Seminar Thesis

Machine Learning Methods in Asset Pricing: Factor Timing

Johann Wolfgang Goethe University Frankfurt am Main
Faculty of Economics and Business Administration

Submitted by
Lukas Körber (7738078)
13.07.2023

Lecturer: Dr. Fabio Girardi

Contents

1 Introduction

Two fundamental questions in the field of finance, particularly within asset pricing, are: Are asset returns predictable? If predictability exists, which variables serve as effective predictors?

A prominent strain of academic literature investigates factor models, a category of models linking firms' characteristics and their expected returns, such as the famous Fama and French 3-factor model (Fama and French, 1993). These models usually consist of constant average factor returns and asset sensitivities to those risk factors to estimate the expected return of an asset. Factors, other than the aggregate market, are generally market-neutral and thus promise a return premium in the cross-section of assets, which motivates factor investing. Due to academic research and the explosion of data availability, a broad range of relevant pricing factors has been identified, leading to a dimensionality problem for traditional statistical tools such as linear regressions (Nagel and Singleton, 2011, p.4). In a previous and related paper, Kozak et al. (2018) employed PCA to tackle the dimensionality problem. PCA is designed to produce a synthetic, lower-dimensional dataset from the original data while maintaining explanatory power of the variables. Kozak et al. (2018) showed that factor models building only on the largest PCs of the large number of pricing factors are successful in explaining asset prices.

Haddad et al. (2020) propose an extension of the work by Kozak et al. (2018) and add time-varying factor premia to the pricing model. The extension is motivated by economic intuition: Investors aversion to various sources of factor risk might vary over time, creating variation in factor risk premia. The economic relevance is twofold: First, it creates the possibility for investors to profit from the variations by adjusting their portfolios accordingly (factor timing). Second, since the optimal factor timing portfolio equals the SDF, it informs us about its properties (Haddad et al., 2020, p.1980-1981). Haddad et al. (2020) propose valuation metrics as predictors for factor returns. However, the literature remains divided regarding this approach: For instance, studies by Arnott et al. (2016), Kelly and Pruitt (2013), andCohen et al. (2003) provide evidence in support of valuation-based factor timing. In contrast, Asness et al. (2017), Asness (2016), and Ilmanen and Nielsen (2015) challenge this notion. Haddad et al. (2020) counter these criticisms by asserting that the detractors often overlook the importance of dimensionality reduction, instead attempting to individually predict a multitude of significant pricing factors' average returns. To address the issue, Haddad et al. (2020) extract the PCs from anomaly returns and predict their time variation (Haddad et al., 2020, p.1990-1991). The authors report robust predictability of large PCs, which makes it feasible to empirically describe the SDF with time-varying risk prices. The predictability translates to individual factors (Haddad et al., 2020, p.1997) and is strong enough to enhance portfolio performance vs. static factor investing (Haddad et al., 2020, p.2003-2006).

This seminar thesis replicates the empirical findings of Haddad et al. (2020) and is structured in the following way: First, the data and the machine learning method PCA are explained. Afterwards, the authors' theoretical argument is stated in detail and the empirical results are discussed.

2 Data and Methodology

2.1 Data

For the replication of the empirical results of Haddad et al. (2020), the authors provide the dataset on the website of Seriyh Kozak(Kozak). The dataset comprises monthly time series data on returns, book-to-market ratios, market capitalization, and number of constituents for 10 portfolios, each sorted based on a distinct characteristic. This procedure is carried out for fifty unique characteristics, resulting in data for a total of 500 portfolios. The sample period ranges from January 1974 to December 2017. Though the dataset lacks direct market return and book-to-market (bm) ratio data, we can derive these by market-capitalization-weighting the returns and bm ratios of 10 portfolios per characteristic, in aggregate representing the full stock universe. Variations in the stock universe exist depending on the selected characteristic; we used "Size" for our final analysis.

Factor returns are calculated by subtracting Portfolio 1's return from Portfolio 10's. The log difference in bm-ratios between these portfolios yields a factor valuation measure—the anomaly bm-ratio. These anomaly returns and bm-ratios are then market adjusted: We estimate market betas for each anomaly, then subtract the products of these betas and market returns or bm-ratios from respective time series. Finally, anomaly returns and bm-ratios are rescaled for variance equal to market variance. Expressed formally, the following operations are performed:

$$\begin{aligned} F_{t+1} &= \frac{F'_{t+1} - \beta \ R_{MKT, \quad t+1}}{std\left(F'\right)} var\left(R_{MKT}\right) \\ bm_{F,t+1} &= \frac{bm_{F,t+1}}{std\left(bm_F\right)} var\left(bm_{MKT}\right) \end{aligned} \tag{2.1}$$

Betas and market variance are estimated using the first half of the sample data only.

2.2 Principle Component Analysis (PCA)

The section is based on the Jollife (2002, p.1-10) and Sanguansat (2012, p.181-206). PCA was invented by Pearson (1901) and later independently developed and named by Hotelling (1933). It aims to simplify a complex dataset with numerous interrelated variables and to reduce dimensionality while preserving the maximum amount of the dataset's inherent variation. This is achieved by transitioning to a new set of variables—principal components (PCs)—which are both uncorrelated and ordered such that the initial few explain most of the variation found across all original variables. The PCs are constructed via linear orthogonal transformation: Assume that we are informed about the true covariance matrix Σ (normally we can only estimate Σ) of our dataset X and perform an eigenvalue decomposition on the matrix: $\Sigma = Q^T \Lambda Q$. Then, the k-th PC is given by $PC_k = q_k^T X$ where q_k is an eigenvector of Σ and corresponding to the k-th largest eigenvalue λ_k. As a covariance matrix is always symmetric and positive semi-definite, we know that all eigenvalues λ_k are positive and that an orthogonal basis of eigenvectors exist. Moreover, since the dimension of the (quadratic) covariance matrix equals the number of features in the dataset one can also construct an equal number of PCs. If the dataset is centered, the estimation of Σ, named S, can be easily calculated via $S = X^T X$. The ratio of eigenvalue k and the sum of all other eigenvalues $\frac{\lambda_k}{\sum_i \lambda_i}$ captures the amount of percentage variance explained by the k-th PC.

The choice of optimal PCs is empirical, yet the initial few typically capture a substantial portion of total variance. This offers computational efficiency and enhances the performance of classical statistical methods in handling high-dimensional data with interrelated features.

3 Replication of Results

Establishing a robust SDF description, under the premise of time-varying risk prices, presents a practical challenge due to the necessity of predicting numerous risk premiums over time. This results in an empirical problem of high dimensionality. The authors employ PCA, a common machine learning technique, to mitigate this issue. Instead of attempting to predict a multitude of pricing factors—which proves impracticable—they aim to identify a limited number of significant drivers of risk premiums – the PCs of the anomaly returns which offer a more robust predictability. This chapter is structured in the following way: The subsequent section presents the authors' theoretical reasoning for using the PCs of factor returns to represent the SDF with time-varying risk prices. Afterwards, the empirical results are replicated and analyzed.

3.1 Theoretical Argument

This section elaborates in more detail on the theoretical arguments the authors use to motivate their empirical methodology (Haddad et al., 2020, p.1985-1991). An important initial result, to which the authors refer later in the argument, is the following relation 3.1 which holds under the assumption of uncorrelated and homoscedastic asset returns:

$$E\left[SR_t^2\right] = \sum_{i=1}^{N} \frac{E\left[R_{i,t+1}\right]^2}{\sigma_i^2} + \sum_{i=1}^{N} \left(\frac{R_i^2}{1 - R_i^2}\right) \tag{3.1}$$

The proof is detailed in the appendixA.1.2. From this dissection one can see that the average squared Sharpe ratio in the economy is driven by two terms: The first one is the sum over the squared unconditional mean returns of the single assets divided by their respective variance (constant since homoscedastic returns). Looking at a single asset i, this quantity is similar to its unconditional Sharpe ratio:

$$SR_i = \frac{E\left[R_{i,t+1}\right] - r_f}{\sigma_i} \approx \frac{E\left[R_{i,t+1}\right]}{\sigma_i} = \frac{E\left[R_{i,t+1}\right]^2}{\sigma_i^2} \tag{3.2}$$

The second term of equation 3.1, thus the average Sharpe ratio, is increasing in the R-squared of the single assets, R_i^2, capturing the investment gains stemming from the predictability of returns.

The authors' line of argument aims at dimensionality reduction of the formulation of the SDF in the span of excess returns of N assets, $R_{t+1} = \left(r_{1,t+1}, \ldots, r_{N,t+1}\right)^T$. The SDF is a random variable which satisfies the following two equations in the economy:

$$\begin{aligned} E\left[M_{t+1} R_{t+1}\right] &= 0 \\ E_t\left[M_{t+1}\right] = B_t &= \frac{1}{R_{f,t,t+1}} \end{aligned} \tag{3.3}$$

Hansen and Jagannathan (1991) show that the minimum variance SDF (a random variable with the minimum variance, which fulfills the previous conditions), equals:

$$\begin{aligned} M_{t+1} &= 1 - b_t^T\left(R_{t+1} - E\left[R_{t+1}\right]\right) \\ \text{with: } b_t &= \Sigma_R^{-1} E_t[R_{t+1}] \end{aligned} \tag{3.4}$$

Here, b_t, which is a $N \times 1$ vector, represents the stock-level SDF loadings. Observe also, that in our setting b_t is time-varying. The dimensionality problem is obvious, as b_t needs to be estimated from data for every time point $t \in 1, \ldots, T$, likely leading to high estimation errors.

The authors' first assumption tackles this problem: It is assumed that stock characteristics capture most of pricing relevant information, which formally means that the vector b_t is a linear combination of time-varying coefficients δ_t (vector of dimension $K \times 1$), with $K \ll N$:

$$b_t = C_t \ \delta_t \tag{3.5}$$

C_t (a $N \times K$-matrix) carries the stocks' characteristics at time t. From that matrix, one can construct returns of characteristics-managed portfolios, the factor returns:

$$F_t = C_{t-1}^T \ R_t \tag{3.6}$$

For example, one could construct the Fama-French high-minus-low (HML or "value"-factor(Fama and French, 1993)) return if a column of C_t would take values -1,0,1 depending on if the corresponding stock is in the bottom (-1) or top (+1) decile or in neither of the two (0) when sorted on the stocks' earnings yield. Substituting equation 3.6 into equation 3.3 leads us to a new, dimensionality-reduced formulation of the SDF:

$$\begin{aligned} M_{t+1} &= 1 - \delta_t \ \left(F_{t+1} - E\left[F_{t+1}\right]\right) \\ \text{with: } \delta_t &= \Sigma_F^{-1} E_t[F_{t+1}] \end{aligned} \tag{3.7}$$

Consequently, δ_t can now be interpreted as (time-varying) factor risk prices. Moreover, one can observe that this SDF formulation is equivalent to a conditional factor model of the form:

$$E_t\left[R_{j,t+1}\right] = \beta_{j,t}^T \Sigma_{F,t} \delta_t = \beta_{j,t}^T E_t[F_{t+1}] \tag{3.8}$$

The time-t risk prices in the SDF, δ_t, are directly related to the conditional expected factor returns of the factor model, $E_t\left[F_{t+1}\right] : \Sigma_{F,t}\delta_t = E_t\left[F_{t+1}\right]$. Further details on the derivation can be found in the appendix A.1.3. In this economy and at time t, the maximum Sharpe ratio portfolio and its return are given by:

$$
\begin{aligned}
w_t &= \Sigma_{F,t}^{-1} E_t \left[F_{t+1}\right] \\
R^{opt} &= w_t^T F_{t+1} = E_t \left[F_{t+1}\right]^T \Sigma_{F,t}^{-1}
\end{aligned} \tag{3.9}
$$

Again, the derivation of the maximization problem can be found in the appendix A.1.5. It is not necessary to time individual stocks, as factors capture all relevant risk premia. At this stage, the statistical problem shrinks in dimensionality as now “only” K conditional factor risk prices need to be estimated from the data. However, as academic literature has identified an increasingly large number of potentially relevant risk factors, this still leaves us with a significant number of factors to predict.

