Factor Timing with Portfolio Characteristics*

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

January 13, 2022

Abstract

Factor momentum has formed the basis of factor timing strategies. 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 the number of predictors and portfolios to predict. Characteristic-based models outperform factor momentum 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 Winfried Pohlmeier, Harald Lohre, and participants at the 2nd Frontiers of Factor Investing Conference and the 6th KoLaMaFr - Workshop on Financial Econometrics 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 characteristic sorted portfolios. A zero-investment, long-short (LS) factor portfolio is created by buying and selling stocks with extreme characteristic scores and the excess return of such a portfolio is directly associated with the risk premium of the underlying risk factor (Fama and French, 1993). It is still debatable whether positive returns arising from such LS strategies reflect legitimate investment opportunities or are the result of sample and methodological alternations (Hou et al., 2020). Nevertheless, such zero-investment, market-neutral portfolios have given rise to factor investing as they are easily tradable. 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 consistent across portfolios, allowing for substantial investment gains from timing factor portfolio returns.[1] Hence, from an investor's perspective timing is also 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 utilising the momentum in factor portfolio returns as a way to improve on static factor investing. Avramov et al. (2017), Arnott et al. (2021), Gupta and Kelly (2019), Ehsani and Linnainmaa (2021) and Leippold and Yang (2021), all implement different variations of the factor momentum strategy by going long or short anomalies based on their recent performance. Such strategies are profitable due to the persistence in anomaly returns and their strong autocorrelation

[1]For 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.

structure. Nonetheless, it is still unclear whether the profitability of such strategies is driven by a momentum effect in factor portfolio returns or simply by the difference in the mean returns of the factors. For instance, Leippold and Yang (2021) show that the sizable factor premia associated with factor momentum stem from holding factors with significant average returns in the first place. Hence, the recent critique that has risen on factor momentum has opened the way for alternative ways to time the performance of factor portfolios.

In this paper, we create an optimal factor timing strategy, going over and above what momentum has to offer. 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. Given the characteristic-return relationship that is implied by the significant factor premia, it is only natural to examine this relationship in terms of exact predictive accuracy. 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. Hence, we 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. The use of portfolios instead of individual stocks leads to a more stable risk exposure over time as stocks possess an idiosyncratic component which fades when they are concentrated into diversified portfolios. Establishing return predictability in a factor portfolio context has important implications, not only in terms of timing those portfolios, but also in terms of understanding the dynamic properties of the cross-section. 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.

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). The first five PCs capture around 67% of the variation in factor portfolio returns (see Figure A1 in the 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 would lead to the detection of robust predictive patterns across individual anomalies. Applying Principal Component Analysis (PCA) to a set of factor portfolio returns to reduce their dimensions has recently gained a lot of attention in asset pricing, with the most prominent example perhaps being that of Haddad et al. (2020), who form PC portfolios by running PCA on a set of 50 anomalies and use their own book-to-market ratio to predict their performance. In such a context, PCs are not just statistical factors but have an investable interpretation as well. Specifically, as every PC is a linear combination of the underlying variables, PC portfolios are portfolios of factor portfolios, meaning that their returns as well as their characteristics are calculable. To calculate the returns and characteristics of PC portfolios we use conventional PCA as well as the Risk Premium PCA (RPPCA) proposed by Lettau and Pelger (2020*a*). PCA extracts components that explain the variation in factor portfolio returns, while RPPCA utilises information in the mean returns of the factor portfolios on top of their variance and leads to the extraction of 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 they help us guide 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). Given that returns possess a sizable unpredictable component and many characteristics end up being unimportant in terms of prediction, only a small fraction of the variation in the characteristic has predictive value. Conventional PCA focuses on the variance within the predictors and can lead to components that mix return-relevant and irrelevant variation. As a result, a large number of components may be required to make predictions, where each component only makes a marginal contribution. 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. Apart from reducing the dimensions of the predictors by compressing their variation into a smaller set of factors, the use of PCA and PLS also account for any multicolinearity issues associated with raw characteristics, since the resulting components are uncorrelated to each other.

After rotating characteristics in space using the methods mentioned above, we either use the first characteristic component in standard predictive regressions or apply LASSO on the whole set of components to identify the relevant subset of features for predicting PC portfolio returns. The first case is used to investigate the predictability that arises from observed characteristics even in the simplest case of a single predictive factor. The use of LASSO allows for successive components to be included in the surviving subset of predictors as the importance of each characteristic-based component is 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-based 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 coefficient shrinkage across time allows us to examine the time variation in the strength of the characteristic signal overall.

Our findings show that characteristics are extremely useful in timing PC portfolio returns and ultimately individual anomalies. Using a collection of 72 anomalies documented in the literature for the period 1970 to 2019, we find that the majority of them are highly predictable by the information contained in their characteristics. Moreover, most of the anomalies that are not predictable have low covariance with the rest of the anomaly universe and insignificant average returns, implying a lack of risk premium in the first place. However, the underlying characteristics of these anomalies can still be important in explaining dynamics of other factor portfolios. Hence, we assess the information usefulness of characteristics in a collective way rather than examining each of them individually as other studies have done. Furthermore, we distinguish predictive ability in terms of exact predictive accuracy (predicting individual returns in exact terms) and the ability the predict the cross-sectional dispersion in factor portfolio returns (differentiating winners from losers). We find the characteristic-based models outperform factor momentum in both terms. More concretely, characteristic-based models generate smaller forecasting errors and result in higher cross-sectional correlations between forecasted and realized returns, compared to factor momentum. With regards to factor momentum in particular, we find its performance to be significantly time-varying. Specifically, the momentum signal in factor portfolios was strong during the 90s, though it has diminished considerably in recent years. In addition, although the 1-month momentum signal is the strongest overall, according to prior literature (e.g. Gupta and Kelly (2019)), we find that the 12-month signal actually performs better after 2010.

In terms of the different methods used, the implications of using of PCA or RPPCA to reduce the number of portfolios to predict are minimal. Naturally, PCA focuses on portfolios with higher variance, while RPPCA also focuses on portfolios with high average returns. As a result, the former leads to models with slightly smaller forecasting errors while the later leads to investment strategies with slightly higher Sharpe ratios, though the differences are insignificant. Yet, when it comes to reducing the number of predictors, the dimension reduction technique matters, as PLS outperforms PCA when a single predictor is used. Even though the difference is not visible in terms of total forecasting error, a single PCA-based predictor results in model forecasts that do not capture any cross-sectional dispersion in factor portfolio returns. The reason is that many characteristics that have been documented in the literature end up having no predictive value. Hence, a single PCA-based predictor captures variation that is irrelevant in return prediction.

Nonetheless, no difference is observed between PCA and PLS when multiple components are considered in conjunction with LASSO, suggesting that once we account for the whole information set the rotation method becomes unimportant. After employing LASSO results improve uniformly across models, reflecting the importance of accounting for further components as well as the benefits of regularization in dealing with over-fitting. Combining LASSO with PCA or PLS also accounts for the limitations of LASSO in the presence of correlated predictors, since the resulting variables are orthogonal. Finally, the use of LASSO uncovers some interesting patterns about the characteristics-returns relationship. Specifically, the degree of coefficient shrinkage, or the required number of features, varies significantly across time for all the PC portfolios. Hence, we observe that characteristics work better in predicting returns in certain periods than others, which is expected given the time-variation in factor risk premia. In some extreme cases, the number of features goes down to zero, suggesting that at times characteristics may not convey any information at all. Even so, our factor timing strategies are flexible enough to downgrade information in the characteristics when their informativeness is low. Using three different weighting schemes we find large economic gains from characteristic-based model forecasts in terms of average returns and Sharpe ratios. For example, characteristic-based model deliver an annualized Sharpe ratio of up to 0.73, while the best factor momentum strategy only earns a Sharpe ratio of 0.45. Finally, factor timing strategies based on characteristics show no decay in return performance over time, although many individual anomalies have been found empirically to do so (McLean and Pontiff, 2016).

The rest of the paper is structured as follows: Section 2 provides a comprehensive literature review on asset pricing models, dimension reduction techniques and factor portfolio predictability. Section 3 describes the different models used and our estimation approach. Section 4 provides an assessment of the various models in terms of forecasting ability and investment performance and Section 5 concludes.

2 Literature review

Returns and characteristics

Our paper is related to several strands of the literature. First of all, we build on the literature modeling asset returns as a function of observed characteristics. Traditional methods in the literature of asset return prediction usually involve cross-sectional or time-series regressions of future returns on a small set of lagged stock and aggregate market characteristics. The cross-sectional approach is motivated by evidence provided by Fama and MacBeth (1973), who show that average stock returns are associated with firm characteristics. Empirical applications of the Fama-MacBeth procedure include Fama and French (2008), Lewellen (2015) and Green et al. (2017), among others, who examine the joint predictability of multiple observed characteristics. The time-series approach has been exemplified by Welch and Goyal (2008) and Rapach and Zhou (2013), who use a large collection of economic variables to predict the excess return of the US stock market. Interestingly, both studies find that conventional predictive regressions fail to provide reliable out-of-sample performance. We enrich the current setting by examining the predictability of factor portfolios instead of individual stocks, using a large collection of portfolio characteristics as predictors. After transforming the factor portfolios and the characteristic into components, we find that a predictive regression approach can be fruitful under a regularized framework.

