

Factor Timing with Portfolio Characteristics*

Anastasios Kagkadis[†] Ingmar Nolte[‡] Sandra Nolte[§] Nikolaos Vasilas[¶]

March 24, 2023

Abstract

In a factor timing context, academic research has focused on identifying a set of predictors that can explain the dynamics of factor portfolios. We propose an alternative approach for timing factor portfolio returns by exploiting the information from their portfolio characteristics. Different combinations of dimension reduction techniques are employed to independently reduce both the number of predictors and portfolios to predict. Characteristic-based models outperform existing methods in terms of exact predictability, as well as investment performance.

Keywords: *Return Predictability, Factor Portfolios, Dimension Reduction, Factor Timing, Anomalies*

JEL Classification: *G10, G11, C52, C55*

*We would like to thank Jeffrey Pontiff (the Editor), an anonymous referee, David Buckle, Amit Goyal, Harald Lohre, Alejandro Lopez-Lira, Robert Macrae, Winfried Pohlmeier, as well as participants at the INQUIRE UK Spring 2022 Residential Seminar, the 11th International Conference of the Financial Engineering and Banking Society, the QFFE 2022 International Conference, the 2022 EFMA Annual Meeting, the 3rd Frontiers of Factor Investing Conference and the 2022 FMA Annual Meeting for their valuable comments and suggestions.

[†]Lancaster University Management School, Bailrigg, Lancaster LA1 4YX, Lancaster, email: a.kagkadis@lancaster.ac.uk

[‡]Lancaster University Management School, Bailrigg, Lancaster LA1 4YX, Lancaster, email: i.nolte@lancaster.ac.uk

[§]Lancaster University Management School, Bailrigg, Lancaster LA1 4YX, Lancaster, email: s.nolte@lancaster.ac.uk

[¶]Corresponding author: Lancaster University Management School, Bailrigg, Lancaster LA1 4YX, Lancaster, email: n.vasilas@lancaster.ac.uk

1 Introduction

The asset pricing literature has long been shaped by the idea that observable firm characteristics convey information about the cross-section of expected stock returns. A common practice in the literature is to extract the risk premium associated with these characteristics by constructing long-short (LS) factor portfolios (Fama and French, 1993). Such zero-investment, market-neutral portfolios have given rise to the so-called *factor investing*. Yet, there are benefits over and above static factor investing. Studies such as Stambaugh et al. (2012), Jacobs (2015), Akbas et al. (2016) and Keloharju et al. (2016) show that the performance of LS portfolios, and therefore the benefits from factor investing, are significantly time-varying. More importantly, such time variation in performance is not harmonious across portfolios, allowing for substantial investment gains from timing factor portfolio returns.¹ As such, from an investor’s perspective timing is important and an active factor allocation is needed in order to capitalize on the fluctuations in LS portfolio returns.

In a factor timing context, several studies have emerged utilizing a variety of predictive signals as a way to improve upon static factor investing. Valuation ratios, investor sentiment, issuer-repurchaser spread and technical indicators, such as factor momentum, are the most prominent examples, among others. In this paper, we create an optimal factor timing strategy, going over and above existing methods for predicting factor portfolio returns. In doing so, we extend the predictability of stock returns from observable firm characteristics to a portfolio level and predict factor portfolio returns using a collection of portfolio characteristics. Specifically, the characteristics used to sort stocks into portfolios are subsequently aggregated into portfolio characteristics and used as predictive variables to forecast future factor portfolio returns. The use of multiple characteristics to predict individual factor portfolio returns is motivated by the fact that many stocks

¹For example, Haddad et al. (2020) find that the loadings of a size portfolio on the optimal factor timing portfolio are pro-cyclical while those of a momentum portfolio are counter-cyclical.

coexist in different factor portfolio legs simultaneously.² Hence, it is sensible to assess the joint predictability that arises from characteristics at a portfolio level and examine the possibility that factor portfolios are predictable by characteristics other than their own. A comparison of the collective characteristic-based predictability against alternative sets of predictors documented in the literature highlights the joint importance of characteristics in explaining the dynamics of factor portfolios.

A key aspect of our methodology is the use of different dimension reduction techniques to reduce the dimensions of both sides of the predictability problem. In line with Haddad et al. (2020), we begin by reducing the number of forecasting targets, recognizing the underlying factor structure in factor portfolio returns. Instead of independently predicting individual anomalies, we focus our attention on the main sources of return variation by isolating the first five Principal Components (PCs). These PCs capture around 67% of the variation in factor portfolio returns (see Figure IA.1 in the Internet Appendix), allowing us to greatly reduce the dimensions of the problem at the expense of little return variation foregone. Since the dominant PCs capture common variation in the underlying risk premia, being able to accurately predict their performance leads to the detection of robust predictive patterns across individual anomalies.³ In addition, PCs are not just statistical factors but have an investable interpretation as well. As each PC is a linear combination of the underlying variables, PC portfolios are portfolios of factor portfolios, meaning that their returns and characteristics are calculable. To construct the PC portfolios we use conventional PCA, as well as the Risk Premium PCA (RPPCA) proposed by Lettau and Pelger (2020a).⁴ Unlike PCA, RPPCA utilises information of the mean returns of the factor portfolios in addition to their variances and leads to the extraction of

²For example, the stocks with the highest asset growth are also the ones with the lowest book-to-market ratio, the highest return on assets and the highest accruals (Cooper et al., 2008).

³Applying Principal Component Analysis (PCA) to a set of factor portfolio returns in order to achieve dimension reduction has recently gained a lot of attention in asset pricing. For example, Haddad et al. (2020) form PC portfolios by running PCA on a set of 50 anomalies and use their own book-to-market ratio to predict their performance.

⁴Henceforth, “PC portfolios” refers to the estimation of the Principal Component portfolios using either PCA or RPPCA.

factors that may explain a smaller part of the time-series variation but are important in pricing the cross-section. The resulting PCs have higher Sharpe ratios and in our context help guiding the forecasting study around factor portfolios with higher average returns.

We then proceed by compressing the predictive information from the characteristics of the PC portfolios. To achieve this, we do not only rely on PCA, but employ methods that account for the covariance structure between predictors and forecasting targets, such as Partial Least Squares (PLS) (Wold et al., 1984). Conventional PCA focuses on the variance within the predictors and can lead to components that mix return-relevant and irrelevant variation. By using PLS we aim to capture only the variation in the characteristics that is relevant in predicting returns, potentially resulting in sparser and more accurate models.

After rotating characteristics in space using PCA or PLS, we either use the first characteristic component in standard predictive regressions or apply LASSO on the whole set of characteristic components to identify the relevant subset of features.⁵ The first case is used to investigate the predictability in the simplest case of a single predictive factor, while the use of LASSO allows for successive components to be included in the surviving subset of predictors, with the importance of each characteristic component being assessed based on its contribution to minimizing the forecasting error rather than the magnitude of its eigenvalue. Our procedure is implemented recursively and the optimal degree of coefficient shrinkage is identified separately for each PC portfolio based on a cross-validation step. This approach has two important implications. First, the number of factors can be different across PC portfolios, allowing for different sources of variation in factor portfolio returns to be approximated by models of different complexity. For instance, many characteristic components may be required to predict the first PC portfolio but only a few for the second. Second, allowing for different values for the level of

⁵The combination of LASSO with PCA or PLS is particularly suitable in this case because the PC portfolios are by construction orthogonal.

coefficient shrinkage across time allows us to examine the time variation in the strength of the characteristic signal overall.

In our empirical analysis we use a collection of 72 anomalies spanning the period from 1970 to 2019 and find that characteristics are particularly useful for factor timing purposes. We distinguish factor portfolio predictability in terms of exact predictive accuracy (comparing predicted with future realized returns) and ability to predict the cross-sectional dispersion in returns (differentiating winners from losers). The characteristic-based models that incorporate LASSO are the most successful and consistently outperform existing methods in both terms as they deliver smaller forecasting errors and higher cross-sectional correlations between forecasted and realized returns. They also deliver average monthly returns of up to 1.47% and annualized Sharpe ratios of up to 0.73, while the best benchmark model delivers 1.06% and 0.55, respectively. Importantly, our factor timing strategies show no decay in return performance over time, although many individual anomalies have been found empirically to do so (McLean and Pontiff, 2016).

In terms of the different methods used, the implications of using PCA or RPPCA to reduce the number of portfolios to predict are minimal. Yet, when it comes to reducing the number of predictors down to a single predictive factor, the dimension reduction technique matters. In particular, PCA delivers slightly better exact predictability, but severely underperforms PLS in terms of ranking the anomaly portfolios successfully. Essentially, when a single-factor model is used, it is better to condense the information from the predictors using a tool that is specifically designed for forecasting purposes.

Nonetheless, the difference between PCA and PLS disappears when multiple characteristic components are considered in conjunction with LASSO, suggesting that the exact rotation method of the predictors is less important once we account for the whole information set. After employing LASSO results improve uniformly across models reflecting

the importance of accounting for further components and the benefits of regularization in dealing with overfitting. Furthermore, the cross-validation step reveals that the required number of features varies significantly across time for all the PC portfolios. This implies that characteristics work better in predicting returns in certain periods than others, which is expected given the time-variation in factor risk premia. Our LASSO-based factor timing strategies are flexible enough to downgrade (upgrade) information in the characteristics when their informativeness is low (high).

The rest of the paper is structured as follows: Section 2 provides a discussion of the relevant literature. Section 3 describes the general framework and our estimation approach and Section 4 provides an assessment of the various models in terms of forecasting ability and investment performance. Section 5 modifies the baseline models and examines alternative model specifications. Finally, Section 6 concludes.

2 Literature review

Our paper is related to several strands of the literature. Without attempting a full-scale review, we discuss briefly how we contribute to two main categories, namely studies that utilize dimension reduction techniques in the context of asset pricing and studies that explore factor portfolio predictability.

Dimension reduction in asset pricing

Machine learning has surfaced in recent years in various asset pricing applications due to the limitations of standard methodologies in a high dimensional setting. Gu et al. (2020) compare various machine learning techniques in their effort to forecast stock returns using a large collection of stock characteristics. Similarly, numerous studies attempt to identify the extent to which characteristics are associated with expected returns by regularizing the cross-sectional regressions or the characteristic-based portfolio sorts used in the esti-

mation of risk premia. For instance, DeMiguel et al. (2020), Freyberger et al. (2020) and Feng et al. (2020) employ LASSO regularization to create a stochastic-discount-factor (SDF) with sparse characteristic exposure. However, imposing sparsity in the number of return predictors under a LASSO approach may not be a realistic assumption after all due to the diverse characteristic space (Kozak et al., 2020). Nevertheless, sparse models allow for a parsimonious representation of the cross-section of expected stock returns and an easier interpretation and link to economic theories. In our empirical application, we apply LASSO on a set of characteristic PCs instead of raw characteristics. Hence, our approach still encourages a sparse factor structure, while allowing multiple characteristics to have an effect on expected factor portfolio returns through their exposure to the characteristic PCs.

Another strand of the literature applies PCA on a set of stock or portfolio returns to reduce their dimensions. Examples of PCA applications in asset pricing include Connor and Korajczyk (1988), who apply Asymptotic PCA on asset returns to extract latent factors, and Kozak et al. (2018), who form a low dimensional SDF using the first few PCs of anomaly returns. Kozak et al. (2020) also find that a low dimensional specification in terms of PC portfolios is feasible due to the high degree of common variation in factor portfolio returns. In general, the use of PCA in this context is both economically and empirically motivated. Economically, the existence of arbitrageurs in the economy implies that near-arbitrage opportunities, meaning extremely high Sharpe ratios, are implausible to achieve. Hence, high Sharpe ratios associated with low eigenvalue PCs should make no contribution to explaining returns (Kozak et al., 2018).⁶ Empirically, returns possess a spiked covariance structure, meaning the variance-covariance matrix is dominated by a small number of large eigenvalues, separated from the rest. Combining these facts implies that asset returns should be adequately explained by a small number of dominant PCs. We contribute to this literature by constructing PC portfolios of LS portfolios and

⁶Still, this argument does not explicate whether high eigenvalue PCs reflect risk or mispricing.

examining their predictability.

Several recent studies also focus on modifying conventional PCA with the purpose of making it more suitable for asset pricing applications. Kelly et al. (2019) propose a new method of Instrumental Principal Components, allowing latent factor loadings to be time-varying and partially dependent on firm characteristics.⁷ They find that only a small number of characteristic-based factors are important for identifying a successful latent factor model. Lettau and Pelger (2020*a*) augment standard PCA by a cross-sectional pricing error in order to extract factors that can simultaneously explain the time-series variation and the cross-section of asset returns and Lettau and Pelger (2020*b*) demonstrate the superiority of the estimator compared to standard PCA on a set of 37 factor portfolios. Finally, Giglio and Xiu (2021) account for omitted factors in the estimation of risk premia by combining PCA with two-pass cross-sectional regressions. We exploit the recent advancements in the literature by also using the RPPCA of Lettau and Pelger (2020*a*) to extract factors from LS portfolio returns.

Factor portfolio predictability

In a factor timing context, factor momentum has emerged as a mechanism to time factor portfolio returns. Early contributors to this literature include Grundy and Martin (2001), who document a momentum effect in the factor component of stock returns. The momentum effect in factor portfolio returns is strong and has its own distinctive behaviour, different from that of stock momentum. For example, Arnott et al. (2021) and Gupta and Kelly (2019) find that the effect is the strongest at the 1-month horizon, even though stocks exhibit reversals in such short intervals. Nonetheless, factor momentum captures the effect at its purest form as it subsumes stock, industry momentum as well as momentum found in other well diversified portfolios (Arnott et al., 2021). Further-

⁷The method is an extension of the Projected-PCA by Fan et al. (2016) and can be thought of as standard PCA on characteristic sorted portfolios.

more, factor momentum is concentrated in the highest eigenvalue PCs of factor portfolio returns, which implies that momentum is intertwined with the covariance structure of factor portfolios (Ehsani and Linnainmaa, 2022). Whether looking at PC portfolios or individual factors, factor momentum can accommodate factor timing simply by buying (selling) portfolios that have performed well (poorly) in the recent past or relative to their peers. Such strategies deliver strong return performance and are not susceptible to crashes, as stock momentum (Gupta and Kelly, 2019). Nevertheless, using exactly the same investment rule we show that characteristic-based forecasts provide superior information and result in more profitable investment strategies compared to factor momentum.

Outside factor momentum, numerous studies attempt to predict the performance of individual factor portfolios using a collection of potential predictors. Daniel and Moskowitz (2016) forecast stock momentum using market indicators and volatility proxies in an effort to explain momentum crashes. Similarly, Huang (2022) finds that the return spread between winners and losers negatively predicts stock momentum returns. Baba-Yara et al. (2021) analyse the ability of the value spread to forecast the returns of the value-minus-growth portfolio across asset classes. They find that the first principal component of the value spread captures most of the variation in expected value returns. In a similar manner, we also use the first principal component of multiple characteristics to predict PC portfolio returns, even though we examine the possibility that further characteristic components are required. In contrast to previous studies targeting only specific anomalies, we examine factor portfolio predictability across a large set of factor portfolios.

Other studies also examine the predictability of multiple portfolios at once, using either a single or multiple predictors. Asness et al. (2017) use the value spread to construct timing strategies for value, momentum and betting-against-beta portfolios, though they observe little improvement upon a constant multi-style strategy. Greenwood and Hanson (2012) show that corporate share issuance can be used to forecast the performance of

factor portfolios related to size and value. Stambaugh et al. (2012) find that LS strategies appear to be stronger following periods of high investor sentiment. They find the sentiment effect to be concentrated on the short leg of anomalies, which they base on the short sale impediments that results in relatively higher overpricing compared to underpricing. On a much larger scale, Jacobs (2015) confirms the findings of Stambaugh et al. (2012) by examining the role of sentiment in a large set of 100 anomalies. Kelly and Pruitt (2013) forecast four sets of characteristic-sorted portfolios using the cross-section of book-to-market ratios and observe higher predictability at lower frequencies. Dichtl et al. (2019) attempt to predict 20 equity factors using fundamental and technical indicators. They distinguish between cross-sectional and time-series predictability which results in factor-tilting and factor timing portfolio allocations, respectively. Haddad et al. (2020) construct PC portfolios by running PCA on the time-series of 50 anomalies and find that the largest eigenvalue PCs are the most predictable by their own book-to-market ratio.

We extend Haddad et al.'s (2020) framework by incorporating information across a large set of observable characteristics to predict a large set of factor portfolio returns. Furthermore, we allow the effect of characteristics to be independently identified for every PC portfolio, examining the possibility that different characteristics affect different sources of variation in factor portfolio returns.

3 Methodology

This section begins by setting out the general framework, followed by our forecasting procedure and the benchmark models employed. Section IA.2 of the Internet Appendix introduces the statistical methods used in this study and provides a comprehensive overview of their functional form and statistical properties.

3.1 General framework

The main objective is to predict a large set of factor portfolio returns using a large set of portfolio characteristics. Let R be a $(T \times N)$ matrix of N factor portfolio returns for T periods. Equivalently, let $R_{t,.} = (R_{t,1}, \dots, R_{t,N})$ be a $(1 \times N)$ vector of portfolio returns at time t and C_t , a $(N \times M)$ matrix of M characteristics for N factor portfolios at time t . The base case arises from a conditional version of Cochrane's (2011) framework for modeling returns as a function of characteristics:

$$R_{t+1,n} = a_{t,n} + \sum_{m=1}^M b_{t,n}^m C_{t,n}^m + \varepsilon_{t+1,n}, \quad (1)$$

where $a_{t,n}$ and $b_{t,n}^m$ denote the conditional alpha and beta at time t , and $\varepsilon_{t+1,n}$ is the pricing error at time $t + 1$. Entertaining time variation in $b_{t,n}^m$ and $a_{t,n}$ due to changes in portfolio attributes is the essence of factor timing.⁸ By combining different dimension reduction techniques, we essentially investigate the possibility that the conditional alphas and betas are a function of the covariance of returns, the covariance of the characteristics, or even the covariance of returns with the characteristics. The covariance of returns comes into play by focusing on the dominant components of factor portfolio returns instead of predicting each factor portfolio separately. More concretely, assuming a linear latent factor specification, excess asset returns can be expressed as:

$$R_{t+1,.} = Z_{t+1,.} W_t' + \Theta_{t+1}, \quad (2)$$

where $Z_{t+1,.} = (z_{t+1,1}, z_{t+1,2}, \dots, z_{t+1,K})$ is a $(1 \times K)$ vector of factor returns with $K \ll N$, $W_t = (w_{t,1}, w_{t,2}, \dots, w_{t,K})$ is a $(N \times K)$ matrix of factor loadings and Θ_{t+1} is a $(1 \times N)$ vector of idiosyncratic errors. The time dimension in this context arises by the recursive estimation of eigenvectors and principal components. The first term of the right-hand-

⁸Cochrane (2011) uses the formulation in Equation (1) to model the returns of an individual stock in excess of the risk-free rate. In our setting, we model factor returns, i.e., the returns of a long portfolio in excess of the returns of a short portfolio. Getting from individual stocks to factor portfolios is straightforward and hence we focus directly on the latter to simplify the exposition of our framework.

side reflects compensation for the exposure on systematic risk factors while the second term reflects asset specific risk. Under the assumption that the factors and the errors are uncorrelated, the variance-covariance matrix of asset returns can be decomposed into a systematic and idiosyncratic part. A common practice is to estimate $Z_{t+1,\cdot}$ and W_t directly, by applying PCA on the variance-covariance matrix of R and retaining the dominant components (e.g., Connor and Korajczyk (1986) and Kozak et al. (2018)). Provided that time variation in asset risk premia is driven by exposure to time-varying aggregate risk, being able to accurately predict the dominant components $Z_{t+1,\cdot}$, allows to form forecasts for individual anomalies through W_t . By only focusing on $Z_{t+1,\cdot}$, we isolate common sources of predictability across factor portfolios and ignore spurious predictability associated with smaller PCs.

In order to forecast $Z_{t+1,\cdot}$, we model PC portfolio returns as a function of observable characteristics. Specifically, lagged characteristics are used to predict next-period PC portfolio returns. The characteristics of the PC portfolios are computed by combining factor portfolio characteristics according to their weights given by the i^{th} eigenvector $w_{t,i}$. The cross-section of characteristics for the i^{th} $i = (1, \dots, K)$ PC portfolio is calculated as $H_{t,i} = w'_{t,i} C_t$. Repeating the process for every t and every i results in a $(T \times M)$ matrix H_i of characteristics for each PC portfolio.