This motivates the authors’ second assumption: There are usually no near-arbitrage opportunities available in the economy. Put differently, the average conditional squared Sharpe ratio is bounded above by a constant.
Adding this assumption implies that to describe the SDF one does not need to predict the conditional means of all factors but only those of their largest PCs. The result follows from combining the following two statements: First, recall equation 3.1 which can be reformulated for the PCs of the factor returns:

$$
E\left[SR_t^2\right] = \sum_{i=1}^{K} \frac{E\left[PC_{i,t+1}\right]^2}{\lambda_i} + \sum_{i=1}^{K} \left(\frac{R_i^2}{1 - R_i^2} \right) \tag{3.10}
$$

K is again equal to the dimensionality of the number of factors included in the analysis. The reformulation works, as Q, resulting from the eigenvalue decomposition of the anomaly returns $cov\left(F_{t+1}\right) = Q^T \Lambda Q$, is an orthogonal matrix, thus it only rotates vectors in the vector/asset space. In the appendix it is proven that the squared Sharpe ratio is invariant against rotations in the asset space A.1.7. The PCs can be interpreted as portfolio returns as they are a linear combination of factor returns.

$$
PC_{t+1} = \left(PC_{1,\ t+1}, \ldots,\ PC_{K,t+1}\right)^T = Q^T F_{t+1} \tag{3.11}
$$

The equation shows again that the mean squared Sharpe ratio is composed of two terms: The first term captures the investment benefits of static factor investing, while the second term captures the additional gains stemming from the predictability of the factors.

Moreover, the authors start from the definition of the total amount of predictability of the anomaly returns, R^2_{total}, to derive how much each PC contributes to R^2_{total}:

$$R^2_{total} = \frac{tr[covEtFt + 1]}{tr[covFt + 1]} \tag{3.12}$$

The derivation is detailed in the appendix A.1.4 and leads to the following result:

$$\begin{aligned} R^2_{total} &= \sum_{i=1}^{K} \left(\frac{R_i^2}{1 - R_i^2} \right) \frac{\lambda_i}{\lambda} \\ \text{with: } \lambda &= \sum_i \frac{\lambda_i}{1 - R_i^2} \approx \sum_i \lambda_i \end{aligned} \tag{3.13}$$

The total predictability of factor returns can be expressed as a weighted sum of each PC's predictability, where the weights correspond to their respective contributions to total variance of anomaly returns. Combining the two equations, one can see that if PC i explains only a small share of the total variance (small eigenvalue λ_i), PC i cannot contribute significantly to the total amount of predictability of the factors: To illustrate that, imagine a "small PC", PC_j, contributes significantly to R^2_{total}. In that case, as $\frac{\lambda_i}{\lambda}$ is small, the term $\frac{R_j^2}{1-R_j^2}$ would need to be large, so large that it would render $E\left[SR_t^2\right]$ too high to be in accordance with assumption 2. In short, the assumption implies that small PCs cannot contribute meaningfully to the predictability of factor returns, thus the large PCs must capture time-series variation in average factor returns.

This conclusion adds to the one of Kozak et al. (2018) who proof in the unconditional case (with constant factor risk prices) that under the assumption of anomaly returns following an intense factor structure and the absence of no-arbitrage the largest principal components are sufficient to explain most of the cross-sectional variation in assets' average returns. Put differently, the SDF (with constant risk prices)

can be approximated using only the largest PCs of factor returns or, equivalently, a factor model using only the largest PC portfolios' average returns is successful in pricing assets.
Taking the results together, the large PCs must thus capture both the cross-sectional and time-series variation in factor risk premia, implying that the SDF can be approximated using vector $Z_{t+1} = (R_{mkt,\ t+1}, PC_{1,t+1}, \ldots,\ PC_{j,t+1})^T$ with $j \ll K$:

$$M_{t+1} \approx 1 - E_t\left[Z_{t+1}\right]^T \Sigma_{Z,t}^{-1}(Z_{t+1} - E_t\left[Z_{t+1}\right]) \tag{3.14}$$

The same holds true for the optimal Sharpe ratio portfolio:

$$\begin{aligned} w_t &\approx \Sigma_{Z,t}^{-1} E_t\left[Z_{t+1}\right] \\ R^{opt} &\approx w_t^T Z_{t+1} = E_t\left[Z_{t+1}\right]^T \Sigma_{Z,t}^{-1} \end{aligned} \tag{3.15}$$

The empirical part deals with the estimation of this formulation of the SDF.

3.2 Empirical Part

The empirical examination in the study is conducted in a three-step process. The first step involves the computation of the PCs from the 50-dimensional anomaly returns, specifically focusing on the top five PCs with the greatest explanatory power. In the second step, the PC strategies' monthly returns and valuation ratios are constructed by interpreting each PC as a linear combination of anomaly returns. Consequently, each PC and the aggregate market are predicted using linear regressions, with the lagged valuation ratio serving as a predictor for returns. The final step involves the formulation of diverse portfolio strategies based on the estimated returns, enabling the estimation of the SDF and the interpretation of its properties.

3.2.1 Principal Component Analysis

We apply PCA to the market-adjusted and rescaled returns of the 50 long-short strategies, denoted as F_{t+1}. In our empirical replication, we carry out an eigenvalue

Table 3.1: Percentage of variance explained by anomaly PCs

This table reports the percentage of variance explained of the fifty anomaly strategies by the ten most explaining PCs.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
% var. explained	25.5	12.6	10.4	6.6	4.8	4.0	3.6	2.8	2.2	2.1
Cumulative	25.5	38.1	48.5	55.1	59.9	63.9	67.5	70.3	72.5	74.6

decomposition on the covariance matrix derived from these preprocessed anomaly returns:

$$cov\,(F_{t+1}) = Q^T \Lambda\ Q \tag{3.16}$$

The PCA yields the eigenvectors, denoted as q_i which are the columns of matrix Q, alongside the eigenvalues, λ_i which comprise the diagonal of Λ. The proportion of variance explained by PC_i can be calculated by dividing its corresponding eigenvalue by the aggregate sum of all eigenvalues, as expressed in the formula $\frac{\lambda_i}{\sum_{i=1}^{50} \lambda_i}$. These computations are subsequently tabulated in Table 3.1.

The market-adjusted long-short anomaly returns show exhibit a noticeable factor structure, with the initial five PCs accounting for approximately two-thirds of the total variance. Most strikingly, the largest PC holds a considerable share of the variance, succeeded by a progressively reduced variance explanation for the subsequent PCs. In this analysis, the authors adopt a theoretical methodology to establish the optimal number of PCs, which they show to be five. Empirical evaluation, as visually represented in Figure 3.1, along with the robustness test displayed in Table 3.3, collectively support the assertion that neither an increase nor a reduction in the PC count enhances the employed methodology.

3.2.2 Predictive Regression on the PCs

The central part of the empirical section is the predictive regression of the most significant principal components. Each of the five eigenvectors serves as a combination of anomalies, leading to strategy returns defined by $PC_{i,t+1} = q\prime_i\ F_{t+1}$, where q_i stands for the i-th eigenvector. A similar approach is implemented for the valuation

Table 3.2: Predicting dominant equity components with BE/ME ratios

We report results from predictive regressions of excess market returns and five PCs of long-short anomaly returns. The market is forecasted using the log of the aggregate book-to-market ratio. The anomaly PCs are forecasted using a restricted linear combination of anomalies log book-to-market ratios with weights given by the corresponding eigenvector of pooled long-short strategy returns. The first row shows the coefficient estimate. The second row shows asymptotic t-statistics. The third and fourth rows show the out-of-sample results. The fifth and sixth rows shows the in-sample and out-of-sample monthly R^2.

	MKT	PC1	PC2	PC3	PC4	PC5
Own bm (IS)	0.71	3.78	3.04	1.96	4.81	1.43
t-stat. (IS)	1.09	4.13	3.13	1.72	4.52	1.41
Own bm (OOS)	1.30	2.78	2.56	2.70	5.56	2.88
t-stat. (OOS)	1.09	2.05	1.47	1.95	2.82	2.48
R^2	0.26	3.15	1.83	0.56	3.74	0.38
OOS R^2	0.96	4.10	2.59	0.18	4.08	-0.07

ratio, denoted by $bm_{PCi,t} = q\prime_i \ bm_t$, where bm_t are scaled valuation ratios. The structure of the predictive regressions is as follows:

$$PC_{i,t+1} = \alpha_{PCi} + \beta_{PCi} \ bm_{PCi,t} + \epsilon_{i,t} \tag{3.17}$$

And for the market returns

$$R_{MKT,t+1} = \alpha_{MKT} + \beta_{MKT} \ bm_{MKT,t} + \epsilon_{i,t} \tag{3.18}$$

In this paper, the predictive regressions are executed with a one-month holding period serving as a distinct timestep. A dual-method approach is employed, comprising an in-sample regression that utilizes the full dataset from 1974 through 2017, and an out-of-sample regression, which employs data from 1974 up to 1996 to predict values from 1996 until 2017. A summarization of the results from the predictive regression is found in Table 3.2.

The table reports the coefficients, converted into percentages, along with their corresponding t-statistics for the in-sample regression. Neither the market nor the fifth PC coefficients are statistically significant, while the other coefficients display analogous directions and are of comparable magnitude. Slight variations are observed in the out-of-sample coefficients compared to their in-sample counterparts, leading to distinctive prediction patterns. Corresponding t-statistics are presented in the fourth row, with a decline noticed for most PCs, resulting in non-significance for PC2. However, an increase is noted for the t-statistic of PC 5. The R^2 values for both the in-sample and out-of-sample regressions are detailed in the fifth and sixth rows, with the out-of-sample $\mathbf{R^2}$ often exceeding the in-sample R^2.

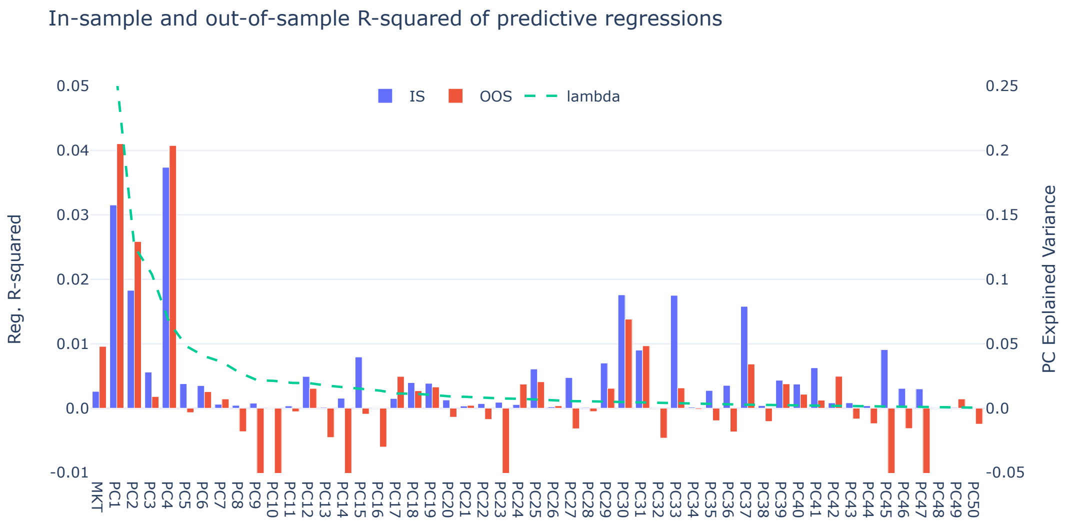

Figure 3.1: PC and market predictability with own bm ratios in and out-of-sample The plot shows the in-sample (IS) and out-of-sample monthly R^2 of predicting each PC of anomaly portfolio returns with its own bm ratio. The dashed green line (using the right axis) represents the fraction of total variance explained by each PC.