Dimension reduction

Secondly, we build on an emerging strand of the asset pricing literature that employs machine learning methods to deal with the high dimensional zoo of factors. Machine learning has surfaced in recent years in various asset pricing applications due to the limitations of the standard methodologies in a high dimensional setting. Gu et al. (2020) compare various machine learning techniques in their effort to forecast US stock returns using a large collection of stock characteristics. Similarly, a large number of studies try to identify the extent to which stock characteristics are associated with expected returns by regularizing the cross-sectional regressions or the characteristic-based portfolio sorts used in the estimation of risk premia. For instance, DeMiguel et al. (2020), Freyberger et al. (2020) and Feng et al. (2020) employ LASSO (L^1 penalty) regularization to create a stochastic-discount-factor (SDF) with sparse characteristic exposure, with all confirming a high degree of redundancy among characteristics. 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 on the characteristic PCs.

We also build on a strand of the literature that is applying PCA on a set of stock or portfolio returns to reduce their dimensions. Early empirical contributors to this literature include Connor and Korajczyk (1988), who apply asymptotic PCA of Connor and Korajczyk (1986) on asset returns to extract the latent factors. More recent examples of PCA applications in asset pricing include 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 in explaining returns (Kozak et al., 2018). Still, this argument does not explicate whether high eigenvalue PCs reflect risk or mispricing. 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 on this literature by constructing PC portfolios of LS portfolios and 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.[2] They find that only a small number of characteristic-based factors are important in identifying a low dimensional 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 to extract five factors from LS portfolio returns.

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

Factor portfolio predictability

Finally, we build on studies that look into the potential predictability of the factor portfolios. Recently, factor momentum has been a workhorse in timing factor portfolio 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). Furthermore, 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, 2021). 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 momentum, numerous studies have attempted to predict the performance of factor portfolios using a collection of potential predictors. Daniel and Moskowitz (2016) forecast stock momentum using market indicators and volatility proxies, in their effort to explain momentum crashes. 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. Similarly, 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 with previous studies that target only specific anomalies, we examine factor portfolio predictability across a large set of factor portfolios.

Other studies have also examined the predictability of multiple portfolios at once. Stambaugh et al. (2012) find that LS strategies appear to be stronger following periods of high sentiment, with the effect being concentrated on the short leg. Kelly and Pruitt (2013) forecast four sets of characteristic sorted portfolios using the cross-section of book-to-market ratios and observe higher predictability for 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 expand the existing framework by incorporating information across a large collection 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 presents our forecasting approach and the statistical methods used in this study. We begin by explaining our general forecasting procedure and each subsequent subsection introduces a new method and provides a comprehensive overview of its functional form and statistical properties.

General forecasting procedure

The main objective is to predict a large set of factor portfolio returns using a large set of portfolio characteristics. Instead of separately predicting each factor portfolio by its collection of characteristics we focus on the dominant components of factor portfolio returns. Let R be a $(T \times N)$ matrix of N factor portfolio returns. Equivalently, let $R_{t,\cdot} = (R_{t,1}, \ldots, R_{t,N})$ be a $(1 \times N)$ vector of portfolio returns at time t and $R_{\cdot,n} = (R_{1,n}, \ldots, R_{T,n})$ be a time-series of excess returns of the n^{th} factor portfolio. Assuming a linear latent factor specification, excess asset returns can be expressed as:

$$R = \mathrm{Z}_K W_K' + E, \tag{1}$$

where Z_K is a $(T \times K)$ matrix of factor returns, W_K is a $(N \times K)$ matrix of factor loadings and E is a is a $(T \times N)$ matrix of idiosyncratic errors. The first term of the right-hand-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 in current finance literature is to estimate Z_K and W_K 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_K allows us to form forecasts for individual anomalies through W_K. By only focusing on Z_K, we isolate common sources of predictability across factor portfolios and ignore spurious predictability associated with smaller PCs. Hence, consider the eigenvalue decomposition of the variance-covariance matrix of factor portfolio returns $Var(R) = W \Lambda W'$, where W is a $(N \times N)$ matrix of eigenvectors and Λ is a $(N \times N)$ diagonal matrix of eigenvalues in decreasing order. The i^{th} $i = 1, \ldots, K$ PC portfolio is then calculated as $z_i = R w_i$, where w_i is the i^{th} column of W.

Applying PCA to the variance-covariance matrix of factor portfolio returns implicitly assumes that the mean of R is equal to zero (zero intercept no arbitrage restriction). However, this assumption can be restrictive as the mean returns of factor portfolios can contain valuable information about the underlying factor structure. More broadly, low variance components may still be important in an asset pricing context and using a variance-based criterion may not result in the extraction of true factors. We therefore also use RPPCA to extract the latent asset pricing factors Z_K. A comparison is made between the two methods in terms of predictability and investment performance.

The first decision being made is on the optimal number of factors in (1). Specifying the optimal number of factors 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 that perform PCA on a set of factor portfolios, e.g. Haddad et al. (2020) and Lettau and Pelger (2020*b*). Hence, let $Z_5 = (z_1, z_2, \ldots, z_5)$ be the set of the 5 largest PC portfolios.

In order to forecast $z_{t+1,i}$, we model PC portfolio returns as a function of observable characteristics. Specifically, lagged characteristic scores 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_i. Let C^t be a $(N \times M)$ matrix of M characteristics for N factor portfolios at time t. The cross-section of characteristics for the i^{th} $i = 1, \ldots, 5$ PC portfolio is calculated as $H^i_{t,.} = w_i' C^t$. Repeating the process for every t 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 components using PCA and PLS. This enables us to filter out any multicolinearity associated with raw characteristics and reduce the dimensions of the problem even further by compressing the information contained in the characteristics into a handful of latent factors. Hence, let X^i be a $(T \times M)$ matrix of linear combinations of characteristics of the i^{th} PC portfolio. For PCA, H^i is rotated into X^i based on the eigendecomposition 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 X^i under the different methods is provided later in the section. 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. As long as most of the variation in the characteristics explains PC portfolio returns, no significant difference should arise between the two.

The next step is to identify the most informative characteristic components for predicting PC portfolio returns. Here, we examine two different cases. In the first case, we take a simple stance by using only the first characteristic component of each PC portfolio in standard bivariate predictive regressions. Though this is the sparsest specification possible, multiple characteristics can have an effect on PC portfolio returns through their weight on the first characteristic PC. As an alternative, we apply LASSO on the whole set of characteristic components for each PC portfolio to identify a subset that is useful for our forecasting objective. The optimal amount of coefficient shrinkage is selected by conducting cross-validation on a rolling basis. It is important to highlight that LASSO will not necessarily retain high eigenvalue characteristic PCs. Low eigenvalue characteristic PCs can have non-zero coefficients as long as they contribute to minimizing the forecasting error in the validation period. More details about our LASSO procedure can be found later in this section. Hence, the general forecasting model can be written as:

$$\hat{z}_{t+1,i} = \beta_{i,0} + \beta_{i,m} X^{i}_{t,m} + \epsilon_{t+1,i}, \tag{2}$$

where in the first case we use OLS and a single predictive factor ($m = 1$) and in the second case the βs are obtained through LASSO for $m = 1, \ldots, M$. PC portfolio forecasts are then extended to individual anomalies using their loadings on the dominant components such as,

$$\hat{R}_{t+1,.} = \sum_{i=1}^{5} w_i \hat{z}_{t+1,i}. \tag{3}$$

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 PCA or RPPCA and in the number of predictors with the use of PCA or PLS. All model combinations are estimated using either a single or multiple predictors (via LASSO), resulting in a total of eight forecasting models. 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.

