefficients, an MVE portfolio can be constructed for every decile. The Sharpe Ratios of these portfolios can be compared via the prewhitened HAC inference of Ledoit and Wolf (2008). This results in the following hypothesis:

$$H_0: \qquad max(SR_{\tau*}) = SR_{\,\zeta \cap \tau*^c}, \tag{30}$$

where $\tau*$ is the decile with the maximum Sharpe Ratio and ζ is defined as in Equation (10).

For a further evaluation across deciles, canonical correlation analysis introduced by Hotelling (1992) is considered. In essence, this method tries to find linear combinations of two matrices for which their correlation is maximized and hence whether the factors span the same linear space, that is

$$(a, b) = argmax_{(a,b)} corr(a'X, b'Y), \tag{31}$$

where X and Y correspond to some chosen matrices with equal dimension.

The correlations are sorted (high to low) and can then be tested to be significantly different from 0, which yields

$$H_0: \qquad \hat{\rho}_i = 0, \tag{32}$$

where $\hat{\rho}_i$ is the estimated correlation of the factors at index i.

In this paper, this implies that every combination of decile factor matrices, within its data subset, can be compared. For clearance, the first element of the vector of the retrieved canonical correlations corresponds to the maximum attainable correlation of a linear combination between the factors of different deciles at the same index. Since this method only considers correlation, it is also an indirect comparison method in the sense that the underlying factors do not necessarily have to have the same interpretation. The test statistic for each characteristic i, following Martin and Maes (1979), equals:

$$-(T - 0.5 - H) ln \prod_{j=1}^{H} (1 - \hat{\rho}_j^2) \sim \chi^2(1), \tag{33}$$

under the assumption of a large number of independent observations T, which is true for the considered data subsets.

5 Data

The factor returns F_t used in the analysis of this research consist of different characteristic-based portfolio returns. These portfolios are divided into two subsets based on different characteristics and well-known anomalies in the same spirit as Kozak et al. (2020). The data is retrieved from the website of S. Kozak.[1]

The first subset of portfolios considered is made up of the daily data of **the 25 ME-BM-sorted Fama and French portfolios** from July 1926 to December 2017, provided in the French (2021) online data library. As explained by French (2021), the data underlying these portfolios contains all stocks listed

[1] Data and variable descriptions have been made available by S. Kozak on https://sites.google.com/site/serhiykozak/data

on the NYSE, AMEX, and NASDAQ with two constraints. Firstly, at time t, for a portfolio from July of year t until June of year $t+1$, information about the market equity should be available for December of year $t-1$ until June of year t. Secondly, there should be available information about the book equity of year $t-1$, where the BE/ME ratio in June of year t is defined as the book equity of the last fiscal year end in $t-1$ divided by the market equity in December of year $t-1$. All available stocks are then simultaneously divided into 5 different size and 5 different BE/ME classes, for which the NYSE quintiles serve as breakpoints. The 25 portfolios are then constructed as the intersects of the 5 size and 5 book-to-market groups. Lewellen et al. (2010) show that these portfolio returns have an extremely high factor structure, such that factors considered can be the portfolio returns themselves. Then, the membership of one of these portfolios is the underlying characteristic.

The other subset of portfolios is based on the universe of US firms in the Center for Research in Security Prices (CRSP) database. This subset constructs the factor returns corresponding to one of **50 different stock characteristics that underlie certain anomalies** in the period November 1973 to December 2017. This paper uses the same 50 definitions as Kozak et al. (2020), which are based on the standard anomaly definitions of Novy-Marx and Velikov (2016), Hou et al. (2015), Kogan and Tian (2015), and McLean and Pontiff (2016).

For both portfolio subsets, the factor returns are orthogonalized with respect to the value-weighted index returns of CRSP, where the βs estimated over the full sample are used. After this, the standard deviation of the Fama and French portfolios is further rescaled to have a standard deviation equal to the in-sample excess return of the market index.

In the spirit of Kozak et al. (2020), the focus of this paper is on the cross-sectional aspect of the predictability in return. To remove certain outliers and to keep constant leverage, normalizations are performed that define the characteristic-based factors F_t for the 50 anomalies.

First, as by Asness et al. (2019) and Freyberger et al. (2020) a rank transformation for each characteristic is performed. Let $c^i_{s,t}$ denote characteristic i for stock s at time t. Across all stocks s, the values of a characteristic are sorted and assigned a rank from 1 to n_t, where n_t corresponds to the number of stocks that have this certain characteristic available at time t. Note, the rank 1 is assigned to the lowest and rank n_t to the highest value. When a tie in rank would occur, average rank is assigned to both. Then, the ranks are normalized to get the value of the rank transformation as

$$rc^i_{s,t} = \frac{rank(c^i_{s,t})}{n_t + 1}. \tag{34}$$

Second, the rank-transformed characteristics are normalized by taking the deviation from the cross-sectional mean and dividing it by the absolute sum of deviations from this mean for all stocks as follows:

$$z^i_{s,t} = \frac{rc^i_{s,t} - \bar{rc}^i_t}{\sum_{s=1}^{n_t} |rc^i_{s,t} - \bar{rc}^i_t|}, \tag{35}$$

where $\bar{rc}^i_t = \frac{1}{n_t}\sum_{s=1}^{n_t} rc^i_{s,t}$. All the elements $z^i_{s,t}$ are combined into the instrument matrix Z_t that is used to construct the factor returns as $F_t = Z'_{t-1}R_t$, such that each characteristic has its own factor. Now the factor returns can be seen as characteristic-based portfolio returns. Finally, stocks with a market cap lower than 0.01% of the aggregate market value are excluded from the data at each point in time.

Appendix C shows the descriptive statistics for both sets of portfolios. The mean returns equal the results in the Internet Appendices of Kozak et al. (2020). For every portfolio, it is visible that there exists excess skewness and kurtosis, which hints at a heavy-tailed underlying distribution. This gives an indication that QFA models can be successfully used in this context for dimension-reduction purposes.

6 Results

This section discusses the results obtained by applying the estimation as described in Section 4.1. Again, this section is split up into two main parts that contain the replication and extension. Firstly, the main results of Kozak et al. (2020) are recreated, extended and evaluated. After that, the results of the extension on QFA models are presented in a similar way. All results are obtained with a modified version of the codes provided by S. Kozak and L. Chen[2].

6.1 Replication

This subsection presents a recreation and evaluation of the results of Kozak et al. (2020).

6.1.1 Fama and French portfolios

The first considered set of portfolios contains 25 Fama and French BE/ME sorted portfolios. Table 1 shows the estimated SDF coefficients and their corresponding t-statistics for the raw portfolio returns and the PCs at the optimal level of shrinkage for the dual penalty estimation.

Table 1: *Largest SDF factors for the 25 Fama and French ME/BM sorted portfolios.*
Coefficient estimates and their absolute t-statistics at the optimal prior root expected squared Sharpe Ratio (shrinkage). The left part of the table depicts the raw return portfolios. In the right part the returns are pre-rotated in the PC space and the corresponding estimates are shown. The sample period is daily from July 1926 to December 2017.

Portfolio	b	t-stat	Portfolio	b	t-stat
SMALLHiBM	0.357	1.310	PC 1	-0.534	3.077***
ME3BM4	0.270	0.993	PC 6	0.305	1.136
ME1BM4	0.266	0.979	PC 7	0.222	0.823
ME2BM4	0.226	0.828	PC 11	-0.210	0.764
ME3BM3	0.213	0.784	PC 5	0.185	0.696
ME2BM5	0.209	0.768	PC 2	0.126	0.564
ME4BM4	0.207	0.765	PC 14	-0.148	0.533
ME3BM2	0.207	0.764	PC 10	0.139	0.507
ME5BM4	-0.180	0.672	PC 19	-0.140	0.501
ME2BM1	-0.176	0.655	PC 16	0.139	0.499

Note. * = $p < .10$, ** = $p < 0.05$,*** = $p < 0.01$, where the p-value is two-sided based.

[2]The code provided by S. Kozak is accessible on https://sites.google.com/site/serhiykozak/data and the code of L. Chen is available at https://onlinelibrary-wiley-com.eur.idm.oclc.org/doi/10.3982/ECTA15746?af=R.

This table is not included in the work of Kozak et al. (2020), so the results cannot be compared directly. It is noteworthy that only the first PC is significantly different from 0 at regular statistical significance levels. None of the raw portfolios has a coefficient significantly different from 0, but, as Kozak et al. (2020) mention, the joint significance of linear combinations matters most in the estimation of the SDF.

The OOS R^2 results for the combined L^1- and L^2-Norm are shown in Figure 1.

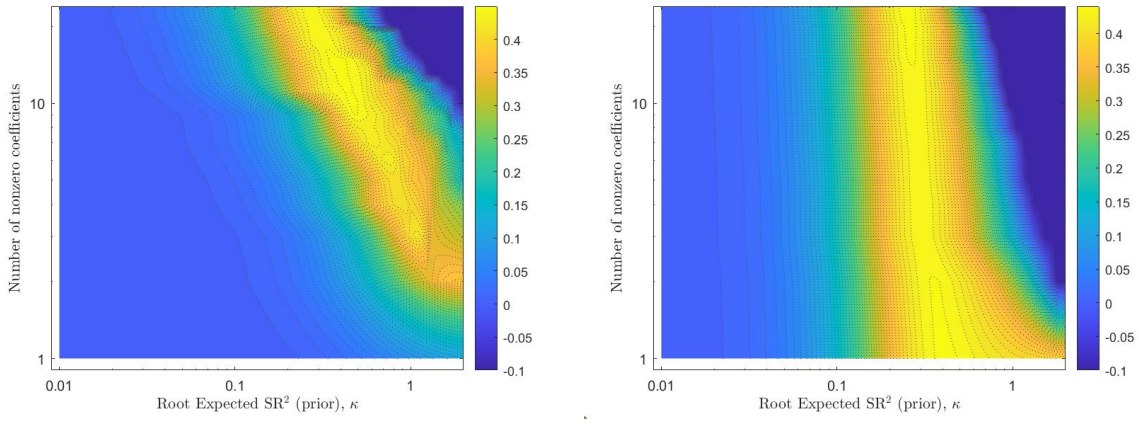

Figure 1: OOS R^2 values for the combined L^1- and L^2-Norm specification for the 25 Fama and French ME/BM sorted portfolios in the period July 1926 to December 2017 (daily). The left panel depicts the OOS cross-sectional R^2 of the dual-penalty specification for the raw portfolio returns, whereas the right panel shows these values for their corresponding PCs. The strength of shrinkage (L^2-Norm) is quantified on the x-axis as the prior root expected squared Sharpe Ratio (κ). The amount of sparsity (L^1-Norm) is quantified on the y-axis as the number of retained variables in the SDF. Note that both axis are on the logarithmic scale and there is a hard minimum of -0.1 imposed on the OOS R^2.