Figure 3.1 illustrates the in-sample and out-of-sample R^2 values, each computed for all PCs and the market according to the regression described in equation X. In addition, the figure reports the explained variance (lambda) for each PC. The analysis suggests a direct correlation between higher explained variance and increased predictability of PCs, when based on their valuation ratio. This observation supports the authors' theoretical reasoning to constrain the number of PCs used in the analyses. By the following operation, predictions for the individual factors can be backed out from the predictions of the PC returns:

$$E\left[F_{t+1}\right] = \sum_{i}^{5} PC_{i,t+1}\; q_{i,F} \tag{3.19}$$

Overall, the results show that a multitude of anomalies present positive R^2 values, with the 'long run reversal' anomaly boasting the maximum value, realizing an R^2 of 5% in both in-sample and out-of-sample environments. However, exceptions exist, especially with regards to in-sample R^2, such as the bad performance of the "Industry Relative Reversal (Low Volatility)" anomaly. The disparity in performance between the in-sample and out-of-sample may be due to the shift in the anomaly's average returns, descending from 1.72% to 0.56% in monthly returns between the in-sample and out-of-sample periods, thereby closely aligning with the mean anomaly return during the out-of-sample period.

3.2.3 Robustness Tests

The authors implement their pre-established return prediction method, as explained in previous sections, to derive an array of SDF estimates. To validate the robustness of this approach, they carry out a set of robustness tests, which can be broadly classified into two primary categories. The first category is centered on the preprocessing of the data and the setup of the regression for the PC returns. In our replication effort, we undertook a selection of these robustness tests, the outcomes of which are reported in Table 3.3.

The evaluation of diverse methodologies involves investigating the aggregate out-of-sample R^2 and the quantity of PCs with a positive out-of-sample R^2. In addition to sorting the anomaly portfolios into ten bins, we executed tests with five clusters by merging two groups on market equity weighting. This modification led to a minor reduction in the total out-of-sample R^2 but enhanced the number of PCs with a positive out-of-sample R^2. In a separate experiment, we expanded the number of utilized PCs to seven, prompting a decline in the total out-of-sample R^2 primarily due to the poor performance of PC_5 and PC_7. In the following variant, the returns and valuation ratios are not market-adjusted, giving rise to a marginal increase in

Table 3.3: Various data choices

The table reports summary statistics of predictive regressions in Table 2 for various data construction choices. Specifically, we report the OOS total R^2 and the number of PC portfolios for which the OOS R^2 is above zero. The first column reports the number of portfolios used for the underlying characteristic sorts. The second column reports the holding period in months. For holding periods longer than one month, the third column reports whether principal components are estimated using monthly holding period returns. The fourth column reports whether the anomaly returns are orthogonalized relative to the aggregate market. The fifth column reports whether the anomaly returns and book-to-market values are normalized to have equal variance compared to the market.

Anomaly portfolio sort	Holding period	# of PCs	Monthly PC	Market adjusted returns	Scaled market variance	OOS Total R^2	Significant PCs
Deciles	1	5	X	X	X	2.18	4
Quintiles	1	5	X	X	X	2.24	5
Deciles	1	7	X	X	X	2.01	6
Deciles	1	5	X			2.42	4
Deciles	6	5		X	X	8.97	3

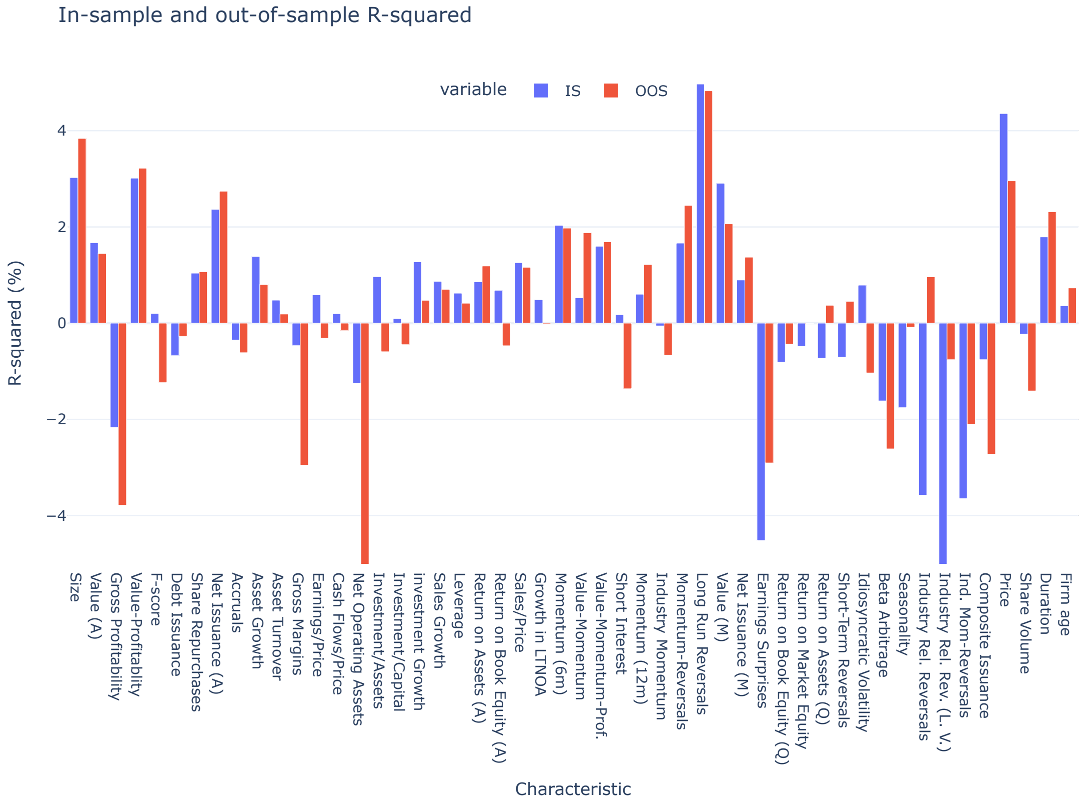

Figure 3.2: Factor predictability based on PC predictions in and out-of-sample
Monthly predictive R^2 of individual anomalies returns using implied fitted values based on PC forecasts. IS provides estimates in the full sample. OOS shows out-of-sample R^2.

the total out-of-sample R^2. For the last robustness test, the data is aggregated into six-month intervals, yielding a six-month return and a six-month average valuation ratio. This provokes a notable increase in the total out-of-sample R^2 but diminishes the number of informative PCs. This reduction could potentially be ascribed to the reduction of available data points resulting from the data aggregation process.

In the subsequent series of robustness analyses, diverse regression techniques for return forecasting are juxtaposed. Echoing the initial series of evaluations, we carried out a selection of the tests enacted by the original authors, with the outcomes succinctly represented in Table 3.4.

Table 3.4: Out-of-sample R^2 of various forecasting methods

The table presents the monthly out-of-sample (OOS) total R^2, along with the mean, median, and standard deviation of OOS R^2, for different forecasting methods applied to individual anomaly portfolios. The first column provides information on the assets being forecasted, the predictive variables used, and the specific forecasting method employed. If not specified, the method defaults to ordinary least squares.

	OOS total R^2	Mean	Median	Std.
50 Anom, BM of Anom, OLS	-123.87	-224.40	-197.79	153.47
5 PCs, Own BM	2.18	1.29	1.65	2.67
5 PCs, BM of PCs, Ridge 1DoF	0.56	-0.04	-0.36	0.74
5 PCs, BM of PCs, Lasso-OLS 1DoF	0.59	-0.23	-0.28	1.15
50 Anom, Own BM	0.52	-0.70	-0.03	2.58

Drawing upon the data configuration described in the initial row of Table 3.4, we compare five forecasting methodologies based on their out-of-sample R^2 values and the total out-of-sample R^2. The first row represents the baseline approach, in which the returns of the 5 PCs are forecasted based on their individual valuation ratio. The second row leverages all anomaly valuation ratios and conducts an ordinary least squares (OLS) regression to predict each return. This strategy showcases poor out-of-sample performance, likely a consequence of overfitting. The third technique engages all valuation ratios of the PCs to anticipate each PC return, and to mitigate overfitting, a ridge regression with a single degree of freedom is employed. Nevertheless, this technique lags behind in comparison to the baseline approach. The fourth method introduces lasso-OLS regularization, an alternate form of regularization. In contrast to ridge regression, lasso-OLS results in a superior total R^2 but a lower mean R^2, resulting in an unclear outcome. Both regularization techniques, however, fall short when compared with the baseline. In the final method, each anomaly is predicted using its unique valuation ratio. While the total R^2 is positive, the mean R^2 is negative once more. This inconsistency might be due to pronounced outliers, such as the "Industry Relative Reversal (Low Volatility)" anomaly, which exhibits poor performance as noticed in the preceding section (Figure 3.2).

3.2.4 Optimal Factor Timing Portfolio

Applying the method suggested in section 3.2.2 and validated through a series of robustness checks in section 3.2.3, the authors build multiple portfolio strategies. The primary strategy, referred to as factor timing, uses the predictive regressions from Table 3.2 to generate projections for future returns. Complementary strategies represent variations of this approach and are designed to distinguish the impacts of timing and the factor investing elements inherent in factor timing. Broadly, the strategies have a common approach of investing in either the market or the PCs. The weight vector for each strategy is constructed using the subsequent equations:

$$w_{F.I.} = \Sigma_Z^{-1} \left[E(R_{mkt,t+1}), E(PC_{1,t+1}), \ldots, E(PC_{5,t+1})\right]', \tag{3.20}$$

$$w_{M.T.,t} = \Sigma_Z^{-1} \left[E_t(R_{mkt,t+1}), E(PC_{1,t+1}), \ldots, E(PC_{5,t+1})\right]', \tag{3.21}$$

$$w_{F.T.,t} = \Sigma_Z^{-1} \left[E_t(R_{mkt,t+1}), E_t(PC_{1,t+1}), \ldots, E_t(PC_{5,t+1})\right]', \tag{3.22}$$

$$w_{A.T.,t} = \Sigma_Z^{-1} \left[E(R_{mkt,t+1}), E_t(PC_{1,t+1}), \ldots, E_t(PC_{5,t+1})\right]', \tag{3.23}$$

$$w_{P.A.T.,t} = \Sigma_Z^{-1} \left[0, \left[E_t - E\right](PC_{1,t+1}), \ldots, \left[E_t - E\right](PC_{5,t+1})\right]' \tag{3.24}$$

The parameter Sigma, invariant across all strategies, denotes the risk from prediction errors stemming from return predictions. It is calculated as the conditional covariance of the in-sample forecasting errors and remains constant throughout the time:

$$\Sigma_Z = cov\left(\left[E_t(R_{mkt,t+1}) - R_{mkt,t+1}, E_t(PC_{1,t+1}) - PC_{1,t+1}, \ldots\right]\right) \tag{3.25}$$

The strategies from equation 3.24 primarily differ based on their estimation methodologies for future returns. Factor Investing employs unconditional means to approximate both market and PC returns. Factor timing uses conditional expectations for all investment returns, while market timing exclusively adopts conditional expectations for market returns. In contrast, Anomaly timing employs conditional expectations only for PC returns. Pure anomaly timing underscores the merits of timing by singularly investing in the time-series variation of expected returns around the unconditional mean of PC returns.