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

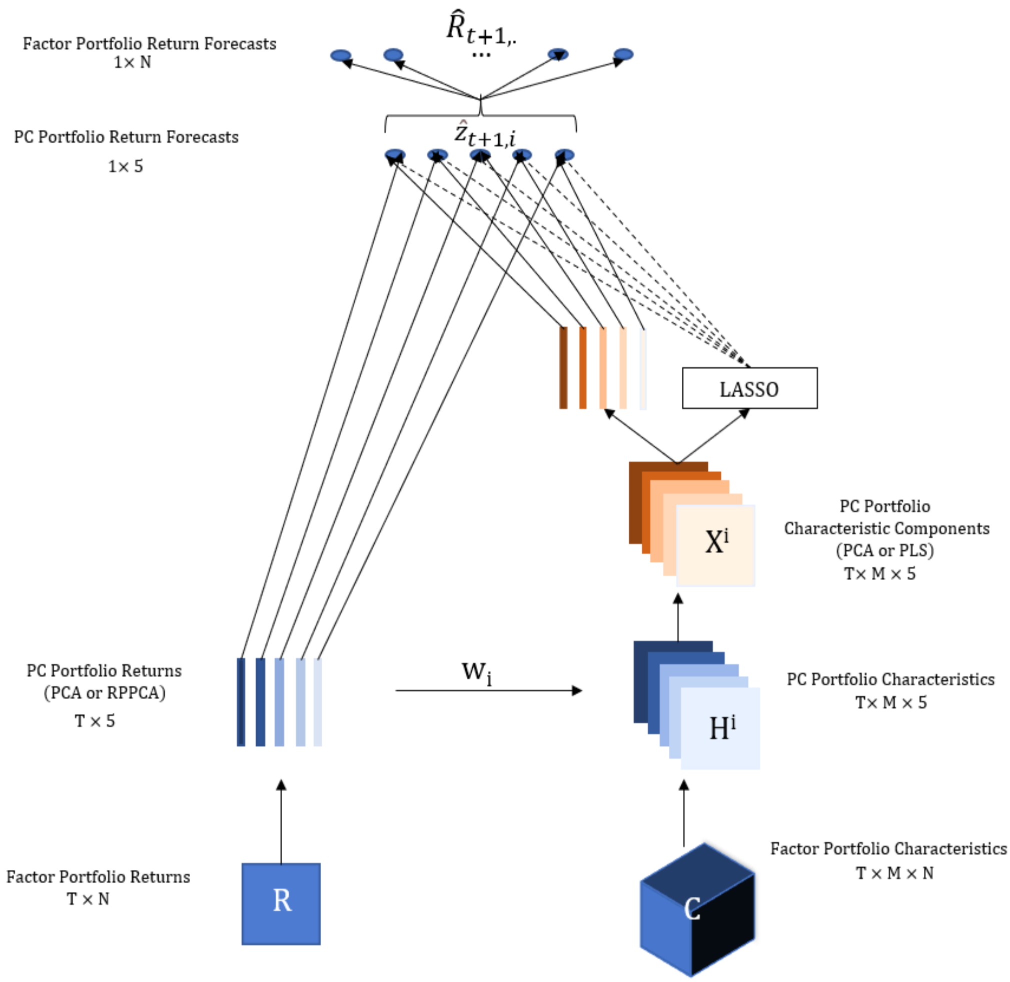

Figure 1: Visual depiction of our modelling procedure. The figure presents the process of forecasting factor portfolio returns using their portfolio characteristics. PC portfolios are calculated as linear combinations of factor portfolios. The same weighting vectors are used to decompose the three-dimensional 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 the 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.

Model construction and number of factors via validation

We use at least 20 years (240 months) of information to estimate the (RP) PC portfolios and their characteristics and 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_i $i = 1, \ldots, 5$ based on the in-sample variance-covariance matrix of factor portfolio returns. Notice that PC portfolio characteristics H^i do not only change because of the change in the underlying factor portfolio characteristics C^m but because of the change in the weighting vectors w_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 predictors is obtained as follows; for PCA, which only utilizes information contained in the characteristics to extract the latent factors, characteristic scores 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 βs in (2) 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 (2) to obtain forecasts for each PC portfolio returns at $t+1$. Hence, our forecasts are completely out of sample and do not suffer from any look ahead bias.

Another important aspect of our estimation procedure is the use of LASSO to account for over-fitting 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,

$$\min_{\beta \in \mathbb{R}^M} \left\{ \frac{1}{T} \|z_i - X^i \beta\|_2^2 - \lambda \|\beta\|_1 \right\}. \tag{4}$$

Identifying the correct value for the penalty parameter λ is critical to the performance of our models. We select λ through a validation sample by conducting cross-validation on a rolling basis. Specifically, before moving to the forecasting step, we further 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 should deliver reliable out-of-sample performance. The validation sample covers the last five years (60 months) of the in-sample period while the training sample increases by one at each iteration. At the start, the training sample is used to forecast the first period in the validation sample subject to a geometric sequence of λ values. The sequence of λ values is strictly positive and terminates at a value for which all coefficients are equal to zero. Hence, our approach examines the possibility that none of the characteristic components is relevant in predicting PC portfolio returns, in which case returns forecasts shrink down to a constant term. 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 60 times for every iteration, we pick the value of λ that minimizes the mean-squared error in the validation sample. We then re-estimate the PC portfolios and characteristic PCs using the whole in-sample period (estimation and validation) and apply LASSO using the fixed value of λ to estimate $\beta_{i,m}$ and predict PC portfolios at $t+1$.

Notice, that LASSO is applied separately on each PC portfolio, meaning that λ and hence 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 characteristics are only useful in predicting some of them. Furthermore, since LASSO is applied iteratively, λ can also vary across time for each PC portfolio, allowing for a time-varying number of factors depending on how strong the characteristic signal has been in the recent past.

Evidently, there are some clear benefits from combining LASSO with PCA or PLS. First and foremost, applying LASSO on a set of components instead of raw portfolio characteristics removes any multicolinearity concerns that would result in inconsistent solutions as it uses orthogonal variables. Secondly and perhaps more importantly, even though the method imposes a parsimonious specification, it is still flexible enough to incorporate information from multiple characteristics through their loadings on the characteristic components.

3.1 Dimension reduction techniques

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 variance-covariance matrix of the factor portfolio returns R. Consider the eigendecomposition of Σ:

$$\Sigma = W\Lambda W' = \sum_{i=1}^{N} \lambda_i w_i w_i', \tag{5}$$

The i^{th} eigenvector w_i, solves:

$$\begin{gathered} w_1 = \underset{\|w_1\|=1}{\text{argmax}} \left\{ w_1' \Sigma w_1 \right\}, \\ w_2 = \underset{\|w_2\|=1}{\text{argmax}} \left\{ w_2' \Sigma w_2 \right\} \text{ s.t. } w_1' \Sigma w_2 = 0, \\ \vdots \\ w_N = \underset{\|w_N\|=1}{\text{argmax}} \left\{ w_N' \Sigma w_N \right\} \quad \text{s.t.} \quad w_M' \Sigma w_N = 0 \quad \forall \quad M < N. \end{gathered} \tag{6}$$

Practically, the solution in (6) is obtained via a singular value decomposition (SVD) of R. The PCs are 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. Let $Q^i = q^i_1, \ldots, q^i_M$ be the set of eigenvectors of the variance-covariance matrix of H^i and $X^i = H^i Q^i$ be a $(T \times M)$ matrix of characteristic PCs of the i^{th} PC portfolio. Since characteristics are of different scale, running raw PCA on H^i 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^m cross-sectionally before calculating H^i and ultimately X^i.

Risk Premium PCA (RPPCA)

In general, PCA can be used to obtain factors that best explain time-series variation in the data. The variance-covariance matrix of factor portfolio returns can also be rewritten 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. However, some factors may have weak explanatory power in terms of variance if they only affect a small proportion of assets, but still be important in an asset pricing context. In this case, conventional PCA is unable to detect the true factors (Onatski, 2012). Under an APT framework, exposure to systemic risk factors should be able to explain the cross-section of expected asset returns (Ross, 1976). Hence, latent factors should be able to simultaneously capture time-series variation and explain the cross-section of average returns.

Lettau and Pelger (2020a) 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:

$$\frac{1}{T}R'R + \gamma\bar{R}\bar{R}'. \tag{7}$$

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.[3] 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. Factor portfolios with insignificant returns are ultimately unimportant as forecasting their performance cannot translate into investment gains. However, PCA-based PC portfolios can load heavily on factor portfolios with insignificant average returns, provided that they have high variance. Hence, the use of RPPCA allows us concentrate on factor portfolios with superior return performance, even if they possess low volatility.

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, $i = 1, \ldots, 5$. 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. Instead, we apply SDV on each $\frac{1}{T}H^{i\prime}H - \bar{H}^i\bar{H}^{i\prime}$, which converges back to conventional PCA.

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

Partial Least Squares (PLS)

As already discussed, 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. The combination weights for the i^{th} PC are estimated recursively by solving:

$$
\begin{aligned}
q_1^i &= \underset{\|q_1^i\|=1}{\operatorname{argmax}} \left\{ {q_1^i}' H^{i\prime} z_i zi' H^i q_1^i \right\}, \\
q_2^i &= \underset{\|q_2^i\|=1}{\operatorname{argmax}} \left\{ {q_2^i}' H^{i\prime} z_i z_i' H^i q_2^i \right\} \text{ s.t. } {q_1^i}' H^{i\prime} z_i z_i' H^i q_2^i = 0, \\
&\vdots \\
q_N^i &= \underset{\|q_N^i\|=1}{\operatorname{argmax}} \left\{ {q_N^i}' H^{i\prime} z_i z_i' H^i q_N^i \right\} \quad \text{s.t.} \quad {q_M^i}' H^{i\prime} z_i z_i' H^i q_N^i = 0 \quad \forall \quad M < N.
\end{aligned} \tag{8}
$$

Equation (8) highlights the distinction between PLS and PCA. Specifically, by making a comparison between (6) and (8) it is clear that PCA finds linear combinations that maximize the variance of H^i while PLS finds combination weights that maximize the squared covariance between z_i and H^i. 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 (8) 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 $\hat{z}_{t+1,i}$.