To start, it is important to note how these figures can be interpreted in terms of sparsity and shrinkage. The x-axis of both figures represents the quantity of shrinkage applied to the model defined as the root expected squared Sharpe Ratio or κ. The left side of the figure indicates high levels of shrinkage and the right side corresponds to no shrinkage. The y-axis of both figures designates the levels of sparsity in terms of the number of retained variables in the model. The top of the figure indicates no sparsity, whereas the bottom indicates extreme sparsity.

The figures match the figures of Kozak et al. (2020) exactly. First, consider the left figure, which represents the OOS R^2 values for the raw 25 Fama and French BE/ME sorted portfolio returns. The yellow colours indicate a relatively higher OOS R^2. Shrinkage and sparsity are to some extent substitutes for this data set in terms of a relatively high OOS R^2. This substitution effect is clearly visible as the diagonal yellow stripe. In the top-left of this strip, there is no sparsity and a high level of shrinkage applied, whereas on the bottom-right no shrinkage and high levels of sparsity are applied. As Kozak et al. (2020) state, this result is in line with the expectation for this data set. Due to its extremely high factor structure (Lewellen et al., 2010), it is natural that a high explanatory power can be achieved by taking into account only two or three portfolios. When more portfolios are added, the performance becomes worse unless there is more shrinkage applied to overcome the problem of overfitting.

The right figure considers the PCs of the 25 Fama and French BE/ME sorted portfolios. This figure

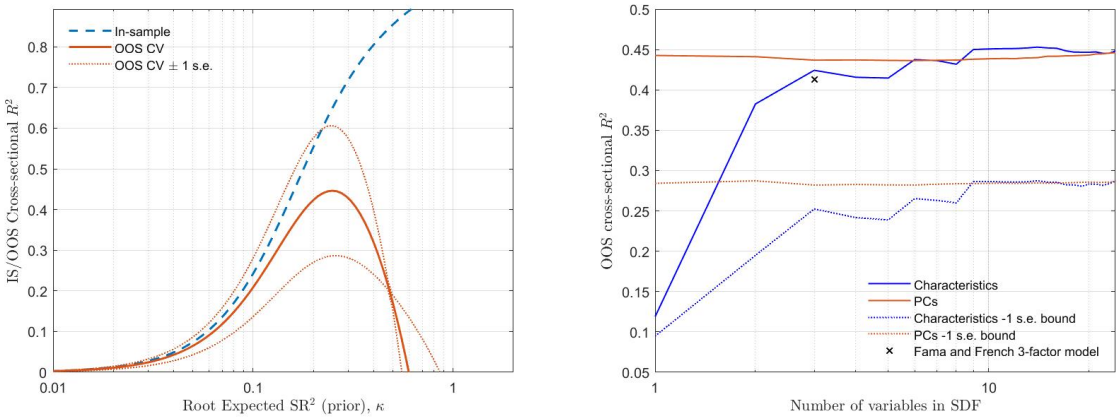

Figure 2: L^2 model selection and the degree of sparsity for the 25 Fama and French ME/BM sorted portfolios in the period July 1926 to December 2017 (daily). The left panel shows the in-sample R^2 as a dashed blue line, together with the cross-validated (CV) OOS cross-sectional R^2 as a solid orange line when no sparsity is imposed. The dotted orange lines indicate the CV OOS cross-sectional R^2 $\pm$ 1 standard error bound of the estimator. The right panel shows the maximum OOS R^2 across all possible L^2 values for models with n factors as specified on the x-axis. The orange dashed line indicates the PCs, the solid blue line indicates the raw characteristics, the dotted lines depict -1 standard error bound for the CV estimator, and the 'x' indicates the performance of the Fama and French three-factor model.

implies that high levels of sparsity can attain a relatively good OOS performance. When the degree of sparsity is decreased, this performance stays stable as long as some shrinkage is applied. However, the substitution effect of the left figure is not present. This is also in line with expectations because PCs inherently down-weight low-eigenvalue PCs. Hence, to maintain a high OOS R^2, a higher level of shrinkage is not required when retaining more variables in the model.

Kozak et al. (2020) restrict these figures to have a hard minimum bound on the OOS R^2 of -0.1. Figure D1 in Appendix D shows the contour maps when this bound is removed. Clearly, Kozak et al. (2020) impose this bound because the image is heavily distorted by the relatively extreme low outliers. That is, this figure contains almost exactly the same information as Figure 1, but the main conclusion has become invisible. The reason, I added these graphs is that it highlights how serious the problem of overfitting is in this high-dimensional setting. When no shrinkage is applied in combination with no sparsity, the relative performance of the models becomes substantially inferior.

Figure 2 presents two cuts of the contour plot. The left figure matches the figure of Kozak et al. (2020) precisely. It illustrates the horizontal cut of the raw 25 portfolio returns contour plot at the top edge, i.e. where no sparsity is imposed. Most interestingly to note about this figure is that the difference between the orange and blue lines shows that an in-sample R^2 can be very misleading about the performance of an SDF model. Especially, when no shrinkage is applied. The orange line shows that the maximum OOS R^2 is attained at $\kappa \approx 0.23$, whereas maximising the in-sample R^2 would imply no shrinkage at all.

The right panel of Figure 2 shows the maximum OOS R^2 for different levels of sparsity over all possible values of shrinkage. This graph confirms the results that are visible in Figure 1. The raw characteristics can attain almost identical maximum OOS R^2 values to the PCs when more than 3

variables are included, whereas they perform worse than the PCs when less than 3 variables are included. This indicates that PCs can successfully achieve a characteristic-sparse SDF specification. The graph also shows a cross that indicates the performance of the Fama and French three-factor model. Except for this cross, this figure is identical to the figure of Kozak et al. (2020). In their research, the cross is on the same level, but one point to the left of the cross in this figure. This difference comes from the fact that I decided to include all three factors instead of only the HML and SMB. With this, I have chosen to do no orthogonalization with respect to the market instead of a conditional orthogonalization to the market. The reason for this is that conditional orthogonalization is not documented in their research or code. However, the introduction of the excess market return as a factor yields an equivalent result.

6.1.2 Anomaly portfolios

The second considered set of portfolios contains the 50 anomaly portfolios, as described by Kozak et al. (2020). Table 2 contains the estimated SDF coefficients and their corresponding t-statistics for the raw returns and the PCs for the optimal level of shrinkage in the dual penalty specification. The results

Table 2: *Largest SDF factors for the 50 anomaly portfolios.*
Coefficient estimates and their absolute t-statistics at the optimal prior root expected squared Sharpe Ratio (shrinkage). The left part of the table depicts the raw return portfolios. In the right part the returns are pre-rotated in the PC space and the corresponding estimates are shown. The sample period is daily from November 1973 to December 2017.

Portfolio	b	t-stat	Portfolio	b	t-stat
Industry rel. rev. (L.V.)	-0.879	3.527***	PC 4	1.014	4.249***
Ind. mom-reversals	0.484	1.945**	PC 1	-0.537	3.081***
Industry rel. reversals	-0.425	1.705*	PC 2	-0.556	2.653***
Seasonality	0.322	1.292	PC 9	0.635	2.514**
Earnings surprises	0.323	1.291	PC 15	-0.324	1.265
Value-profitability	0.297	1.184	PC 17	0.303	1.182
Return on market equity	0.299	1.183	PC 6	-0.287	1.176
Investment/Assets	-0.238	0.948	PC 11	0.189	0.744
Return on equity	0.238	0.947	PC 13	0.166	0.653
Composite issuance	-0.240	0.947	PC 23	0.146	0.564

Note. * = $p < .10$, ** = $p < 0.05$,*** = $p < 0.01$, where the p-value is two-sided based.

in this table are equivalent to the results from Kozak et al. (2020), except for the number of decimals. As Kozak et al. (2020) state, it is not surprising that the characteristics shown in the table have the most significant SDF coefficients, as they are among the most robust anomalies found in the literature. Note that, only 'Industry relative reversals (low vol)', has a coefficient that significantly differs from 0 at conventional significance levels. Kozak et al. (2020) mention that the joint, and not the individual, significance of linear combinations matters most for the SDF. On the other hand, there are four significant SDF coefficients for the PCS, namely PC4, PC1, PC2, and PC9.

Figure 3 shows the OOS R^2 values for the combined L^1- and L^2-Norm specification, which bears

exactly the same interpretation as Figure 1 in terms of shrinkage and sparsity.

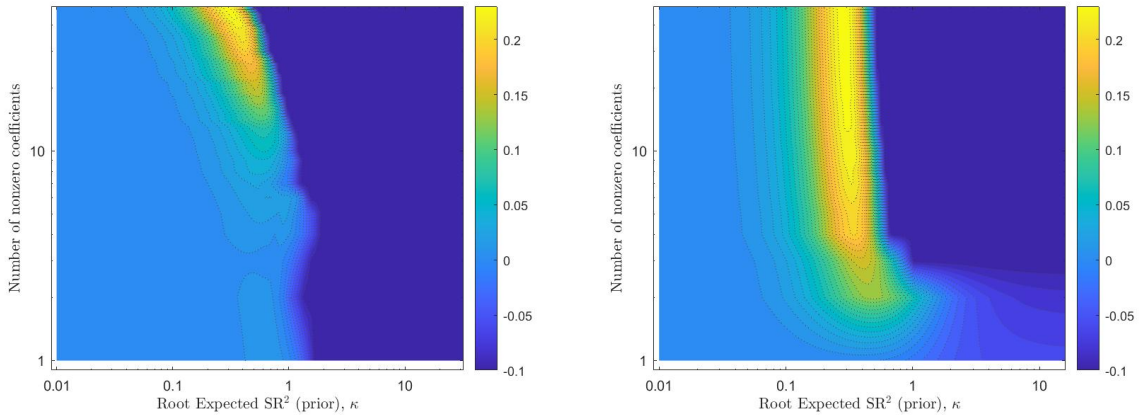

Figure 3: OOS R^2 values for the combined L^1- and L^2-Norm specification for the 50 anomaly portfolios in the period November 1973 to December 2017 (daily). The left panel depicts the OOS cross-sectional R^2 of the dual-penalty specification for the raw portfolio returns, whereas the right panel shows these values for their corresponding PCs. The strength of shrinkage (L^2-Norm) is quantified on the x-axis as the prior root expected squared Sharpe Ratio (κ). The amount of sparsity (L^1-Norm) is quantified on the y-axis as the number of retained variables in the SDF. Note that both axis are on the logarithmic scale and there is a hard minimum of -0.1 imposed on the OOS R^2.