Table 3.5: Performance of various portfolio strategies

The table presents the unconditional Sharpe ratio, information ratio, and average mean-variance utility for five different strategies. The strategies include: (i) static factor investing based on unconditional estimates; (ii) market timing using forecasts of the market return but with unconditional values for PC returns; (iii) full factor timing incorporating predictability of both PCs and the market; (iv) anomaly timing using forecasts of the PCs while setting market returns to unconditional values; and (v) pure anomaly timing investing in anomalies based on the deviation of their forecast from the unconditional average. The strategies assume a homoskedastic conditional covariance matrix estimated from forecast residuals. Information ratios are calculated relative to the static strategy, and out-of-sample (OOS) values are obtained through a split-sample analysis with parameter estimation using the first half of the data.

	Factor investing	Market timing	Factor timing	Anomaly timing	Pure anom. timing
IS Return	1.38	1.39	2.50	2.47	1.12
OOS Return	1.25	1.10	2.50	2.66	1.41
IS Sharpe ratio	1.13	1.09	1.18	1.17	0.70
OOS Sharpe ratio	0.75	0.59	0.90	0.99	0.70
IS Inf. ratio		0.16	0.73	0.70	-0.13
OOS Inf. ratio		-0.19	0.61	0.70	0.07
Expected utility	0.16	0.16	0.28	0.27	0.11

The strategies shown in formula 3.24 are subsequently tested against historical raw return data. For in-sample strategies, the model estimations presented in the first row of Table X are utilized. Conversely, for the out-of-sample period, models outlined in the third row of the same table are employed.

Table 3.5 reports the performance metrics across strategies. The baseline approach, known as static factor investing, realizes moderate returns and Sharpe ratios both during the in-sample and out-of-sample periods, thus providing a reference point for further comparison. The dynamic strategy of market timing displays comparable in-sample performance to factor investing, albeit with weaker out-of-sample results, largely due to its short positions in the aggregate market. Conversely, the dynamic strategies of factor timing and anomaly timing, leveraging PC predictions, showcase potent returns and Sharpe ratios, thereby surpassing factor investing and producing positive information ratios. While the pure anomaly timing portfolio achieves appealing Sharpe ratios, it does not manage to outperform factor investing in the in-sample period. The unscaled utilities correspond with the Sharpe ratio results, with the strategies incorporating factor and anomaly timing achieving twice the utility compared to the other strategies. Comparing these strategy measurements with those reported in Haddad et al. (2020, p.2004), it becomes evident that the timed strategies outperform the static factor investment, which could be credited to the marginally elevated total R^2 for the five PC predictions, reinforcing the notion that higher R^2 values translate into improved Sharpe ratios.

3.2.5 Properties of the SDF

This chapter delves into an examination of several properties of the SDF. To examine these properties, we use the strategies outlined in the preceding section as proxies for the SDF. This methodology is viable due to the equivalence between the optimal portfolio and the SDF. The SDFs are formulated based on the in-sample strategies extracted from the optimal portfolios described in the prior section. The variance of the SDF has been a point of significant interest for economists, most notably underscored by Hansen and Jagannathan (1991). They established that the variance of the SDF equals the maximum attainable squared Sharpe ratio in a complete market, and functions as an upper bound in an incomplete market. Hence, by analyzing the variance of the SDF derived from asset prices, we acquire insights into the required volatility of the SDF in an economic model to potentially account for such empirical evidence. The conditional variance of the SDF is expressed by the following equation:

$$var_t(m_{t+1}) = w_t' \Sigma_Z w_t \tag{3.26}$$

The characteristics of this equation of the variance in the SDF are reported in Table 3.6, corresponding to the three distinct SDF formulations presented in equation 3.26.

Contrasting with an SDF formulation dependent on a static model, the second

Table 3.6: Variance of the SDF

We report the average conditional variance of the SDF and its standard deviation constructed under various sets of assumptions. “Factor timing” is our full estimate, which takes into account variation in the means of the PCs and the market. “Factor investing” imposes the assumption of no factor timing: conditional means are replaced by their unconditional counterpart. “Market timing” only allows for variation in the expectation of the market return

	Factor Investing	Market Timing	Factor Timing
E[var(m)]	1.92	1.95	3.30
std[var(m)]		0.16	2.02

and third rows of the previously mentioned formula induce variation in the SDF. Increasing the amount of conditioning information inherently broadens the spectrum of investment opportunities, consequently amplifying the variance of the SDF.

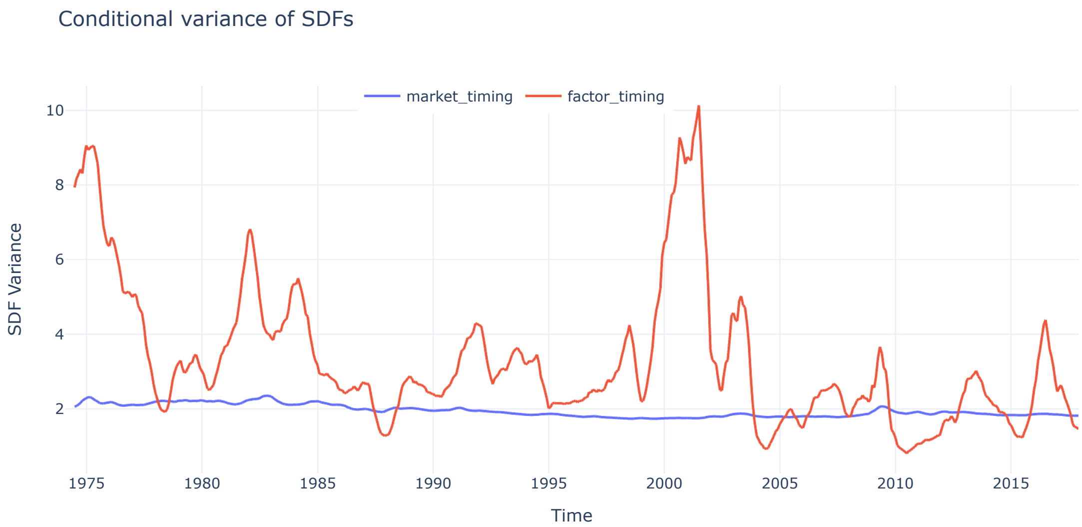

Figure 3.3: Conditional variance of SDFs
In the figure the conditional variances of two estimates of the SDF are presented. The red line represents the SDF volatility obtained from using factor timing as estimate. The blue line presents the SDF volatility using market timing as estimate. The strategies are defined in formula 3.24

Figure 3.4 further highlights the variances of the market timing and factor timing

SDFs, particularly emphasizing the significant spikes in variance during periods of financial crises. To delve deeper into the interplay between crisis periods and the dynamic factor timing based SDF, we carried out a comparative analysis with U.S. consumer price inflation(FRED), as displayed in Figure 3.4. The analysis uncovers a correlation between inflationary peaks and peaks in the SDF, suggesting that crisis periods are indeed characterized by increased variances in the SDF.

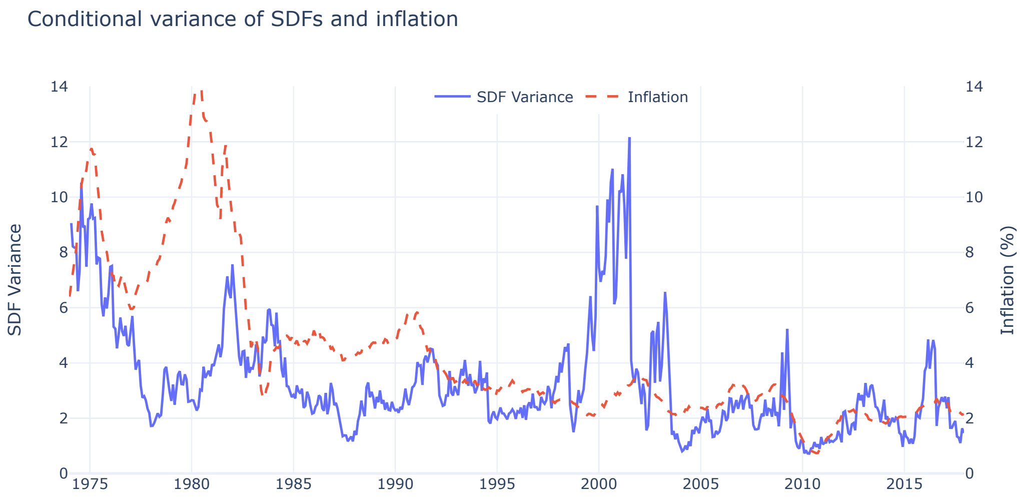

Figure 3.4: SDF variance and U.S. consumer price inflation
This figure compares the variance of the SDF estimate using factor timing with the U.S. consumer price inflation in %(FRED).

Given the observed characteristics of the SDF, it's reasonable to explore what the model suggests about the pricing of risks over time. However, the aforementioned model falls short in investigating such dynamics as it encompasses multiple potential sources of risk, which may be correlated. Nevertheless, the various models outlined previously can inform the development of models capable of identifying priced risks. Notably, as Table 3.6 shows, a static model, such as factor investing, fails to capture the majority of the SDF's variability. To tackle this issue, one possible approach might involve the integration of a time-dependent risk component, adjusting the strategy's exposure based on various risk factors. To evaluate the feasibility of this proposition, Haddad et al. (2020) analyzes the correlation between factor investing and factor timing. Should these two strategies consistently demonstrate a high and stable correlation over time, it would validate the idea that the introduction of a time-dependent risk scaling might be a promising strategy. The temporal correlation of the strategies is as follows:

$$corr_t(m_{F.T.,t+1}, m_{F.I.,t+1}) = \frac{w'_{F.T.,t}\Sigma_Z w_{F.I.}}{\sqrt{\left(w'_{F.T.,t}\Sigma_Z w_{F.T.,t}\right)\left(w'_{F.I.}\Sigma_Z w_{F.I.}\right)}} \tag{3.27}$$

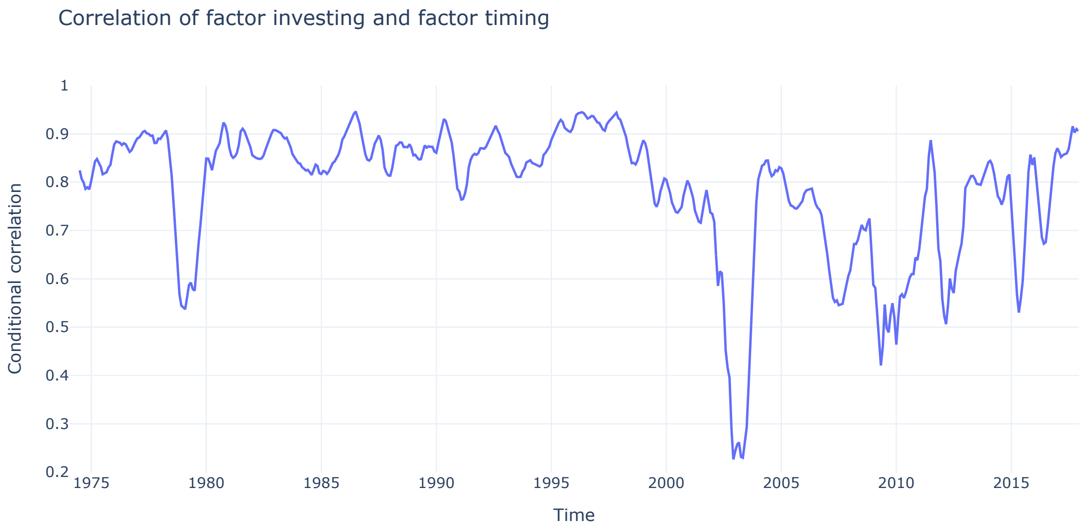

Figure 3.5: Conditional correlation of static and dynamic SDF estimates
This figure plots the conditional correlation of the factor timing estimated SDF and the factor investing estimated SDF, which sets conditional factor means to their sample averages. Reported values are 6-month averages.