3.2 Benchmark models

To examine whether characteristic-based models provide superior information compared to factor momentum, we use two different variations of the momentum signal to compare our models against. 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 is 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 12 months. In this case, the prediction for the return at time $t + 1$ is the average monthly return of the previous twelve months. Although there are multiple variations of factor momentum in terms of the formation period, we believe that these two cases are the most prominent examples. Still, the above methods provide separate anomaly forecasts, while the proposed models only do so indirectly by predicting PC portfolio returns. In order to improve consistency across characteristic and momentum models, we also apply both momentum strategies to the (RP) PC portfolios and then extend the forecasts to individual anomalies as in (3). Hence, we also examine the possibility of a stronger momentum effect on the main sources of variation of factor portfolio returns.[4]

4 Empirical results

4.1 Data

We replicate 72 characteristics that have also been considered by Green et al. (2017). The characteristics are entirely calculated from the Center of Research on Securities (CRSP) and Compustat data. Our data set covers the 50-year period from January 1970 to December 2019. The sample includes most of the well-documented anomalies and therefore minimizes the risk of portfolio selection bias. The stock universe includes common

[4] For instance, Ehsani and Linnainmaa (2021) observe that momentum is highly concentrated among the first five PC-portfolios.

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. We do not consider any I/B/E/S-related anomalies due to the large volume of missing data in the early years of the sample period. Additional information about the characteristics can be found in Table A1 in the Appendix, including origination and characteristic description.

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 10 and short decile 1 even if the characteristic 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. Hence, strategies with a negative risk premium, such as return reversal, should on average be allocated in the short side of our factor timing portfolio. Nonetheless, in some cases, there may be not enough diversity in characteristic values to group stocks into deciles. This turns out to be the case for discrete variables, like firm age in the early years of the sample period, or accounting variables that have a high number of zero entries, such as characteristics based on research and development expenses. To account for this, we allow the number of quantiles to be less than 10 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 on a particular month. Depending on the frequency of newly available information, portfolios are rebalanced either monthly, quarterly or yearly. 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 decile from the top. Notice that the portfolio that was constructed based on a particular characteristic sort will also have the highest characteristic score by construction. For example, the momentum portfolio will always have the highest momentum score compared to all the other factor portfolios.

Figure 2 displays the average monthly returns of the factor portfolios together with the 95% confidence intervals (CIs). Out of all the factor portfolios, 12-month momentum (mom12m) has the highest average returns followed by 6-month momentum (mom6m). Yet, 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 A2 in the Appendix.

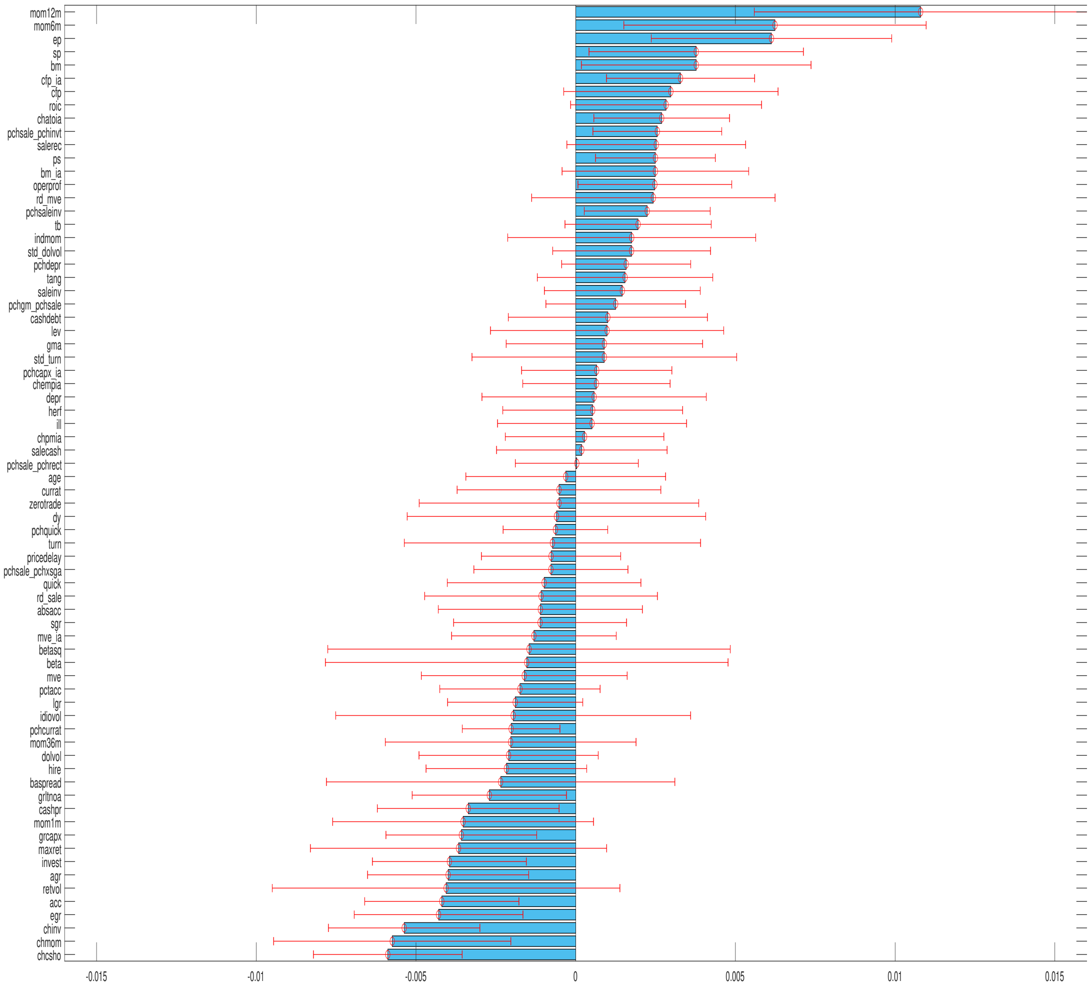

Figure 2: Average monthly returns of factor portfolios with 95% CIs for the period 01/1970-12/2019.

4.2 Performance evaluation

We examine the out-of-sample performance of our predictive models using standard forecast evaluation measures. 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 first estimate the mean-squared-error (MSE) for individual PC portfolios as,

$$\text{MSE} = \frac{1}{T-240}\sum_{t=240}^{T-1}\left(z_{i,t+1}-\widehat{z}_{i,t+1}\right)^2, \tag{9}$$

as well as Total MSE, which pools squared errors across factor portfolios and across time:

$$\text{Total MSE } = \frac{1}{N(T-240)} \sum_{i=1}^{N} \sum_{t=240}^{T-1} \left(R_{i,t+1} - \widehat{R}_{i,t+1} \right)^2. \tag{10}$$

Total MSE 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 MSE results for individual PC portfolios as well as the Total MSE under the various models. Results show that PC portfolios associated with larger eigenvalues possess overall higher MSEs. This is due to the fact that higher eigenvalue PCs posses higher variance and does not imply that those PCs are less predictable. Apropos Panel A, characteristic-based models, even with one predictive factor, deliver on average superior forecasts compared to factor momentum as indicated by the lower Total MSE. With regards to the different dimension reduction techniques used in the characteristic-based models, no significant difference is observed across methods in terms of MSE. The best performing strategy for the single factor case is RPPCA, which outperforms standard PCA as well as PLS-based models. Given that we only use the first characteristic PC, the first PLS PC should condense more return-relevant information than that of PCA. More concretely, one would expect PLS to outperform PCA as it utilizes the covariation of the predictors with the forecasting target, while PCA factors capture variation among returns-related and unrelated variables. However, no outperformance of PLS over PCA is observed out-of-sample in terms of Total MSE.

Moving to Panel B, the combination of dimension reduction techniques with LASSO significantly improves results for all models. Predictive performance improves almost uniformly across all PCs, highlighting the importance of accounting for further characteristic components and the benefits of regularization in out-of-sample performance. It is 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. Here, the best performing model is the combination of PCA with PLS, though it outperforms the rest by only a small margin. 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 MSE for the momentum strategies. Previous month returns provide unreliable forecasts for next period returns as suggested by significantly higher Total MSE. Nonetheless, applying the 1-month momentum signal on the 5 (RP) PC portfolios and then expanding the forecasts to individual anomalies slightly improves performance. When returns are averaged over the past 12 months, results improve significantly though they still fall behind the characteristic-based model forecasts. Interestingly, when momentum is applied on the (RP) PC portfolios instead of individual anomalies results improve for both the 1-month and the 12-month case. Hence, the momentum effect is concentrated in the main sources of variation in factor portfolio returns and applying the momentum signal on the PC portfolios results in more accurate forecasts.