Both figures are again equivalent to the figures of Kozak et al. (2020). The left panel of Figure 3 shows the contour map of the OOS R^2 for the raw 50 anomaly portfolio returns. It is immediately visible that this set of portfolios leads to different conclusions than the Fama and French portfolios. This is not strange, as these anomalies do not have an inherently strong factor structure. The figure shows that a maximum OOS R^2 can only be attained with a high degree of shrinkage combined with the inclusion of almost all variables. This indicates that almost none of the 50 characteristics are redundant for explaining the variation in the SDF.

The right panel of Figure 3 shows the contour map of the OOS R^2 for the PCs of the 50 anomaly portfolios. The PCs can attain a much higher degree of sparsity for a relatively high OOS R^2 values than the raw characteristics. However, there still needs to be a substantial amount of shrinkage to obtain these high values. Also, the degree of sparsity that can result in an optimal performance is not as high as for the Fama and French portfolios, i.e. still ten PCs are required to come close to the maximum OOS R^2. Figure D2 in Appendix D shows the contour maps without the hard minimum bound on the OOS R^2 of -0.1. This figure again shows the seriousness of the problem of overfitting even more than Figure D2. The OOS R^2 values for models with no shrinkage and no sparsity are extremely low.

Figure 4 shows two cuts of the contour plots, which are defined in exactly the same way as Figure 2. Both figures, again, match the figures of Kozak et al. (2020), except that the figure in the right panel has an extended $y-axis$. The left panel of Figure 4 shows the horizontal cut of the OOS R^2 contour map at the point of no sparsity. It is once more visible that the in-sample R^2 can lead to a misleading conclusion when no shrinkage is imposed. This graph also shows the OOS R^2 when the more commonly used prior of $\eta = 1$, by Pástor and Stambaugh (2000) would have been used. There is a slight difference with the same line in the graph of Kozak et al. (2020). This is due to the fact that it is impossible to

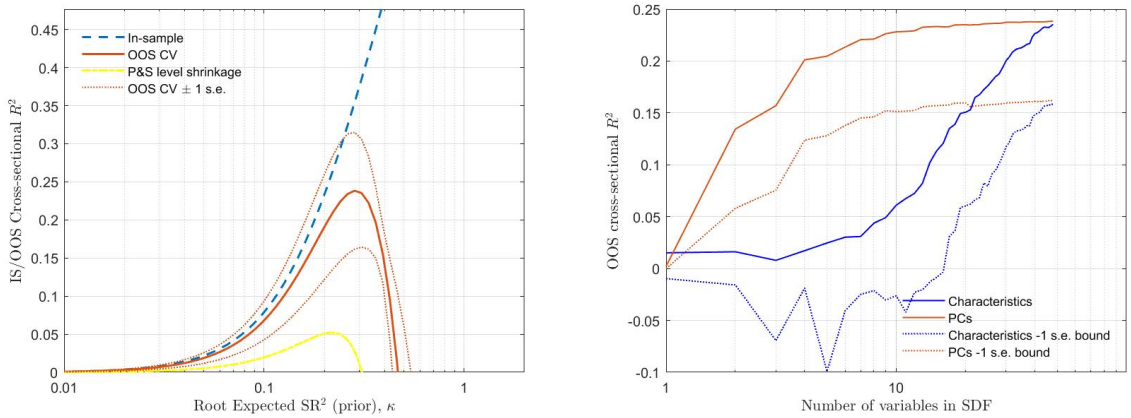

Figure 4: L^2 model selection and the degree of sparsity for the 50 anomaly portfolios in the period November 1973 to December 2017 (daily). The left panel shows the in-sample R^2 as a dashed blue line, together with the cross-validated (CV) OOS cross-sectional (CS) R^2 as a solid orange line when no sparsity is imposed. The dotted orange lines indicate the CV OOS CS $R^2 \pm 1$ standard error bound of the estimator. The yellow line indicates the CV OOS CS R^2 when equal shrinkage, as by Pástor and Stambaugh (2000), with $\eta = 1$ is imposed. The right panel shows the maximum OOS R^2 across all possible L^2 values for models with n factors as specified on the x-axis. The orange dashed line indicates the PCs, the solid blue line indicates the raw characteristics, and the dotted lines depict -1 standard error bound for the CV estimator.

recreate the exact settings of the positioning of the line with respect to the $x - axis$ because this is not documented in the paper or in the code. To recreate the graph, I have chosen to position the line with the help of the level of κ that achieves the overall optimum for this model. With this line, it can be concluded that the chosen prior specification improves the OOS R^2 substantially. Note that, the x-axis for this line does not bear the interpretation of κ anymore, since, under the prior of $\eta = 1$, κ does not equal the root expected squared Sharpe Ratio.

The right panel of Figure 4 shows the maximum OOS R^2 for different levels of sparsity across all possible levels of shrinkage. This graph yields a similar conclusion as Figure 3. The raw return portfolios can only attain an optimal OOS R^2 when almost all characteristics are included. The PCs have a relatively better performance for more characteristic-sparse SDF specifications. The maximum OOS R^2 stays at a rather constant level when more than ten PCs are retained in the model.

6.2 Extension

This section presents an extension of the analysis of Kozak et al. (2020) by considering a QFA type factor model by Chen et al. (2019).

6.2.1 General analysis

The analysis is split up into two parts. First, every decile is considered as its own separate universe. Secondly, all factors of all deciles are combined into one large set of factors. This combination takes the full distribution, split up in deciles, of the returns. It can show what the effect is of taking into account a lot of information without considering the explanatory structure of the individual models and correlation between factors first.

Figure 5 presents the contour maps for the OOS R^2 of the dual penalty specification for each decile for the 25 ME/BM sorted Fama and French portfolios for the period July 1926 to December 2017, similar to Figure 1. The interpretation of these figures is equivalent to Figures 1 and 3. The QFA factors are not orthogonal since they follow the standardization of Equation (23). There are clear differences present between the OOS R^2 patterns for different deciles. Overall, the deciles further away from the median perform better in terms of achieving a characteristic-sparse model with a relatively high OOS R^2. The earlier discussed relation between the median quantile and PCA now gives a useful insight. Namely, relative outliers seem to have a higher explanatory power in the SDF. This is also visible from the fact that the tails have a better performance. Still, there is quite some degree of shrinkage necessary to obtain a relatively high OOS R^2. Interestingly, the tails have a better optimal OOS R^2 than the raw characteristics and the PCs.

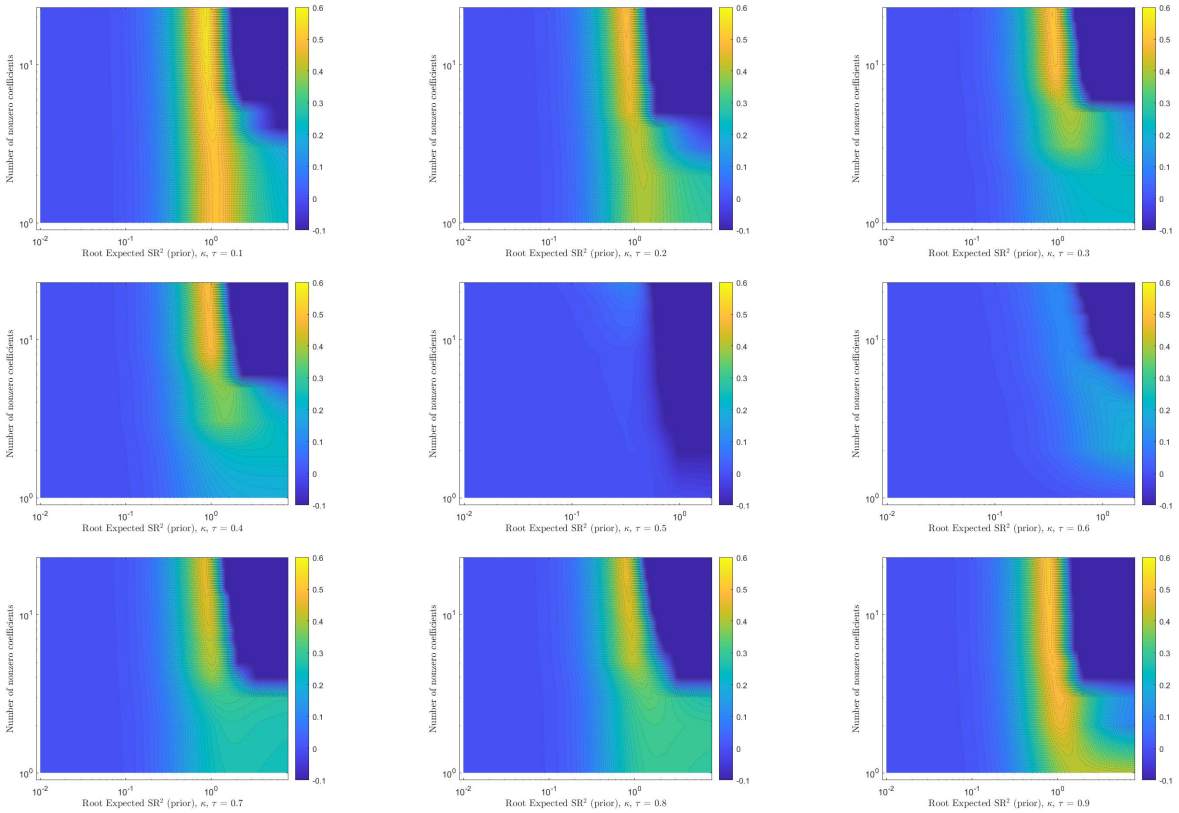

Figure 5: OOS R^2 values for the combined L^1- and L^2-Norm specification for the 25 Fama and French ME/BM sorted portfolios in the period July 1926 to December 2017 (daily). The strength of shrinkage (L^2-Norm) is quantified on the x-axis as the prior root expected squared Sharpe Ratio (κ). Also, the decile τ is specified on the x-axis. The amount of sparsity (L^1-Norm) is quantified on the y-axis as the number of retained variables in the SDF. Note that both axis are on the logarithmic scale, there is a hard minimum of -0.1 imposed on the OOS R^2 and the legend is standardised to be equal across figures.