Graph 3.5 illustrates the time-dependent correlation between the SDF generated from the factor investing strategy (a static formulation) and the SDF derived from the factor timing strategy. The correlation presents an average value significantly less than 1 and varies over time. This suggests that the SDF is not simply a scaled version of the initial estimate that ignores factor timing evidence. Haddad et al. (2020, p.2014) further delves into this evidence by broadening the analysis to incorporate conditional factors, estimated using the factor timing strategy. This expansion supports their claim that factor timing adds valuable information to the SDF. Further details of this methodology are provided in the Appendix.

4 Conclusion

In our replication of Haddad et al. (2020), our aim was to investigate the development of a strategy that integrates long-short factor investing and market timing. This strategy augments not only the discourse on factor investing but also sheds light on the characteristics of the SDF. Nevertheless, the integration of factor timing brings forward challenges related to dimensionality, due to the multitude of possible factors to be timed. To tackle this issue, we utilize a technique to reduce dimensionality, specifically PCA. The outcomes of timing the PCs of factors display an encouraging performance, surpassing market timing. These strategies are subsequently applied to compute the SDF, unveiling that the variance of the SDF is considerably larger and more fluctuating than previously recorded. Moreover, this variation in the SDF presents cyclical patterns that were not previously detected to this degree. Haddad et al. (2020) concludes that the time-varying properties of the cross-section hold equal significance as the average properties of the cross-section in comprehending asset pricing.

In our replication study, we achieved very similar outcomes for the explained variances in the PCA, suggesting that the data preprocessing techniques used are in close alignment. The parameters acquired from the predictive regressions on the PCs and the market also closely align, despite slightly higher average R^2 values. The observed differences could possibly be due to discrepancies in the sources of the risk-free rate(NASDAQ). On reviewing the plotted predictive regression, we noticed a similar pattern but distinct scaling. Subsequent exploration revealed that the scaling parameter for each PC return must be unique, leading us to infer that they are not adjusted by a single scalar. When analyzing the portfolio strategies, we found comparable performance outcomes but with slightly lower Sharpe ratios. This slight divergence could potentially be due to variations in risk-free returns. In terms of relative performance, the timed strategies that utilized PC predictions exhibited superior Sharpe ratios. Given the elevated average R^2 of the predictive regressions, this deviation seems plausible and underscores the value of conditional information. The properties of the SDF derived from the optimal portfolios closely resemble the original findings. Overall, our findings are in close alignment with those of Haddad et al. (2020), and the few observed deviations point in the same direction as the authors' findings.

A Appendix

A.1 Proofs and derivations

A.1.1 Fundamentals

This section relies on the insights presented in Chapter 4 of Cochrane and Piazzesi (2005).

Time is modeled as a discrete variable $t \in \{1, \ldots, T\}$. Moreover, looking from t, the world can take S states in $t+1$. Definition Stochastic Discount Factor (SDF): The SDF is a random variable $M_{t+1} \in R^S$ which fulfills the following to relations in the economy:

$$p_t = E_t[M_{t+1} x_{t+1}] \tag{A.1}$$

$$B_t = E_t\left[M_{t+1}\right] = \frac{1}{R_{f,t,t+1}} \tag{A.2}$$

p_t is a vector of security prices at time t and X_{t+1} the vector of respective asset payoffs at time $t+1$. B_t denotes the price at time t of a risk-free asset that pays 1 at time $t+1$. $R_{f,t,t+1}$ is the risk-free interest rate between time t and $t+1$.

The law of one price implies the existence of an SDF. Refer to Cochrane and Piazzesi (2005, p.62-66) for the proof. Following his algebraic proof, Cochrane and Piazzesi (2005) shows that a unique $x^* \in X$ can be found that prices all assets, thus is a discount factor by: $x^* = p^T E_t \left[x\ x^T\right] x$. X represents the payoff space, or space of traded payoffs, which is the subspace of R^S (S equals the number of states in $t+1$ that contains payoffs which can be constructed by the assets in the economy. Next it is shown that the SDF is orthogonal to the assets' excess returns: From the definition of the SDF, it follows:

$$p_t = E_t\left[M_{t+1}X_{t+1}\right] \tag{A.3}$$
$$1 = E_t\left[M_{t+1}\frac{X_{t+1}}{p_t}\ \right] = E_t\left[M_{t+1}R_{t+1}\ \right] \tag{A.4}$$

In the same way it also holds that:

$$B_t = E_t\left[M_{t+1}\right] \tag{A.5}$$
$$1 = E_t\left[M_{t+1}\frac{1}{B_t}\right] = E_t\left[M_{t+1}R_{f,t,t+1}\right] \tag{A.6}$$

Subtracting relation A.4 and A.6 and the result follows:

$$0 = 1 - 1 = E_t\left[M_{t+1}R_{t+1}\ \right]E_t\left[M_{t+1}R_{f,t,t+1}\right] \tag{A.7}$$
$$= E_t\left[M_{t+1}(R_{t+1} - R_{f,t,t+1})\right] \tag{A.8}$$
$$= E_t\left[M_{t+1}R^e_{t+1}\right] \tag{A.9}$$

Hansen and Jagannathan (1991) formulation of the SDF: Cochrane and Piazzesi (2005, p.72-73) shows that there exists an alternative formula for the SDF, the one used in section 3.1 (reference) of this paper. Start with the assumption of the existence of an SDF that is a linear function of the shocks to the payoffs:

$$x^* = E\left[x^*\right] + (x - E\left[x\right])^T \tag{A.10}$$

x^* needs to price all assets, thus we need to solve the pricing equation for b. Remember that the following relation holds for the covariance matrix $E\left[x\ x^T\right] - E\left[x\right]E\left[x\right]^T = \Sigma$.

$$p = E[x\ x^*] = E\left[x\left(E[x^*] + (x - E[x])^T b\right)\right] \tag{A.11}$$
$$= E[x]\,E[x^*] + E\left[x(x - E[x])^T\right] b \tag{A.12}$$
$$= E[x]\,E[x^*] + \left(E\left[xx^T\right] - E[x]\,E[x]^T\right) b \tag{A.13}$$
$$= E[x]\,E[x^*] + \Sigma b \tag{A.14}$$
$$\Rightarrow b = \Sigma^{-1}(p - E[x]\,E[x^*]) \tag{A.15}$$

Plugging b back into equation A.10

$$x^* = E[x^*] + (p - E[x]\,E[x^*])^T\,\Sigma^{-1}\,(x - E[x]) \tag{A.16}$$

Since x^* is an SDF and excess returns are used (thus $p = 0$), we can reformulate:

$$x^* = \frac{1}{R_{f,t,t+1}} - \frac{1}{R_{f,t,t+1}} E\left[R^e_{t+1}\right]^T \Sigma^{-1}_{R^e}\left(R^e_{t+1} - E\left[R^e_{t+1}\right]\right) \tag{A.17}$$

Finally, under normalization of $E[x^*] = E[M_{t+1}] = 1$, one arrives at the formulation of the SDF:

$$M_{t+1} = 1 - E\left[R^e_{t+1}\right]^T \Sigma^{-1}_{R^e}\left(R^e_{t+1} - E\left[R^e_{t+1}\right]\right) = 1 - b_t^T\left(R^e_{t+1} - E\left[R^e_{t+1}\right]\right) \tag{A.18}$$

A.1.2 Proof: SR^2 decomposition

The first result states that, under the assumption of uncorrelated returns, the expected conditional squared Sharpe Ratio can be decomposed to in the following way:

$$E\left[SR^2_t\right] = \sum_{i=1}^{N} \frac{E\left[r_{i,t+1}\right]^2}{\sigma^2_{i,t}} + \sum_{i=1}^{N}\left(\frac{R^2_i}{1 - R^2_i}\right) \tag{A.19}$$

To see that, start with a vector of returns $R_{t+1} = (r_1, \ldots, r_N)^T$. Without the assumption of uncorrelated asset returns, the maximum conditional SR equals:

$$SR_t^2 = E_t \left[R_{t+1}\right]^T \Sigma_{R,t}^{-1} E_t \left[R_{t+1}\right] \tag{A.20}$$

Now, under the assumption of conditionally uncorrelated asset returns, the variance-covariance matrix is a diagonal matrix, with the assets' respective variances as the diagonal elements. This simplifies SR_t^2 to:

$$SR_t^2 = \begin{pmatrix} E_t \left[r_{1,t+1}\right] \\ \cdots \\ E_t \left[r_{N,t+1}\right] \end{pmatrix}^T \begin{pmatrix} \frac{1}{\sigma_{1,t}^2} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\sigma_{N,t}^2} \end{pmatrix} \begin{pmatrix} E_t \left[r_{1,t+1}\right] \\ \cdots \\ E_t \left[r_{N,t+1}\right] \end{pmatrix} \tag{A.21}$$

$$SR_t^2 = \sum_{i=1}^{N} \frac{E_t \left[r_{i,t+1}\right]^2}{\sigma_{i,t}^2} \tag{A.22}$$

Under the assumption of homoscedastic returns, $\sigma_{i,t}^2$ further simplifies to σ_i^2, leading to:

$$SR_t^2 = \sum_{i=1}^{N} \frac{E_t \left[r_{i,t+1}\right]^2}{\sigma_i^2} \tag{A.23}$$

Taking the unconditional expectation of both sides leads to:

$$E\left[SR_t^2\right] = \sum_{i=1}^{N} \frac{E\left[E_t \left[r_{i,t+1}\right]^2\right]}{\sigma_i^2} \tag{A.24}$$

Using the law of total variance $E\left[E_t \left[r_{i,t+1}\right]^2\right] = E\left[r_{i,t+1}\right]^2 + var\left(E_t \left[r_{i,t+1}\right]\right)$ brings us to:

$$E\left[SR_t^2\right] = \sum_{i=1}^{N} \frac{E\left[r_{i,t+1}\right]^2}{\sigma_i^2} + \sum_{i=1}^{N} \frac{var\left(E_t\left[r_{i,t+1}\right]\right)}{\sigma_i^2} \tag{A.25}$$