	PC1	PC2	PC3	PC4	PC5	Total
Panel A: Single factor						
PCA	56.189	15.193	9.816	5.029	4.304	1.946
PCA-PLS	57.466	15.114	9.967	5.017	4.357	1.965
RPPCA	56.512	12.506	10.291	3.823	5.258	1.943
RPPCA-PLS	57.774	12.392	10.614	3.929	5.303	1.965
Panel B: Time-varying number of factors using LASSO						
PCA	54.129	14.796	9.569	4.884	4.350	1.907
PCA-PLS	54.691	14.417	9.515	4.852	4.155	1.906
RPPCA	55.643	12.543	10.102	3.764	5.082	1.925
RPPCA-PLS	55.434	11.739	10.183	3.772	5.100	1.912
Panel C: Momentum strategies						
1mMOM						3.657
1mMOM-PCA	103.187	26.401	22.147	10.104	8.720	3.057
1mMOM-RPPCA	104.050	22.923	19.562	9.263	10.932	3.031
12mMOM						2.063
12mMOM-PCA	60.559	15.375	10.694	5.184	4.520	2.026
12mMOM-RPPCA	60.983	12.655	10.948	4.125	5.514	2.024

Table 1: Out-of-sample MSE for PC portfolios and Total MSE across all anomalies for the period 01/1990-12/2019 multiplied by 1000. 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 of momentum strategies where 1m is based on the past period return and 12m is based on the past 12 month average.

Whereas Table 1 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 adapted 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 models (1) and (2) is defined as $DM_{12} = \bar{d}_{12}/\hat{\sigma}_{\bar{d}_{12}}$, where $\bar{d}_{12}$ and $\hat{\sigma}_{\bar{d}_{12}}$ are the mean and standard deviation of the error differential, defined as:

$$d_{12,t+1} = \frac{1}{N}\sum_{i=1}^{N}\left(\left(\hat{e}^{(1)}_{n,t+1}\right)^2 - \left(\hat{e}^{(2)}_{n,t+1}\right)^2\right), \tag{11}$$

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

Table 2 reports the results from the DM-test for pairwise comparisons between the different models. For conciseness, we only consider the models that employ LASSO from the proposed models since they outperform the single-factor models in terms of MSE. 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. The first result from Table 2 is that characteristic-based models provide fairly similar return estimates, resulting in statistically insignificant differences across model forecasts. The second result is that characteristic-based models provide superior predictions compared to factor momentum and the difference is statistically significant in all cases. Hence, we can conclude the underlying information conveyed in the characteristics is the main driver of the outperformance of the proposed models over factor momentum while the different dimension reduction techniques only play a complementary role.

	PCA	PCA-PLS	RPPCA	RPPCA-PLS	1mMOM	1mMOM-PCA	1mMOM-RPPCA	12mMOM	12mMOM-PCA
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***					
1mMOM-PCA	5.20***	5.33***	5.09***	5.30***	−14.52 ***				
1mMOM-RPPCA	5.08***	5.20***	4.97***	5.18***	−14.76 ***	−2.61 ***			
12mMOM	2.73***	2.99***	2.40**	2.93***	−6.73 ***	−4.70 ***	−4.58 ***		
12mMOM-PCA	2.24**	2.47**	1.88*	2.39**	−6.85 ***	−4.86 ***	−4.74 ***	−4.57 ***	
12mMOM- RPPCA	2.17**	2.40**	1.83*	2.34**	−6.86 ***	−4.87 ***	−4.75 ***	−4.52 ***	−0.81

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

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 on obtaining accurate predictions for individual returns, but rather on constructing portfolios with good risk-return properties (Nagel, 2021). Put differently, we are more interested in predicting cross-sectional differences in returns rather than predicting individual returns in exact terms. In that sense, Total MSE is just a distance measure that does not reflect whether models can distinguish strong from weak performers. Consequently, models that yield smaller Total MSE do not necessarily yield better portfolios in terms of average returns or Sharpe ratios. This argument also explains why 1-month factor momentum has been found empirically to be the most profitable even though our results show that it has the highest Total MSE.[5] Hence, momentum strategies may fail to predict factor portfolio returns in exact terms but may be able to rank anomalies better than the proposed models do.

To shed more light into this claim, Table 3 shows the percentage of times that the sign of factor portfolio returns was identified correctly as well as the average cross-sectional correlation between forecasted and realised returns. The first measure examines the ability of the models to predict the direction of factor portfolio returns and the second measure examines whether model-based forecasts capture the cross-sectional dispersion in factor portfolio returns. First, Table 3 demonstrates the superiority of PLS over PCA for the

[5]See, for example, Gupta and Kelly (2019).

RHS in the single factor case. Although PCA and RPPCA possess smaller Total MSEs than their PLS counterparts, they actually fail to predict any cross-sectional dispersion in factor portfolio returns or even identify the return direction. Hence, results show the ability of PLS to extract a single characteristic-based factor that is more informative about next period returns and raise questions about the suitability of aggregate forecast accuracy metrics in asset pricing. Nevertheless, accounting for further components under a LASSO approach harmonizes performance across models, with PLS slightly outperforming PCA. Factor momentum displays similar performance with the proposed models in terms of proportion of correct sign, although average cross-sectional correlations are lower. The correlation is higher for the 1-month signal, while it diminishes when the signal (either 1month or 12-month) is applied on the PC portfolios instead of individual anomalies, contradicting the results in Table 1. Overall, results confirm that characteristic-based models can better distinguish winners from losers compared to factor momentum.

	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: Momentum strategies		
1mMOM	51.88	5.98
1mMOM-PCA	51.61	5.70
1mMOM-RPPCA	51.87	5.81
12mMOM	52.40	5.43
12mMOM-PCA	51.45	3.74
12mMOM-RPPCA	51.61	3.92

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 of momentum strategies where 1m is based on the past period return and 12m is based on the past 12 month average.

So far, forecasting performance evaluation is based on composite measures that do not explicate the degree to which individual anomalies are predictable. 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 relative mean-squared-errors (r-MSE) for all anomalies under the different models. We define r-MSE as the sum of return differences between the squared error and the realized squared returns:

$$
\begin{aligned}
\text{r-MSE} &= \frac{\sum_{t=240}^{T-1}\left(\left(R_{i,t+1}-\widehat{R}_{i,t+1}\right)^2 - R_{i,t+1}^2\right)}{\sum_{t=241}^{T-1} R_{i,t+1}^2} \\
&= \frac{\sum_{t=240}^{T-1}\left(R_{i,t+1}-\widehat{R}_{i,t+1}\right)^2}{\sum_{t=240}^{T} R_{i,t+1}^2} - 1. \qquad (12)
\end{aligned}
$$

One way to understand r-MSE is as a measure that benchmarks model forecasts against a return forecast of zero. The measure is capped at –1, with negative values indicating that the underlying model takes a perfect stance in predicting next period returns. Intuitively, as long as the underlying model can predict the sign of returns correctly and does not overshoot above $2 \times R_{i,t}$, r-MSE will be negative. Figure 3 displays a heat-map with the r-MSE for individual anomalies under the proposed models that employ LASSO. For comparison, we also show the r-MSE for 12mMOM-RPPCA, which had the smallest Total MSE and hence the best r-MSE out of the benchmark models. Negative values are highlighted in green while positive values are in red. 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 size (mve), value (bm) and momentum (mom12m) to be highly predictable by observed characteristics. For factor momentum, the number of unpredictable anomalies is significantly larger and the degree of predictability diminishes even for the anomalies that remain predictable. Furthermore, factor portfolios that are unpredictable by characteristics remain unpredictable under factor momentum, implying that the momentum signal does not provide any distinct information that cannot be found in other characteristics.