However, using raw characteristics as inputs in standard predictive regressions would be suboptimal due to high correlations and lack of predictive information for some of them. Therefore, we transform the characteristics of PC portfolios into scores by using PCA and PLS. This is achieved by multiplying the matrix of characteristics H_i with a matrix of eigenvectors, such as:

$$X_i = H_i Q_{t,i}, \quad (3)$$

where X_i is a $(T \times M)$ matrix of component scores, or linear combinations of the un-

derlying characteristics, of the i^{th} PC portfolio. Similarly, $Q_{t,i} = (q_{t,i}^1, q_{t,i}^2, \dots, q_{t,i}^M)$ is an $(M \times M)$ matrix of eigenvectors estimated at time t and sorted by their corresponding eigenvalues. For PCA, $Q_{t,i}$ is estimated based on the eigenvalue decomposition of $Var(H_i)$, while for PLS it is based on the eigendecomposition of $cov(Z_{.,i}, H_i)$; more information on how to obtain $Q_{t,i}$ and X_i under the different methods is provided in the Internet Appendix. Dominant PCA components capture most of the variation within the characteristics, while dominant PLS components capture most of the covariation between lagged characteristics and next-period returns. Making a connection to the conditional betas, the use of PCA for the characteristics introduces their covariance into the conditional beta function, while the use of PLS introduces the covariance of PC portfolio returns with their characteristics.

Next, we model PC portfolio returns using the characteristic components:

$$z_{t+1,i} = \beta_{t,i}^0 + \sum_{m=1}^M \beta_{t,i}^m X_{t,i}^m + \epsilon_{t+1,i}, \quad (4)$$

where $X_{t,i}^m$ is the m^{th} characteristic component of the i^{th} PC portfolio at time t , and $z_{t+1,i}$ is the one-month ahead return of the same portfolio. Equations (1) to (4), lead to:

$$R_{t+1,.} = \sum_{i=1}^K w'_{t,i} \beta_{t,i}^0 + \sum_{i=1}^K \sum_{m=1}^M w'_{t,i} (\beta_{t,i}^m w'_{t,i} C_t q_{t,i}^m) + \eta_{t+1,.}, \quad (5)$$

where $\eta_{t+1,.}$ is a $(1 \times N)$ vector of composite errors capturing both the unexplained return variation from the characteristics, as well as the variation from potentially omitting higher-order PC components. Equation (5) shows that $a_{t,n}$ and $b_{t,n}^m$ from Equation (1) end up being functions of the eigenvectors of the covariance of factor portfolio returns, $w_{t,i}$, the eigenvectors of the covariance of characteristics, $q_{t,i}^m$, and the betas, $\beta_{t,i}^m$, from regressing PC portfolio returns on their characteristic components. Specifically, multiplying the $(1 \times N)$ vector $w'_{t,i}$ with the $(N \times M)$ matrix C_t gives the cross-section of PC portfolio

characteristics while multiplying that product with the $(M \times 1)$ vector $q_{t,i}^m$ gives the m^{th} characteristic component of the i^{th} PC portfolio. Further multiplying that new product with $\beta_{t,i}^m$ and summing over M gives the value for the i^{th} PC. Finally, multiplication with $w'_{t,i}$ and summation over K generates a return vector for the whole cross-section of factor portfolio returns. Note that all the objects have time t subscripts since they can be estimated recursively.

3.2 Forecasting procedure

We use at least 20 years (240 months) of information to estimate the PC portfolios and their characteristics to then make return predictions at $t + 1$. Our forecasts employ an expanding estimation window, with the estimation sample always starting at the beginning of the sample period and incorporating additional observations as they become available. PC portfolios are recursively re-estimated at each point in time, using an updated $w_{t,i}$ with $i = 1, \dots, K$ based on the in-sample variance-covariance matrix of factor portfolio returns.⁹ Notice that PC portfolio characteristics H_i do not only change because of the change in the underlying factor portfolio characteristics C_t , but because of the change in the weighting vectors $w_{t,i}$ as well. Overall, our approach is flexible enough to account for a potentially unstable correlation structure in the factor portfolio returns.

In a similar fashion, the matrix of characteristic components is obtained as follows; for PCA, which only utilizes information contained in the characteristics to extract the latent factors, characteristics up to t are used to estimate X_i . For PLS, which uses information in both characteristics and returns, characteristics up to $t - 1$ and PC portfolio returns up to t are used to estimate X_i . The β s in Equation (4) are always estimated using returns up to t and values in X_i up to $t - 1$. Values of X_i at t are then plugged into Equation (4) to obtain forecasts for each PC portfolio return at $t + 1$. Hence, our forecasts are

⁹When RPPCA is used for the left-hand side, we use a constant value of $\gamma = 10$ for the weight on the cross-sectional pricing error.

completely out of sample and do not suffer from any look-ahead bias.

Another subtle but important detail is the cross-sectional standardization of C_t to account for the difference in the scale of the characteristics. Running raw PCA or PLS on C_t would tilt the PCs towards the larger characteristics, as those will have significantly higher variance. For this reason, we standardize the matrix of factor portfolio characteristics C_t cross-sectionally before calculating H_i and ultimately X_i by subtracting the cross-sectional characteristic mean and dividing by the cross-sectional characteristic standard deviation at each time t . Apart from ensuring a reasonable covariance matrix for the characteristics, such an approach allows us to focus on the cross-sectional differences in the data. As long as factor portfolio characteristics coincide with factor portfolio returns in cross-sectional terms, PC portfolio characteristics should coincide with returns across time, as they are both linear combinations of the cross-section and thus making a predictive regression approach sensible.¹⁰

The first decision being made is on the optimal number of factors in Equation (2). Specifying the optimal number of PCs is ultimately an empirical question as it depends on the underlying factor structure. Bai and Ng (2002), Onatski (2010) and Haddad et al. (2020), all develop critical value thresholds for determining the number of factors. We follow a simple approach and focus on the first five PCs as they capture about 67% of the variation in factor portfolio returns. Selecting the first five PC portfolios is also consistent with similar studies performing PCA on a set of factor portfolios, e.g., Haddad et al. (2020) and Lettau and Pelger (2020b). Hence, let $Z_{t,5} = (z_{t,1}, z_{t,2}, \dots, z_{t,5})$ and $W_{t,5} = (w_{t,1}, w_{t,2}, \dots, w_{t,5})$ be the set of the 5 largest PC portfolios and eigenvectors.

The second decision to be made is on how to estimate $\beta_{t,i}^m$ in Equation (4). Here, we examine two different cases, one that imposes sparsity and one that is data-driven. In

¹⁰In Section IA.3 of the Internet Appendix, we provide a detailed description of the standardization approach and explain the drivers of variation in the PC characteristics across time.

the first case, we only use the first characteristic component of each PC portfolio (i.e., the first column of X_i) in standard bivariate predictive regressions. Although this is the sparsest specification possible, multiple characteristics can have an effect on PC portfolio returns through their weights on the first characteristic PC. As an alternative, we apply LASSO on the whole set of characteristic components for each PC portfolio (i.e., the whole matrix X_i) to identify a subset that is useful for our forecasting objective. Hence, β s in Equation (4) for the first case are obtained through OLS for a single predictive factor ($m = 1$) and in the second case the β s are obtained through LASSO for $m = 1, \dots, M$.

When performing LASSO the optimal amount of coefficient shrinkage is selected by conducting cross-validation on a rolling basis. In particular, before every forecasting step we separate the in-sample period into a training and a validation sample. The training sample is used to estimate the PC portfolios and characteristic PCs and the validation sample is used to identify the degree of model complexity that delivers reliable out-of-sample performance.¹¹ At the start, the training sample is used to forecast the first period in the validation sample subject to a geometric sequence of shrinkage values.¹² The actual value of the forecasted data point is then used as part of the next training set to forecast the subsequent point in the validation sample. After repeating this procedure for every period in the validation sample, we pick the level of shrinkage that minimizes the mean-squared error. We then re-estimate the PC portfolios and characteristic PCs using the whole in-sample period (training and validation) and apply LASSO using the fixed value for the shrinkage parameter to estimate $\beta_{t,i}^m$ and predict PC portfolios at $t + 1$. Depending on the magnitude of the shrinkage, our approach examines the possibility that none of the characteristic components are relevant in predicting PC portfolio returns, in which case returns forecasts shrink down to a constant term.

¹¹The validation sample covers the last five years (60 months) of the in-sample period while the training sample increases by one at each forecasting step.

¹²The sequence of shrinkage values is strictly positive and terminates at a value for which all coefficients are equal to zero.

As already discussed, LASSO is applied separately on each PC portfolio, meaning that the number of features can be different across PC portfolios. Essentially, our method allows for different sources of variation in factor portfolio returns to be approximated by models of different complexity, examining the possibility that characteristic importance varies across the main sources of return variation. Furthermore, since LASSO is applied iteratively, the number of features can also vary across time for each PC portfolio depending on how strong the characteristic signal has been in the recent past. Lastly, it is important to highlight that LASSO can select low eigenvalue characteristic PCs, as long as they contribute to minimizing the forecasting error in the validation period.

To summarize, we attempt to regularize both the left (LHS) and the right-hand side (RHS) of the predictability problem by combining different dimension reduction techniques. Regularization in the number of forecasting targets is achieved with the use of PCA or RPPCA and in the number of predictors with the use of PCA or PLS, resulting in four base models that we define as PCA, RPPCA, PCA-PLS, RPPCA-PLS.¹³ Figure 1 provides a visual depiction of our procedure that can be summarized in the following steps:

1. Reduce a set of factor portfolios to their first five components using PCA or RPPCA.
2. Estimate the characteristics of the PC portfolios using their loadings from the first step.
3. Rotate PC portfolio characteristics using either PCA or PLS.
4. Either select the first characteristic PC or apply LASSO on the whole set of characteristic PCs of each PC portfolio.
5. Produce separate forecasts for each PC portfolio using the selected number of features.

¹³All models are estimated using either single or multiple predictors (via LASSO), resulting in a total of eight forecasting models. Panel A of Table IA.2 includes the listing of the four main models which are estimated using either a single predictor or multiple predictors in combination with LASSO.

6. Expand these forecasts to individual factor portfolios using their loadings on each PC portfolio.

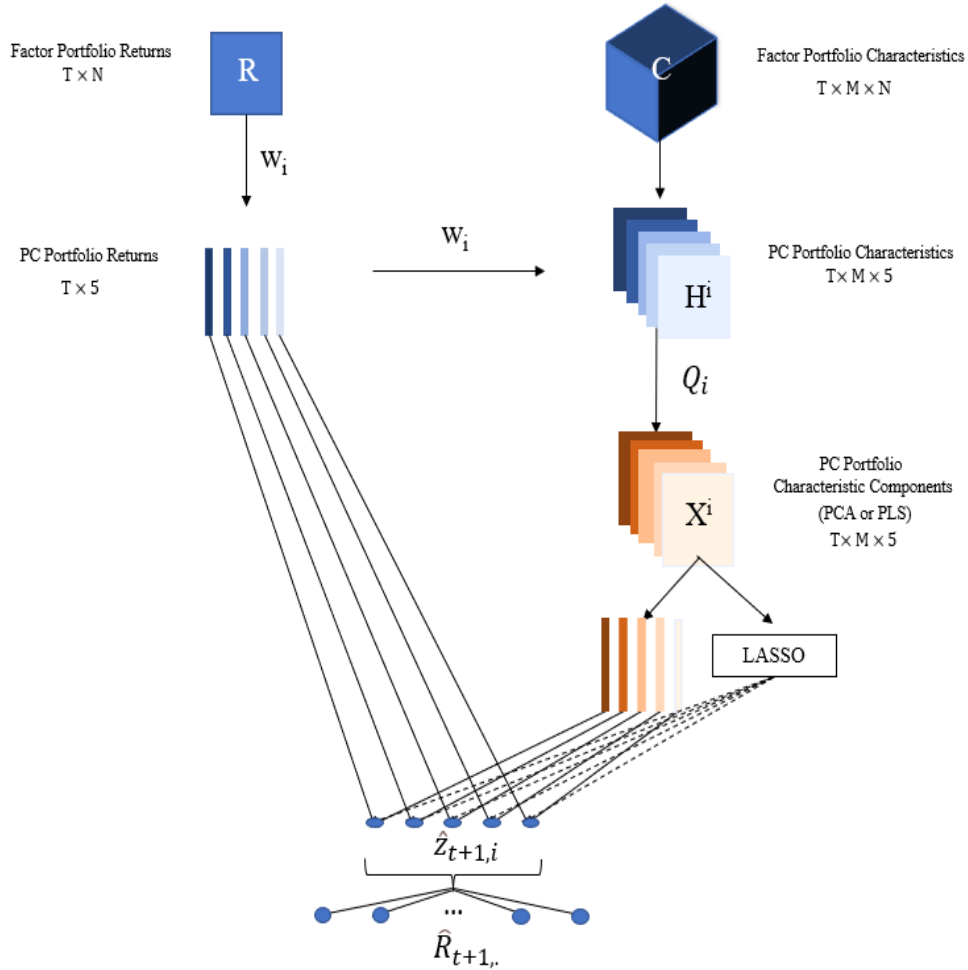


Figure 1: Visual depiction of our modeling procedure. The figure presents the process of forecasting factor portfolio returns using their portfolio characteristics. PC portfolios are calculated as linear combinations of factor portfolios. The same weighting vectors are used to decompose the three-dimensional set of characteristics into 5 independent matrices of characteristics (one for each PC portfolio). The matrices of predictors are transformed to components and either the first component is retained or LASSO is applied on the whole set of components to pick those that are the most informative. Individual forecasts for each PC portfolio are produced and those forecasts are aggregated into factor portfolio return forecasts using the weighting vectors that were used to aggregate factor portfolios into PC portfolios.

3.3 Benchmark Models

To examine whether characteristic-based models provide superior information compared to different approaches, we employ alternative information sets to predict factor portfolio returns. Panel B of Table IA.2 includes a listing of all the benchmark models used. In

a general setting, we form the baseline benchmark models following the methodological framework proposed by the original authors. Section 5 modifies the original models in various ways in order to examine the robustness of our results.

Factor Momentum

The first benchmark, is the 1-month momentum strategy (1mMOM), which forms the momentum signal based on a look-back-window of 1 month. Essentially, the return at time t becomes the prediction for the return at time $t + 1$. The second benchmark is the 12-month momentum strategy (12mMOM), which forms the momentum signal based on a look-back-window of twelve months. In this case, the prediction for the return at time $t + 1$ is the average monthly return of the previous twelve months. In order to improve consistency across characteristic and momentum models, in Section 5 we also apply both momentum strategies to the PC portfolios and then extend the forecasts to individual anomalies as in Equation (5). Hence, we also examine the possibility of a stronger momentum effect on the main sources of variation of factor portfolio returns.¹⁴

Valuation Ratios

As a third benchmark, we use only the book-to-market ratio of factor portfolios as a return predictor. Specifically, we follow Haddad et al. (2020) in predicting the first five PCs by their own book-to-market ratio and then extending the forecasts to individual anomalies. In order to keep things consistent with our framework, we estimate the PC portfolios recursively rather than using the first half of the sample. In Section 5, we simultaneously use the book-to-market ratio of all dominant PC portfolios in combination with LASSO as an alternative to the baseline model.

¹⁴For instance, Ehsani and Linnainmaa (2022) observe that momentum is highly concentrated among the first five PC portfolios.

Issuer-Repurchaser Spread

Following Greenwood and Hanson (2012), we estimate the issuer-repurchaser spread of each portfolio and use it to predict next period factor portfolio returns. The issuer-repurchaser spread is defined as the average characteristic decile difference between issuers and repurchasers. Repurchasers are defined as firms that have reduced their shares outstanding by more than 0.5% during the fiscal year and issuers are firms that have increased their shares outstanding by more than 10% during the fiscal year. The metric can take values from -9 to 9 , with low values implying that issuers are located in the low leg and repurchasers in the high leg of each factor portfolio (and vice versa). In Section 5, we generalize this approach by considering the issuer-repurchaser spreads of the PC portfolios.

Investor Sentiment

We explore the role of investor sentiment in predicting factor portfolio returns. Stambaugh et al. (2012) and Jacobs (2015) find that anomaly performance is stronger following periods of high sentiment. To examine the effect of sentiment, we use the investor sentiment index of Baker and Wurgler (2006), which captures the common component in five sentiment proxies, with each proxy being orthogonalized with respect to six macroeconomic indicators. Specifically, next period factor portfolio returns are regressed on the lagged values of the index and forecasts for individual anomalies are formed based on a standard regression setting. In Section 5, we also form forecasts for individual PC portfolios and employ LASSO to examine potential time variability in the sentiment signal.

Historical Sample Mean

Finally, we use the in-sample average of factor portfolio returns as a forecast for the next period, as in Campbell and Thompson (2008). Such a simple non-parametric technique utilizes information in the returns only, allowing us to examine the incremental effect of sophisticated statistical techniques and different information sets.

4 Empirical results

4.1 Data

We replicate a large set of 72 characteristics, also considered by Green et al. (2017). The characteristics are calculated using data from the Center of Research on Securities (CRSP) and Compustat. Our dataset covers the 50-year period from January 1970 to December 2019. The stock universe includes common stocks listed on NYSE, AMEX, and NASDAQ that have a record of month-end market capitalization on CRSP and a non-missing and non-negative common value of equity on Compustat. Additional information about the characteristics, including origination and characteristic description, can be found in Table IA.3 of the Internet Appendix.

For every month in our sample, stock returns at month t are matched against their respective characteristics at month $t - 1$. For accounting data, we allow at least six months to pass from the firms' fiscal year end before they become available and at least four months to pass for quarterly data. We also winsorize characteristics cross-sectionally at a 99% confidence level to account for extreme outliers. Finally, to isolate the effect of microcaps, we remove stocks with price below \$5 at the portfolio formation period and use NYSE-breakpoints to split stocks into deciles, following Fama and French (2008). These adjustments help us robustify our inferences, since many anomalies have been found to work better on small stocks (Fama and French, 2008).

We then move to the construction of the factor portfolios. For each anomaly, we first group stocks into value-weighted deciles based on their characteristic exposure in the previous month and then go long decile ten and short decile one,¹⁵ even if the characteristic

¹⁵In the early years of the sample period, there are few characteristics, such as characteristics based on research and development expenses, which do not have enough variation in order to form ten separate portfolios. To account for this, we allow the number of quantiles to be less than ten for months in which the required number of cut-off points is not reached. In other words, LS portfolio returns are calculated as long as there are at least two different values for the same characteristic in a particular month.

is negatively related to future returns. Such an approach requires no ex-ante information about the relationship between characteristics and returns and results in the highest dispersion in factor portfolio returns. Furthermore, given that factor timing strategies can take long and short positions on factors, the sign of factor portfolio returns is irrelevant.¹⁶ Similarly to computing factor portfolio returns, the characteristics of factor portfolios can be computed by value-weighting characteristics of stocks within each decile portfolio and then subtracting the value of the bottom from the top decile. Notice that the portfolio constructed based on a particular characteristic sort will also have the highest characteristic score by construction.¹⁷

Figure 2 displays the average monthly returns of the factor portfolios together with the 95% confidence intervals. Out of all the factor portfolios, 12-month momentum (mom12m) has the highest average return, followed by 6-month momentum (mom6m). Out of the 72 portfolios, only 22 have significant average returns, confirming a high degree of redundancy among the documented factors (Hou et al., 2020). When we focus on the out-of-sample period only, this number goes down to 10, reflecting the decay in the performance of the anomalies over time (McLean and Pontiff, 2016). Further descriptive statistics for the factor portfolios can be found in Table IA.4 in the Internet Appendix.

¹⁶Hence, strategies with a negative risk premium, such as asset growth, should on average be allocated in the short side of our factor timing portfolio.

¹⁷For example, the momentum portfolio will always have the highest momentum score compared to all the other factor portfolios.