Now, recall the definition of the R-squared of asset i : $R_i^2 = 1 - \frac{\sigma_i^2}{var(E_t[r_{i,t+1}])+\sigma_i^2}$. Then the following holds:

$$\begin{aligned}
1 - R_i^2 &= \frac{\sigma_i^2}{var\left(E_t\left[r_{i,t+1}\right]\right) + \sigma_i^2} \\
\frac{1}{1 - R_i^2} &= \frac{var\left(E_t\left[r_{i,t+1}\right]\right) + \sigma_i^2}{\sigma_i^2} \\
\frac{R_i^2}{1 - R_i^2} &= \left(1 - \frac{\sigma_i^2}{var\left(E_t\left[r_{i,t+1}\right]\right) + \sigma_i^2}\right) \frac{var\left(E_t\left[r_{i,t+1}\right]\right) + \sigma_i^2}{\sigma_i^2} \\
\frac{R_i^2}{1 - R_i^2} &= \frac{var\left(E_t\left[r_{i,t+1}\right]\right) + \sigma_i^2}{\sigma_i^2} - 1 \\
\frac{R_i^2}{1 - R_i^2} &= \frac{var\left(E_t\left[r_{i,t+1}\right]\right)}{\sigma_i^2} + \frac{\sigma_i^2}{\sigma_i^2} - 1 \\
\frac{R_i^2}{1 - R_i^2} &= \frac{var\left(E_t\left[r_{i,t+1}\right]\right)}{\sigma_i^2}
\end{aligned} \tag{A.26}$$

Substituting this into the previous result concludes the proof:

$$E\left[SR_t^2\right] = \sum_{i=1}^{N} \frac{E\left[r_{i,t+1}\right]^2}{\sigma_{i,t}^2} + \sum_{i=1}^{N} \left(\frac{R_i^2}{1 - R_i^2}\right) \tag{A.27}$$

A.1.3 Proof Lemma 1

Conditional factor model of form:

$$E\left[R_{j,t+1}\right] = \beta_{j,t}^T \Sigma_{F,t} \delta_t = \beta_{j,t}^T E_t\left[F_{t+1}\right] \tag{A.28}$$

is equivalent to the formulation of the SDF:

$$M_{t+1} = 1 - (F_{t+1} - E[F_{t+1}])^T \delta_t$$
$$E_t[M_{t+1} R_{j,t+1}] = 0 \tag{A.29}$$

For simplicity, $E_t[M_{t+1}]$ is normalized to 1. $R_{j,t+1}$ is an excess return. This proof is adapted from Cochrane and Piazzesi (2005, p.106-107). Proof : Start with the fundamental pricing equation of the SDF:

$$\begin{aligned} E_t[M_{t+1} R_{j,t+1}] &= 0 \\ E_t[R_{j,t+1}] &= cov_t\left(R_{j,t+1}, 1 - (F_{t+1} - E[F_{t+1}])^T\right)\delta_t \\ E_t[R_{j,t+1}] &= cov_t\left(R_{j,t+1}, F_{t+1}^T\right)\delta_t \\ E_t[R_{j,t+1}] &= cov_t\left(R_{j,t+1},\ F_{t+1}^T\right)\Sigma_{F,t}^{-1}\Sigma_{F,t}\delta_t \end{aligned} \tag{A.30}$$

Now substitute $\Sigma_{F,t}\delta_t = E_t[F_{t+1}]$ to arrive at the result:

$$E_t[R_{j,t+1}] = \beta_{j,\ t}^T E_t[F_{t+1}] \tag{A.31}$$

Where the betas with respect to the single factors are given by:

$$\beta_{j,\ t}^T = cov_t\left(R_{j,t+1},\ F_{t+1}^T\right)\Sigma_{F,t}^{-1} \tag{A.32}$$

A.1.4 Proof of the R_{total}^2 derivation

The proof starts with the definition of R_{total}^2 for a set of factor returns $F_{t+1} = (f_1, \ldots, f_K)^T$

$$R_{total}^2 = \frac{tr[cov(E_t[F_{t+1}])]}{tr[cov(F_{t+1})]} \tag{A.33}$$

The trace function, tr, simply takes the sum of the diagonal elements of the matrix. It is invariant under similarity transformation. This means that if we have a square matrix A (or $cov\left(F_{t+1}\right)$) and an invertible matrix Q, the trace of A and QAQ^{-1} are the same. If matrix Q is orthogonal, it's inverse equals it's transpose, thus the following holds:

$$R_{total}^{2}=\frac{tr\left[Q^{T}cov\left(E_{t}\left[F_{t+1}\right]\right)Q\right]}{tr\left[Q^{T}cov\left(F_{t+1}\right)Q\right]} \tag{A.34}$$

Under the assumption of homoscedastic factor returns (constant covariance matrix of factor returns $\Sigma_{F,t}$), we can perform the following eigenvalue decomposition: $\Sigma_{F,t}=Q\Lambda Q^{T}$. One can now construct the PCs: $PC_{t+1}=Q^{T}F_{t+1}$. Also remember that$Q^{T}\Sigma_{F,t}Q=\Sigma_{PC,t}$. It follows:

$$R_{total}^{2}=\frac{tr\left[cov\left(E_{t}\left[PC_{t+1}\right]\right)\right]}{tr\left[cov\left(PC_{t+1}\right)\right]} \tag{A.35}$$

Moreover, under the law of total covariance it holds: $cov\left(PC_{t+1}\right)=\Lambda+cov\left(E_{t}\left[PC_{t+1}\right]\right)$, By the additivity of the trace function, this leads to:

$$R_{total}^{2}=\frac{tr\left[cov\left(E_{t}\left[PC_{t+1}\right]\right)\right]}{tr\left[\Lambda\right]+tr\left[cov\left(PC_{t+1}\right)\right]} \tag{A.36}$$

From Proof A1, recall that $\frac{R_{i}^{2}}{1-R_{i}^{2}}=\frac{var\left(E_{t}\left[PC_{i,t+1}\right]\right)}{\lambda_{i}}$ holds.

$$\begin{aligned}R_{total}^{2}&=\sum_{i}\frac{var\left(E_{t}\left[PC_{i,t+1}\right]\right)}{tr\left[\Lambda\right]+tr\left[cov\left(PC_{t+1}\right)\right]}\\R_{total}^{2}&=\sum_{i}\frac{var\left(E_{t}\left[PC_{i,t+1}\right]\right)}{tr\left[\Lambda\right]+tr\left[cov\left(PC_{t+1}\right)\right]}\,\frac{\lambda_{i}}{\lambda_{i}}\end{aligned} \tag{A.37}$$

Define $\lambda=tr\left[\Lambda\right]+tr\left[cov\left(PC_{t+1}\right)\right]$ to arrive at:

$$R^2_{total} = \sum_i \frac{var\left(E_t\left[PC_{i,t+1}\right]\right)}{\lambda_i}\ \frac{\lambda_i}{\lambda}$$

$$R^2_{total} = \sum_i \left(\frac{R_i^2}{1-R_i^2}\right)\ \frac{\lambda_i}{\lambda} \tag{A.38}$$

Also note that because $1 - R_i^2 = \frac{\lambda_i}{var(E_t[PC_{i,t+1}])+\lambda_i}$

$$\lambda = \ tr\left[\Lambda\right] + tr\left[cov\left(PC_{t+1}\right)\right]$$

$$\lambda = \ \sum_i \lambda_i + var\left(E_t\left[PC_{i,t+1}\right]\right)$$

$$\lambda = \sum_i \frac{(\lambda_i + var\left(E_t\left[PC_{i,t+1}\right]\right))\lambda_i}{\lambda_i}$$

$$\lambda = \sum_i \frac{\lambda_i}{1-R_i^2} \approx \sum_i \lambda_i \tag{A.39}$$

A.1.5 Highest Sharpe Ratio Portfolio

The portfolio resulting in the maximum Sharpe ratio can be found by solving the following maximization problem:

$$\max_{w_t} \frac{w_t^T E_t\left[F_{t+1}\right]}{\sqrt{w_t^T \Sigma_{F,t} w_t}} \tag{A.40}$$

We solve this problem using Lagrangian method.

$$L\left(w,\lambda\right) = \left(w_t^T E_t\left[F_{t+1}\right]\right)\left(w_t^T \Sigma_{F,t} w_t\right)^{-\frac{1}{2}} + \lambda\left(w_t^T e - 1\right) \tag{A.41}$$

This leads us to the following FOCs:

$$\frac{\partial L(w_t,\lambda)}{\partial w_t} = E_t[F_{t+1}]\left(w_t^T \Sigma_{F,t} w_t\right)^{-\frac{1}{2}} - \left(w_t^T E_t[F_{t+1}]\right)\left(w_t^T \Sigma_{F,t} w_t\right)^{-\frac{3}{2}} \Sigma_{F,t} w_t + \lambda e = 0$$

$$\frac{\partial L(w_t,\lambda)}{\partial \lambda} = t^T e = 1 \tag{A.42}$$

Solving this system of linear equations leads us to the final result:

$$w_t = \frac{\Sigma_{F,t}^{-1} E_t[F_{t+1}]}{e^T \Sigma_{F,t}^{-1} E_t[F_{t+1}]} \tag{A.43}$$

where denominator $e^T \Sigma_{F,t}^{-1} E_t[F_{t+1}]$ is a scaling factor that ensures that portfolio weights sum up to 1. Note that we derive the same direction of the weight vector w_t as Haddad et al. (2020), which is given by the nominator $\Sigma_{F,t}^{-1} E_t[F_{t+1}]$. However, due to the constraint in our optimization method, w_t is scaled by the denominator $e^T \Sigma_{F,t}^{-1} E_t[F_{t+1}]$.

A.1.6 Variance of the SDF

First note that the SDF can be formulated for the weight vector of the optimal portfolio in the following way:

$$M_{t+1} = 1 - w_t^T (Z_{t+1} - E[Z_{t+1}]) \tag{A.44}$$

For the variance of the SDF, it now holds:

$$\begin{aligned} var_t(M_{t+1}) &= var_t(1 - w_t^T (Z_{t+1} - E[Z_{t+1}]) \\ &= var_t\left(w_t^T Z_{t+1}\right) \\ &= w_t^T \Sigma_Z w_t \end{aligned} \tag{A.45}$$

A.1.7 Squared Sharpe ratio: invariance against rotations in asset space

Start with the definition of the squared Sharpe ratio. It follows from matrix transpose ($(AB)^T = B^T A^T$) and inverse properties $(AB)^{-1} = B^{-1}A^{-1}$:

$$
\begin{aligned}
SR_t^2 &= E_t\left[R_{t+1}\right]^T \Sigma_{R,t}^{-1} E_t\left[R_{t+1}\right] \\
SR_t^2 &= E_t\left[R_{t+1}\right]^T Q Q^T \Sigma_{R,t}^{-1} Q Q^T E_t\left[R_{t+1}\right] \\
SR_t^2 &= E_t\left[R_{t+1}^T Q\right] Q^T \Sigma_{R,t}^{-1} Q E_t\left[Q^T R_{t+1}\right] \\
SR_t^2 &= E_t\left[Q^T R_{t+1}\right]^T Q^{-1} \Sigma_{R,t}^{-1} Q E_t\left[Q^T R_{t+1}\right] \\
SR_t^2 &= E_t\left[PC_{t+1}\right]^T \left(Q^T \Sigma_{R,t} Q\right)^{-1} E_t\left[PC_{t+1}\right] \\
SR_t^2 &= E_t\left[PC_{t+1}\right]^T \Sigma_{PC,t}^{-1} E_t\left[PC_{t+1}\right] \qquad \text{(A.46)}
\end{aligned}
$$

Remember that Q is an orthogonal matrix, thus: $Q^T = Q^{-1}$. This proof assumes that Q results from the eigenvalue decomposition of $\Sigma_{R,t}$.