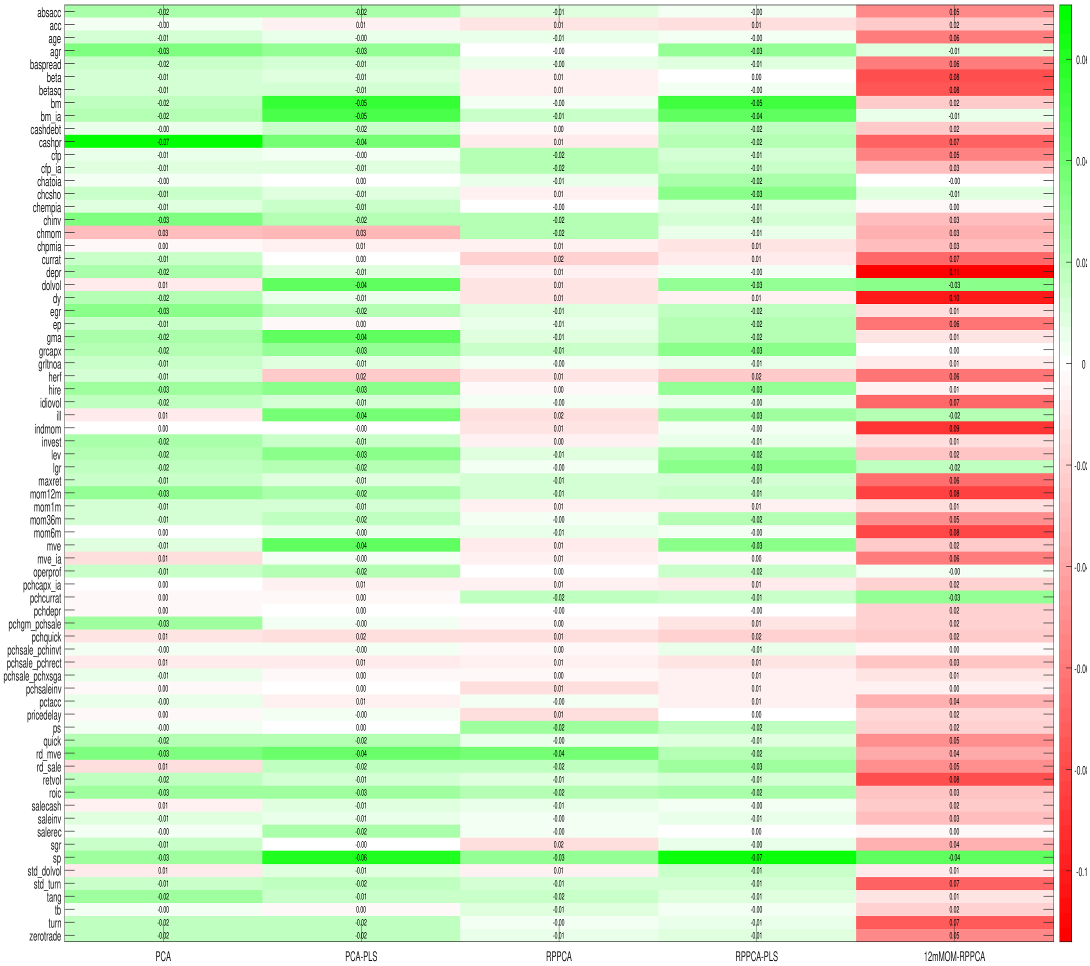

Figure 3: Relative mean-squared-errors for individual anomalies under the characteristic-based models that employ LASSO and 12mMOM-RPPCA. Positive values (in red) show lack of predictive ability while negative values (in green) show predictive ability of the underlying model for a given factor portfolio.

Overall, results show that anomalies are predictable to a high extent. Depending on the method used individual anomaly predictability changes, with PCA working better for the LHS and PLS for the RHS. Consistent with Total MSE results, PCA-PLS has the lowest r-MSE while RPPCA has the highest. Still, RPPCA delivers negative r-MSE for anomalies with high average returns (in absolute terms) in the out of sample period. In fact, r-MSE estimates are significantly more correlated with absolute average returns when RPPCA is used for the LHS, reflecting the ability of the method to extract components that account for the difference in average returns of the factors. When PCA is used for the LHS, r-MSE is more correlated with factor portfolio volatility, reflecting the variance maximization objective. Finally, almost all models fail to predict anomalies that are based on a % change in accounting variables such as % change in the quick ratio (pchquick) and % change in sales minus % change in inventory (pchsale_pchinvt), among others, located in the lower half of the heat-map. These portfolios have returns 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.

We then 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 PC and RP PC portfolios 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 MSE in the validation period without specifying which these components are. 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 inconstant, 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 inconstant 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 variation in the number of components being used.

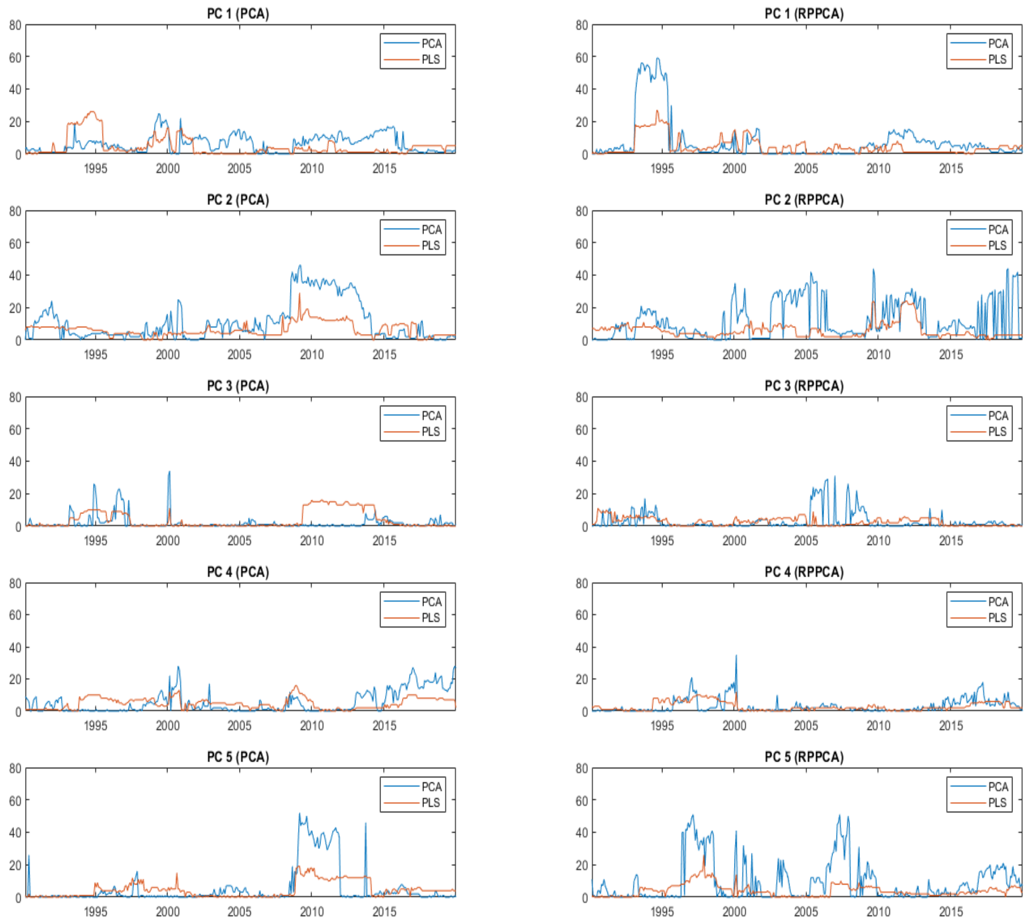

Figure 4: Number of features for each PC portfolio under the different models. The number of features is identified by recursively applying LASSO on the set of components and picking the penalty factor that minimizes the mean-squared-error in the validation period.

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 a LS portfolio of factor portfolios. Factor portfolios are grouped into equally-weighted deciles based on their return forecasts and a long-short strategy (P10-P1) 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 only 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 returns at each period even if they are indecisive about anomalies with returns close to zero. It is also important to highlight that characteristics that are negatively associated with expected returns, like size or asset growth, will usually be located in the short leg, while characteristics that are positively associated with returns, like momentum, will be located in the long leg. Still, depending on the underlying change in their characteristics, portfolios with positive average returns can be left out of the long leg or even move to the short leg. For example, momentum may stay out of the investible portfolios in one month if the signal is weak or even be shorted if expected returns for momentum are negative. Hence, our timing strategy allows for a dynamic selection of portfolios and accounts for variations in factor portfolio performance.

The second investment strategy is similar to the time-series factor momentum (TSFM) strategy by Gupta and Kelly (2019). TSFM scales factor portfolio returns $R_{t+1,\cdot}$ according to return forecasts $\hat{R}_{t+1,\cdot}$. 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:

$$s_{t,n} = \min\left(\max\left(\frac{1}{\sigma_{t,n}}\hat{R}_{t+1,n}, -2\right), 2\right). \tag{13}$$

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

$$\text{TSFM}_{t+1} = \frac{\sum_n 1_{\{s_{t,n}>0\}} R_{t+1,n} \times s_{t,n}}{\sum_n 1_{\{s_{t,n}>0\}} s_{t,n}} - \frac{\sum_n 1_{\{s_{t,n}\leq 0\}} R_{t+1,n} \times s_{t,n}}{\sum_n 1_{\{s_{t,n}\leq 0\}} s_{t,n}}. \tag{14}$$

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 be different for TSFM while they remain 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 return determines the relative weight. Hence, under TSFM the long and the short leg can have a disproportional number of constituents and in extreme cases the strategy can converge to a long or short only. Lastly, individual factor volatility can have an effect on portfolio weights as higher volatility will result in smaller z-scores other things equal. Nevertheless, this approach still disregards the covariance structure of the factor portfolios, possibly resulting in excessive volatility.