Figure 6 presents the contour maps for the OOS R^2 of the dual penalty specification for each decile for the 50 anomaly portfolios for the period November 1973 to December 2017, similar to Figure 3. The interpretation of these figures is equal to Figures 1, 3 and 5. Again, the deciles that lie further away from the median can obtain a higher optimal OOS R^2 value, which hints that there is more explanatory

power in the relative outliers. However, there is not a clear cut conclusion about which decile can obtain a characteristic-sparse SDF most successfully. Overall, all QFA specifications show that there is almost no redundancy among the anomalies as explanatory factors, similar to the raw characteristics case as in the left panel of Figure 3. Even though there is no characteristic-sparse specification that can obtain a relatively high OOS R^2 within the model, the optimal values are still higher than the optimal values for the raw characteristics and PCs.

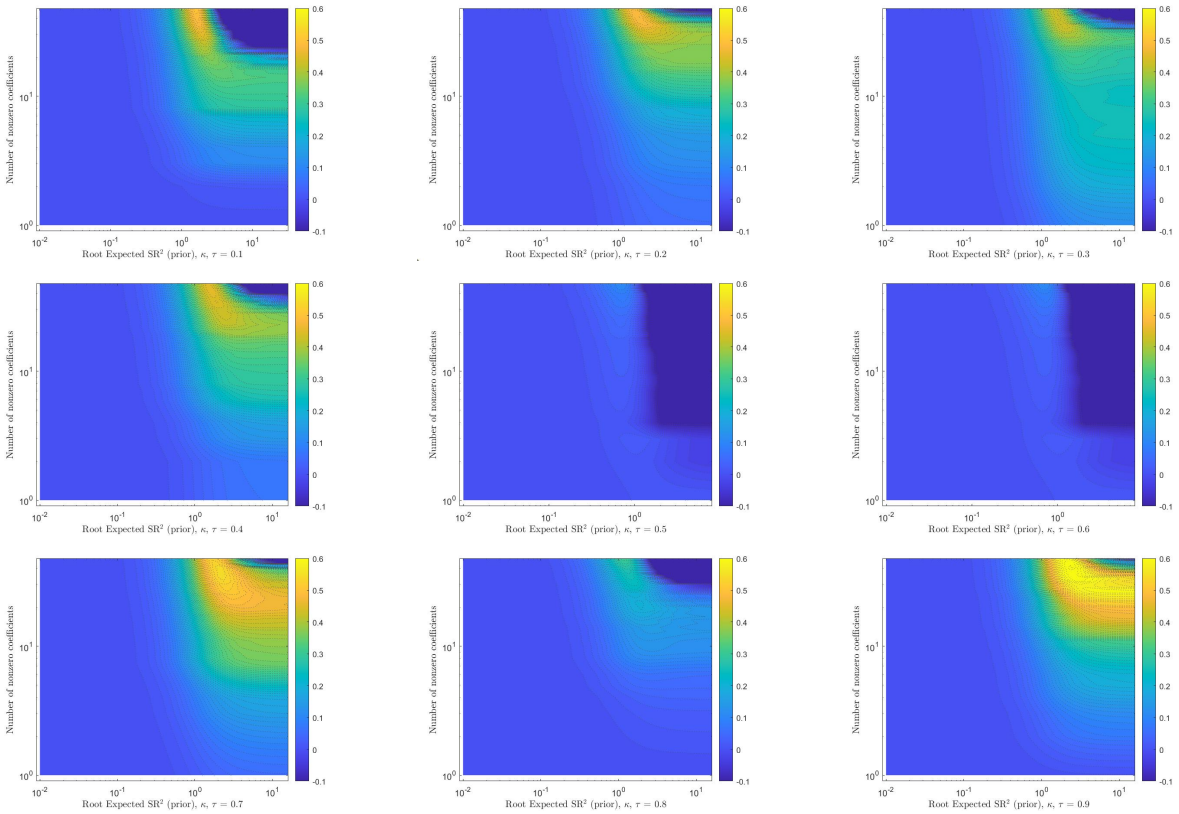

Figure 6: OOS R^2 values for the combined L^1- and L^2-Norm specification for the 50 anomaly portfolios in the period November 1973 to December 2017 (daily). The strength of shrinkage (L^2-Norm) is quantified on the x-axis as the prior root expected squared Sharpe Ratio (κ). Also, the decile τ is specified on the x-axis. The amount of sparsity (L^1-Norm) is quantified on the y-axis as the number of retained variables in the SDF. Note that both axis are on the logarithmic scale, there is a hard minimum of -0.1 imposed on the OOS R^2 and the legend is standardised to be equal across figures.

Since it is difficult to compare the deciles precisely from these contour maps, Figure 7 shows the maximum attainable OOS R^2 values for all levels of possible shrinkage and for every QFA decile. These figures confirm the results of the contour plots of Figures 5 and 6. The deciles in the tails of the distribution can attain high levels of OOS R^2, whereas the median decile and the deciles closest to the median perform substantially worse. Note that, in the right panel, the values for the 5^{th} and 6^{th} are so similar that the lines overlap. It is striking that, compared to Figures 2 and 4, the maximum achievable values for QFA models are approximately 12% and 36% higher for the Fama & French and anomaly portfolios, respectively. Figure 7 also shows that the contour maps actually paint a distorted image because all aforementioned analysis states relative results within models. The contour maps of Section 6.1 are for the purpose of reproduction not standardized. To get a better idea of the comparison across

different techniques, Appendix E shows the contour maps of the reproduction on the same standardised scale as Figures 5, 6, and 8. Overall, for both subsets, there always exists a QFA model that outperforms both the raw characteristics and PCs for every possible number of retained variables in terms of maximum OOS R^2. This leads to the conclusion that the reasoning of Kozak et al. (2020) is only consistent within the framework of their chosen models and is not robust to changes in the statistical construction of the factors. The results suggest that the SDF of the 50 anomaly portfolios cannot adequately be described by a characteristic-sparse specification, even when the PCs are considered as factors.

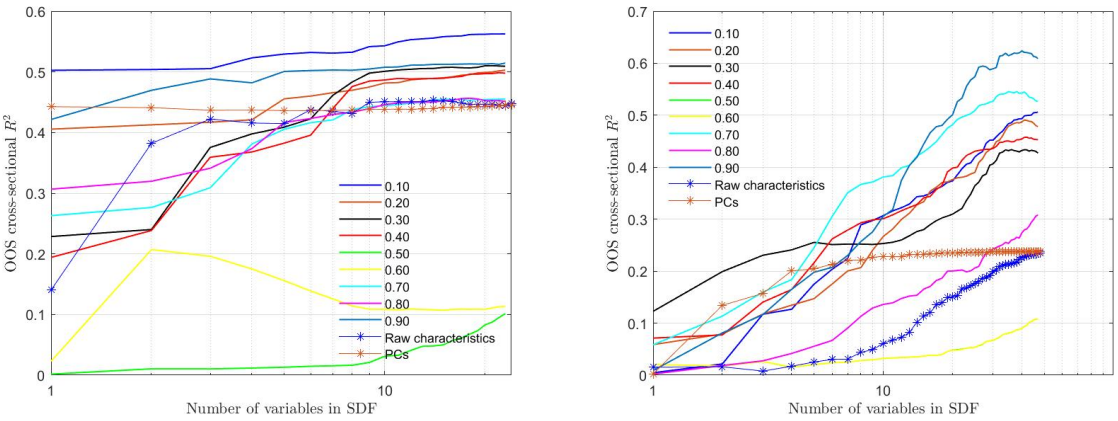

Figure 7: The maximum OOS R^2 across all possible L^2 values for models with n factors as specified on the x-axis. The left panel depicts the values of all QFA deciles for the 25 Fama and French ME/BM sorted portfolios in the period July 1926 to December 2017 (daily), whereas the right panel depicts this information for the 50 anomaly portfolios in the period November 1973 to December 2017 (daily). The graphs of right panels of Figures 2 and 4 are indicated with asterisks.

Figure 8 shows the contour maps of the OOS R^2 for both data sets when combining all QFA factors of each decile into one large set. The overall conclusions already drawn from the contour maps in Figures 5 and 6 do not change. However, the optimal OOS R^2 values are lower than in some of the separate decile models. Note that this comparison can be made because the OOS R^2 is a metric that tests predictive ability for out-of-sample data, dimension has no role in this. This shows that a random combination of models does not per se outperform the individual models. The methods and algorithm have two drawbacks that lead to this result. Firstly, by creating this space of factors, one is not only extending the explanatory variables but also the variables to be explained, so the results do not have to be at least as good as some individual models. The second drawback comes from the fact that the algorithm selects the next most significant factor to be retained in the model when increasing the total number of retained variables. This in itself does not take into account any sort of correlation or joint significance between the factors. It may very well be the case that for any number of retained variables there is a combination possible that outperforms the results presented here in terms of OOS R^2. However, this algorithm is not designed to be optimal in terms of OOS R^2, only to be optimal for the minimization of the penalty specification. It does prove that using the full distribution is not the optimal strategy and that the explanatory power of the tails is substantially higher. This re-confirms what is also visible in Figures 5, 6, and 7. Note that this conclusion can only be drawn because all factors of one combined set are based on the same underlying data set and that this does not prove that the chosen model methodology is not

robust to changes in the data.