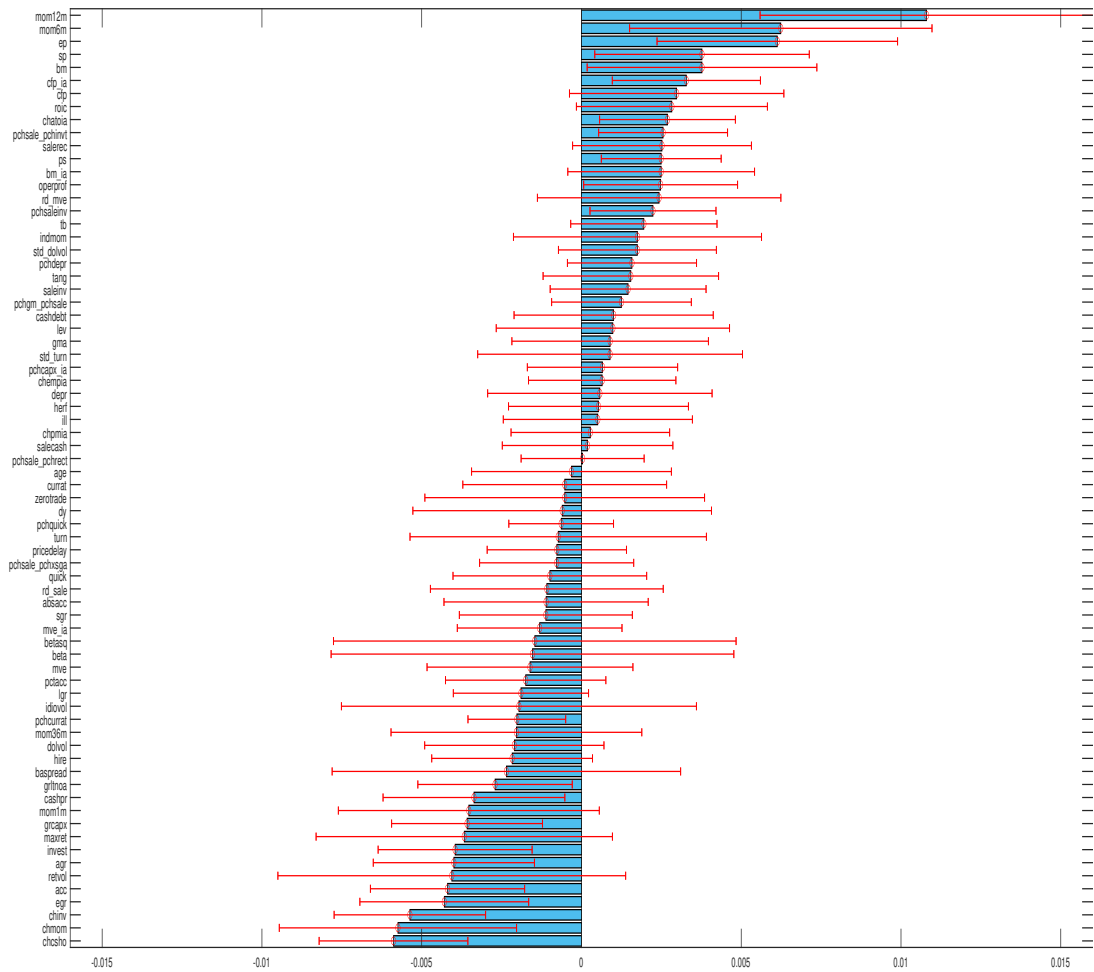


Figure 2: Average monthly returns of factor portfolios with 95% confidence intervals for the period 01/1970-12/2019.

As already discussed, we proceed by constructing recursively five PC portfolios, i.e., linear combinations of the 72 factor portfolios using either PCA or RPPCA. These PC portfolios are by construction affected by all factor portfolios in a time-varying fashion; as a result, at a first glance they might look as if they do not have any economic interpretation. In order to tackle this, we regress recursively each PC portfolio return on each of the 72 anomalies and estimate the monthly time-series of R^2 values for each anomaly. The analysis, which is detailed in Section IA.6 of the Internet Appendix, shows that the constructed PC portfolios have in fact a quite clear economic interpretation. For example, the first PC portfolio (based on PCA) loads heavily on volatility characteristics, the second one loads more on value characteristics, while the third one is mostly driven by momentum characteristics. Moreover, despite the recursive construction procedure of the

PC portfolios, these economic relations are very stable over time.

4.2 Predictive performance

We examine the out-of-sample performance of our predictive models using standard forecast evaluation measures and a monthly holding period as in Campbell and Thompson (2008). We use an in-sample window of at least 240 months, with the initial in-sample period covering the period 01/1970-12/1989 and forecasts being obtained out-of-sample for the period 01/1990-12/2019. As a first indication of the out-of-sample fit of our models, we estimate the out-of-sample R^2 for each individual PC portfolio as:

$$\text{OOS } R^2 = 1 - \frac{\sum_{t=240}^{T-1} (z_{i,t+1} - \widehat{z}_{i,t+1})^2}{\sum_{t=240}^{T-1} (z_{i,t+1} - \bar{z}_{i,t+1})^2}, \quad (6)$$

where $\widehat{z}_{i,t+1}$ is the PC portfolio return forecast at time $t + 1$ and $\bar{z}_{i,t+1}$ is the average PC portfolio return using information up to period t . We also estimate a Total OOS R^2 , which pools squared errors across factor portfolios and across time:

$$\text{Total OOS } R^2 = 1 - \frac{\sum_{i=1}^N \sum_{t=240}^{T-1} (R_{i,t+1} - \widehat{R}_{i,t+1})^2}{\sum_{i=1}^N \sum_{t=240}^{T-1} (R_{i,t+1} - \bar{R}_{i,t+1})^2}. \quad (7)$$

Total OOS R^2 assesses the predictive ability of each model under a grand panel framework and therefore is a bulk measure of the accuracy of the model-based predictions of future factor portfolio returns. Table 1 presents the OOS R^2 results for individual PC portfolios, as well as the Total OOS R^2 under the various models. Apropos Panel A, characteristic-based models with one predictive factor deliver negative OOS R^2 , with only the second PC portfolio being predictable. With regards to the different dimension reduction techniques used, models that use PCA for the RHS outperform their PLS counterparts in terms of Total OOS R^2 , although someone would expect the opposite given that PCA factors capture variation among returns-related and unrelated variables.

	PC1	PC2	PC3	PC4	PC5	Total
Panel A: Single factor						
PCA	-1.00	0.24	-1.33	0.36	0.28	-0.55
PCA-PLS	-3.29	0.76	-2.89	0.60	-0.95	-1.55
RPPCA	-0.75	0.90	-0.13	-1.15	-0.74	-0.38
RPPCA-PLS	-3.00	1.80	-3.28	-3.94	-1.60	-1.54
Panel B: Time-varying number of factors using LASSO						
PCA	2.71	2.85	1.23	3.24	-0.78	1.46
PCA-PLS	1.70	5.34	1.78	3.87	3.75	1.53
RPPCA	0.80	0.60	1.71	0.41	2.64	0.52
RPPCA-PLS	1.17	6.97	0.92	0.20	2.30	1.17
Panel C: Benchmark models						
1mMOM						-88.98
12mMOM						-6.61
PCA-BM	0.40	3.47	-0.20	0.89	-0.62	0.43
IR Spread						0.13
Sentiment						0.42

Table 1: OOS R^2 for PC portfolios and Total OOS R^2 across all anomalies for the period 01/1990-12/2019 in percentage terms. Panel A displays results using a single latent factor to predict PC portfolio returns. Panel B shows the results where the optimal number of factors is selected by applying LASSO on the set of latent factors. Panel C displays results for the benchmark models.

Moving to Panel B, the combination of dimension reduction techniques with LASSO significantly improves results for all models delivering positive Total OOS R^2 s. The predictive performance improves almost uniformly across all PCs, highlighting the importance of accounting for further characteristic components and the benefits of regularization on out-of-sample performance. The use of LASSO in particular allows the models to underweight (overweight) information in the characteristics in periods where the characteristic signal diminishes (becomes stronger).¹⁸ Overall, results in Panel B confirm that imposing a sparse or constant factor structure may not be a realistic assumption in the context of asset return prediction.

Finally, Panel C displays the Total OOS R^2 for the benchmark models.¹⁹ With regards to factor momentum, previous month returns provide unreliable forecasts in exact terms, as

¹⁸It is also important to highlight that LASSO may select characteristic components other than the first, potentially resulting in considerably different forecasts compared to the single factor case.

¹⁹The historical sample mean is not included as it has a zero Total OOS R^2 by construction.

implied by the highly negative Total OOS R^2 . When returns are averaged over the past twelve months, results improve significantly, although the Total OOS R^2 remains on the negative side. Conversely, models based on the book-to-market ratio, issuer-repurchaser spread and investor sentiment deliver positive Total OOS R^2 , though they still fall behind the characteristic-based models that employ LASSO.

Ultimately, we are interested in the predictability of individual factor portfolios based on PC portfolio forecasts. As a measure of individual factor portfolio predictability, we estimate the individual OOS R^2 for all anomalies under the different models. Apropos Figure 3, expanding PC portfolio return forecasts to individual anomalies reveals predictive patterns in a robust and systematic way. In line with Haddad et al. (2020), we observe substantial anomaly predictability and find many predominant anomalies, such as value (bm) and sales-to-price ratio (sp) to be highly predictable by observed characteristics. However, almost all characteristic-based models fail to predict anomalies that are based on a % change in accounting variables, such as % change in sales minus % change receivables (pchsale_pchrect) and % change in the current ratio (pchcurrat) among others, located in the lower half of the heat-map. These portfolios have returns statistically indistinguishable from zero and low covariance with the rest of the anomaly universe. As a result, they do not load heavily on the first five components and their performance is not adequately captured by PC portfolio forecasts. With regards to the benchmark models, only factor momentum results in high forecasting errors and therefore negative OOS R^2 for almost all anomalies. The remaining benchmark models perform sufficiently well, delivering positive OOS R^2 for the majority of the anomaly universe.

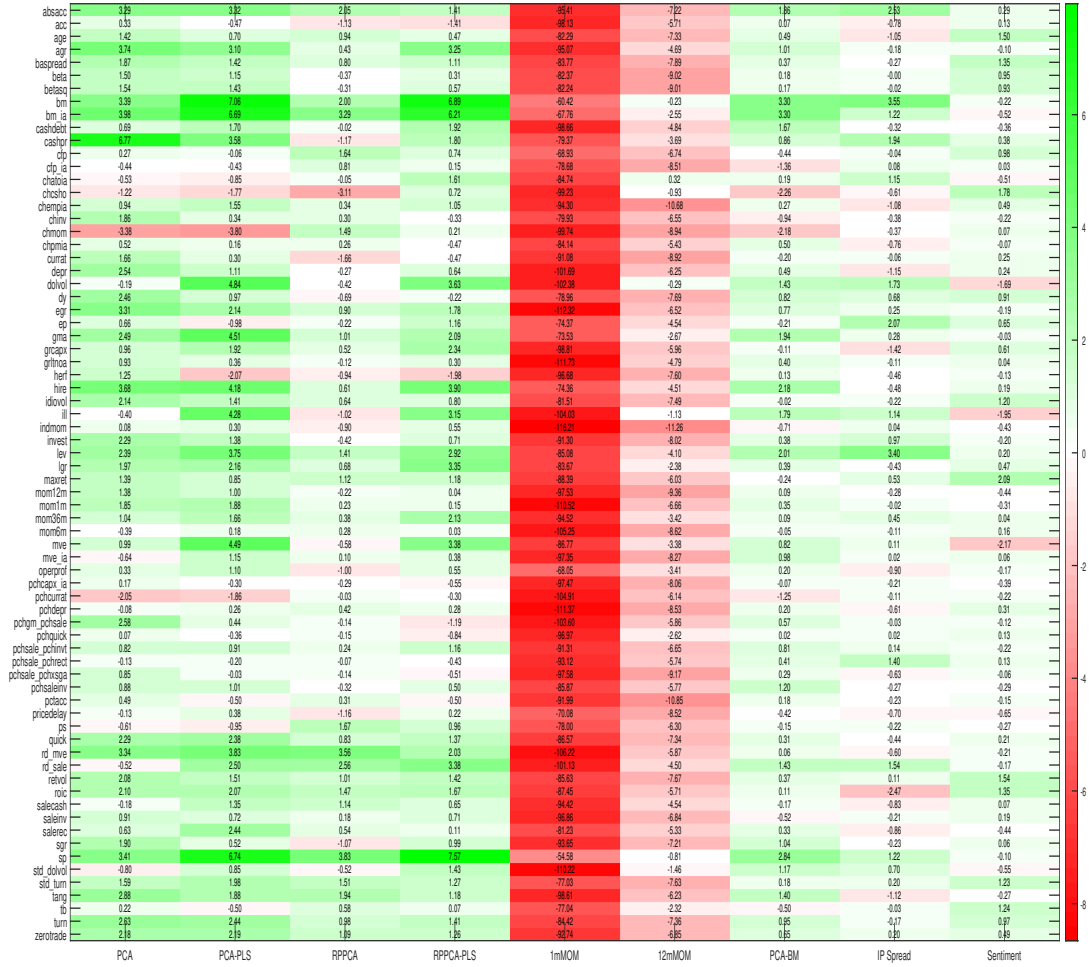


Figure 3: OOS R^2 for individual anomalies under the characteristic-based models that employ LASSO and benchmark models (historical sample mean is inferred by the R^2 metric). Negative values (in red) show lack of predictive ability while positive values (in green) show predictive ability of the underlying model for a given factor portfolio.

Whereas OOS R^2 accommodates a general quantitative comparison of the predictive performance of the various models, it is also important to assess the statistical significance of the differences among model forecasts. To make pairwise comparisons of the out-of-sample predictive accuracy we use the modified Diebold and Mariano (DM) test by Gu et al. (2020), which compares the cross-sectional average error differential between two models. The DM test statistic between two models (1) and (2) is defined as $DM_{1,2} = \bar{d}_{1,2} / \hat{\sigma}_{\bar{d}_{1,2}}$, where $\bar{d}_{1,2}$ and $\hat{\sigma}_{\bar{d}_{1,2}}$ are the mean and standard deviation of the error differential, defined as:

$$d_{1,2;t+1} = \frac{1}{N} \sum_{n=1}^N \left(\left(\hat{e}_{n,t+1}^{(1)} \right)^2 - \left(\hat{e}_{n,t+1}^{(2)} \right)^2 \right), \quad (8)$$

where $\left(\hat{e}_{n,t+1}^{(1)} \right)^2$ and $\left(\hat{e}_{n,t+1}^{(2)} \right)^2$ denote the prediction error of factor portfolio return n at time $t + 1$ under models (1) and (2), respectively.

	PCA	PCA-PLS	RPPCA	RPPCA-PLS	1mMOM	12mMOM	PCA-BM	IR Spread	Sentiment
PCA-PLS	-0.10								
RPPCA	1.35*	1.26							
RPPCA-PLS	0.34	0.61	-0.93						
1mMOM	7.05***	7.19***	6.94***	7.17***					
12mMOM	2.73***	2.99***	2.40***	2.93***	-6.73				
PCA-BM	1.17	1.06	0.13	0.78	-6.88	-2.21			
IR Spread	1.40*	1.32*	0.56	1.25	-6.85	-2.32	0.39		
Sentiment	1.16	1.08	0.19	0.94	-6.72	-2.30	0.02	-0.74	
Historical Sample Mean	1.75**	1.56*	0.89	1.35*	-6.83	-2.08	1.65*	0.31	0.94

Table 2: Modified Diebold-Mariano test for models that employ LASSO and benchmark models. The table displays the modified DM-statistic that compares the predictive performance of the column model with the row model. A positive value indicates that the column model outperforms the row model. The asterisks, indicate statistical significance at a 10% (single), 5% (double) and 1% (triple) level.

Table 2 reports the results from the DM-test for pairwise comparisons between the different models. For conciseness, we only consider the characteristic-based models that employ LASSO as they outperform the single-factor models in terms of Total OOS R^2 . A positive value for the DM test statistic indicates that the column model outperforms the row model and the asterisks indicate statistical significance at a 10% (single), 5% (double) and 1% (triple) level, respectively. We observe that the characteristic-based models provide significantly higher predictive accuracy than the factor momentum models and the historical sample mean model, even though the results are less strong in the latter case. In contrast, the higher predictive accuracy compared to the other benchmark model is not translated into statistical significance.

Nevertheless, predicting anomaly returns is of interest as long as it accommodates the construction of a profitable investment strategy. Specifically, in asset pricing the focus of interest is not so much on obtaining accurate predictions for individual returns, but rather on constructing portfolios with good risk-return properties (Nagel, 2021). Put dif-

ferently, we are more interested in predicting cross-sectional differences in returns rather than predicting individual returns in exact terms. In that sense, Total OOS R^2 is just a distance measure that does not reflect whether models can distinguish strong from weak performers. Consider, for example, a stylized hypothetical scenario with three factor portfolios and a forecasting period of only one month. If the realized returns of the portfolios are 3%, 2% and 1%, the estimated historical samples means are 0%, 1% and 2%, and the model-implied predictions are 6%, 5% and 4%, respectively, then the predictive model will end up having a very negative OOS R^2 (-145.45%) even though it will be able to rank the portfolios perfectly. Consequently, models that yield higher Total OOS R^2 do not necessarily yield better portfolios in terms of average returns or Sharpe ratios. This argument explains, for example, why the 1-month factor momentum has been found empirically to be particularly profitable even though our results show that it has a very negative Total OOS R^2 . The disconnect between OOS R^2 and investment performance is discussed in detail both theoretically and empirically in Kelly et al. (2022).

Given that predictive accuracy in relative terms might be more important than predictive accuracy in exact terms, we proceed by exploring two alternative measures, namely the percentage of times that the sign of future factor portfolio returns is identified correctly and the average cross-sectional correlation between forecasted and realized returns. The former measure examines the ability of the models to predict the direction of individual factor portfolio returns and the latter measure examines whether model-based forecasts capture the cross-sectional dispersion in factor portfolio returns. Table 3 presents the results.

	Proportion of correct sign	Average cross-sectional correlation
Panel A: Single factor		
PCA	49.90	0.70
PCA-PLS	52.74	9.12
RPPCA	50.40	2.31
RPPCA-PLS	52.75	9.17
Panel B: Time-varying number of factors using LASSO		
PCA	52.19	7.83
PCA-PLS	51.95	9.04
RPPCA	52.23	7.74
RPPCA-PLS	52.24	8.51
Panel C: Benchmark models		
1mMOM	51.88	5.98
12mMOM	52.40	5.43
PCA-BM	51.45	5.81
IR Spread	51.22	4.26
Sentiment	51.27	4.71
Historical Sample Mean	50.50	3.24

Table 3: Percentage of correct sign identifications and average cross-sectional correlation. Panel A displays results using a single latent factor to predict PC portfolio returns. Panel B shows the results where the optimal number of factors is selected by applying LASSO on the set of latent factors. Panel C displays results for the different benchmark models.

When considering the single factor predictive models in Panel A, we observe that the models that use PLS for the RHS are far superior to the models that use PCA for the RHS, even though Table 1 shows that they exhibit worse OOS R^2 . In order to understand this discrepancy, we can take the example of the PCA-PLS model. This model has a forecasting error that is lower than that of the historical sample mean model in 54% of the times. In those cases, it exhibits an OOS R^2 of 12.24% and an average cross-sectional correlation of 41%. In the remaining 46% of the cases, it exhibits an OOS R^2 of -17.17% and an average cross-sectional correlation of -28% . This means that, while the model's low overall OOS R^2 is driven by some large forecasting errors, its high overall cross-sectional correlation is due to the fact that in the majority of the cases it is particularly informative for the ranking of next period portfolio returns. Importantly, Panel B reveals that accounting for further components under a LASSO approach harmonizes

the performance across all four characteristic-based models. Finally, Panel C shows that the benchmark models display slightly lower proportions of correct sign and markedly lower average cross-sectional correlations compared to the models in Panel B.²⁰ Overall, results confirm that characteristic-based models can better distinguish anomaly performance compared to alternative approaches.

Recall from Section 3.2 that our predictive approach entails cross-sectional standardization of each characteristic in each month. Therefore, given the success of the approach, a natural question that arises is what is the source of variation in the characteristics of the factor portfolios that leads to predictability. This issue is discussed in detail in Section IA.3 of the Internet Appendix. We show that the main source of time variation comes from the higher moments of the cross-sectional distribution of the characteristics. This is intuitive given that the literature with respect to stock return predictability already establishes that the predictive power of several characteristics is closely related to their non-normal distribution. For example, Cooper et al. (2008) show that asset growth is highly positively skewed and accordingly its predictive power is mainly driven by the high- rather than the low-asset growth stocks. Another source of variation comes from the time-varying correlations across the different characteristics. For example, it is possible that for a given month the correlation between stock momentum and value is high and hence the standardized momentum score of the respective factor portfolios is similar, while in another month the correlation might be low and hence the momentum score of the respective factor portfolios will be completely different. In the latter case, there is additional information content that can be exploited.²¹

²⁰It is noteworthy that, similar to the case of the PLS single factor models, the factor momentum models perform reasonably well despite their negative OOS R^2 values. In fact, the 1-month factor momentum has a forecasting error that is lower than that of the historical sample mean model in only 20% of the times. In those cases, it exhibits an OOS R^2 of 38.93% and an average cross-sectional correlation of 60%. In the remaining 80% of the cases, it exhibits an OOS R^2 of -153.32% and an average cross-sectional correlation of -7%. This means that the good overall performance of the 1-month factor momentum is driven by only a small subsample of observations during which it can predict future factor portfolio returns particularly well in terms of both exact and relative terms.