(a) Market

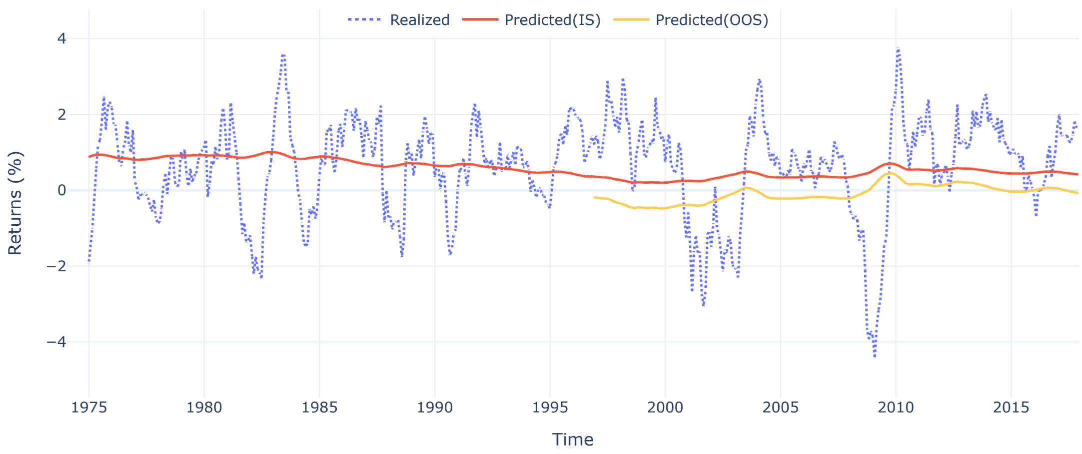

Figure A.1: Market predictability in and out-of-sample

(b) PC1

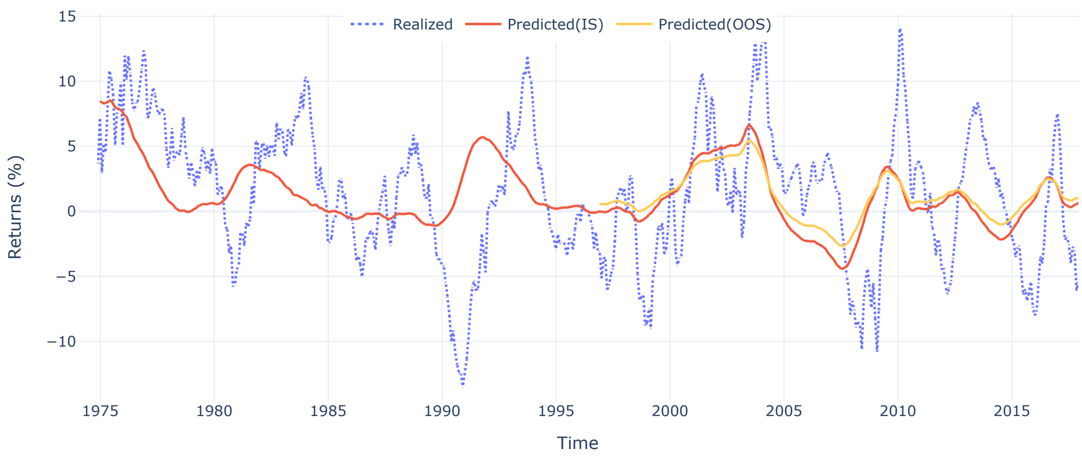

Figure A.2: PC1 predictability in and out-of-sample

A.2 Further results

A.2.1 Predictive regression plots

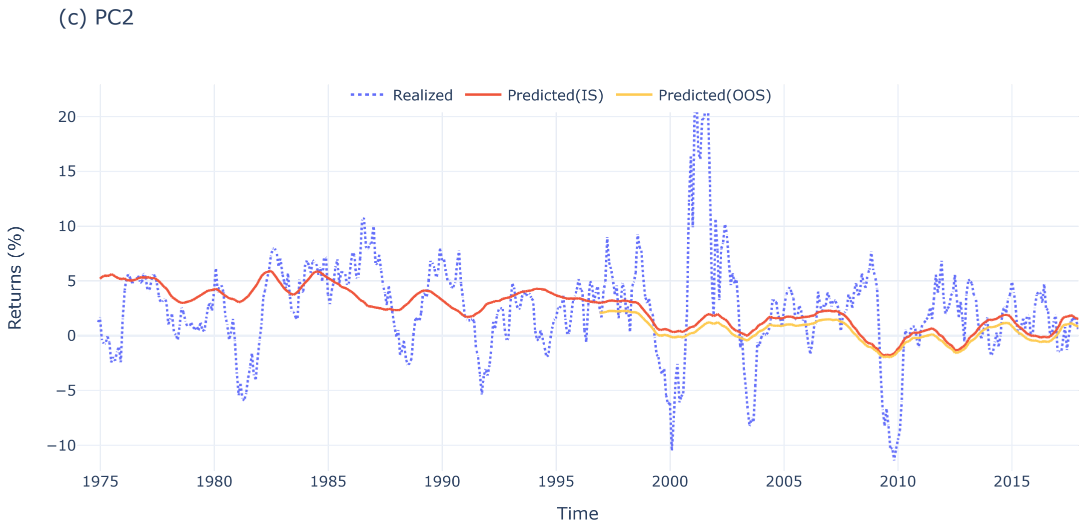

Figure A.3: PC2 predictability in and out-of-sample

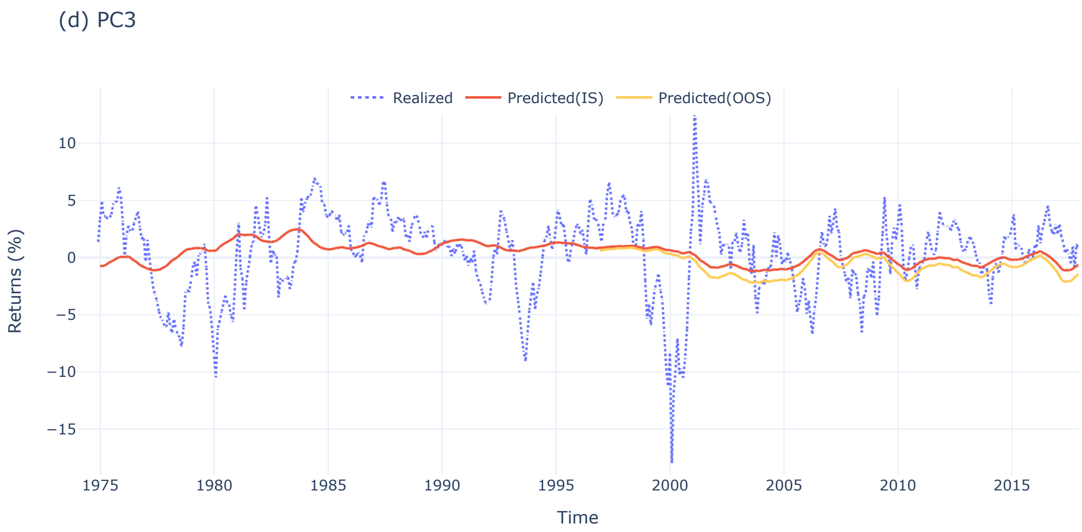

Figure A.4: PC3 predictability in and out-of-sample

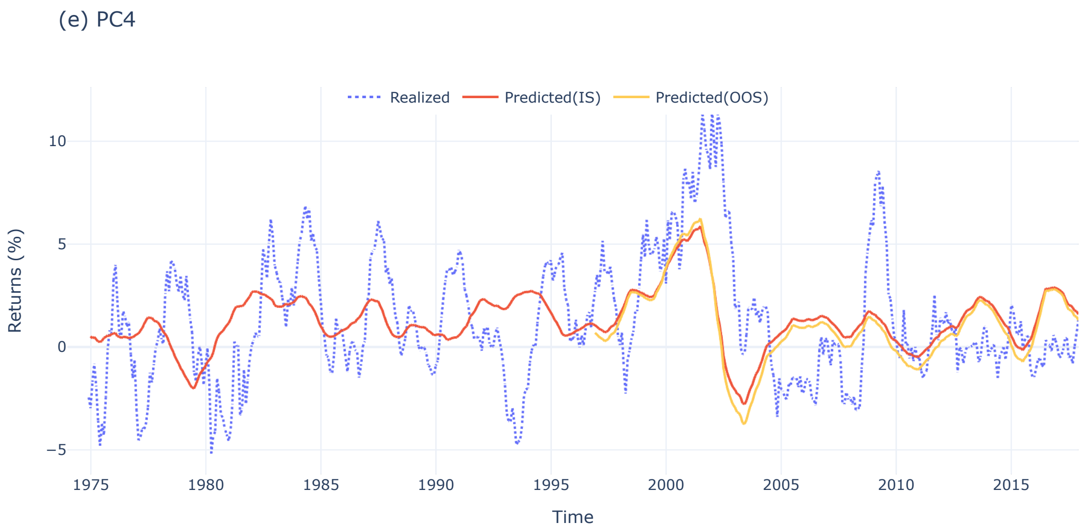

Figure A.5: PC4 predictability in and out-of-sample

(f) PC5

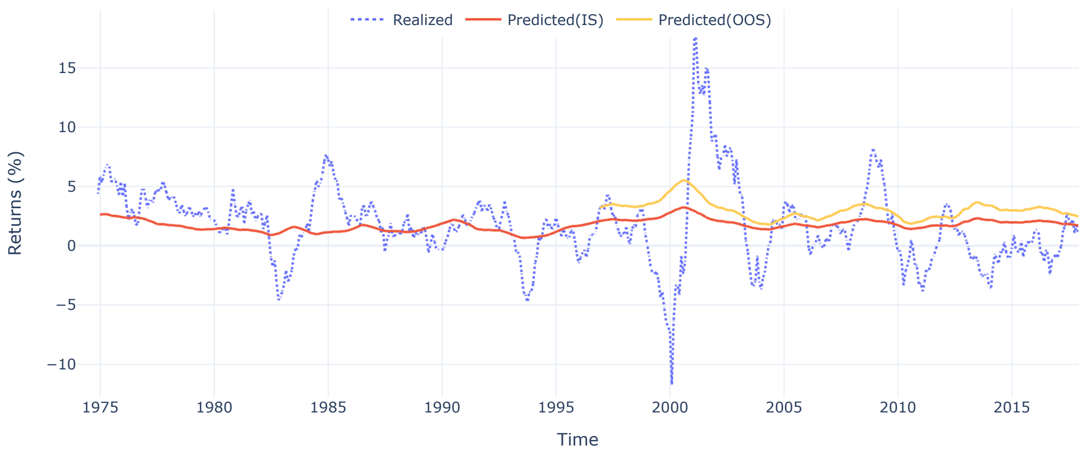

Figure A.6: PC5 predictability in and out-of-sample

A.2.2 Factor predictability

Table A.1: Predicting individual anomaly returns: R^2 (%)

Monthly predictive R^2 of individual anomalies returns using implied fitted values based on PC forecasts. IS provides estimates in full sample. OOS shows out-of-sample R^2