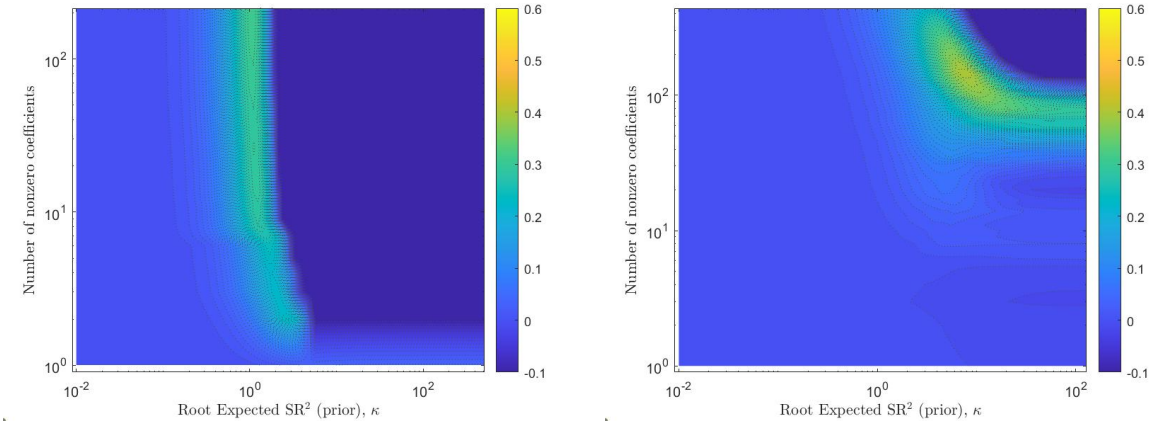

Figure 8: OOS R^2 values for the combined L^1- and L^2-Norm specification. The left panel depicts the OOS cross-sectional R^2 of the dual-penalty specification for all QFA factors for each decile combined into one set for the 25 Fama and French ME/BM sorted portfolios in the period July 1926 to December 2017 (daily), whereas the right panel shows these values for the 50 anomaly portfolios in the period November 1973 to December 2017 (daily). The strength of shrinkage (L^2-Norm) is quantified on the x-axis as the prior root expected squared Sharpe Ratio (κ). The amount of sparsity (L^1-Norm) is quantified on the y-axis as the number of retained variables in the SDF. Note that both axis are on the logarithmic scale, there is a hard minimum of -0.1 imposed on the OOS R^2 and the legend is standardised to be equal across figures.

To gain more in-depth insights into the correlation of different deciles, canonical correlation analysis is considered. It can be seen from Figures 5 and 6 that some deciles follow patterns that very much look alike. Hence, it can be expected that there is some kind of relation between the factors of different deciles. This is also the case. The canonical correlation shows extremely high levels of correlation between factors of any combination of deciles (i.e. there are 36 unique combinations per data subset). Firstly, for the Fama and French portfolios, the average correlation of factors from different deciles at the same index equals 0.959. Only a few correlations differ absolutely from 1, that is, on average, 1.806 per combination. The number of correlations that absolutely deviates from 1 never exceeds 3. From the total number of 864 tested correlations, only 4 cannot be said to be significantly different from 0. Secondly, for the anomaly portfolios, the average correlation equals 0.978, with an average number of correlations that deviate absolutely from 1 being equal to 2.361 per combination. Again, this number never exceeds 3. From the total number of 1764 tested correlations, only 5 do not differ significantly from 0.

These results show that only a few of the factors make the difference between the deciles in terms of their out-of-sample explanatory power. If factors are perfectly correlated (within a linear combination framework), it should not matter which one is used in the pricing of the SDF and hence, this does not account for any difference. Ultimately, this also explains why the combination of all deciles does not perform well. As previously mentioned, the algorithm does not account for any correlation but only absolute significance.

6.2.2 Statistical comparison of QFA specifications

The individual factors in all QFA models can be seen as characteristic-based portfolios. Furthermore, by construction, they are linear with respect to the deciles of the characteristics. Table 3 shows the maximum Sharpe Ratio for each decile and the proportion of Sharpe Ratios of other factors for that respective decile that is significantly lower at a 5% level. Thus, the second and fourth columns show the fraction of the other factors within each decile that has a significantly lower Sharpe Ratio than the maximum. The significance is determined by the prewhitened HAC inference of Ledoit and Wolf (2008). This paper makes use of an adaptation of the code of Ledoit and Wolf (2008) to obtain the results.[3]

Table 3: *Maximum attainable Sharpe Ratio for the individual factors of each quantile and the corresponding proportion of factors that have a significantly lower Sharpe Ratio at a 5% level. The significance is determined based on pre-whitened HAC test statistic of Ledoit and Wolf (2008). The maximum Sharpe Ratio per subset is indicated in bold.*

	Fama and French		Anomalies	
τ	Max. SR	Proportion sign. higher	Max. SR	Proportion sign. higher
0.10	0.066	0.96	0.251	1.00
0.20	0.038	0.78	0.219	1.00
0.30	0.361	1.00	**0.322**	1.00
0.40	0.351	1.00	0.206	1.00
0.50	0.014	0.26	0.039	0.56
0.60	0.024	0.65	0.097	0.98
0.70	0.029	0.78	0.157	1.00
0.80	0.026	0.70	0.130	0.90
0.90	**0.580**	1.00	0.267	0.98

There is no clear pattern visible regarding which decile can outperform in terms of Sharpe Ratio. However, it can be stated that, for the anomaly subset, when a factor achieves the highest Sharpe Ratio, this value is almost always significantly higher than all other Sharpe Ratios. This image is a bit different for the Fama and French subset, but the maximum Sharpe Ratios are still significantly higher than the majority. For a more in-depth view of these Sharpe Ratios, Figure 9 presents the efficient frontiers and scatter plots of the individual QFs as characteristic-based portfolios.

[3]The code of Ledoit and Wolf (2008) is available at https://www.econ.uzh.ch/en/people/faculty/wolf/publications.html#9.

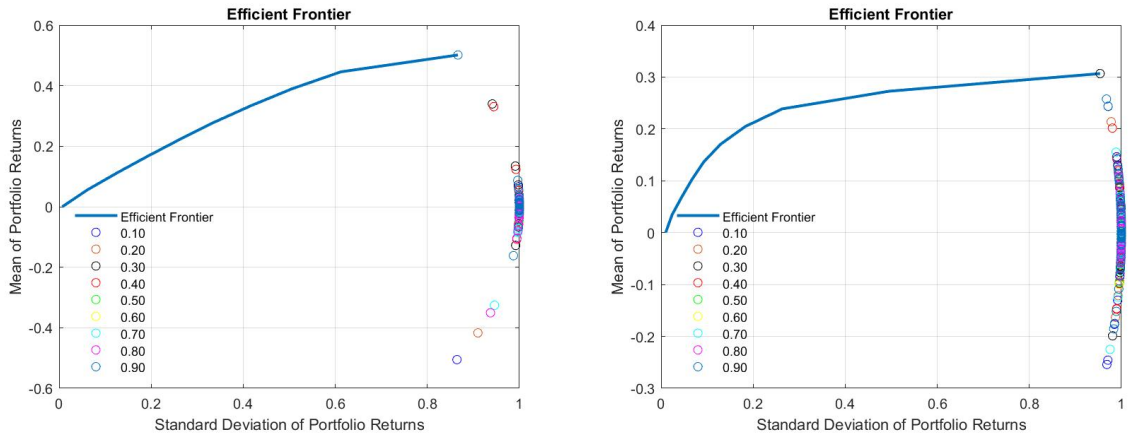

Figure 9: Efficient frontiers combined with a scatter plot of the individual characteristic-based portfolios in the QFA universe. The left panel depicts the frontier and scatter plot for QFs of the 25 Fama and French ME/BM sorted portfolios in the period July 1926 to December 2017 (daily), whereas the right panel shows these values for the QFs of the 50 anomaly portfolios in the period November 1973 to December 2017 (daily). All deciles are included.

There is a sideways parabola structure visible for both subsets of portfolios in the standard deviations. The factors with a standard deviation that deviate most from 1 correspond to the deciles furthest away from the median, that is $\tau = 0.10, 0.20, 0.80$ and 0.90. This result is in accordance with the empirical findings of Chen et al. (2019). They do not focus on SR, but the results for the volatility are similar. Namely, they find that the tail distribution of financial data is more stable than the median. It has to be noted that Chen et al. (2019) find a much greater difference in volatility, but this is due to the fact that they use more extreme quantiles, i.e. $\tau = 0.01$ and 0.99. There exists one portfolio that lies exactly on the efficient frontier and that is the one with the highest Sharpe Ratio as indicated in bold in Table 3. This implies that for both data sets the factor with the highest Sharpe Ratio is an efficient portfolio.

Since the optimal SDF coefficients can be seen as the weights of the MVE portfolio, the returns of these portfolios can be constructed. Table 4 shows the Sharpe Ratios of the MVE portfolios per decile. The main difference when compared to the individual Sharpe Ratios of Table 3 is that the Sharpe Ratios of the MVE portfolios are always larger, sometimes even substantially, than the individual maximum Sharpe Ratios. In this portfolio universe, it is again visible that the tails furthest away from the median have the best relative performance in terms of Sharpe Ratio. The decile for which the maximum SR is attained differs between the data subsets, but the SR of the first and ninth decile have no significant difference for the Fama and French portfolios. This gives the indication that the relative outliers are more useful for the construction of MVE portfolios.

Table 4: *Sharpe ratios of the full SDF MVE portfolios. The highest value is indicated in bold. Asterisks indicate a significant difference with the maximum Sharpe Ratio as by the prewhitened HAC inference of Ledoit and Wolf (2008).*

τ	Fama and French	Anomalies
0.10	0.616	**0.769**
0.20	0.505**	0.556**
0.30	0.445**	0.595**
0.40	0.418**	0.454**
0.50	0.047**	0.226**
0.60	0.047**	0.226**
0.70	0.388**	0.501**
0.80	0.428**	0.390**
0.90	**0.651**	0.736*

Note. * $= p < 0.05$,** $= p < 0.01$, where the p-value is based on prewhitened HAC inference of Ledoit and Wolf (2008)

Figure 10 shows the efficient frontiers combined with a scatter plot of the MVE portfolios. Even though all the separate portfolios are mean-variant efficient within the deciles, only one of them lies exactly on the efficient frontier for both subsets. This is in both cases the portfolio of the 9^{th} decile. Once again, the combined data set under-performs the individual factors, since the efficient frontiers in Figure 10 lie below the frontiers in Figure 9.