²¹Obviously, another source of variation stems from the recursively estimated weighting vector $w_{i,t}$. However, we show that this vector remains relatively stable across time.

Finally, we examine the implications of applying LASSO on the sets of characteristic components in terms of model complexity. Our approach allows for the number of features to vary across factor portfolios and across time, enabling us to see when the characteristic signal is strong and when it diminishes. Figure 4 displays the number of non-zero coefficients for the characteristic PCs of each anomaly PC portfolio when PCA and PLS are used for the RHS in the out-of-sample period. Each line chart shows the number of characteristic-based components that minimize the mean-squared error in the validation period.

Results from Figure 4 confirm the existence of significant time variability in the required number of features across time and across PC portfolios. The time variation in the number of features by itself implies that the predictive ability of characteristics is not constant, something that is expected given the time variation in factor portfolio risk premia. Interestingly, at certain periods the number of features falls down to zero, implying that at times characteristics provide no predictive information at all and the PC return forecasts shrink down to an intercept term. Conversely, a high number of features implies that a lot of the variation in the characteristics is useful in predicting PC portfolio returns. Such peaks and troughs in the number of features are observed at different points in time for the different PC portfolios, which implies that the importance of characteristics is also unstable across the main sources of variation and that each source should be approached independently in terms of model specification. Finally, with regards to the different methods used for the RHS, it is evident that PCA uses on average more features and has higher variability in the number of features across time compared to PLS. PCA components mix return-relevant and irrelevant information, making the selection of the optimal number of features more sensitive to the validation sample and as a result less stable. PLS condenses the characteristic information into fewer PCs than PCA and is more stable over time, although there is still significant time variability in the number of components being used.

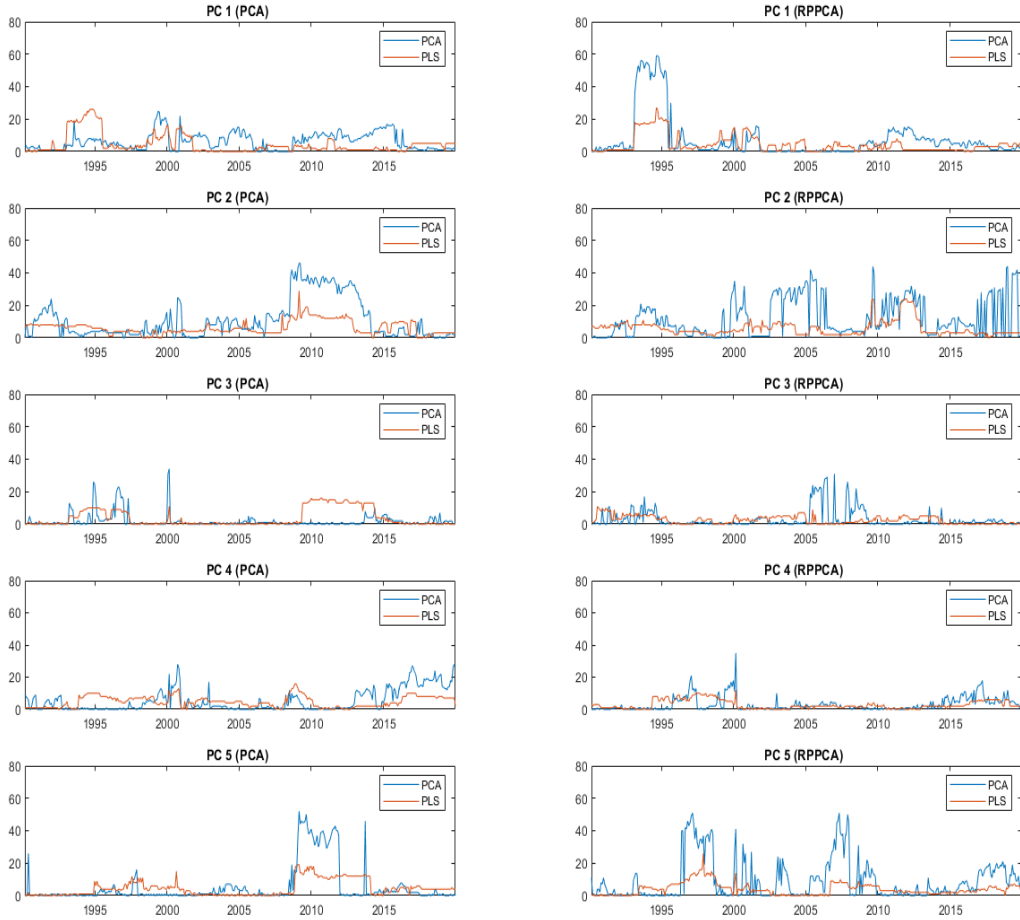


Figure 4: Number of features for each PC portfolio under the different models. The number of features is identified by recursively applying LASSO on the set of components and picking the penalty factor that minimizes the mean-squared-error in the validation period.

4.3 Investment performance

In this section, we assess the performance of each model in terms of economic rather than statistical contribution and examine how return forecasts can be translated into factor timing strategies. We construct three different strategies and assess their performance using a monthly holding period and standard portfolio evaluation measures. The first strategy is a simple long-short strategy (LSS), or an LS portfolio of factor portfolios. Factor portfolios are grouped into equally-weighted deciles based on their return forecasts and a long-short strategy is constructed that goes long the top 10% and short the bottom

10% of the anomalies. Such a strategy focuses on the extremes of the conditional returns distribution and neglects factor portfolios that lie in the middle. Hence, LSS will work well as long as the models can identify anomalies with very high or very low expected returns at each period, even if they are indecisive about anomalies with conditional returns close to zero.

The second investment strategy is similar to the time-series factor momentum (TSFM) strategy by Gupta and Kelly (2019). TSFM scales factor portfolio returns R_{t+1} , according to return forecasts \hat{R}_{t+1} . The scaling vector $s_{t,n}$ is obtained by dividing return forecasts by individual factor in-sample monthly volatility and capping them at ± 2 , as shown below:

$$s_{t,n} = \min \left(\max \left(\frac{1}{\sigma_{t,n}} \hat{R}_{t+1,n}, -2 \right), 2 \right). \quad (9)$$

The strategy goes long in factors with positive scores and short in factors with negative scores. The scores are rescaled to form unit dollar weights for the long and the short leg.²² Multiplying next period factor portfolio returns by their respective weights reveals the return of the strategy:

$$\text{TSFM}_{t+1} = \frac{\sum_n 1_{\{s_{t,n}>0\}} R_{t+1,n} \times s_{t,n}}{\sum_n 1_{\{s_{t,n}>0\}} s_{t,n}} - \frac{\sum_n 1_{\{s_{t,n}\leq 0\}} R_{t+1,n} \times s_{t,n}}{\sum_n 1_{\{s_{t,n}\leq 0\}} s_{t,n}}. \quad (10)$$

The main difference between LSS and TSFM is that, while both are technically long-short, TSFM invests in the whole universe of factor portfolios and not in factor portfolios with extreme return forecasts only. Furthermore, the number of factor portfolios in each leg, as well as the relative weights, can differ for TSFM while remaining constant under LSS. More concretely, the sign of the return forecast determines whether the anomaly will be bought or sold, while the magnitude of the forecast determines the relative weight. Hence, under TSFM the long and the short legs can have a disproportional number of constituents and in extreme cases, the strategy can converge to long or short only.

²²Specifically, positive scores are divided by the sum of the positive scores and the negative scores are divided by the sum of negative scores.

The last strategy, also in Gupta and Kelly (2019), is the cross-sectional version of TSFM (CSFM). The main difference between CSFM and TSFM is that the cross-sectional median is subtracted from the return forecasts before scaling with volatility. This strategy takes positions in factor portfolios that have outperformed or underperformed relative to their peers. For example, if return forecasts are positive for all factor portfolios then TSFM will take a long position in all of them, while CSFM will go long only in those with above median return forecasts and short the rest. Hence, even if the models cannot identify the sign correctly, this strategy will still be profitable if forecasts are consistent in relative terms, similarly to LSS:

$$s_{t,n} = \min \left(\max \left(\frac{1}{\sigma_{t,n}} \hat{R}_{t+1,n} - \text{median}(\hat{R}_{t+1,.}), -2 \right), 2 \right). \quad (11)$$

Table 4 presents the portfolio evaluation measures for the various models under the three strategies. First, we find that LSS delivers the highest average return among the three strategies across almost all the models, while CSFM and TSFM tend to have higher Sharpe ratios. Turning to Panel A, results confirm the superiority of PLS over PCA for the RHS in the single factor case, as also presented in Table 3. Evidently, models based on a single factor that concentrates the variation among multiple characteristics are

	Average Return			Standard Deviation			Sharpe Ratio			t-statistic			Hit-Rate			Max Drawdown		
	LSS	TSFM	CSFM	LSS	TSFM	CSFM	LSS	TSFM	CSFM	LSS	TSFM	CSFM	LSS	TSFM	CSFM	LSS	TSFM	CSFM
Panel A: Single factor																		
PCA	-0.10	-0.03	-0.01	7.23	4.14	4.09	-0.01	-0.01	-0.00	-0.26	-0.13	-0.02	52.37	52.92	52.65	80.99	47.64	43.94
PCA-PLS	1.16	0.74	0.76	8.57	5.27	5.25	0.13	0.14	0.14	2.55	2.65	2.75	57.66	59.05	59.61	41.09	33.83	33.01
RPPCA	0.20	0.20	0.21	5.64	3.11	3.15	0.03	0.06	0.07	0.66	1.21	1.29	54.32	54.32	56.55	46.54	27.73	26.68
RPPCA-PLS	1.12	0.73	0.74	8.28	4.81	4.81	0.13	0.15	0.15	2.55	2.87	2.93	55.15	58.22	60.17	37.46	30.31	29.79
Panel B: Time-varying number of factors using LASSO																		
PCA	1.47	0.97	0.95	8.16	5.01	4.96	0.18	0.19	0.19	3.40	3.65	3.64	55.71	56.82	55.99	16.76	13.93	12.16
PCA-PLS	1.38	0.96	0.97	8.22	4.99	4.98	0.17	0.19	0.19	3.18	3.66	3.68	61.00	62.67	61.56	15.00	13.19	12.70
RPPCA	1.21	0.84	0.83	7.06	4.01	4.00	0.17	0.21	0.21	3.26	3.99	3.93	57.66	60.72	59.89	38.11	22.24	22.25
RPPCA-PLS	1.23	0.84	0.86	6.89	4.04	4.09	0.18	0.21	0.21	3.39	3.96	3.98	60.72	61.84	61.00	26.77	17.26	17.16
Panel C: Benchmark models																		
1mMOM	1.06	0.56	0.58	8.81	4.95	4.96	0.12	0.11	0.12	2.28	2.13	2.22	57.10	56.82	57.10	18.45	16.94	17.32
12mMOM	0.84	0.67	0.67	8.69	5.08	5.12	0.10	0.13	0.13	1.84	2.51	2.48	55.43	55.99	56.82	25.96	17.89	18.28
PCA-BM	0.79	0.59	0.61	6.16	3.79	3.79	0.13	0.16	0.16	2.43	2.96	3.03	57.10	57.66	57.66	31.65	21.59	20.61
IR Spread	0.87	0.54	0.54	6.49	3.92	3.92	0.13	0.14	0.14	2.49	2.63	2.61	52.87	52.65	52.92	34.74	25.51	25.71
Sentiment	0.74	0.47	0.46	5.29	3.13	3.22	0.14	0.15	0.14	2.60	2.82	2.72	56.32	53.48	52.92	21.80	20.15	19.80
Historical Sample Mean	0.48	0.35	0.35	3.47	2.59	2.61	0.14	0.14	0.13	2.64	2.59	2.56	58.22	57.94	58.22	19.19	19.22	19.03

Table 4: Portfolio evaluation measures for long-short (LSS), time-series (TSFM) and cross-sectional (CSFM) strategies under the different models for the sample period 1990-2019. Panel A displays results using a single latent factor to predict PC portfolio returns. Panel B shows the results where the optimal number of factors is selected by applying LASSO on the whole set of latent factors. Panel C displays results for the benchmark models. Average Return: average monthly return, Standard Deviation: monthly standard deviation, Sharpe ratio: monthly Sharpe ratio, t-statistic: t-statistic on H_0 : Average Return = 0, Hit-Rate: percentage of the total number of occasions that the strategy resulted in positive returns, Maxdrawdown: maximum cumulative loss. The best performing model for each metric under each strategy is highlighted in bold.

unable to predict the cross-sectional dispersion of factor portfolio returns, implying again that a lot of variation in the characteristics is irrelevant in asset return prediction. As a result, strategies based on PCA and RPPCA deliver returns indistinguishable from zero, with returns for PCA even becoming negative. Conversely, when PLS is used for the RHS all strategies deliver positive and significant returns, reflecting the ability of the method to concentrate return-relevant variation into a single predictor.

Panel B of Table 4 shows that the use of further components in combination with LASSO uniformly improves investment performance across all models. This result is again consistent with the results reported in Table 3. In particular, all strategies deliver highly significant returns, surpassing the t-value threshold of three by Harvey et al. (2016). Turning to the specifics, the use of PCA for the LHS leads to investment strategies with higher average returns while the use of RPPCA leads to higher Sharpe ratios, irrespective of the strategy or the RHS model. Furthermore, although strategies that utilize PCA for the LHS have higher volatility, they also exhibit a higher hit-rate and a lower max drawdown, reflecting higher consistency and lower downside risk. Overall, results are now similar across models, suggesting that once further components are considered, no significant difference arises across methods.

Panel C displays the results for the benchmark models. In line with prior literature (e.g., Gupta and Kelly (2019)), factor momentum using a 1-month formation period achieves the highest return among the benchmark models for the LSS strategy, while the 12-month signal delivers higher returns for TFMS and CFMS. Using the book-to-market ratio, issuer-repurchaser spread or investor sentiment as predictors results in strategies with moderate return performance and Sharpe ratios. The historical sample average strategy delivers low average returns, albeit statistically significant. Such a strategy produces conservative return forecasts and as a result, takes more static positions compared to the rest of the models. Comparing results across panels, characteristic-based models

that employ LASSO outperform all benchmark models under all three strategies, demonstrating the benefits of conditioning factor portfolio returns on observable characteristics under a regularized framework.

In order to compare the performance of the various models across time, Figure 5 presents the cumulative return performance of the factor timing portfolios under the three investment strategies. For conciseness, we only display the performance for the characteristic-based models employing LASSO together with the benchmark models. Graphs to the left show the cumulative performance over the whole out-of-sample period and graphs to the right focus on the last ten years. As it can be seen from the graphs, the 1-month factor momentum outperforms the characteristic-based models in the early years of the out-of-sample period, up until the late 90s. A spike in performance occurs for all strategies around 2000, i.e., during the buildup of the dot-com bubble. Unlike the majority of the benchmarks, characteristic-based models do not plummet after the burst and continue to outperform thereafter. Furthermore, the performance of the characteristic-based models is relatively unaffected by the 2008 financial crisis and a second spike in performance is observed as the economy enters the recovery phase in 2009. Hence, our strategies work well in periods of financial turmoil while still enjoying the upside potential of a bull market.

Finally, factor timing portfolios based on characteristics exhibit strong return performance in the post-2010 period. Looking at the graphs in the right panel of Figure 5, characteristic-based models display a positive trend in later years, while strategies based on issuer-repurchaser spread, investor sentiment, factor momentum and in-sample average remain relatively stagnant. Out of all benchmark models, the historical sample mean remains the most stagnant, especially throughout the later years.

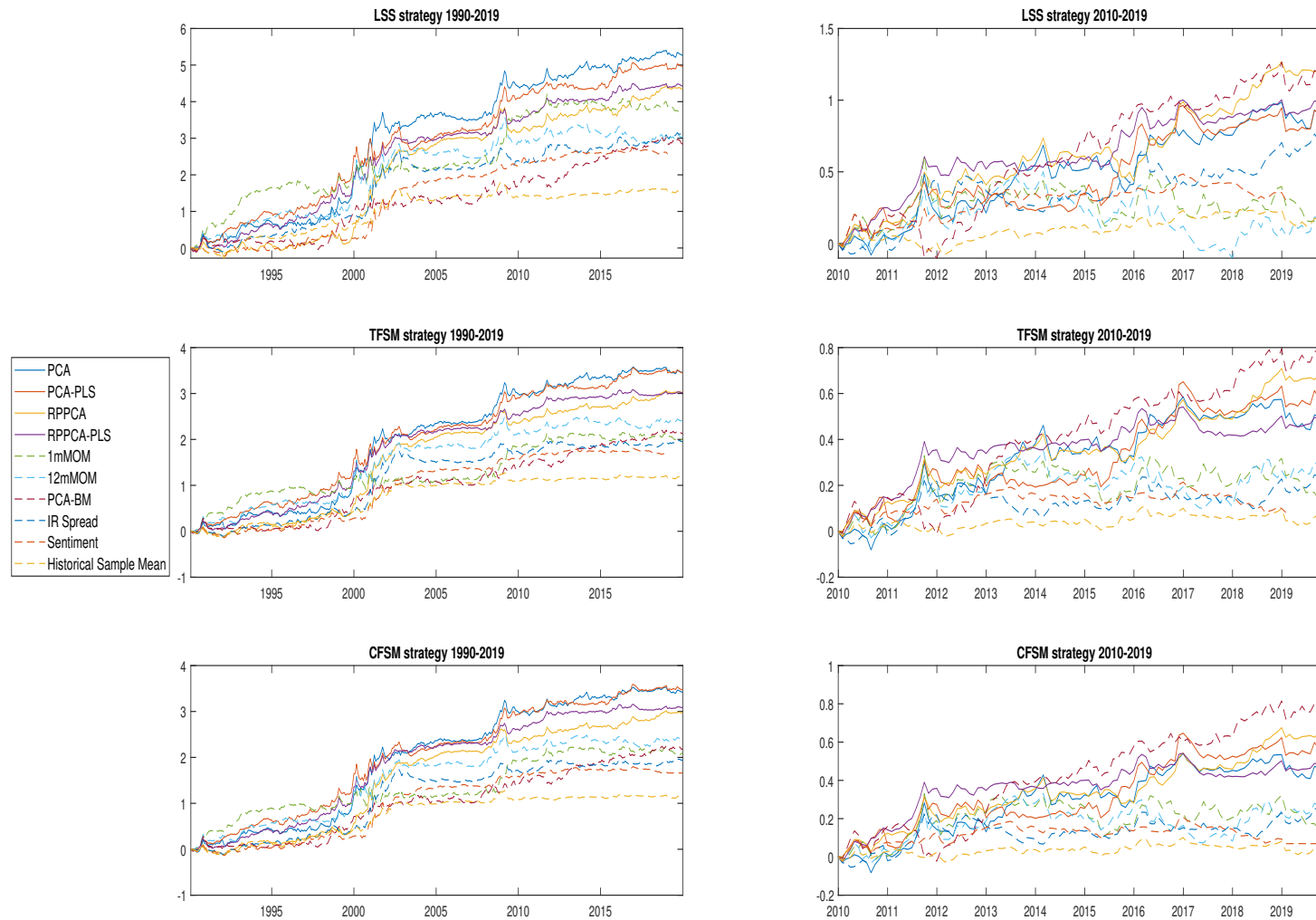


Figure 5: Cumulative return performance of factor timing strategies. The figure displays the performance of LSS, TFSM and CFM for characteristic-based models using LASSO and the benchmark models. Graphs to the left display the cumulative return performance over the whole sample period (1990-2019) and graphs to the right display the cumulative performance over the last ten years of the sample period (2010-2019). All strategies begin with a zero dollar investment.

Notably, the book-to-market approach works equally well with the characteristic-based models in later years. Both approaches isolate the first five PCs of factor portfolio returns and use a characteristic-based measure to create forecasts. As such, results highlight the importance of focusing on the main sources of variation and the ability of characteristics to explain the dynamics of factor portfolios. Characteristic-based models outperform the rest of the benchmarks under all three strategies, with the difference being more pronounced for the LSS strategy, as it focuses solely on the most prominent subset of factor portfolios. Overall, the profitability of the benchmark strategies erodes significantly in later years, suggesting that the informativeness of alternative predictors about future factor portfolio returns has faded.

It is also important to note that factor timing strategies based on observed characteristics yield positive returns in the most recent period, even though most factors have been found empirically to die out over time (Chordia et al., 2014; McLean and Pontiff, 2016; Green et al., 2017). Corroborating this evidence, a comparison between Table 4 and Table IA.4 reveals that characteristic-based factor timing strategies exhibit investment performance superior to that of unconditional factor portfolios. In that sense, our paper acknowledges the fact that unconditional risk premia lack robustness and shows that focusing on the predictability of conditional risk premia can help an investor expand her investment opportunity set. In a similar vein, Haddad et al. (2020) find that strong factor portfolio predictability implies a stochastic discount factor that is much more volatile than previously thought.