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 return forecasts before they are being transformed into z-scores. This strategy takes positions in factor portfolios that have out/under-performed 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 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.

$$s_{t,n} = \min\left(\max\left(\frac{1}{\sigma_{t,n}}\hat{R}_{t+1,n} - \text{median}(\hat{R}_{t+1,.}), -2\right), 2\right) \tag{15}$$

Table 4 presents the portfolio evaluation measures for the various models under the three strategies. Out of the three strategies, LSS delivers the highest average return while CSFM and TSFM have higher Sharpe ratios. The high average monthly return of the LSS strategy confirms that the models correctly identify the factor portfolios in the extremes of the conditional return distribution. Even though model forecasts do not result in the exact classification of anomalies across the whole spectrum of the distribution, as implied by the relatively low average cross-sectional correlations, they are able to distinguish anomalies with high returns in absolute terms. Nonetheless, factor portfolios with average returns close to zero are unimportant from an investing perspective and little is sacrificed by not being able to properly rank them or forecast their performance. The higher Sharpe ratios for TSFM and CSFM are due to the lower volatility of these strategies as they invest in a higher number of factor portfolios and therefore enjoy a larger diversification benefit. Furthermore, TSFM and CSFM strategies have a higher hit-rate and a lower max drawdown compared to LSS, implying more consistent performance over time. However, these strategies inevitably end up taking positions in anomalies with weak performance, resulting in lower average returns.

Looking at Panel A, results confirm the superiority of PLS over PCA for the RHS in the single factor case. Evidently, models based on a single factor that concentrates the variation among multiple characteristics are unable to predict the cross-sectional dispersion in factor portfolio returns, implying 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 going to the negative side. Conversely, when PLS is used for the RHS all strategies deliver the positive and significant returns, reflecting the ability of the method to concentrate return-relevant variation into a single predictor. Comparing the two PLS-based models, PCA-PLS delivers higher average returns while RPPCA-PLS has higher Sharpe ratios.

The first conclusion from Panel B is that the use of further components in combination with LASSO uniformly improves investment performance across all models. This result is also consistent with the improvement in forecasting performance, although the best performing model investment-wise is not necessarily the one with the smallest Total MSE. All strategies deliver highly significant returns, surpassing the t-value threshold of three by Harvey et al. (2016). Still, depending on the strategy, relative model performance changes. For example, PCA delivers the highest return under LSS and TFSM, while PCA-PLS delivers the highest return under CFSM. Similarly, RPPCA and RPPCA-PLS have the highest Sharpe ratios under TFSM and CFSM, although PCA has a marginally higher Sharpe ratio under LSS. More generally, 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 have a higher hit-rate and a lower max drawdown, reflecting higher consistency and lower downside risk. Overall, results are now similar across methods, suggesting that once further components are considered no significant difference arises across methods.

In line with prior literature (e.g. Gupta and Kelly (2019)), factor momentum results in Panel C show that the effect is the strongest for the 1-month formation period. The 1-month momentum signal outperforms the 12-month, whether applied to individual anomalies or (RP) PC portfolios. The 12-month factor momentum even underperforms the 12-month stock momentum (mom12m), as it produces low average returns and in many cases insignificant. Surprisingly, applying the momentum signal to PC portfolios has an inconsistent effect on investment performance. For the 1-month momentum, return performance improves slightly while for the 12-month momentum performance deteriorates. Comparing results across panels, 1-month momentum displays similar return performance with the PLS-based single factor models, although the former have higher Sharpe ratios and the latter lower max drawdown. Nonetheless, characteristic-based models that employ LASSO outperform factor momentum under all three strategies, demonstrating the benefits of conditioning factor portfolio returns on observable characteristics under a regularized framework. Even more importantly, by comparing the results in Table 4 to Table A2, it is evident that characteristic-based factor timing strategies deliver investment performance over that of any individual factor portfolio.

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 that employ LASSO and factor momentum. 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, 1-month momentum outperforms characteristic-based models in the early years of the out-of-sample period, up until the late 90s. A spike in performance occurs around 2000 for all strategies, during the buildup of the dot-com bubble. Interestingly, unlike factor momentum, characteristic-based models do not plummet after the burst and remain above thereafter. Furthermore, the performance of all strategies 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 even after the 2010 period. Looking at graphs in the right panel of Figure 5, characteristic-based models display a positive trend in later years, while factor momentum strategies remain relatively stagnant. Even the worst characteristic-based model outperforms the best momentum model under all three strategies, with the difference being more pronounced for the LSS strategy, as it focuses on the most prominent subset of factor portfolios only. With regards to factor momentum, no significant difference is observed between the 1-month and the 12-month signal in later years of the out-of-sample period. In fact, for TSFM and CSFM 12mMOM-PCA has the highest cumulative returns out of all momentum variations, suggesting that the superiority of the 1-month signal has faded in recent years. Out of all models, RPPCA has the highest cumulative returns in the last 10 years whilst it was underperforming in the early years of the out-of sample period. Overall, factor timing strategies based on observed characteristics yield positive returns in later periods, even though most factors have been found empirically to die out over time (McLean and Pontiff, 2016).

	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: Momentum strategies																		
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
1mMOM-PCA	1.13	0.64	0.65	9.24	5.53	5.54	0.12	0.12	0.12	2.32	2.20	2.22	57.38	55.71	54.60	25.74	19.91	19.82
1mMOM-RPPCA	1.12	0.64	0.66	9.45	5.47	5.48	0.12	0.12	0.12	2.25	2.21	2.27	55.99	55.71	56.82	24.97	19.06	18.69
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
12mMOM-PCA	0.76	0.59	0.58	9.40	5.92	5.85	0.08	0.10	0.10	1.53	1.88	1.89	52.65	53.76	53.20	36.46	28.23	27.61
12mMOM-RPPCA	0.76	0.60	0.59	9.31	5.63	5.53	0.08	0.11	0.11	1.54	2.03	2.01	54.60	51.81	51.81	31.10	24.77	24.86

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 of factor momentum strategies. 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.

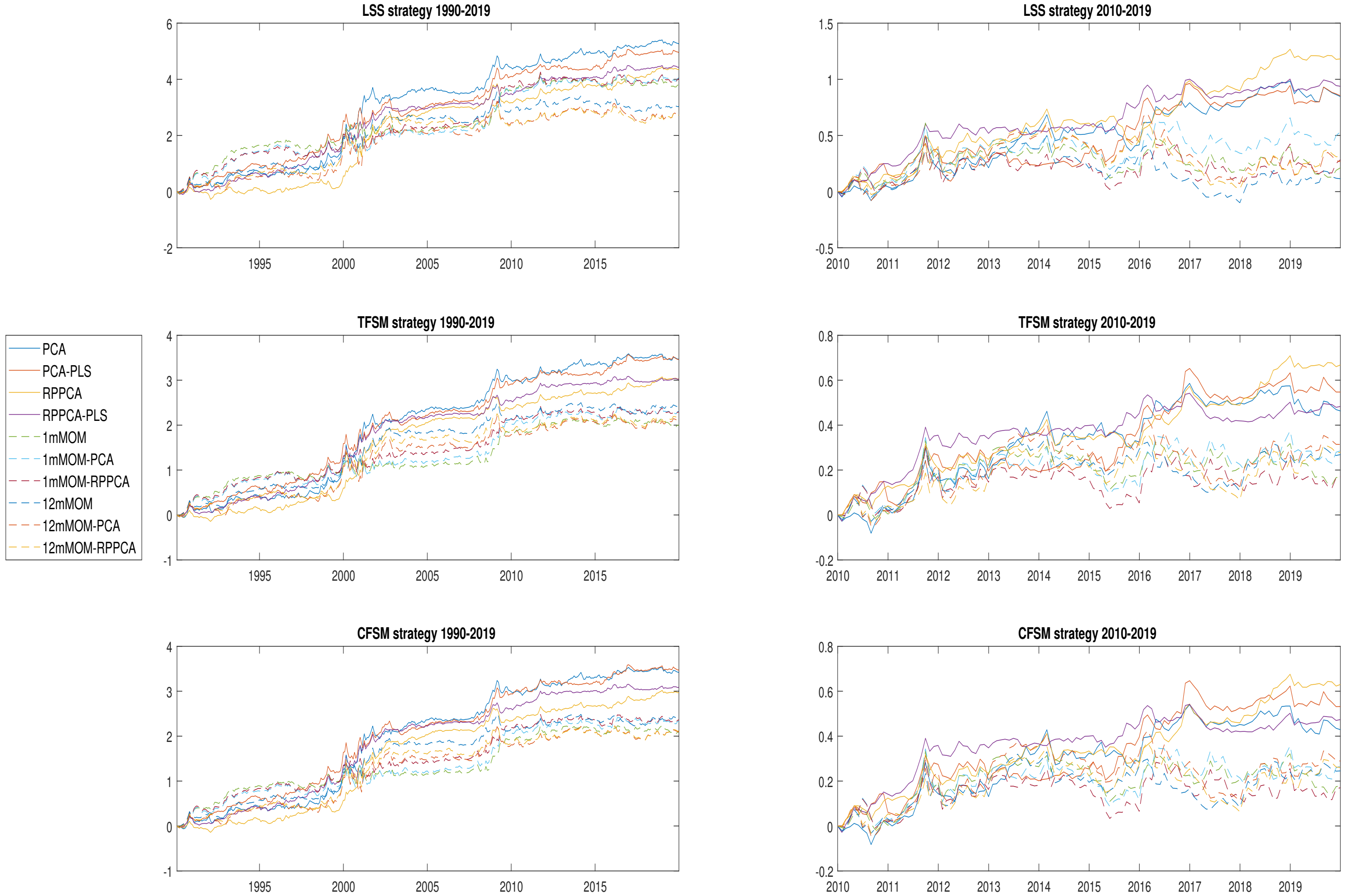

Figure 5: Cumulative return performance of factor timing strategies. The figure displays the performance of LSS, TSFM and CSFM for characteristic-based models using LASSO and factor momentum. Graphs to the left display the cumulative return performance over the whole sample period (1990-2019) and graphs 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.