All in all, the results of the statistical comparison share a similar conclusion with the contour maps of Figures 5 and 6. The tail deciles can produce the most stable and best-performing output in terms of OOS R^2 values and portfolio properties for both data subsets. This is in line with the empirical results on financial data of Chen et al. (2019).

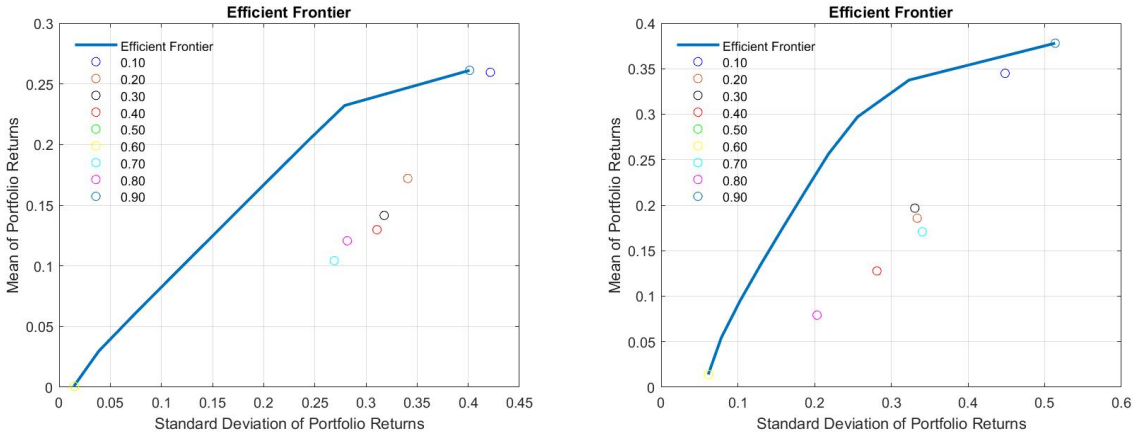

Figure 10: Efficient frontiers combined with a scatter plot of the MVE portfolios in the QFA universe. The left panel depicts the frontier and scatter plot for MVE portfolios based on the 25 Fama and French ME/BM sorted portfolios in the period July 1926 to December 2017 (daily), whereas the right panel shows these values for the MVE portfolios based on the 50 anomaly portfolios in the period November 1973 to December 2017 (daily).

7 Discussion and Conclusion

The goal of this paper is to answer two main research questions. The first research question aims to reproduce, verify, and evaluate the main results found by Kozak et al. (2020) as follows:

To what extent can a characteristic-sparse stochastic discount factor representation successfully summarize the cross-sectional variation in stock returns?

The second research question extends the research of Kozak et al. (2020) by considering a novel type of factor model introduced by Chen et al. (2019). This results in the following research question:

In what manner can quantile factor analysis be applied to characteristic-based factors to summarize the cross-sectional variation in stock returns?

To answer both questions, two sets of data are considered. These consist of the daily data of the 25 ME-BM-sorted Fama and French portfolios from July 1926 to December 2017, provided in the French (2021) online data library, and daily data of 50 different stock characteristics that underlie certain anomalies in the period November 1973 to December 2017. There are 11 different factor specifications examined: raw characteristics, principal components, and nine decile specifications of quantile factors. For the estimation of the stochastic discount factor, this paper uses the same principle as Kozak et al. (2020), which consists of maximising the out-of-sample R^2 values for a dual penalty specification based on a novel prior that penalizes with an L^1- and L^2-norm penalty. This estimation is executed via a three-fold cross-validation.

This research confirms the main results found by Kozak et al. (2020). The Fama and French portfolios have a strong inherent factor structure. For the raw characteristics of this data set, there exists a trade-off between sparsity and shrinkage for achieving a high out-of-sample R^2, i.e. more shrinkage is necessary when less sparsity is applied. For the principle components and the quantile factors of the tails of the distribution, this trade-off is not present. With a relatively high amount of shrinkage, these specifications can relatively, within the models themselves, yield a successful characteristic-sparse SDF specification.

The anomaly portfolios have no inherently strong factor structure and hence yield a different conclusion. The raw characteristics and all quantile factor deciles show that there is relatively, within the model, no redundancy among the factors to obtain a high out-of-sample R^2. The principal components, on the other hand, can achieve a characteristic-sparse SDF specification within its own universe.

I have found that the relative performances within the model can result in a distortion of the appropriate conclusion. For any number of retained variables in the model, there always exists a quantile factor decile specification that can outperform the raw characteristics and principal components in terms of out-of-sample R^2. This implies that the results of Kozak et al. (2020) are not robust to changes in the statistical construction of the factors.

Furthermore, the quantile factor analysis shows that the tails of the distribution provide more stable results in terms volatility when the factors are treated as characteristic-based portfolios. Additionally, the factors across different deciles show extremely high levels of canonical correlation, i.e. there exist linear

combinations of the factors that are almost perfectly correlated with each other. This indicates only a few of the factors are accountable for the difference in results across deciles. Next to that, the factors and the full mean-variant efficient portfolios of the tail deciles provide a significantly higher Sharpe Ratio and are sometimes part of the efficient frontier. In the out-of-sample analysis, the tail deciles also provide the highest out-of-sample R^2.

All in all, it can be stated that, for data sets with no inherently strong factor structure, a characteristic-sparse SDF specification cannot adequately capture the variation in this SDF. In this conclusion, there is no distinction between raw data, principle components or quantile factors. Quantile factors may not be able to provide a characteristic-sparse specification successfully, but the results show that there is a great explanatory power for the variation in the SDF in moments higher than the mean and that relative outliers (tails) have an important role in this. This result is in accordance with the original quantile factor model research of Chen et al. (2019). Lastly, it is also shown that quantile factors create beneficial properties for the construction of characteristic-based portfolios.

The main limitation of this paper is that the results are not robust to changes in the statistical construction of the factors. There may or may not exist certain factor specifications which can yield even higher out-of-sample R^2. These models could be able to prove that a successful characteristic-sparse SDF specification can be created. The canonical correlation analysis has shown an inherently strong linear correlation between different deciles. It can be expected that a model consisting of only, relatively, uncorrelated factors can outperform all other models in terms of OOS R^2. Thus, a suggestion for future research is to try and find such a specification. One should approach this type of research with great care because data-mining or lucky exceptions lie in wait.

Another limitation of this research to consider is the rather narrow scope of data. The US market, on which the data is based, is a well-developed and highly researched area. It could very well be the case that, for example, emerging market data can lead to different conclusions. It would be worthwhile to conduct this research for different markets or financial products and to qualitatively and quantitatively investigate any differences in results.

The last limitation of this research to discuss is whether the applied Bayesian structure and shrinkage is a suitable method for quantile factor analysis. Since the factors are linear with respect to the quantiles but non-linear with respect to the full set of data, it is unclear if all methods applied behave as they are expected to in this framework. To investigate this goes beyond the scope of this research and is a suggestion for further research.

References

Asness, C. S., Frazzini, A., & Pedersen, L. H. (2019). Quality minus junk. *Review of Accounting Studies*, *24*(1), 34–112.

Bai, J., & Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, *70*(1), 191–221.

Barnes, M., & Hughes, A. (2002). A quantile regression analysis of the cross section of stock market returns. *SSRN Electronic Journal*.

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

Chen, L., Dolado, J. J., & Gonzalo, J. (2019). Quantile factor models. *arXiv preprint arXiv:1911.02173*.

Chiang, T. C., & Li, J. (2012). Stock returns and risk: Evidence from quantile regression analysis. *Journal of Risk and Financial Management*, *5*(1), 20–58.

Cochrane, J. H. (1996). A cross-sectional test of an investment-based asset pricing model. *Journal of Political Economy*, *104*(3), 572–621.

DeMiguel, V., Garlappi, L., Nogales, F. J., & Uppal, R. (2009). A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Management science*, *55*(5), 798–812.

DeMiguel, V., Martin-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.

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

Feng, G., Giglio, S., & Xiu, D. (2017). Taming the factor zoo. *Chicago Booth research paper*, (17-04).

French, K. R. (2021). Kenneth r. french - data library. *Kenneth R. French - Data Library*. https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

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

Green, J., Hand, J. R., & Zhang, X. F. (2017). The characteristics that provide independent information about average us monthly stock returns. *The Review of Financial Studies*, *30*(12), 4389–4436.

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

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, *24*(6), 417–441.

Hotelling, H. (1992). Relations between two sets of variates. *Breakthroughs in statistics* (pp. 162–190). Springer.

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

Koenker, R., & Bassett, G. J. (1978). Regression quantiles. *Econometrica*, *46*(1), 33–50.

Koenker, R., & Hallock, K. F. (2001). Quantile regression. *Journal of economic perspectives*, *15*(4), 143–156.

Kogan, L., & Tian, M. H. (2015). Firm characteristics and empirical factor models: A model-mining experiment. *FRB International Finance Discussion Paper*.

Kozak, S., Nagel, S., & Santosh, S. (2018). Interpreting factor models. *The Journal of Finance*, *73*(3), 1183–1223.

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

Ledoit, O., & Wolf, M. (2008). Robust performance hypothesis testing with the sharpe ratio. *Journal of Empirical Finance*, *15*(5), 850–859.

Ledoit, O., & Wolf, M. (2004). Honey, i shrunk the sample covariance matrix. *The Journal of Portfolio Management*, *30*(4), 110–119.

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

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

Markowitz, H. M. (1959). *Portfolio selection: Efficient diversification of investments*. Yale University Press. http://www.jstor.org/stable/j.ctt1bh4c8h

Martin, N., & Maes, H. (1979). *Multivariate analysis*. Academic press London.

McLean, R. D., & Pontiff, J. (2016). Does academic research destroy stock return predictability? *The Journal of Finance*, *71*(1), 5–32.

Novy-Marx, R., & Velikov, M. (2016). A taxonomy of anomalies and their trading costs. *The Review of Financial Studies*, *29*(1), 104–147.

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

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

Pearson, K. (1901). Liii. on lines and planes of closest fit to systems of point in space. *Philosophical Magazine*, *2*(11), 559–572.

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.

Roll, R., & Ross, S. A. (1980). An empirical investigation of the arbitrage pricing theory. *The journal of finance*, *35*(5), 1073–1103.