Lastly, a question that arises is what are the trading positions that our characteristic-based models take over time. The analysis presented in Section IA.7 of the Internet Appendix provides some interesting insights. First, even though prominent anomalies such as `mom12m` and `retvol` are heavily traded, the factor timing strategies rotate among multiple anomalies and do not focus on only a small subset with high unconditional

returns. Second, several anomalies appear almost equally often in the long and the short legs. Finally, anomalies that have only a small impact on the PC portfolios are hardly considered by our factor timing strategies, which is expected given that their return forecasts are by construction tilted towards zero.

5 Alternative approaches

In this section, we examine different estimation approaches. Our method uses a large collection of characteristics and combines different dimension reduction and regularization techniques to achieve robust out-of-sample predictability. As such, it is important to examine where the predictability stems from by evaluating the incremental effect of each contributor on the out-of-sample performance. Furthermore, it is important to assess whether the benchmark models can beat our characteristic-based models once dimension reduction and regularization techniques are also used in their cases.

Starting with the characteristics, the simplest approach is to forecast each anomaly using the time-series of its own characteristic spread. Alternatively, one can forecast each anomaly individually using the whole collection of characteristics and can further employ a dimension reduction technique, such as PLS, or a regularization technique, such as LASSO, for the RHS. Finally, one can create PC portfolios on the LHS without using any dimension reduction technique (but potentially using LASSO) for the RHS. Moreover, the benchmark models can also be modified in various ways. For instance, factor momentum, issuer-repurchaser spread and investor sentiment can be applied on the PC portfolios. For book-to-market ratio and issuer-repurchaser spread, the dataset can be expanded by using all the ratios and spreads simultaneously to predict each PC portfolio or individual anomaly. Finally, LASSO can be applied on these richer information sets to account for overfitting. A detailed description of the models discussed in this section can be found in Panel C of Table IA.2.

Table 5 reports the Total OOS R^2 and average cross-sectional correlation for the modified forecasting methods. Panel A shows that, in line with Haddad et al. (2020), predicting each anomaly by its own spread is not particularly successful as it provides negative OOS R^2 and relatively low average cross-sectional-correlation. When we incorporate the full set of characteristics for each anomaly, the OOS R^2 metric worsens possibly due to overfitting, but the average cross-sectional correlation improves in two out of the three cases (the exception being the model that uses PLS on the RHS). When we further condense the information content of the anomalies into five PC portfolios the average cross-sectional correlation increases even more. Nevertheless, the OOS R^2 remains negative and the cross-sectional correlation is still at the levels of 6%, clearly lower than the 8%-9% provided by our main models using dimension reduction also on the RHS (Panel B of Table 3). Overall, the results of Panel A corroborate the importance of using the full set of characteristics for factor timing purposes, while they further highlight the additional benefits that arise when incorporating dimension reduction and regularization techniques on both sides of the forecasting exercise.

Panel B shows the results for the alternative specifications of the benchmark models. Applying the momentum signal on the PCs instead of individual anomalies has an inconsistent effect on forecasting performance as it improves the OOS R^2 but reduces the cross-sectional correlation. Using the book-to-market ratios of all PC portfolios to predict each single of them individually is not particularly fruitful with both performance measures worsening compared to the main PCA-BM model. Predicting each anomaly by its own book-to-market ratio delivers a positive OOS R^2 and higher cross-sectional correlation, but it still falls behind the baseline BM model. In terms of the issuer-repurchaser spread, using the spreads of all portfolios instead of the spread of each individual portfolio, as in the baseline IR Spread model, reduces the OOS R^2 but improves considerably

the cross-sectional correlation.²³ Nevertheless, the use of LASSO does not make any substantial contribution in this case, with the OOS R^2 still being on the negative side and the cross-sectional correlation decreasing. Finally, using the investor sentiment index to predict the PC portfolios or in combination with LASSO to predict individual anomalies has little effect compared to the baseline Sentiment model. Overall, we find mixed results for the modified benchmark models, with the dimension reduction and regularization additions improving the models only occasionally. In any case, even the best modified models exhibit clearly worse performance than our main characteristic-based models.

	Total OOS R^2	Average cross-sectional correlation
Panel A: Modified characteristic-based models		
Anom-Own Characteristic	-0.78	3.77
Anom-1 PLS	-2.51	2.94
Anom-All Characteristics	-93.60	4.48
Anom-All Characteristics LASSO	-2.09	5.82
5 PCs-All Characteristics	-53.93	6.19
5 PCs-All Characteristics LASSO	-1.14	6.26
Panel B: Modified benchmark models		
5 PCs-1mMOM	-57.99	5.77
5 PCs-12mMOM	-4.72	3.73
5 PCs-5 BMs	-0.75	2.61
5 PCs-5 BMs LASSO	-0.44	1.97
Anom-Own BM	0.20	4.29
Anom-All IR Spreads	-31.73	7.11
Anom-All IR Spreads LASSO	-2.91	3.06
5 PCs-IR Spread	-3.04	0.91
5 PCs-All IR Spreads	-1.19	2.31
5 PCs-Sentiment	0.39	3.55
Sentiment-LASSO	0.27	4.70

Table 5: Total OOS R^2 and average cross-sectional correlation of factor portfolio return forecasts based on various forecasting methods.

²³Using the collection of issuer-repurchaser spreads can be justified by considering the different factor portfolios as substitutes for the same investor. In that sense, time-varying characteristic mispricing can propagate from one factor to the rest due to changes in demand and supply for the different factors, making the issuer-repurchaser spreads of other factors important.

	Average Return			Standard Deviation			Sharpe Ratio			t-statistic			Hit-Rate			Max Drawdown		
	LSS	TSFM	CSFM	LSS	TSFM	CSFM	LSS	TSFM	CSFM	LSS	TSFM	CSFM	LSS	TSFM	CSFM	LSS	TSFM	CSFM
Panel A: Modified characteristic-based models																		
Anom-Own Characteristic	0.46	0.34	0.32	4.04	2.48	2.53	0.11	0.14	0.13	2.17	2.57	2.40	60.72	60.72	61.00	33.44	21.02	21.76
Anom-1 PLS	0.46	0.30	0.30	4.33	2.96	2.97	0.11	0.10	0.10	2.03	1.95	1.93	56.82	57.66	58.50	29.95	25.56	25.41
Anom-All Characteristics	0.74	0.52	0.51	4.71	2.76	2.77	0.16	0.19	0.18	2.98	3.54	3.49	57.38	58.77	59.05	11.65	8.74	9.16
Anom-All Characteristics LASSO	1.02	0.81	0.81	5.54	4.05	4.06	0.18	0.20	0.20	3.49	3.78	3.79	58.77	59.05	57.38	19.40	16.62	16.50
5 PCs- All Characteristics	1.02	0.67	0.65	7.98	4.65	4.72	0.13	0.14	0.14	2.42	2.72	2.60	54.04	54.60	54.60	33.36	25.60	26.61
5 PCs- All Characteristics LASSO	1.09	0.70	0.72	7.59	4.64	4.63	0.14	0.15	0.16	2.71	2.88	2.95	54.60	55.99	55.99	19.23	16.62	16.32
Panel B: Modified benchmark models																		
5 PCs-1mMOM	1.13	0.64	0.65	9.24	5.52	5.54	0.12	0.12	0.12	2.32	2.20	2.22	57.38	55.71	54.60	25.74	19.89	19.82
5 PCs-12mMOM	0.76	0.58	0.58	9.40	5.90	5.85	0.08	0.10	0.10	1.53	1.88	1.89	52.65	53.76	53.20	36.46	28.19	27.61
5 PCs-5 BMs	0.28	0.22	0.25	7.37	4.86	4.81	0.04	0.05	0.05	0.73	0.86	0.98	55.99	52.37	55.15	47.72	37.71	36.62
5 PCs-5 BMs LASSO	0.35	0.26	0.27	7.85	5.01	5.00	0.04	0.05	0.05	0.85	0.99	1.03	53.76	53.48	53.76	46.79	37.78	37.21
Anom-Own BM	0.70	0.48	0.48	4.37	2.60	2.69	0.16	0.18	0.18	3.05	3.50	3.37	57.10	56.27	55.99	28.28	21.53	21.61
Anom-All IR Spreads	1.24	0.80	0.81	7.70	4.47	4.49	0.16	0.18	0.18	3.06	3.38	3.43	60.17	58.22	59.33	21.60	14.53	14.61
Anom-All IR Spreads LASSO	0.67	0.50	0.51	6.88	4.85	4.87	0.10	0.10	0.10	1.84	1.95	1.98	56.27	55.99	56.27	24.21	20.57	20.46
5 PCs-IR Spread	0.07	-0.14	-0.09	6.28	4.52	4.22	0.01	-0.03	-0.02	0.22	-0.59	-0.39	48.19	47.08	47.08	77.69	99.97	87.40
5 PCs-All IR Spreads	0.51	0.29	0.28	7.99	4.89	4.82	0.06	0.06	0.06	1.20	1.12	1.11	52.65	52.65	52.65	57.43	33.41	32.55
5 PCs-Sentiment	0.56	0.35	0.33	6.86	3.89	3.97	0.08	0.09	0.08	1.54	1.70	1.59	54.87	53.76	52.37	57.10	27.64	29.55
Sentiment-LASSO	0.67	0.46	0.46	5.37	3.22	3.25	0.12	0.14	0.14	2.35	2.72	2.69	55.15	56.27	56.55	17.90	19.22	19.13

Table 6: Portfolio evaluation measures for long-short (LSS), time-series (TSFM) and cross-sectional (CSFM) strategies under the alternative specifications for the sample period 1990-2019. Average Return: average monthly return, Standard Deviation: monthly standard deviation, Sharpe ratio: monthly Sharpe ratio, t-statistic: t-statistic on H_0 : Average Return = 0, Hit-Rate: percentage of the total number of occasions that the strategy resulted in positive returns, Maxdrawdown: maximum cumulative loss.

Table 6 presents the portfolio evaluation results for the modified models. Starting with Panel A, the investment performance of the modified characteristic-based models is broadly in line with the cross-sectional correlations from Table 5. In particular, all strategies exhibit good investment performance, while the average returns and Sharpe ratios tend to improve when, on top of using the whole set of portfolio characteristics, we further incorporate PCA and/or LASSO in the forecasting exercise. Still, even the best performing modified models fall behind the main ones presented in Panel B of Table 4. Therefore, it is confirmed again that using multiple portfolio characteristics is indispensable for forming a successful factor timing strategy, but the dimension reduction and regularization techniques provide additional benefits. Turning to Panel B, the PCA-based momentum models, the model that uses all issuer-repurchaser spreads, the model that uses each portfolio’s book-to-market ratio, and the model that employs LASSO together with market sentiment appear to be the strongest ones. This is unsurprising given that these models also deliver the highest cross-sectional correlations in Table 5. Importantly, even these alternative benchmark models exhibit weaker investment performance than our preferred characteristic-based models in Panel B of Table 4. Overall, the alternative information sets have lower factor timing ability compared to the set of portfolio characteristics even if they are enhanced by employing dimension reduction or LASSO.

6 Conclusion

We investigate the predictability of factor portfolios from their own portfolio characteristics, going over and above existing methods for predicting factor portfolio returns and examining the possibility that factor portfolios are predictable by characteristics other than their own. Our approach offers a natural continuation to the stock return predictability problem and our findings shed light on the evolution of the underlying return drivers over time. Under our empirical framework, a large collection of stock characteristics is used to initially construct the LS portfolios and subsequently predict their

performance. A key aspect of our methodology is the reduction of the dimensions of the predictability problem, which we achieve by independently shrinking the number of predictors and forecasting targets. Our approach provides a new framework for dealing with panel data, allowing each source of variation to be approximated by models of different complexity. By using a flexible model specification that combines LASSO with dimension reduction techniques, we allow the number of predictors to vary across PC portfolios and over time. We find this approach to be especially fruitful, as it considerably improves results over a static single latent factor model.

In terms of factor portfolio predictability, we observe significant benefits from timing factor portfolio returns using observed characteristics. These benefits go over and above existing methods documented in the literature, highlighting the importance of considering the information in the characteristics in a collective way. Specifically, the dominant PC portfolios are highly predictable by the information contained in their characteristics and this predictability can be easily extended to individual anomalies. In that sense, dimension reduction techniques not only accommodate the computational tractability of the estimation problem, but also improve forecasting and investment performance by enabling us to focus on the sources of variation that are most predictable. The performance of our factor timing strategies is superior to that of any individual anomaly and persistent over the later years of the sample period, demonstrating the benefits of timing over static factor investing. Hence, in the context of anomaly return prediction it is important to (1) account for the information contained in multiple characteristics, (2) focus on the main sources of variation in factor portfolio returns since those are the most predictable and (3) apply some kind of time-varying regularization on the set of predictors to account for the time variability in characteristic informativeness. Overall, our findings have important implications for the use of machine learning methods in asset pricing applications and help justify the importance of observable characteristics in explaining the dynamics of factor portfolios.

References

- Akbas, F., Armstrong, W. J., Sorescu, S. and Subrahmanyam, A. (2016), ‘Capital market efficiency and arbitrage efficacy’, *Journal of Financial and Quantitative Analysis* **51**(2), 387–413.
- Arnott, R. D., Clements, M., Kalesnik, V. and Linnainmaa, J. T. (2021), ‘Factor momentum’, Available at SSRN: <https://ssrn.com/abstract=3116974> or <http://dx.doi.org/10.2139/ssrn.3116974> .
- Asness, C., Chandra, S., Iltanen, A. and Israel, R. (2017), ‘Contrarian factor timing is deceptively difficult’, *Journal of Portfolio Management* **43**(5), 72–87.
- Baba-Yara, F., Boons, M. and Tamoni, A. (2021), ‘Value return predictability across asset classes and commonalities in risk premia’, *Review of Finance* **25**(2), 449–484.
- Bai, J. and Ng, S. (2002), ‘Determining the number of factors in approximate factor models’, *Econometrica* **70**(1), 191–221.
- Baker, M. and Wurgler, J. (2006), ‘Investor sentiment and the cross-section of stock returns’, *Journal of Finance* **61**(4), 1645–1680.
- Campbell, J. Y. and Thompson, S. B. (2008), ‘Predicting excess stock returns out of sample: Can anything beat the historical average?’, *Review of Financial Studies* **21**(4), 1509–1531.
- Chordia, T., Subrahmanyam, A. and Tong, Q. (2014), ‘Have capital market anomalies attenuated in the recent era of high liquidity and trading activity?’, *Journal of Accounting and Economics* **58**(1), 41–58.
- Cochrane, J. H. (2011), ‘Presidential address: Discount rates’, *Journal of Finance* **66**(4), 1047–1108.
- Connor, G. and Korajczyk, R. A. (1986), ‘Performance measurement with the arbitrage pricing theory: A new framework for analysis’, *Journal of Financial Economics* **15**(3), 373–394.
- Connor, G. and Korajczyk, R. A. (1988), ‘Risk and return in an equilibrium APT: Application of a new test methodology’, *Journal of Financial Economics* **21**(2), 255–289.
- Cooper, M. J., Gulen, H. R. and Schill, M. J. (2008), ‘Asset growth and the cross-section of stock returns’, *Journal of Finance* **63**(4), 1609–1651.
- Daniel, K. and Moskowitz, T. J. (2016), ‘Momentum crashes’, *Journal of Financial Economics* **122**(2), 221–247.
- DeMiguel, V., Martin-Utrera, A., Nogales, F. J. and Uppal, R. (2020), ‘A transaction-cost perspective on the multitude of firm characteristics’, *Review of Financial Studies* **33**(5), 2180–2222.

- Dichtl, H., Drobetz, W., Lohre, H., Rother, C. and Vosskamp, P. (2019), ‘Optimal timing and tilting of equity factors’, *Financial Analysts Journal* **75**(4), 84–102.
- Ehsani, S. and Linnainmaa, J. T. (2022), ‘Factor momentum and the momentum factor’, *Journal of Finance* **77**(3), 1877–1919.
- Fama, E. F. and French, K. R. (1993), ‘Common risk factors in the returns on stocks and bonds’, *Journal of Financial Economics* **33**(1), 3–56.
- Fama, E. F. and French, K. R. (2008), ‘Dissecting anomalies’, *Journal of Finance* **63**(4), 1653–1678.
- Fan, J., Liao, Y. and Wang, W. (2016), ‘Projected principal component analysis in factor models’, *Annals of Statistics* **44**(1), 219.
- Feng, G., Giglio, S. and Xiu, D. (2020), ‘Taming the factor zoo: A test of new factors’, *Journal of Finance* **75**(3), 1327–1370.
- Freyberger, J., Neuhierl, A. and Weber, M. (2020), ‘Dissecting characteristics nonparametrically’, *Review of Financial Studies* **33**(5), 2326–2377.
- Giglio, S. and Xiu, D. (2021), ‘Asset pricing with omitted factors’, *Journal of Political Economy* **129**(7), 1974–1990.
- Green, J., Hand, J. R. and Zhang, X. F. (2017), ‘The characteristics that provide independent information about average U.S. monthly stock returns’, *Review of Financial Studies* **30**(12), 4389–4436.
- Greenwood, R. and Hanson, S. G. (2012), ‘Share issuance and factor timing’, *Journal of Finance* **67**(2), 761–798.
- Grundy, B. D. and Martin, J. S. (2001), ‘Understanding the nature of the risks and the source of the rewards to momentum investing’, *Review of Financial Studies* **14**(1), 29–78.
- Gu, S., Kelly, B. and Xiu, D. (2020), ‘Empirical asset pricing via machine learning’, *Review of Financial Studies* **33**(5), 2223–2273.
- Gupta, T. and Kelly, B. (2019), ‘Factor momentum everywhere’, *Journal of Portfolio Management* **45**(3), 13–36.
- Haddad, V., Kozak, S. and Santosh, S. (2020), ‘Factor timing’, *Review of Financial Studies* **33**(5), 1980–2018.
- Harvey, C. R., Liu, Y. and Zhu, H. (2016), ‘... and the cross-section of expected returns’, *Review of Financial Studies* **29**(1), 5–68.
- Hou, K., Xue, C. and Zhang, L. (2020), ‘Replicating anomalies’, *Review of Financial Studies* **33**(5), 2019–2133.
- Huang, S. (2022), ‘The momentum gap and return predictability’, *Review of Financial Studies* **35**(7), 3303–3336.

- Jacobs, H. (2015), ‘What explains the dynamics of 100 anomalies?’, *Journal of Banking & Finance* **57**, 65–85.
- Kelly, B. and Pruitt, S. (2013), ‘Market expectations in the cross-section of present values’, *Journal of Finance* **68**(5), 1721–1756.
- Kelly, B. T., Malamud, S. and Zhou, K. (2022), ‘The virtue of complexity in return prediction’, *Journal of Finance*, forthcoming.
- Kelly, B. T., Pruitt, S. and Su, Y. (2019), ‘Characteristics are covariances: A unified model of risk and return’, *Journal of Financial Economics* **134**(3), 501–524.
- Keloharju, M., Linnainmaa, J. T. and Nyberg, P. (2016), ‘Return seasonalities’, *Journal of Finance* **71**(4), 1557–1590.
- Kozak, S., Nagel, S. and Santosh, S. (2018), ‘Interpreting factor models’, *Journal of Finance* **73**(3), 1183–1223.
- Kozak, S., Nagel, S. and Santosh, S. (2020), ‘Shrinking the cross-section’, *Journal of Financial Economics* **135**(2), 271–292.
- Lettau, M. and Pelger, M. (2020a), ‘Estimating latent asset-pricing factors’, *Journal of Econometrics* **218**(1), 1–31.
- Lettau, M. and Pelger, M. (2020b), ‘Factors that fit the time series and cross-section of stock returns’, *Review of Financial Studies* **33**(5), 2274–2325.
- McLean, R. D. and Pontiff, J. (2016), ‘Does academic research destroy stock return predictability?’, *Journal of Finance* **71**(1), 5–32.
- Nagel, S. (2021), *Machine Learning in Asset Pricing*, Princeton University Press.
- Onatski, A. (2010), ‘Determining the number of factors from empirical distribution of eigenvalues’, *Review of Economics and Statistics* **92**(4), 1004–1016.
- Stambaugh, R. F., Yu, J. and Yuan, Y. (2012), ‘The short of it: Investor sentiment and anomalies’, *Journal of Financial Economics* **104**(2), 288–302.
- Wold, S., Ruhe, A., Wold, H. and Dunn, III, W. J. (1984), ‘The collinearity problem in linear regression. the partial least squares (pls) approach to generalized inverses’, *SIAM Journal on Scientific and Statistical Computing* **5**(3), 735–743.