5 Conclusion

We investigate the predictability of factor portfolios from their own portfolio characteristics, trying to go over and above factor momentum. 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 to predict their performance. Instead of focusing on a single predictive signal to time factor portfolio returns, we simultaneously use the whole universe of characteristics, examining the possibility that factor portfolios are predictable by characteristics other than their own.

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. Dimension reduction in the number of forecasting targets is attained by focusing on the main sources of return variation in terms of PC portfolios. The information in the characteristics is then compressed into a handful of latent factors with the use of different dimension reduction techniques, such as PCA and PLS. Such an approach allows us to combine information across a large collection of characteristics, while still accounting for potential multicolinearity or redundancy among the predictors. Hence, we conduct a comparative analysis not only of factor momentum and characteristic-based models but also of different machine learning methods in general.

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. Results show that indeed the characteristics' signal strength differs across PC portfolios and even within the same PC portfolio the characteristic-return relationship is significantly time-varying. However, our models account for this by under- or over-weighting information contained in the characteristics through adjusting the degree of coefficient shrinkage. 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. Specifically, we find that the dominant PC portfolios are highly predictable by the information contained in their characteristics and that this predictability can be easily extended to individual anomalies. Under all methodological alternations, forecasts based on characteristics yield far smaller MSEs and result in factor timing strategies with higher average returns and Sharpe ratios, compared to factor momentum. Furthermore, the investment performance of our factor timing strategies is superior to that of any individual anomaly, demonstrating the benefits of timing over static factor investing. Finally, the use of factor momentum as a benchmark also reveals a relative instability in the strength of the momentum signal, as the optimal look-back window varies across time.

In terms of different methods used, for the LHS, PCA results in smaller MSEs and higher returns when combined with LASSO, while RPPCA results in higher Sharpe ratios, even though the differences are marginal. For the RHS, PLS outperforms PCA when a single predictor is used as PCA ends up capturing variation in the characteristics that is irrelevant in the forecasting objective. Still, after accounting for further components the difference between PCA and PLS becomes insignificant, suggesting once the whole information set is considered the method that is used to construct the components becomes inconsequential. 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., Ilmanen, A. and Israel, R. (2017), 'Contrarian factor timing is deceptively difficult', *Journal of Portfolio Management* **43**(5), 72–87.

Avramov, D., Cheng, S., Schreiber, A. and Shemer, K. (2017), 'Scaling up market anomalies', *Journal of Investing* **26**(3), 89–105.

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

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.

Daniel, K. and Moskowitz, T. J. (2016), 'Momentum crashes', *Journal of Financial Economics* **122**(2), 221–247.

De Jong, S. (1993), 'Simpls: an alternative approach to partial least squares regression', *Chemometrics and intelligent laboratory systems* **18**(3), 251–263.

DeMiguel, V., Martin-Utrera, A., Nogales, F. J. and Uppal, R. (2020), 'A transactioncost 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. (2021), 'Factor momentum and the momentum factor', *Journal of Finance, Forthcoming* .

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.

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

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.

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.

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., 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.

Leippold, M. and Yang, H. (2021), 'The anatomy of factor momentum', *Available at SSRN: https://ssrn.com/abstract=3517888 or http://dx.doi.org/10.2139/ssrn.3517888* .

Lettau, M. and Pelger, M. (2020*a*), 'Estimating latent asset-pricing factors', *Journal of Econometrics* **218**(1), 1–31.

Lettau, M. and Pelger, M. (2020*b*), 'Factors that fit the time series and cross-section of stock returns', *Review of Financial Studies* **33**(5), 2274–2325.

Lewellen, J. (2015), 'The cross-section of expected stock returns', *Critical Finance Review* **4**(1), 1–44.

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.

Onatski, A. (2012), 'Asymptotics of the principal components estimator of large factor models with weakly influential factors', *Journal of Econometrics* **168**(2), 244–258.

Rapach, D. and Zhou, G. (2013), Forecasting stock returns, *in* 'Handbook of economic forecasting', Vol. 2, Elsevier, pp. 328–383.

Ross, S. (1976), 'The arbitrage theory of capital asset pricing', *Journal of Economic Theory* **13**(3), 341–360.

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.

Welch, I. and Goyal, A. (2008), 'A comprehensive look at the empirical performance of equity premium prediction', *Review of Financial Studies* **21**(4), 1455–1508.

Wold, S., Ruhe, A., Wold, H. and Dunn, Iii, W. (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.

Yara, F. B., Boons, M. and Tamoni, A. (2021), 'Value return predictability across asset classes and commonalities in risk premia', *Review of Finance* **25**(2), 449–484.

Appendix A

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 (oancf) divided by average total assets (at); if oancf 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.

Acronym	Author(s)	Journal	Definition
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.
chinv	Thomas & Zhang	2002, RAS	Change in inventory (inv) scaled by average total assets (at).

Acronym	**Author(s)**	**Journal**	**Definition**
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).

Acronym	**Author(s)**	**Journal**	**Definition**
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).
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 (ppegt) + annual change in inventories (invt) all scaled by lagged total assets (at).

Acronym	Author(s)	Journal	Definition
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.

Acronym	**Author(s)**	**Journal**	**Definition**
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).
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.

Acronym	**Author(s)**	**Journal**	**Definition**
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_mve	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.

Acronym	Author(s)	Journal	Definition
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 A1: 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

	Average Return	Standard Deviation	Sharpe Ratio	t-statistic
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
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

	Average Return	Standard Deviation	Sharpe Ratio	t-statistic
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_mve	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 A2: 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: test statistic of H_0: Average monthly return=0.

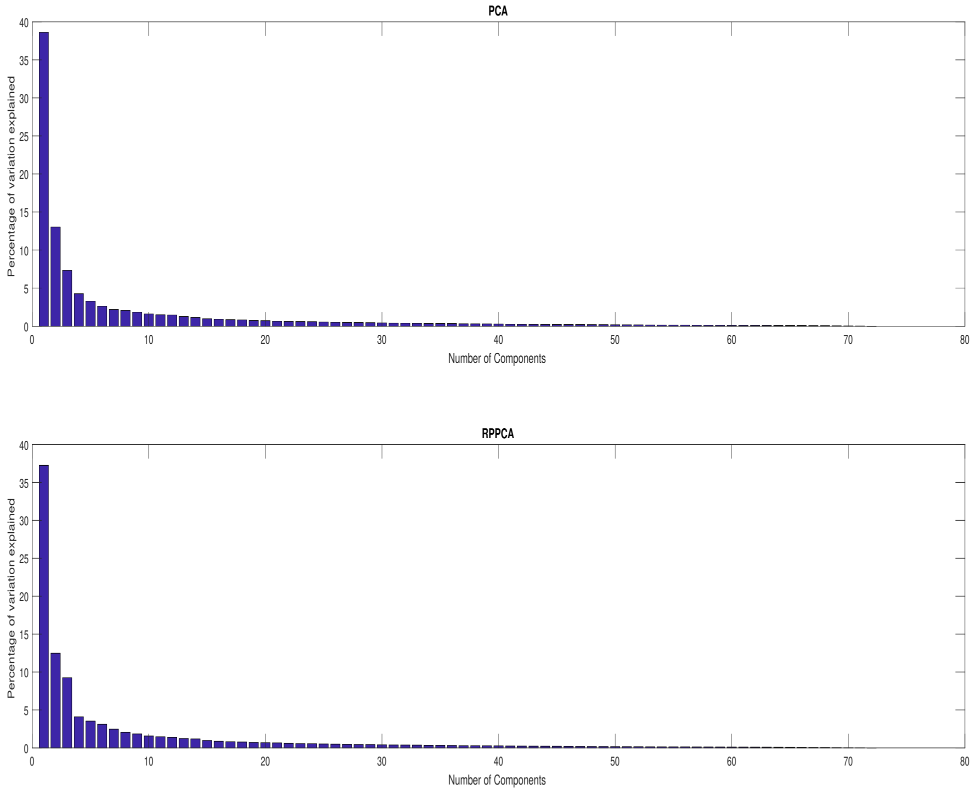

Figure A1: Percentage of the variation explained by each PC of factor portfolio returns for the sample period January 1970 to December 2019.