Sharpe, W. F. (1963). A simplified model for portfolio analysis. *Management science*, *9*(2), 277–293.

Stock, J. H., & Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, *97*(460), 1167–1179.

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

Tibshirani, R. J., & Tibshirani, R. (2009). A bias correction for the minimum error rate in cross-validation. *The Annals of Applied Statistics*, 822–829.

Varma, S., & Simon, R. (2006). Bias in error estimation when using cross-validation for model selection. *BMC bioinformatics*, *7*(1), 1–8.

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 Derivation of characteristic based SDF

The SDF in a linear span of excess returns, in the spirit of Hansen and Jagannathan (1991), can be found by solving

$$M_t = 1 - b'_{t-1}(R_t - \mathbb{E}R_t), \tag{A.1}$$

for the $N \times 1$ vector of SDF loadings b, such that

$$E_{t-1}[M_t R_t] = 0. \tag{A.2}$$

The equation needs be transformed into a characteristic-based SDF, which implies

$$b_{t-1} = Z_{t-1} b, \tag{A.3}$$

where Z_{t-1} is an $N \times H$ matrix of stock characteristics and b equals the $H \times 1$ vector of time-invariant loadings. Now, in the spirit of Kozak et al. (2020), combining the above-mentioned equations, the SDF can be expressed in the linear span of the H stock characteristic returns ($F_t = Z'_{t-1} R_t$) as follows:

$$M_t = 1 - b'(F_t - \mathbb{E}F_t). \tag{A.4}$$

This paper focuses on the unconditional asset pricing equation constraint when solving the above equation. That is,

$$\mathbb{E}[M_t F_t] = 0, \tag{A.5}$$

B Derivation of Bayesian posterior

Firstly, to derive the posterior, as by Kozak et al. (2020), the posterior mean of μ is specified based on a conjugate multivariate normal prior, where the covariance matrix is assumed known. That is, the prior parameters equal

$$\mu_0 = 0 \quad \text{and} \quad \Sigma_0 = \frac{\kappa^2}{\theta}\Sigma^{\eta}. \tag{B.1}$$

Then, the posterior mean of μ can be constructed as

$$\hat{\mu} = (\Sigma_0^{-1} + T\Sigma^{-1})^{-1}(\Sigma_0^{-1}\mu_0 + T\Sigma^{-1}\mu) = (\Sigma + \gamma\Sigma^{(\eta-2)})^{-1}\Sigma\bar{\mu}, \tag{B.2}$$

$$\text{with} \quad \gamma = \frac{\theta}{\kappa^2 T}. \tag{B.3}$$

When the above-mentioned equation is combined with $\hat{b} = \Sigma^{-1}\hat{\mu}$, the following result is obtained:

$$\hat{b} = (\Sigma + \gamma\Sigma^{(\eta-2)})^{-1}\bar{\mu}. \tag{B.4}$$

C Descriptive statistics

Table C5: *Descriptive statistics for the 25 Fama and French ME/BM sorted portfolios. The portfolio returns correspond to a monthly buy-and-hold strategy and are rescaled to have a standard deviation equal to the in-sample standard deviation of the excess returns on the market index. There are 24140 daily observations from July 1926 to December 2017. Note that, the Mean and St. Dev. are annualized numbers for ease of interpretation.*

(N=24140)	Mean	St. Dev.	Skewness	Kurtosis	Min	Max	JB	P-value JB
1	0,100	0,4865	6,820	247,948	-0,484	1,275	62024197,001	0,000
2	0,077	0,312	1,924	51,620	-0,269	0,458	2695010,392	0,000
3	0,100	0,236	0,824	34,937	-0,198	0,371	1230471,839	0,000
4	0,120	0,208	0,580	28,471	-0,142	0,245	816684,963	0,000
5	0,133	0,208	0,590	30,225	-0,187	0,289	920302,637	0,000
6	0,064	0,232	0,535	25,094	-0,155	0,282	634514,369	0,000
7	0,095	0,196	0,637	30,093	-0,133	0,310	912496,415	0,000
8	0,102	0,190	-0,074	15,670	-0,147	0,190	246996,740	0,000
9	0,111	0,193	0,361	28,299	-0,154	0,293	806014,185	0,000
10	0,130	0,232	0,582	23,375	-0,157	0,270	550926,137	0,000
11	0,072	0,201	-0,311	10,629	-0,136	0,138	114032,001	0,000
12	0,096	0,179	-0,384	13,588	-0,138	0,156	186298,864	0,000
13	0,100	0,181	0,188	29,452	-0,194	0,234	872640,352	0,000
14	0,109	0,188	0,296	22,146	-0,152	0,232	493655,478	0,000
15	0,120	0,239	0,506	24,147	-0,191	0,303	587498,283	0,000
16	0,078	0,187	-0,428	16,049	-0,187	0,141	259804,321	0,000
17	0,081	0,174	-0,018	23,307	-0,158	0,218	546367,495	0,000
18	0,093	0,182	0,295	29,197	-0,158	0,269	857809,322	0,000
19	0,105	0,197	0,394	20,302	-0,138	0,214	415214,340	0,000
20	0,112	0,260	0,478	22,420	-0,169	0,336	506519,375	0,000
21	0,072	0,179	-0,132	14,571	-0,184	0,136	213614,613	0,000
22	0,072	0,172	-0,142	20,049	-0,208	0,168	404400,672	0,000
23	0,080	0,181	0,380	22,939	-0,186	0,208	529862,942	0,000
24	0,074	0,214	0,463	24,791	-0,204	0,236	619028,941	0,000
25	0,110	0,280	0,533	19,000	-0,183	0,240	364246,228	0,000

Table C6: *Descriptive statistics for the 50 portfolios linear in one of the stated anomaly characteristics. The portfolio returns correspond to a monthly buy-and-hold strategy and are rescaled to have a standard deviation equal to the in-sample standard deviation of the excess returns on the market index. There are 11141 daily observations from November 1973 to December 2017. Note that, the Mean and St. Dev. are annualized numbers for ease of interpretation.*

(N = 11141)	Mean	St. Dev.	Skewness	Kurtosis	Min.	Max.	JB	P-Value JB
Size	-0,023	0,167	1,328	23,822	-0,099	0,176	266706,650	0,000
Value (A)	0,062	0,167	-0,392	19,483	-0,156	0,101	176498,775	0,000
Gross profitability	0,036	0,167	-0,083	8,433	-0,100	0,098	33023,053	0,000
Value-profitability	0,132	0,165	-0,447	22,034	-0,149	0,119	225748,712	0,000
F-score	0,081	0,168	0,747	25,875	-0,131	0,145	311838,683	0,000
Debt issuance	0,018	0,157	-0,428	22,048	-0,142	0,112	225997,498	0,000
Share repurchases	0,069	0,157	0,681	24,410	-0,126	0,125	277455,166	0,000
Net issuance (A)	-0,095	0,168	0,238	23,780	-0,113	0,154	262605,545	0,000
Accruals	-0,056	0,159	0,051	4,002	-0,082	0,080	7441,185	0,000
Asset growth	-0,086	0,164	0,451	14,497	-0,075	0,155	97940,047	0,000
Asset turnover	0,053	0,164	-0,197	9,876	-0,107	0,111	45348,874	0,000
Gross margins	-0,012	0,163	0,437	9,437	-0,068	0,118	41696,338	0,000
Dividend/Price	0,036	0,161	0,159	8,383	-0,096	0,103	32667,218	0,000
Earnings/Price	0,083	0,166	0,085	16,496	-0,130	0,108	126331,587	0,000
Cash flow/Price	0,079	0,167	-0,419	16,177	-0,154	0,094	121806,873	0,000
Net operating assets	0,019	0,170	0,587	21,784	-0,120	0,137	220935,158	0,000
Investment/Assets	-0,100	0,165	0,112	6,367	-0,069	0,105	18844,327	0,000
Investment/Capital	-0,041	0,170	0,295	18,739	-0,099	0,147	163169,511	0,000
Investment growth	-0,090	0,166	0,196	9,004	-0,074	0,115	37702,704	0,000
Sales growth	-0,057	0,166	0,213	9,717	-0,071	0,128	43914,736	0,000
Leverage	0,049	0,169	-0,080	15,096	-0,133	0,102	105805,691	0,000
Return on assets (A)	0,024	0,166	0,109	6,955	-0,094	0,086	22476,600	0,000
Return on book equity (A)	0,047	0,167	0,505	12,115	-0,110	0,138	68605,131	0,000
Sales/Price	0,094	0,168	-0,435	25,939	-0,157	0,114	312676,336	0,000
Growth in LTNOA	-0,025	0,125	-0,200	7,056	-0,100	0,062	23188,341	0,000
Momentum (6m)	0,021	0,166	-1,321	21,171	-0,133	0,101	211299,666	0,000
Industry momentum	0,056	0,169	-1,346	23,586	-0,157	0,097	261612,413	0,000
Value momentum	0,051	0,166	-1,357	26,697	-0,189	0,099	334262,249	0,000
Value-momentum-prof.	0,065	0,165	-1,046	21,225	-0,167	0,089	211161,666	0,000
Short interest	0,003	0,154	0,057	3,999	-0,074	0,085	7430,229	0,000
Momentum (12m)	0,090	0,166	-0,929	14,496	-0,122	0,082	99143,331	0,000
Momentum-reversals	-0,057	0,166	0,077	20,758	-0,102	0,146	200043,621	0,000