Internet Appendix
for
Factor Timing with Portfolio Characteristics

Anastasios Kagkadis* Ingmar Nolte[†] Sandra Nolte[‡] Nikolaos Vasilas[§]

*Lancaster University Management School, Bailrigg, Lancaster LA1 4YX, Lancaster, email: a.kagkadis@lancaster.ac.uk

[†]Lancaster University Management School, Bailrigg, Lancaster LA1 4YX, Lancaster, email: i.nolte@lancaster.ac.uk

[‡]Lancaster University Management School, Bailrigg, Lancaster LA1 4YX, Lancaster, email: s.nolte@lancaster.ac.uk

[§]Corresponding author: Lancaster University Management School, Bailrigg, Lancaster LA1 4YX, Lancaster, email: n.vasilas@lancaster.ac.uk

IA.1 Factor variation explained by the PC portfolios

One of the key elements of our predictive approach is condensing the information of the factor portfolios using either conventional PCA or the RPPCA of Lettau and Pelger (2020). While the estimation of the PC portfolios is straightforward, the number of PC portfolios to be retained remains an empirical question. Figure IA.1 shows that irrespective of whether we use PCA or RPPCA the variation explained by the principal components is very similarly. In particular, the first component captures around 37% of the total variation, the second component captures around 13%, the third component captures around 8%, while the fourth and fifth components capture around 4%. After the fifth component the decrease in explained variation is quite gradual but the components contribute very little to the total variation. Given the aforementioned pattern and in order to be consistent with Haddad et al. (2020), we focus our analysis on five PC portfolios.

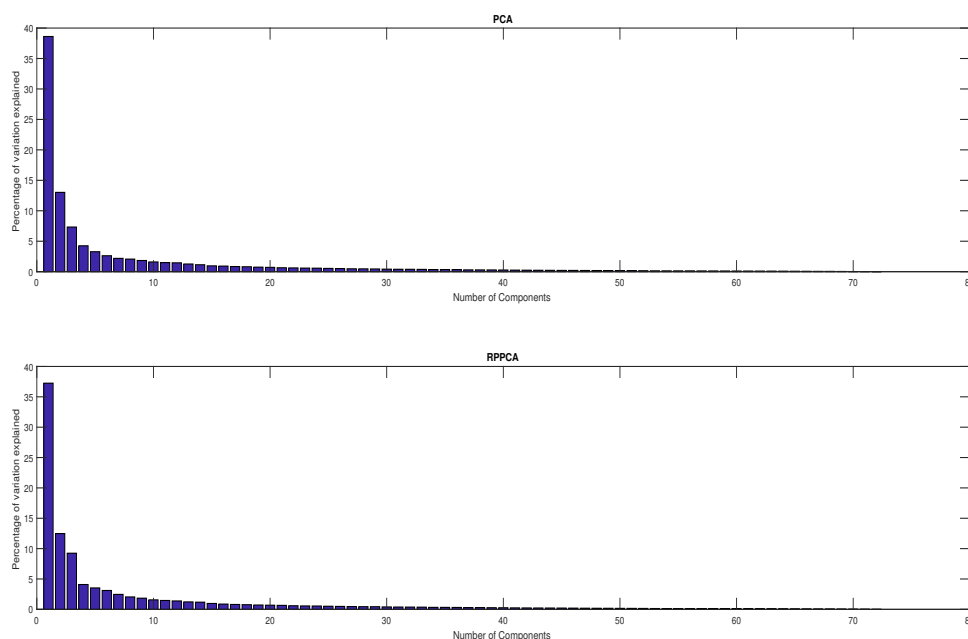


Figure IA.1: Percentage of the variation explained by each PC of factor portfolio returns under PCA and RPPCA for the sample period January 1970 to December 2019.

IA.2 Details of the dimension reduction and regularization techniques

For exposition purposes we omit the time dimension when explaining the different statistical methods. In the methodology section of the main manuscript the time subscript t indicates the recursive estimation of the different objects.

Principal Component Analysis (PCA)

The first and most popular dimension reduction method is PCA. The method produces linear combinations of the original data (PCs) while best preserving the covariance structure among the variables. Each PC successively contains as much new information about the observed variables and dimension reduction can be accommodated by focusing on the first few (dominant) PCs, while omitting the rest which are usually noise-related. Let Σ , be the $(N \times N)$ variance-covariance matrix of the $(T \times N)$ factor portfolio return matrix R . Consider the eigendecomposition of Σ :

$$\Sigma = W\Lambda W' = \sum_{i=1}^N \lambda_i w_i w_i', \quad (1)$$

where W is a $(N \times N)$ matrix whose i^{th} column w_i is the eigenvector of Σ and Λ is a diagonal matrix whose diagonal elements are the corresponding eigenvalues in decreasing order. The i^{th} eigenvector w_i , solves:

$$\begin{aligned} w_1 &= \operatorname{argmax}_{\|w_1\|=1} \{w_1' \Sigma w_1\}, \\ w_2 &= \operatorname{argmax}_{\|w_2\|=1} \{w_2' \Sigma w_2\} \text{ s.t. } w_1' \Sigma w_2 = 0, \\ &\vdots \\ w_N &= \operatorname{argmax}_{\|w_N\|=1} \{w_N' \Sigma w_N\} \text{ s.t. } w_M' \Sigma w_N = 0 \quad \forall \quad M < N. \end{aligned} \quad (2)$$

Practically, the solution in Equation (2) is obtained via a singular value decomposition

(SVD) R . The $(T \times N)$ matrix of PCs is then obtained by multiplying the matrix of factor portfolio returns with the eigenvectors, $Z = RW$. Notice that since W is an orthogonal matrix, this is equivalent to regressing the factor portfolio returns on the eigenvectors. PCA is also used to regularize the characteristics of each PC portfolio, H_i . This logic is identical to Principal Component Regression (PCR) where the predictors are transformed to their PCs and the coefficients of low variance PCs are set to zero.

Risk Premium PCA (RPPCA)

In general, PCA extracts factors that best explain time-series variation in the data. The variance-covariance matrix of factor portfolio returns can also be written as $\Sigma = \frac{1}{T}R'R - \bar{R}\bar{R}'$, where \bar{R} is an $(N \times 1)$ vector of average portfolio returns. Since average returns are subtracted, PCA utilizes information from the second moment while it neglects information from the first moment of the data. Some factors may have weak explanatory power in terms of variance if they only affect a small proportion of assets, but may still be important in an asset pricing context. In this case, conventional PCA is unable to detect the true factors (Onatski, 2012). Under an Arbitrage Pricing Theory framework, exposure to systemic risk factors should be able to explain the cross-section of expected asset returns (Ross, 1976). As such, latent factors should be able to simultaneously capture time-series variation and explain the cross-section of average returns.

Lettau and Pelger (2020) propose a new estimator by augmenting PCA with a penalty term to account for pricing errors in average returns. RPPCA is a generalization of PCA, regularized by a cross-sectional pricing error and can be implemented by simple eigenvalue decomposition of the variance-covariance matrix of asset returns after a simple transformation:

$$\frac{1}{T}R'R + \gamma\bar{R}\bar{R}'. \quad (3)$$

Essentially, the method applies PCA to the variance-covariance matrix with over-weighted

means. The resulting PCs jointly minimize the unexplained variation and the cross-sectional pricing error. The choice of the tuning parameter γ determines the relative weight of the cross-sectional pricing error compared to the time-series error. For conventional PCA $\gamma = -1$, while $\gamma = 0$ is equivalent to applying PCA to the second moment matrix. Values of $\gamma > -1$ can lead to the detection of weak factors with high Sharpe ratios. We opt for a constant value of $\gamma = 10$, as it provides a balance between explaining time-series variation and detecting weak factors.¹ The use of RPPCA should help us focus on factor portfolios with high average returns as by definition those will have a higher weight on dominant components.

Again, we apply SVD on $\frac{1}{T}R'R + 10\bar{R}\bar{R}'$ and retain the first five eigenvectors to calculate the PC portfolios $Z_i \in \mathbb{R}^{(T \times 1)}$, $i = 1, \dots, 5$. Since the purpose of RPPCA is to detect weak factors within asset returns and given that characteristics are standardized due to their difference in scale, it would be insensible to apply it on $H_i \in \mathbb{R}^{(T \times M)}$, $i = 1, \dots, 5$. Instead, we apply SDV on each $\frac{1}{T}H_i'H_i - \bar{H}_i\bar{H}_i'$, which converges back to conventional PCA.

Partial Least Squares (PLS) One of the limitations of PCA is that it focuses on condensing the covariation within the predictors. However, some of the characteristics may have no predictive power, meaning that PCA-based PCs can contain information that is ultimately useless in the forecasting exercise. In contrast, PLS constructs linear combinations of the characteristics based on their relationship with future returns by directly exploiting the covariance between the two. The method can be used to rotate H_i into linear combinations that best explain Z_i while still being orthogonal to each other.

¹A value of $\gamma = 10$ is also consistent with what the authors identify as optimal in their empirical exercise.

The vector of weights for the i^{th} PC is estimated recursively by solving:

$$\begin{aligned}
q_1^i &= \operatorname{argmax}_{\|q_1^i\|=1} \{q_1^{i'} H_i' Z_i Z_i' H_i q_1^i\}, \\
q_2^i &= \operatorname{argmax}_{\|q_2^i\|=1} \{q_2^{i'} H_i' Z_i Z_i' H_i q_2^i\} \text{ s.t. } q_1^{i'} H_i' Z_i Z_i' H_i q_2^i = 0, \\
&\vdots \\
q_N^i &= \operatorname{argmax}_{\|q_N^i\|=1} \{q_N^{i'} H_i' Z_i Z_i' H_i q_N^i\} \text{ s.t. } q_M^{i'} H_i' Z_i Z_i' H_i q_N^i = 0 \quad \forall \quad M < N. \quad (4)
\end{aligned}$$

Equation (4) highlights the distinction between PLS and PCA. Specifically, by making a comparison between Equation (2) and Equation (4), it is clear that PCA finds linear combinations that maximize the variance of H_i while PLS finds combinations of weights that maximize the squared covariance between Z_i and H_i , or the product of the variance of the predictors with the squared correlation with the forecasting target. In other words, PLS diverges from the solution that best describes H_i in order to find components that can better predict future returns. Practically, Equation (4) can be efficiently solved using the SIMPLS algorithm by De Jong (1993). Again, we calculate the PLS components $X_i = H_i Q_i$ and either retain the first component or apply LASSO on X_i to predict each $\hat{Z}_{t+1,i}$.

LASSO

Another important aspect of our estimation procedure is the use of LASSO to account for overfitting and control for model complexity. LASSO imposes sparsity by selecting a subset of features and setting the remaining coefficients to zero. This is achieved by slightly modifying the OLS objective function to incorporate a penalty for the sum of the absolute value of the coefficients. For instance, at each time t we estimate the $\beta \in \mathbb{R}^{1 \times M+1}$ of PC portfolio $Z_i \in \mathbb{R}^{T \times 1}$ on a set of characteristic PCs $X_i \in \mathbb{R}^{T \times M}$ as:

$$\min_{\beta_0, \beta} \sum_{t=1}^T (Z_i - \beta_i^0 - \sum_{m=1}^M \beta_i^m X_i^m)^2 + \delta \sum_{m=1}^M |\beta_i^m|, \quad (5)$$

where δ is a hyperparameter that determines the degree of regularization such as:

$$\sum_{m=1}^M |\beta_i^m| \leq \delta. \quad (6)$$

High values for δ result in solutions that set many coefficients exactly equal to zero, delivering parsimonious models. Using the coordinate decent algorithm by Friedman et al. (2010), we fit many values of δ simultaneously and pick the one that minimizes the forecasting error in the validation period.

IA.3 Sources of variation in the PC characteristics

The characteristics of long-short factor portfolios are initially calculated by value-weighting characteristics of stocks within each decile portfolio and then subtracting the value of the bottom decile from the top. At this stage no standardization is applied, meaning that the average characteristic across factor portfolios still preserves its time series trend and the cross-sectional variance for any given characteristic changes over time. The factor portfolio characteristics are then standardized cross-sectionally by subtracting each month the cross-sectional characteristic mean and dividing by the cross-sectional characteristic standard deviation. These standardized characteristics are then transformed into PC portfolio characteristics by being multiplied with $w_{t,i}$, allowing us to focus on the cross-sectional dispersion in the data.

Table IA.1 presents a stylized hypothetical example with three characteristics and one PC portfolio for one time period. Each factor portfolio has its own characteristic and two other characteristics. As in our empirical exercise, the characteristics are cross-sectionally standardized to have zero cross-sectional mean and unit cross-sectional variance. In this example, the PC portfolio loads positively on momentum and value and negatively on reversal. Momentum and value characteristics are positively related to returns, while reversal is negatively related. As a result, the PC portfolio has high momentum and value

scores and low reversal score, resulting in a high return for that month. Provided that characteristics jointly explain returns and that this relationship holds on average over time, PC portfolio returns will be high when PC momentum and value characteristics are high and PC reversal is low.

	Momentum Portfolio	Value Portfolio	Reversal Portfolio	PC Portfolio
PC Portfolio Weights	0.75	0.5	-0.25	
Momentum Characteristic	1	0	-1	1
Value Characteristic	0	1	-1	0.75
Reversal Characteristic	-1	0	1	-1
Return $t + 1$	1.0%	0.8%	-0.2%	1.2%

Table IA.1: Stylized example of estimating PC portfolio characteristics and returns using three anomalies.

Given our procedure, it is important to understand the sources of variation in the characteristics of each PC portfolio that ultimately lead to PC portfolio return predictability. Let us start with the diagonal elements in the characteristic matrix in Table IA.1 (e.g., the momentum of the momentum portfolio). The diagonal elements will always have the highest scores across rows (1 in our hypothetical example), since factor portfolios have the highest score for their own characteristic by construction. Still, the diagonal elements would remain constant across time only if each characteristic's distribution across the portfolio cross-section remained identical in terms of skewness and kurtosis. However, we observe significant variability in those higher moments for all characteristics over time. In particular, Figures IA.2 and IA.3 display the cross-sectional skewness and kurtosis of the 72 characteristics across all the factor portfolios over the whole sample period. As it can be seen, there is significant variation in the skewness and kurtosis of the characteristics across time, which suggests that the diagonal elements do change affecting also the characteristic scores of the PC portfolio.

With regards to the off-diagonal elements of the characteristic matrix in Table IA.1

(e.g., the momentum characteristic of the value portfolio), those can further change from month to month depending on the characteristics of stocks within each factor portfolio. Provided that different factor portfolios do not contain the exact same stocks, each characteristic can differ significantly across factor portfolios and can vary over time in non-standard ways. For example, it is possible that for a given month the correlation between stock momentum and value is high and hence the standardized momentum score of the two factor portfolios is similar, while in another month the correlation might be low and hence the momentum score of the two factor portfolios will be completely different. To shed light on the time-varying interrelations arising among characteristics, we display in Figures IA.4 and IA.5 the mean and standard deviation of the monthly cross-sectional correlation across all the different characteristic pairs. Apropos Figure IA.4, average cross-sectional correlations across characteristics are fairly low (typically between -0.2 and 0.2), with the exception being market friction and volatility proxies. Figure IA.5 further shows that there is significant time variability on those correlations (standard deviation values mostly range from 0.2 to 0.4), with again volatility proxies being the most profound exception, while, as expected, correlations also remain relatively constant for characteristic pairs with similar economic interpretation. Overall, our results suggest that time variation in correlations among stock characteristics is another reason for which the characteristics of the PC portfolios exhibit significant time-series variability.

The above analyses show that, despite the fact that we use the same weighting vector $w_{t,i}$ for all the months of a given in-sample period, the PC characteristics naturally vary across months *even within the same predictive iteration/in-sample period*. Obviously, an additional source of variation in the characteristics of the PCs arises from the recursive estimation of the weighting vector $w_{t,i}$ across predictive iterations. Because of the variability in the covariance structure of factor portfolio returns (and the change in their average returns in the case of RPPCA), the weighting vectors change across time, affecting the characteristic construction process, as the PC characteristics are calculated by

multiplying factor portfolio characteristics with $w_{i,t}$. To visualise how much the weighting vectors vary, Figures IA.6 and IA.7 display heatmaps that correspond to the (absolute value of the) factor loadings of each PC on the different anomalies across time, for PCA and RPPCA respectively. As expected, there is some time variability in the weights. However, we observe that in most of the cases the loadings (especially the most important ones) remain quite stable throughout the whole sample period.

Overall, the PC characteristics change across months (within the same predictive iteration) because the characteristic distributions across portfolios exhibit time-varying skewness and kurtosis and because the characteristic themselves have time-varying correlations. An additional but less important source of variation stems from the recursive estimation of PC portfolios and consequently weighting vectors.

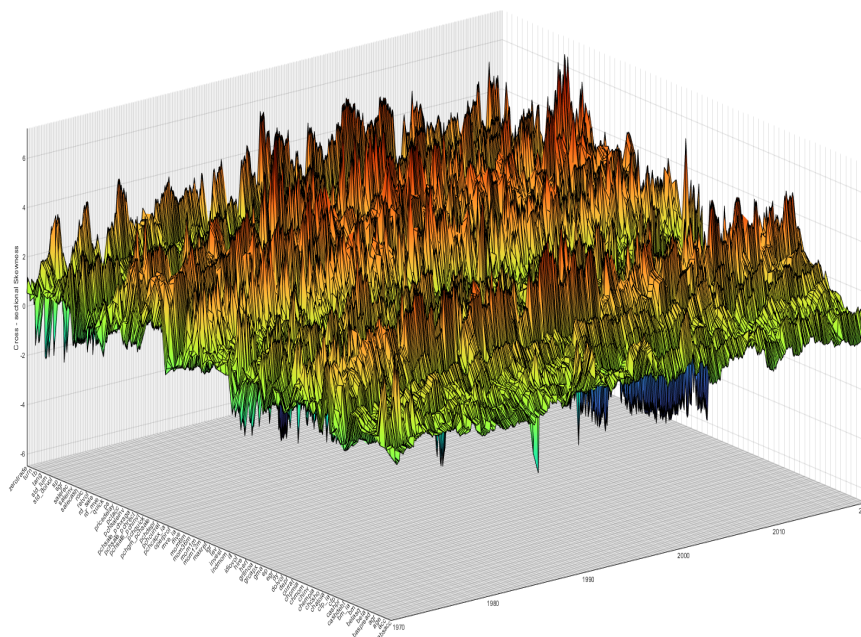


Figure IA.2: Monthly cross-sectional skewness of 72 characteristics across factor portfolios. The sample period is 01/1970-12/2019.

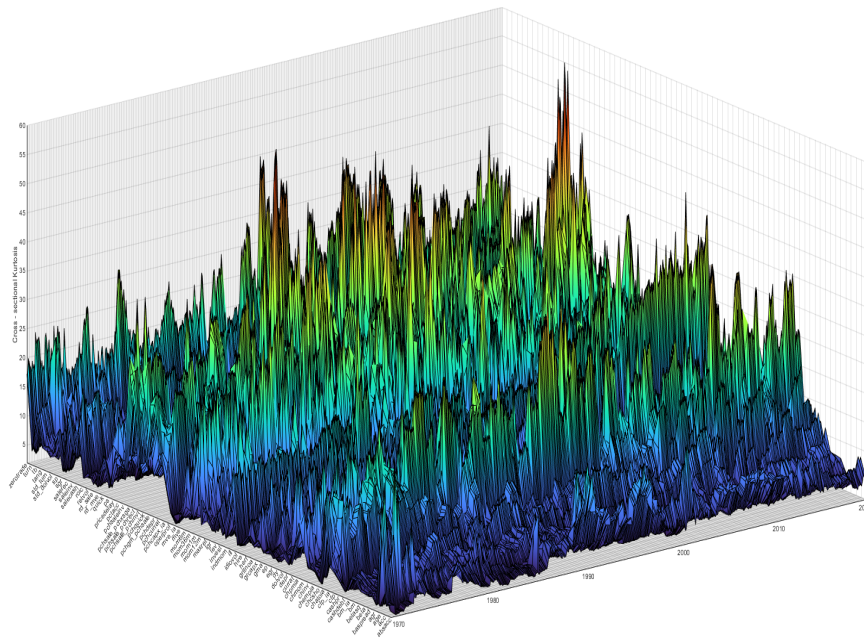


Figure IA.3: Monthly cross-sectional kurtosis of 72 characteristics across factor portfolios. The sample period is 01/1970-12/2019.