#	Characteristic	IS	OOS
1	Size	3.1	3.8
2	Value (A)	1.7	1.4
3	Gross Profitability	-2.2	-3.9
4	Value-Profitablity	3.1	3.2
5	F-score	0.2	-1.3
6	Debt Issuance	-0.6	-0.2
7	Share Repurchases	1.1	1.1
8	Net Issuance (A)	2.4	2.8
9	Accruals	-0.3	-0.6
10	Asset Growth	1.4	0.8
11	Asset Turnover	0.5	0.2
12	Gross Margins	-0.4	-2.8
13	Earnings/Price	0.6	-0.3
14	Cash Flows/Price	0.2	-0.2
15	Net Operating Assets	-1.2	-5.5
16	Investment/Assets	1.0	-0.6
17	Investment/Capital	0.1	-0.5
18	investment Growth	1.2	0.4
19	Sales Growth	0.9	0.7
20	Leverage	0.6	0.4
21	Return on Assets (A)	0.9	1.3
22	Return on Book Equity (A)	0.7	-0.4
23	Sales/Price	1.3	1.2
24	Growth in LTNOA	0.5	0.0
25	Momentum (6m)	2.0	2.0
26	Value-Momentum	0.6	1.9
27	Value-Momentum-Prof.	1.6	1.7
28	Short Interest	0.2	-1.3
29	Momentum (12m)	0.6	1.2
30	Industry Momentum	-0.1	-0.7
31	Momentum-Reversals	1.7	2.4
32	Long Run Reversals	5.0	4.8
33	Value (M)	2.9	2.0
34	Net Issuance (M)	0.9	1.4
35	Earnings Surprises	-4.4	-2.9
36	Return on Book Equity (Q)	-0.8	-0.5
37	Return on Market Equity	-0.4	-0.0
38	Return on Assets (Q)	-0.7	0.4
39	Short-Term Reversals	-0.8	0.5
40	Idiosyncratic Volatility	0.9	-0.9
41	Beta Arbitrage	-1.4	-2.3
42	Seasonality	-1.9	-0.1
43	Industry Rel. Reversals	-3.7	1.0
44	Industry Rel. Rev. (L. V.)	-14.1	-0.9
45	Ind. Mom-Reversals	-3.9	-2.2
46	Composite Issuance	-0.7	-2.4
47	Price	4.4	3.0
48	Share Volume	-0.2	-1.3
49	Duration	1.8	2.3
50	Firm age	0.4	0.8

A.2.3 Strategy weights over time

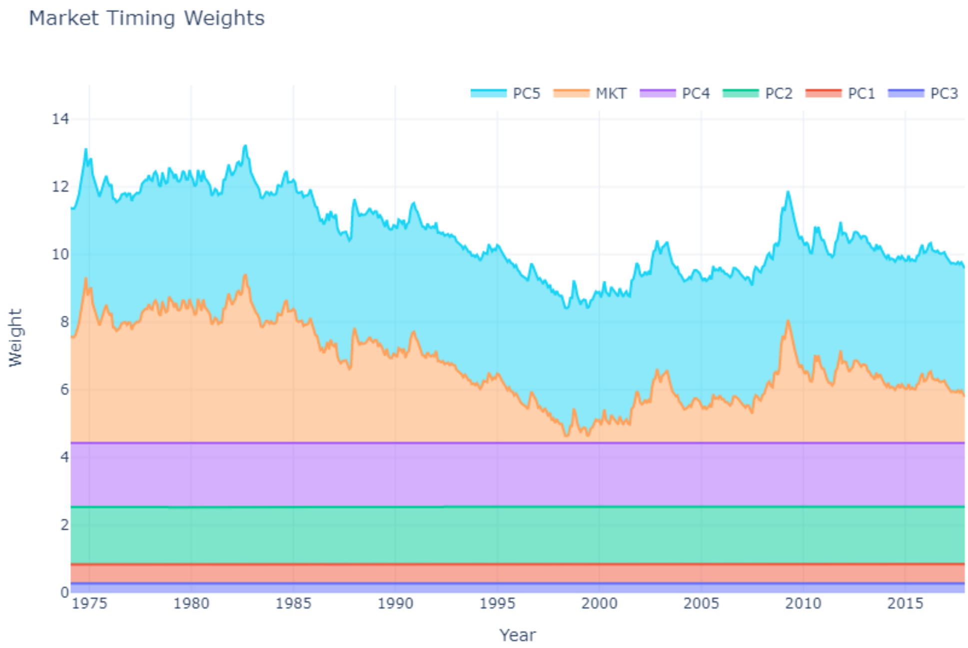

Figure A.7: Market timing weights
The plot shows the time-series of the weight the strategy market timings puts on the different investment opportunities

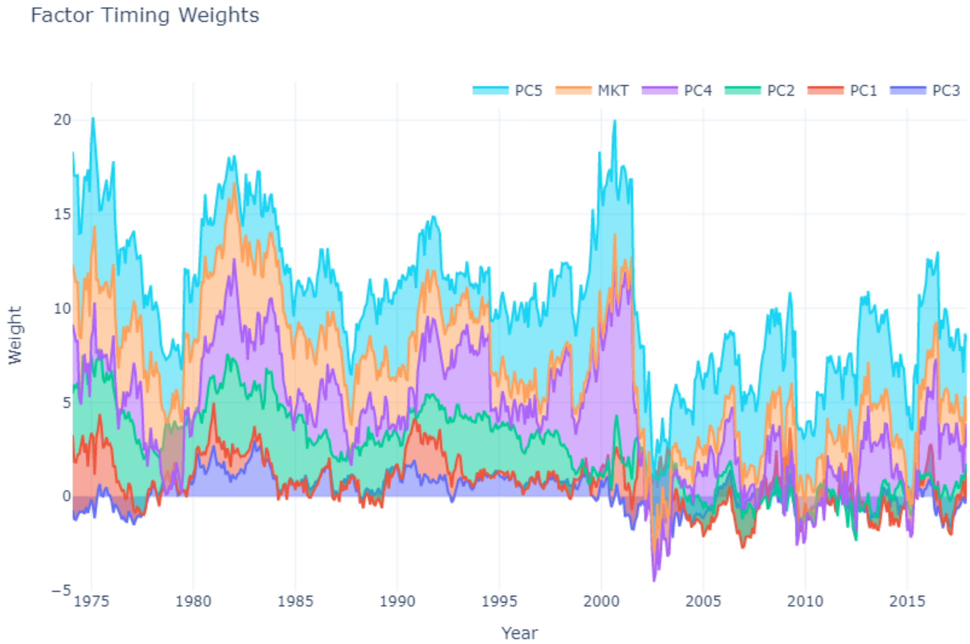

Figure A.8: Factor timing weights
The plot shows the time-series of the weight the strategy factor timings puts on the different investment opportunities

A.2.4 Predicted characteristic returns

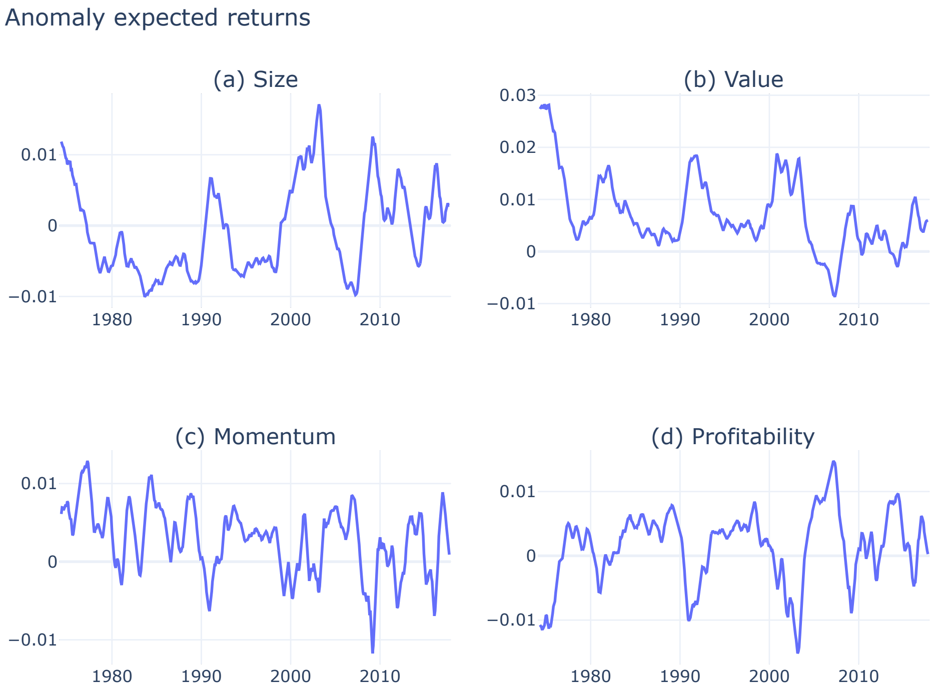

Figure A.9: Expected anomaly returns
The plot shows the time-series of conditional expected returns on four anomaly strategies: size, value, momentum, and return on assets

B List of Figures

C List of Tables

D List of Abbreviations

bm logarithmic boot-to-market ratio. 3

IS in-sample. 13, 16

OLS omitted least squares. 17

OOS out-of-sample. 15–17, 19

PC Principal Component. 1, 2, 4, 5, 8–21, 25, 46

PCA Principal Component Analysis. 1, 2, 4, 5, 10, 11, 25

SDF Stochastic Discount Factor. 1, 2, 5–10, 14, 20–29

E Bibliography

R. D. Arnott, N. Beck, and V. Kalesnik. Timing "smart beta" strategies? of course! buy low, sell high! 2016.

C. Asness, S. Chandra, A. Ilmanen, and R. Israel. Contrarian factor timing is deceptively difficult. *The Journal of Portfolio Management*, 43(5):72–87, 2017.

C. S. Asness. The siren song of factor timing aka "smart beta timing" aka "style timing". *The Journal of Portfolio Management*, 42(5):1–6, 2016.

J. H. Cochrane and M. Piazzesi. Bond risk premia. *American Economic Review*, pages 138–160, 2005.

R. B. Cohen, C. Polk, and T. Vuolteenaho. The value spread. *The Journal of Finance*, 58(2):609–641, 2003.

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

FRED. Sticky price consumer price index. `https://fred.stlouisfed.org/series/STICKCPIM159SFRBATL`.

V. Haddad, S. Kozak, and S. Santosh. Factor timing. (26708), January 2020. doi: 10.3386/w26708. URL `http://www.nber.org/papers/w26708`.

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

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

A. Ilmanen and L. N. Nielsen. Are defensive stocks expensive? a closer look at value spreads. 2015.

I. Jollife. *Principal component analysis*. 2002.

B. Kelly and S. Pruitt. Market expectations in the cross-section of present values. *The Journal of Finance*, 68(5):1721–1756, 2013.

S. Kozak. Data source of the "factor timing". `https://sites.google.com/site/serhiykozak/data?authuser=0`.

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

S. Nagel and K. J. Singleton. Estimation and evaluation of conditional asset pricing models. *The Journal of Finance*, 66(3):873–909, 2011.

NASDAQ. Us treasury zero-coupon yield curve. `https://data.nasdaq.com/data/FED/SVENY`.

K. Pearson. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572, 1901.

S. Sanguansat. Principal component analysis for portfolio optimization. *International Journal of Computer Science Issues*, 9(3):1694–0814, 2012.

Declaration of Honesty,

I hereby declare that I, Lukas Körber, have independently and without using sources or aids other than those specified, written the present work. Directly quoted sentences or parts of sentences are indicated as citations, and other references, in terms of content and scope, are acknowledged. The work has not been submitted to any other examination board in the same or similar form, and it has not been published. It has not been used, either in full or in part, for any other examination or academic achievement.

Frankfurt am Main, 13.07.2023,

L. Körber