Continued on the next page

(N=11141)	Mean	St. Dev.	Skewness	Kurtosis	Min	Max	JB	P-value JB
Long-run reversals	-0,054	0,164	0,501	16,142	-0,085	0,156	121426,076	0,000
Value (M)	0,055	0,168	0,905	21,227	-0,107	0,136	210692,739	0,000
Net issuance (M)	-0,087	0,169	-0,183	27,595	-0,129	0,150	353554,575	0,000
Earnings surprises	0,120	0,156	-0,394	11,654	-0,093	0,097	63338,297	0,000
Return on equity	0,105	0,153	0,125	8,547	-0,077	0,081	33943,033	0,000
Return on market equity	0,122	0,158	0,321	23,555	-0,125	0,120	257750,785	0,000
Return on assets	0,071	0,154	-0,037	7,113	-0,074	0,081	23490,626	0,000
Short-term reversals	-0,080	0,166	-1,016	20,977	-0,145	0,102	206194,234	0,000
Idiosyncratic volatility	-0,031	0,163	-0,184	13,049	-0,105	0,120	79107,230	0,000
Beta arbitrage	-0,007	0,160	0,269	8,358	-0,079	0,123	32565,561	0,000
Seasonality	0,115	0,163	0,120	9,013	-0,071	0,130	37740,220	0,000
Industry rel. reversals	-0,178	0,164	-0,755	17,067	-0,127	0,091	136276,710	0,000
Industry rel. rev. (L.V.)	-0,349	0,163	-0,207	6,646	-0,085	0,082	20580,165	0,000
Ind. mom-reversals	0,201	0,166	-0,572	11,490	-0,108	0,135	61889,897	0,000
Composite issuance	-0,084	0,160	-0,054	7,553	-0,086	0,095	26483,982	0,000
Price	-0,011	0,167	-1,180	20,162	-0,152	0,090	191289,291	0,000
Age	0,035	0,170	0,711	29,817	-0,135	0,150	413651,001	0,000
Share volume	-0,012	0,160	-0,065	11,005	-0,098	0,113	56227,137	0,000

D Contour maps without hard minimum

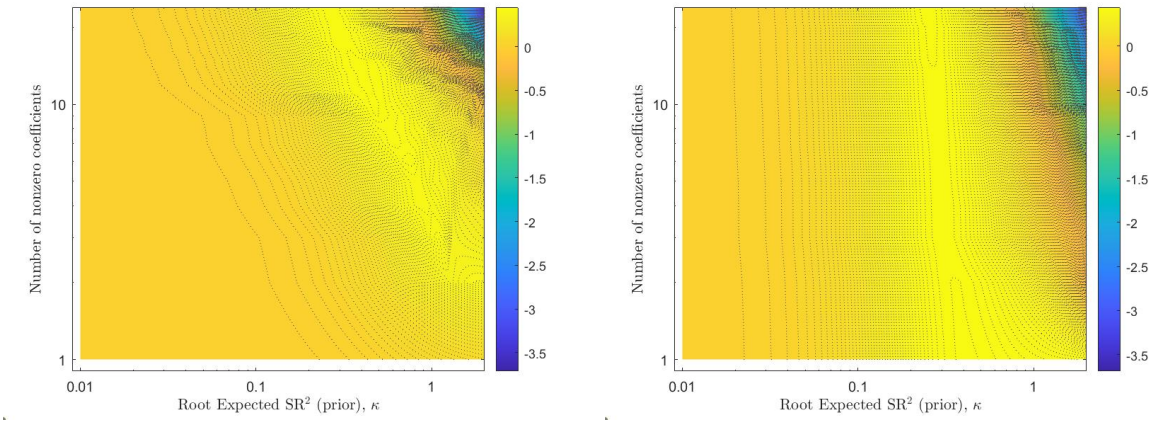

Figure D1: OOS R^2 values for the combined L^1- and L^2-Norm specification for the 25 Fama and French ME/BM sorted portfolios in the period July 1926 to December 2017 (daily) with no hard minimum imposed. The left panel depicts the OOS cross-sectional R^2 of the dual-penalty specification for the raw portfolio returns, whereas the right panel shows these values for their corresponding PCs. The strength of shrinkage (L^2-Norm) is quantified on the x-axis as the prior root expected squared Sharpe Ratio (κ). The amount of sparsity (L^1-Norm) is quantified on the y-axis as the number of retained variables in the SDF. Observe that both axis are on the logarithmic scale.

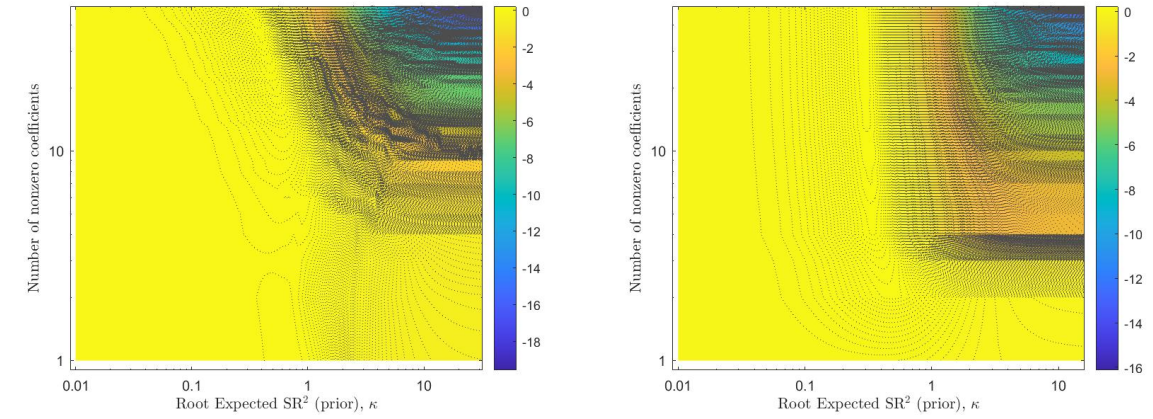

Figure D2: OOS R^2 values for the combined L^1- and L^2-Norm specification for the 50 anomaly portfolios in the period November 1973 to December 2017 (daily) with no hard minimum imposed. The left panel depicts the OOS cross-sectional R^2 of the dual-penalty specification for the raw portfolio returns, whereas the right panel shows these values for their corresponding PCs. The strength of shrinkage (L^2-Norm) is quantified on the x-axis as the prior root expected squared Sharpe Ratio (κ). The amount of sparsity (L^1-Norm) is quantified on the y-axis as the number of retained variables in the SDF. Observe that both axis are on the logarithmic scale.

E Contour maps of replication with a standardised scale

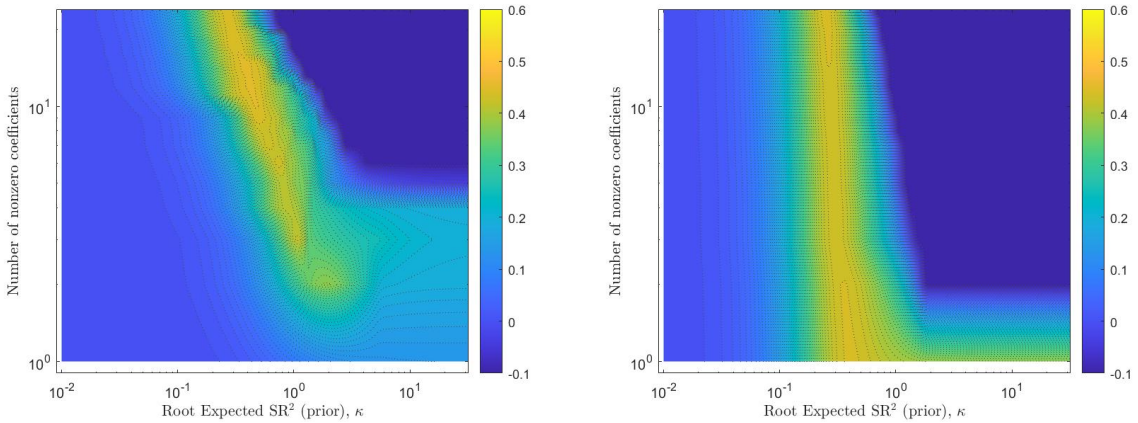

Figure E1: OOS R^2 values for the combined L^1- and L^2-Norm specification for the 25 Fama and French ME/BM sorted portfolios in the period July 1926 to December 2017 (daily). The left panel depicts the OOS cross-sectional R^2 of the dual-penalty specification for the raw portfolio returns, whereas the right panel shows these values for their corresponding PCs. The strength of shrinkage (L^2-Norm) is quantified on the x-axis as the prior root expected squared Sharpe Ratio (κ). The amount of sparsity (L^1-Norm) is quantified on the y-axis as the number of retained variables in the SDF. Note that both axis are on the logarithmic scale, there is a hard minimum of -0.1 imposed on the OOS R^2 and the legend is standardised to be equal across figures.

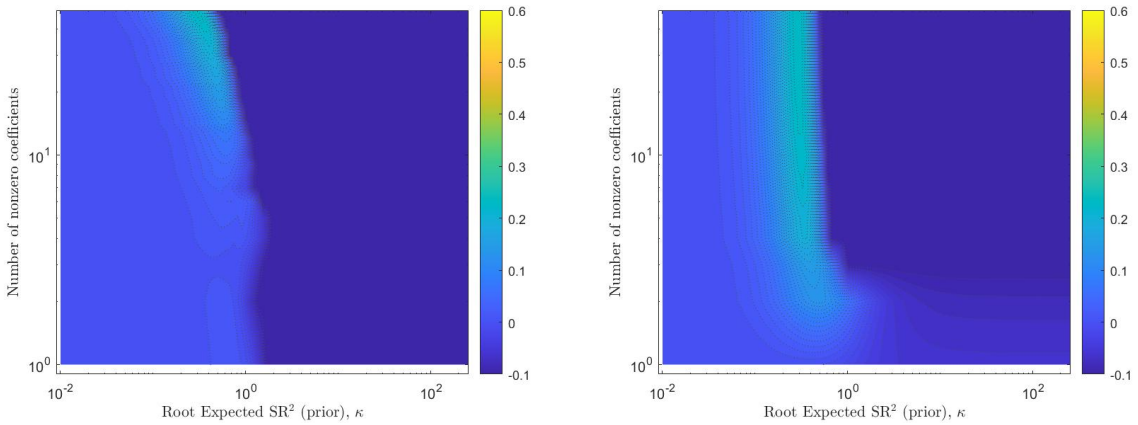

Figure E2: OOS R^2 values for the combined L^1- and L^2-Norm specification for the 50 anomaly portfolios in the period November 1973 to December 2017 (daily). The left panel depicts the OOS cross-sectional R^2 of the dual-penalty specification for the raw portfolio returns, whereas the right panel shows these values for their corresponding PCs. The strength of shrinkage (L^2-Norm) is quantified on the x-axis as the prior root expected squared Sharpe Ratio (κ). The amount of sparsity (L^1-Norm) is quantified on the y-axis as the number of retained variables in the SDF. Note that both axis are on the logarithmic scale, there is a hard minimum of -0.1 imposed on the OOS R^2 and the legend is standardised to be equal across figures.