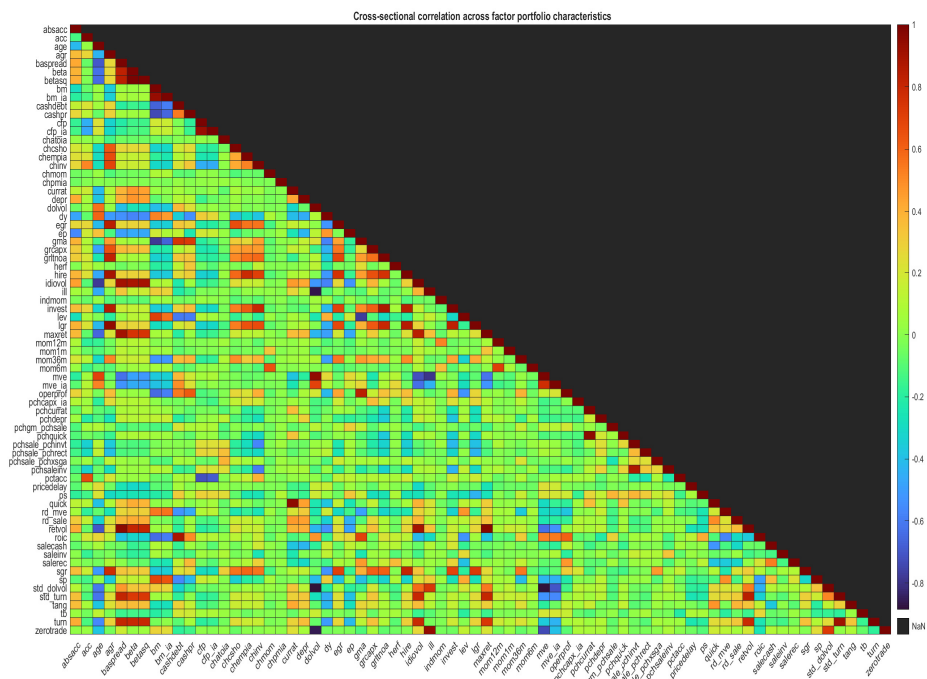


Figure IA.4: Average monthly cross-sectional correlations across 72 characteristics. The sample period is 01/1970-12/2019.

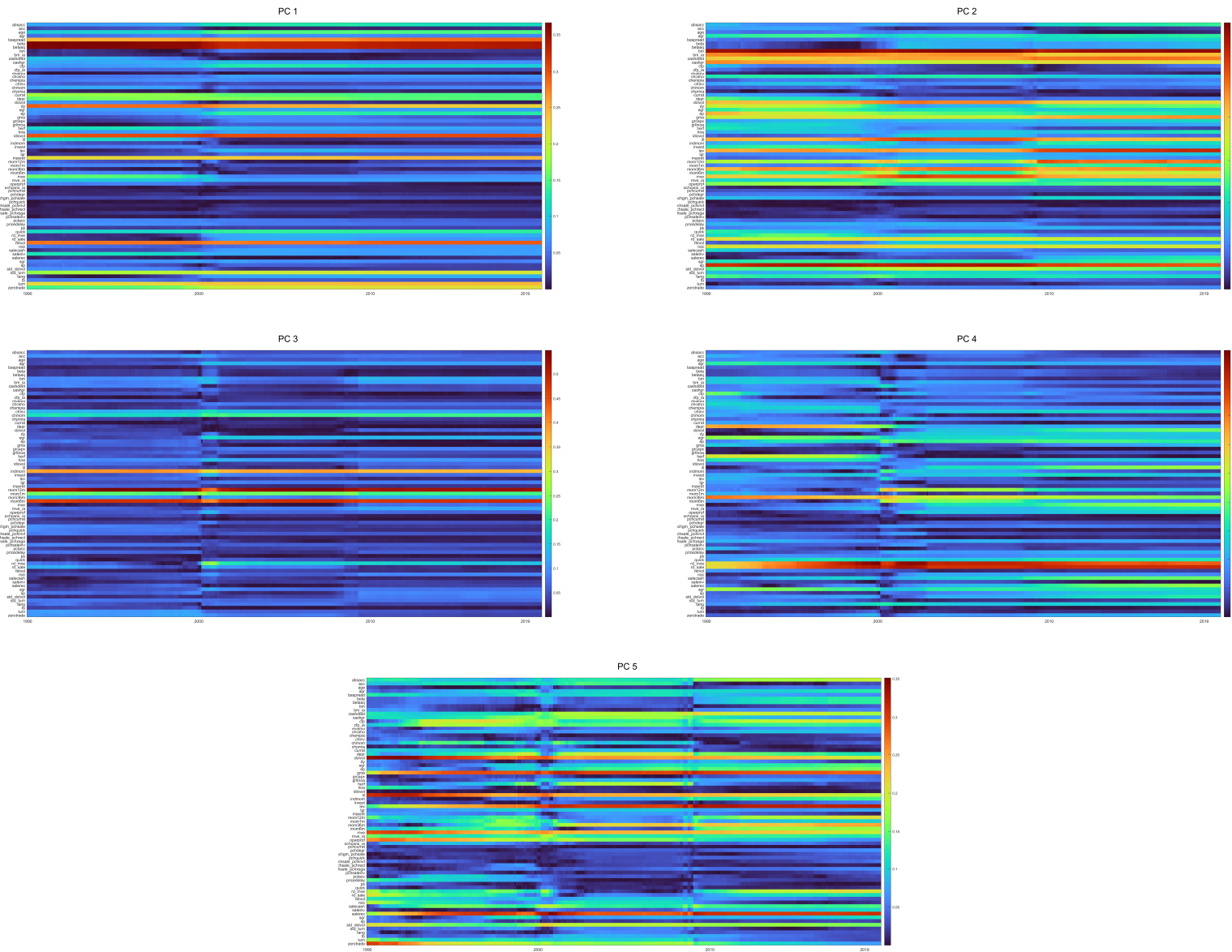


Figure IA.6: Recursively estimated PC portfolio weights on 72 factor portfolios for the out-of-sample period 01/1990-12/2019. PC portfolios are constructed using PCA.

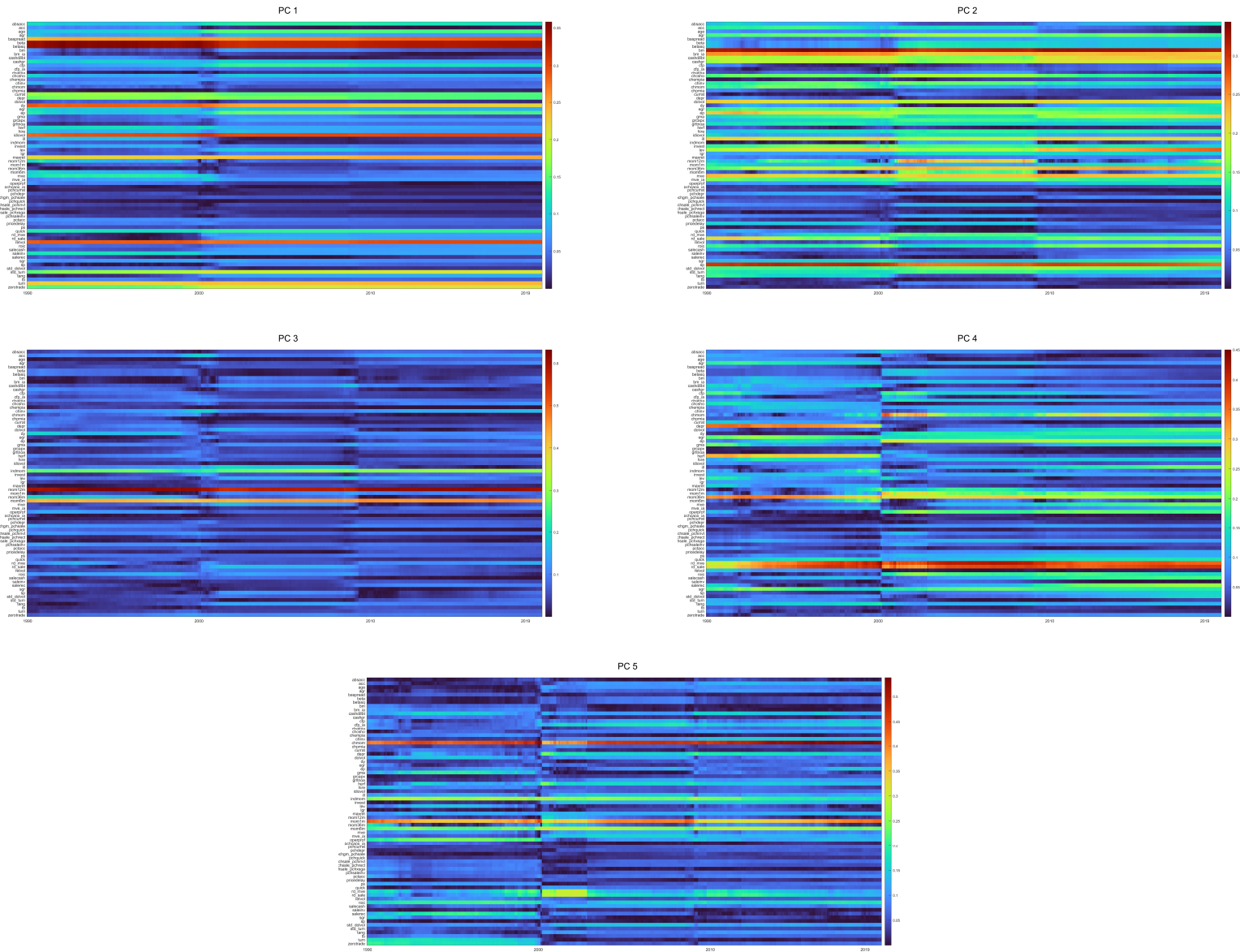


Figure IA.7: Recursively estimated PC portfolio weights on 72 factor portfolios for the out-of-sample period 01/1990-12/2019. PC portfolios are constructed using RPPCA.

IA.4 Details of the predictive models

The main analysis of the paper focuses on four models that incorporate PCA or RPPCA for condensing the variation in factor portfolio returns and PCA or PLS for condensing the variation in the portfolio characteristics. We further consider two cases, one where we retain only a single latent factor from the characteristics and another one where we retain all the characteristic factors but employ LASSO to select the important ones at each forecasting period. In addition, we select a series of alternative benchmark models that rely on predictors such as the past return of the factor portfolios, the issuer-repurchaser spread, market sentiment etc. Finally, we modify our main models in order to investigate what the most important pillars of our successful predictive method are, and also modify the benchmark models in order to examine whether adding additional statistical features to them delivers performance similar to the one provided by the portfolio characteristics. For convenience, Table IA.2 presents the list with all the different models used in the study, together with a detailed description of how we handle the left-hand-side and the right-hand-side of the forecasting problem.

IA.5 Factor database and statistics

Our empirical analysis uses 72 factor portfolios based on stock characteristics considered by Green et al. (2017). We provide a detailed description of each characteristic in Table IA.3. In particular, we present the acronym of the characteristic, the paper that identifies the characteristic, and the exact definition. In Table IA.4, we further report some summary statistics for the 72 factor portfolios that are based on our stock characteristics. Specifically, we report the average monthly return with the respective t-statistic, the monthly volatility, and the monthly Sharpe ratio of each portfolio. As discussed in the main paper, there is high cross-sectional variation in the performance of the portfolios with only 22 of them having an average return significantly different from zero.

Acronym	Left-hand-side	Right-hand-side
Panel A: Characteristic-based models (with and without LASSO)		
PCA	PCA	PCA
PCA-PLS	PCA	PLS
RPPCA	RPPCA	PCA
RPPCA-PLS	RPPCA	PLS
Panel B: Baseline Benchmark models		
1mMOM	No dimension reduction.	No dimension reduction. Each anomaly is predicted by its past month return.
12mMOM	No dimension reduction.	No dimension reduction. Each anomaly is predicted by its 12-month average past return.
PCA-BM	PCA	No dimension reduction. Each PC portfolio is predicted by its own book-to-market ratio.
IR Spread	No dimension reduction.	No dimension reduction. Each anomaly is predicted by its own issuer-repurchaser spread.
Sentiment	No dimension reduction.	No dimension reduction. Each anomaly is predicted by the investors sentiment index.
Historical Sample Mean	No dimension reduction.	No dimension reduction. Each anomaly is predicted by its in-sample average return.
Panel C: Alternative Specifications		
Anom-Own characteristic	No dimension reduction.	No dimension reduction. Each anomaly is predicted by its own characteristic spread.
Anom-1 PLS	No dimension reduction.	PLS. One characteristic-based component is retained.
Anom-All characteristics	No dimension reduction.	No dimension reduction. Each anomaly is predicted by the collection of all its characteristics.
Anom-All characteristics LASSO	No dimension reduction.	LASSO is applied on the collection of characteristics.
5 PCs-All characteristics	PCA	No dimension reduction. Each PC portfolio is predicted by the collection of its characteristics.
5 PCs-All characteristics LASSO	PCA	LASSO is applied on the collection of characteristics.
5 PCs-1mMOM	PCA	No dimension reduction. Each PC portfolio is predicted by its past return.
5 PCs-12mMOM	PCA	No dimension reduction. Each PC portfolio is predicted by its 12-month average past return.
5 PCs-5 BMs	PCA	No dimension reduction. The 5 book-to-market ratios of all PC portfolios are used to predict each one individually.
5 PCs-5 BMs LASSO	PCA	LASSO is applied on the collection of the 5 book-to-market ratios.
Anom-Own BM	No dimension reduction.	No dimension reduction. Each anomaly is predicted its own book-to-market ratio.
Anom-All IR Spreads	No dimension reduction.	No dimension reduction. Each anomaly is predicted by the collection of all issuer-repurchaser spreads.
Anom-All IR Spreads LASSO	No dimension reduction.	LASSO is applied on the collection of all issuer-repurchaser spreads.
5 PCs-IR Spread	PCA	No dimension reduction. Each PC portfolio is predicted by its own IR Spread.
5 PCs-All IR Spreads	PCA	No dimension reduction. Each PC portfolio is predicted by the collection of PC IR Spreads.
5 PCs-Sentiment	PCA	No dimension reduction. Each PC is predicted by the investor sentiment index.
Sentiment-LASSO	No dimension reduction.	LASSO is applied to the investor sentiment index.

Table IA.2: Detailed description of all the predictive models used in the study. The table includes the acronym for each model and the description of the modelling approach for each side of the return predictability exercise.

Acronym	Author(s)	Journal	Definition
absacc	Bandyopadhyay, Huang, & Wirjanto	2010, WP	Absolute value of acc.
acc	Sloan	1996, TAR	Annual income before extraordinary items (ib) minus operating cash flows (oanfc) divided by average total assets (at); if oanfc is missing then set to change in act – change in che – change in lct + change in dlc + change in txp–dp.
age	Jiang, Lee, & Zhang	2005, RAS	Number of years since first Compustat coverage.
agr	Cooper, Gulen & Schill	2008, JF	Annual percentage change in total assets (at).
baspread	Amihud & Mendelson	1989, JF	Monthly average of daily bid-ask spread divided by average of daily spread.
beta	Fama & MacBeth	1973, JPE	Estimated market beta from weekly returns and equal weighted market returns for 3 years ending month t-1 with at least 52 weeks of returns.
betasq	Fama & MacBeth	1973, JPE	Market beta squared.
bm	Rosenberg, Reid, & Lanstein	1985, JPM	Book value of equity (ceq) divided by fiscal year end market capitalization.
bm_ia	Asness, Porter & Stevens	2000, WP	Industry adjusted book-to-market ratio.
cashdebt	Ou & Penman	1989, JAE	Earnings before depreciation and extraordinary items (ib+dp) divided by avg. total liabilities (lt).
cashpr	Chandrashekar & Rao	2009, WP	Fiscal year end market capitalization plus long-term debt (dltt) minus total assets (at) divided by cash and equivalents (che).
cfp	Desai, Rajgopal & Venkatachalam	2004, TAR	Operating cash flows divided by fiscal year end market capitalization.
cfp_ia	Asness, Porter & Stevens	2000, WP	Industry adjusted cfp.
chatoia	Soliman	2008, TAR	2-digit SIC fiscal year mean adjusted change in sales (sale) divided by average total assets (at).
chcsho	Pontiff & Woodgate	2008, JF	Annual percentage change in shares outstanding (csho).
chempia	Asness, Porter & Stevens	1994, WP	Industry-adjusted change in number of employees.

Acronym	Author(s)	Journal	Definition
chinv	Thomas & Zhang	2002, RAS	Change in inventory (inv) scaled by average total assets (at).
chmom	Gettleman & Marks	2006, WP	Cumulative returns from months t-6 to t-1 minus months t-12 to t-7.
chpmia	Soliman	2008, TAR	2-digit SIC fiscal year mean adjusted change in income before extraordinary items (ib) divided by sales (sale).
currat	Ou & Penman	1989, JAE	Current assets / current liabilities.
depr	Holthausen & Larcker	1992, JAE	Depreciation over PPE.
dolvol	Chordia, Subrahmanyam, & Anshuman	2001, JFE	Natural log of trading volume times price per share from month t-2.
dy	Lanstein	1982, JF	Total dividends (dvt) divided by market capitalization at fiscal year end.
egr	Richardson, Sloan, Soliman & Tuna	2005, JAE	Annual percentage change in book value of equity (ceq).
ep	Basu	1977, JF	Annual income before extraordinary items (ib) divided by end of fiscal year market capitalization.
gma	Novy-Marx	2013, JFE	Revenues (revt) minus cost of goods sold (cogs) divided by lagged total assets (at).
grcapx	Anderson & Garcia-Feijoo	2006, JF	Percentage change in capital expenditures from year t-2 to year t.
grltnoa	Fairfield, Whisenant & Yohn	2003, TAR	Growth in long term net operating assets.
herf	Hou & Robinson	2006, JF	2-digit SIC fiscal year sales concentration (sum of squared percentage of sales in industry for each company).
hire	Bazdresch, Belo & Lin	2014, JPE	Percentage change in number of employees (emp).

Acronym	Author(s)	Journal	Definition
idiovol	Ali, Hwang, & Trombley	2003, JFE	Standard deviation of residuals of weekly returns on weekly equal weighted market returns for 3 years prior to month end.
ill	Amihud	2002, JFM	Average of daily (absolute return / dollar volume).
indmom	Moskowitz & Grinblatt	1999, JF	Equal weighted average industry 12-month returns.
invest	Chen & Zhang	2010, JF	Annual change in gross property, plant, and equipment (ppeg _t) + annual change in inventories (invt) all scaled by lagged total assets (at).
lev	Bhandari	1988, JF	Total liabilities (lt) divided by fiscal year end market capitalization.
lgr	Richardson, Sloan, Soliman & Tuna	2005, JAE	Annual percentage change in total liabilities (lt).
maxret	Bali, Cakici & Whitelaw	2011, JFE	Maximum daily return from returns during calendar month t-1.
mom12m	Jegadeesh	1990, JF	11-month cumulative returns ending one month before month end.
mom1m	Jegadeesh & Titman	1993, JF	1-month cumulative return.
mom36m	Jegadeesh & Titman	1993, JF	Cumulative returns from months t-36 to t-13.
mom6m	Jegadeesh & Titman	1993, JF	5-month cumulative returns ending one month before month end.
mve	Banz	1981, JFE	Natural log of market capitalization at end of month t-1.
mve_ia	Asness, Porter, & Stevens	2000, WP	2-digit SIC industry-adjusted fiscal year end market capitalization.
operprof	Fama & French	2015, JFE	Revenue minus cost of goods sold - SG&A expense - interest expense divided by lagged common shareholders' equity.
pchcapx_ia	Abarbanell & Bushee	1998, TAR	2-digit SIC fiscal year mean adjusted percentage change in capital expenditures (capx).

Acronym	Author(s)	Journal	Definition
pchcurrat	Ou & Penman	1989, JAE	Percentage change in currat.
pchdepr	Holthausen & Larcker	1992, JAE	Percentage change in depreciation.
pchgm_pchsale	Abarbanell & Bushee	1998, TAR	Percentage change in gross margin (sale-cogs) minus percentage change in sales (sale).
pchquick	Ou & Penman	1989, JAE	Percentage change in quick.
pchsale_pchinvt	Abarbanell & Bushee	1998, TAR	Annual percentage change in sales (sale) minus annual percentage change in inventory (invt).
pchsale_pchrect	Abarbanell & Bushee	1998, TAR	Annual percentage change in sales (sale) minus annual percentage change in receivables (rect).
pchsale_pchxsga	Abarbanell & Bushee	1998, TAR	Annual percentage change in sales (sale) minus annual percentage change in SG&A (xsga).
pchsaleinv	Ou & Penman	1989, JAE	Percentage change in saleinv.
pctacc	Hafzalla, Lundholm & Van Winkle	2011, TAR	Same as acc except that the numerator is divided by the absolute value of ib; if ib = 0 then ib set to 0.01 for denominator.
pricedelay	Hou & Moskowitz	2005, RFS	The proportion of variation in weekly returns for 36 months ending in month t explained by 4 lags of weekly market returns incremental to contemporaneous market return.
ps	Piotroski	2000, JAR	Sum of 9 indicator variables to form fundamental health score.
quick	Ou & Penman	1989, JAE	(current assets – inventory) / current liabilities.
rd_mv	Guo, Lev & Shi	2006, JBFA	R&D expense divided by end of fiscal year market capitalization.
rd_sale	Guo, Lev & Shi	2006, JBFA	R&D expense divided by sales (xrd/sale).
retvol	Ang et al.	2006, JF	Standard deviation of daily returns from month t-1.

Acronym	Author(s)	Journal	Definition
roic	Brown & Rowe	2007, WP	Annual earnings before interest and taxes (ebit) minus non-operating income (nopi) divided by non-cash enterprise value (ceq+lt-che).
salecash	Ou& Penman	1989, JAE	Annual sales divided by cash and cash equivalents.
saleinv	Ou& Penman	1989, JAE	Annual sales divided by total inventory.
salerec	Ou& Penman	1989, JAE	Annual sales divided by accounts receivable.
sgr	Lakonishok, Shleifer & Vishny	1994, JF	Annual percentage change in sales (sale).
sp	Barbee, Mukherji, & Raines	1996, FAJ	Annual revenue (sale) divided by fiscal year end market capitalization.
std_dolvol	Chordia, Subrahmanyam, & Anshuman	2001, JFE	Monthly standard deviation of daily dollar trading volume.
std_turn	Chordia, Subrahmanyam, & Anshuman	2001, JFE	Monthly standard deviation of daily share turnover.
tang	Almeida & Campello	2007, RFS	Cash holdings + $0.715 \times$ receivables + $0.547 \times$ inventory + $0.535 \times$ PPE/total assets.
tb	Lev & Nissim	2004, TAR	Tax income, calculated from current tax expense divided by maximum federal tax rate, divided by income before extraordinary items.
turn	Datar, Naik, & Radcliffe	1998, JFM	Average monthly trading volume for most recent 3 months scaled by number of shares outstanding in current month.
zerotrade	Liu	2006, JFE	Turnover weighted number of zero trading days for most recent 1 month.

Table IA.3: Listing of firm characteristics used in the study, including the source and the exact definition.

	Average Return	Standard Deviation	Sharpe Ratio	t-statistic
absacc	-0.111	3.982	-0.028	-0.680
acc	-0.419	3.009	-0.139	-3.410
age	-0.031	3.896	-0.008	-0.196
agr	-0.399	3.143	-0.127	-3.110
baspread	-0.235	6.796	-0.035	-0.846
beta	-0.153	7.854	-0.020	-0.478
betasq	-0.146	7.851	-0.019	-0.456
bm	0.377	4.480	0.084	2.061
bm_ia	0.250	3.642	0.069	1.677
cashdebt	0.101	3.884	0.026	0.635
cashpr	-0.336	3.545	-0.095	-2.323
cfp	0.298	4.181	0.071	1.744
cfp_ia	0.328	2.889	0.114	2.781
chatoia	0.269	2.647	0.102	2.490
chcsho	-0.588	2.901	-0.203	-4.963
chempia	0.065	2.874	0.023	0.553
chinv	-0.537	2.955	-0.182	-4.448
chmom	-0.574	4.628	-0.124	-3.037
chpmia	0.028	3.094	0.009	0.220
currat	-0.052	3.975	-0.013	-0.322
depr	0.058	4.376	0.013	0.322
dolvol	-0.210	3.495	-0.060	-1.471
dy	-0.060	5.823	-0.010	-0.254
egr	-0.429	3.292	-0.130	-3.191
ep	0.613	4.691	0.131	3.198
gma	0.090	3.832	0.023	0.573
grcapx	-0.358	2.941	-0.122	-2.979
grltnoa	-0.270	3.010	-0.090	-2.199
herf	0.053	3.508	0.015	0.371
hire	-0.217	3.135	-0.069	-1.694
idiovol	-0.196	6.923	-0.028	-0.692
ill	0.051	3.688	0.014	0.340
indmom	0.175	4.837	0.036	0.887
invest	-0.395	3.003	-0.132	-3.223
lev	0.098	4.550	0.022	0.529
lgr	-0.189	2.638	-0.072	-1.758
maxret	-0.367	5.778	-0.063	-1.554
mom12m	1.080	6.492	0.166	4.071
mom1m	-0.353	5.089	-0.069	-1.696
mom36m	-0.204	4.891	-0.042	-1.019
mom6m	0.624	5.896	0.106	2.590
mve	-0.161	4.015	-0.040	-0.982
mve_ia	-0.131	3.212	-0.041	-0.997
operprof	0.248	3.001	0.083	2.021
pchcapx_ia	0.066	2.932	0.022	0.549

	Average Return	Standard Deviation	Sharpe Ratio	t-statistic
pchcurrat	-0.202	1.906	-0.106	-2.597
pchdepr	0.158	2.518	0.063	1.536
pchgm_pchsale	0.125	2.723	0.046	1.126
pchquick	-0.063	2.042	-0.031	-0.761
pchsale_pchinvt	0.256	2.513	0.102	2.489
pchsale_pchrect	0.004	2.399	0.001	0.036
pchsale_pchxsga	-0.078	3.005	-0.026	-0.632
pchsaleinv	0.224	2.457	0.091	2.230
pctacc	-0.175	3.127	-0.056	-1.366
pricedelay	-0.077	2.717	-0.028	-0.695
ps	0.250	2.337	0.107	2.614
quick	-0.099	3.777	-0.026	-0.641
rd_mv	0.243	4.748	0.051	1.253
rd_sale	-0.109	4.541	-0.024	-0.585
retvol	-0.406	6.782	-0.060	-1.464
roic	0.283	3.725	0.076	1.859
salecash	0.019	3.327	0.006	0.141
saleinv	0.146	3.040	0.048	1.175
salerec	0.252	3.486	0.072	1.770
sgr	-0.112	3.374	-0.033	-0.809
sp	0.377	4.183	0.090	2.208
std_dolvol	0.175	3.080	0.057	1.391
std_turn	0.090	5.163	0.017	0.425
tang	0.155	3.423	0.045	1.105
tb	0.195	2.852	0.068	1.676
turn	-0.073	5.779	-0.013	-0.308
zerotrade	-0.052	5.454	-0.010	-0.235

Table IA.4: Descriptive statistics of factor portfolios for the sample period January 1970 to December 2019. Average Return: Average monthly return, Standard Deviation: Monthly standard deviation, Sharpe Ratio: Monthly Sharpe ratio, t-statistic: t-statistic for the null hypothesis that the average monthly return is equal to zero.

IA.6 Interpreting the PC portfolios

One disadvantage of extracting latent common factors from factor portfolio returns is that the resulting PC portfolios do not have a straightforward economic interpretation. In order to tackle this problem, we regress recursively each PC portfolio return on each of the 72 anomalies and retain the R^2 . Next, we estimate the average R^2 across months for each PC portfolio and each anomaly. As an example, we present in Figure IA.8 the

results of the PCs stemming from PCA. We observe that the first PC portfolio loads more on volatility variables such as `beta`, `idiovol`, `maxret`, `retvol` and `std_turn`. The second PC portfolio loads more on value variables such as `bm`, `bm_ia` and `sp`, as well as on profitability and leverage variables such as `gma`, `cashdebt` and `lev`. The third PC portfolio is clearly affected by the momentum characteristics, while the fourth PC portfolio is mainly driven by r&d variables. The fifth portfolio loads more on `salerec` and to a lesser extent on `gma`.

The above results stem from a time-series aggregation of the estimated R^2 s. Another interesting question that arises is whether the relation between the PCs and the underlying characteristics is stable across time. To this end, we select the most important characteristic for each PC portfolio and plot the R^2 of all of them across time in Figure IA.9. In particular, we plot the R^2 values of `beta`, `bm`, `mom12m`, `rd_sale`, and `salerec`. Each characteristic dominates a respective PC portfolio and it is evident from Figure IA.9 that all the loadings are remarkably stable across time. Overall, we conclude that the PC portfolios extracted with statistical techniques have reasonable economic interpretation and the recursive estimation does not impact negatively this interpretation.

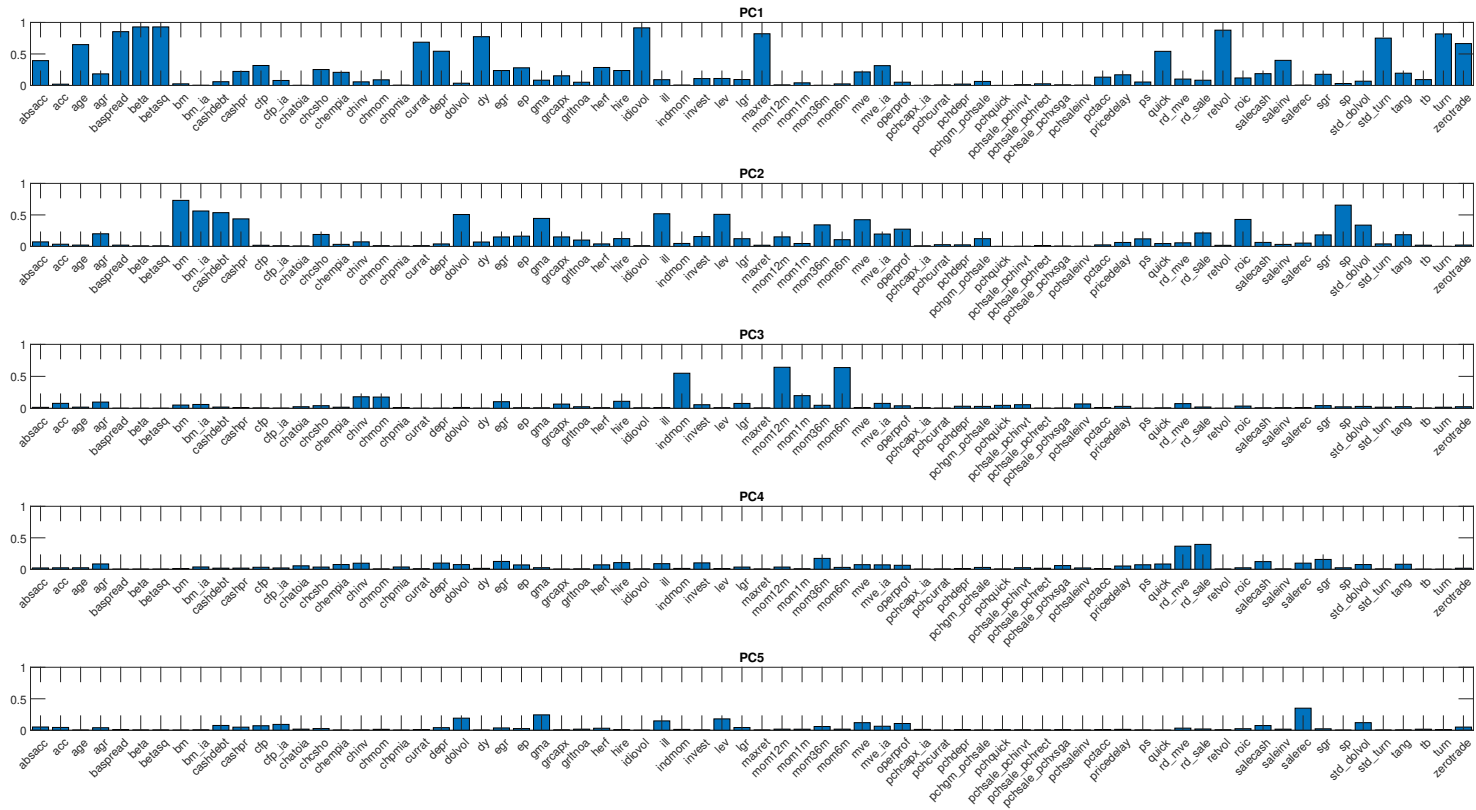


Figure IA.8: Average R^2 across month from recursive regressions of each PC portfolio on each anomaly. The PC portfolios are constructed using PCA.

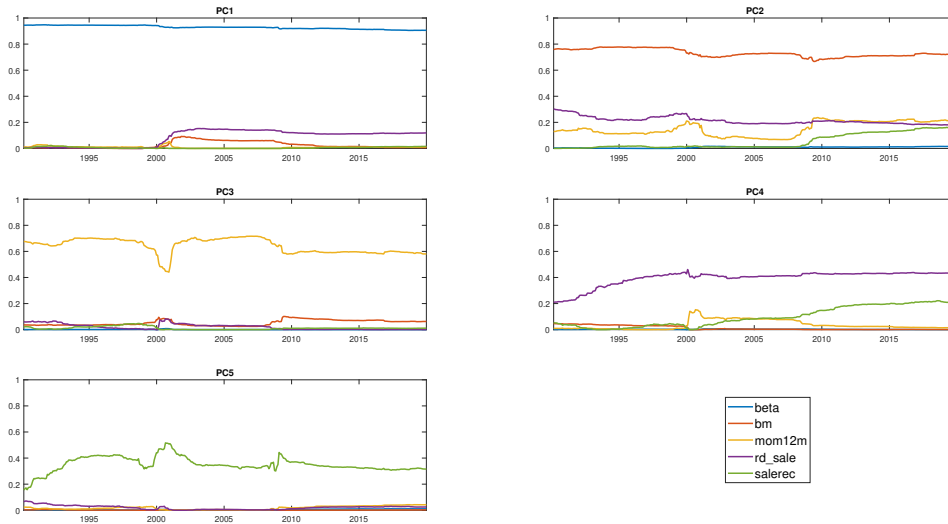


Figure IA.9: R^2 across months from a recursive regression of each PC portfolio on each of the following anomalies: beta, bm, mom12m, rd_sale, and salerec. Each anomaly loads heavily on one PC portfolio and contributes little to the rest. The PC portfolios are constructed using PCA.

IA.7 Factor timing strategy constituents

In this section we examine the trading positions of our factor timing portfolios. Specifically, using the forecasts from the LASSO-based models and the LSS, we investigate how often each anomaly is traded. The anomalies traded under LSS are also the ones with the highest absolute weights under TSFM and CSFM, so focusing on this case only is representative of the general investing approach. Figure IA.10 displays the percentage trading frequency of each anomaly. Blue bars imply long and orange bars imply short positions. Clearly, the benefits of factor timing strategies arise from rotating among multiple anomalies and not by focusing on a handful of picks. Yet, all strategies tend to go long on anomalies with high average returns, such as mom6m and mom12m, and short anomalies with negative average returns, like beta, chmom and retvol. Furthermore, some anomalies appear almost equally often in the short and the long legs of our factor timing portfolio, implying higher volatility in conditional returns. For the models that use PCA, these anomalies are usually related to market frictions, which also have higher volatility and thus load more heavily on the first PCs. When RPPCA is used for the left-hand-side, the anomalies more regularly traded are those with higher absolute aver-

age returns. Finally, anomalies that do not load heavily on the dominant components, using either PCA or RPPCA, stay out of the investable universe as their small loadings compress their individual return forecasts close to zero. Consequently, the use of PCs for the left-hand-side has an impact on the factor timing portfolio formation.

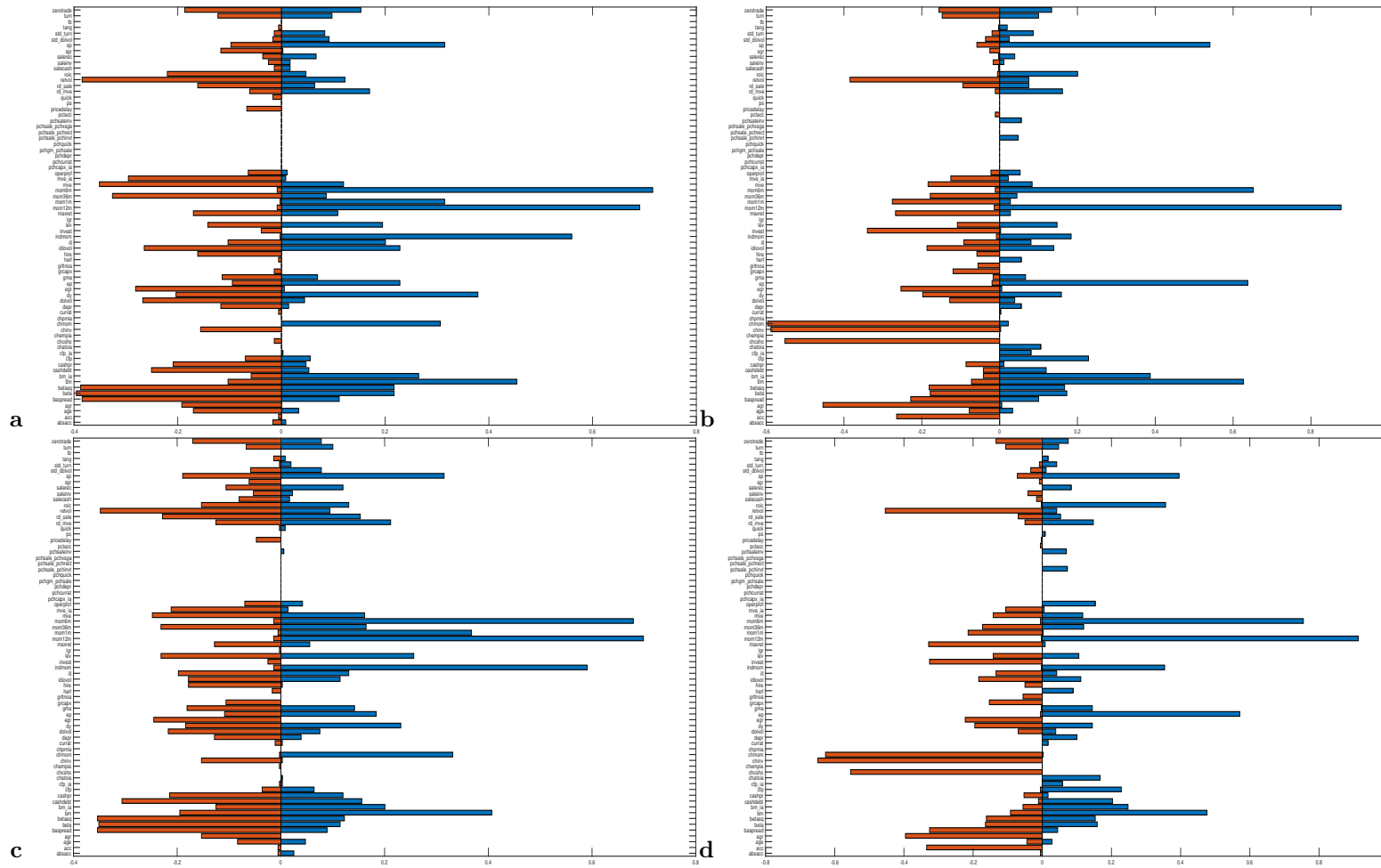


Figure IA.10: Percentage frequency of LSS constituents over the out-of-sample period. Blue bars suggest long and orange bars suggest short positions. (a) PCA (b) RPPCA (c) PCA-PLS (d) RPPCA-PLS.

References

- De Jong, S. (1993), ‘Simpls: an alternative approach to partial least squares regression’, *Chemometrics and intelligent laboratory systems* **18**(3), 251–263.
- Friedman, J., Hastie, T. and Tibshirani, R. (2010), ‘Regularization paths for generalized linear models via coordinate descent’, *Journal of Statistical Software* **33**(1), 1.
- Green, J., Hand, J. R. and Zhang, X. F. (2017), ‘The characteristics that provide independent information about average U.S. monthly stock returns’, *Review of Financial Studies* **30**(12), 4389–4436.
- Haddad, V., Kozak, S. and Santosh, S. (2020), ‘Factor timing’, *Review of Financial Studies* **33**(5), 1980–2018.
- Lettau, M. and Pelger, M. (2020), ‘Estimating latent asset-pricing factors’, *Journal of Econometrics* **218**(1), 1–31.
- Onatski, A. (2012), ‘Asymptotics of the principal components estimator of large factor models with weakly influential factors’, *Journal of Econometrics* **168**(2), 244–258.
- Ross, S. (1976), ‘The arbitrage theory of capital asset pricing’, *Journal of Economic Theory* **13**(3), 341–360.