RESEARCH

Swiss Finance Institute Research Paper Series N°20-67

Principal Portfolios

Bryan T. Kelly
AQR Capital Management, Yale School of Management, and NBER

Semyon Malamud
Ecole Polytechnique Federale de Lausanne, Swiss Finance Institute, and CEPR

Lasse Heje Pedersen
AQR Capital Management, Copenhagen Business School, and CEPR

Principal Portfolios

Bryan Kelly, Semyon Malamud, and Lasse Heje Pedersen*

This version: May 25, 2021

Abstract

We propose a new asset-pricing framework in which all securities' signals are used to predict each individual return. While the literature focuses on each security's own-signal predictability, assuming an equal strength across securities, our framework is flexible and includes cross-predictability—leading to three main results. First, we derive the optimal strategy in closed form. It consists of eigenvectors of a "prediction matrix," which we call "principal portfolios." Second, we decompose the problem into alpha and beta, yielding optimal strategies with, respectively, zero and positive factor exposure. Third, we provide a new test of asset pricing models. Empirically, principal portfolios deliver significant out-of-sample alphas to standard factors in several data sets.

Keywords: Portfolio choice, asset pricing tests, optimization, expected returns, predictability
JEL: C3, C58, C61, G11, G12, G14

*Bryan Kelly is at AQR Capital Management, Yale School of Management, and NBER. Semyon Malamud is at Swiss Finance Institute, EPFL, and CEPR, and is a consultant to AQR. Lasse H. Pedersen is at AQR Capital Management, Copenhagen Business School, and CEPR. We thank David Lmarket-hedged portfoliosando, Brendan Hoffman, Ronen Israel, and Steffen Lauritzen for helpful comments. AQR Capital Management is a global investment management firm, which may or may not apply similar investment techniques or methods of analysis as described herein. The views expressed here are those of the authors and not necessarily those of AQR. Pedersen gratefully acknowledges support from the Center for Financial Frictions (grant no. DNRF102).

1 Introduction

The starting point for much of asset pricing is a signal, $S_{i,t}$, that proxies for the conditional expected return for a security i at time t. In the context of an equilibrium asset pricing model, $S_{i,t}$ may represent the conditional beta with respect to the market (or the pricing kernel). Or it may be a predictor that is agnostic of equilibrium considerations, such as each asset's valuation ratio or its recent price momentum. Standard analyses, such as evaluating characteristic-sorted portfolios or asset pricing tests in the spirit of Gibbons et al. (1989), focus on own-asset predictive signals; that is, the association between $S_{i,t}$ and the return on only asset i, $R_{i,t+1}$.

We propose a new approach to analyzing asset prices through the lens of what we call the "prediction matrix." This approach shows (1) how to optimally invest in light of cross-predictability, where optimal refers to the return-maximizing strategy among a class of linear trading strategies; (2) how to find optimal alpha and beta strategies; and (3) a new test of asset pricing models.

To explain our approach in more detail, we define the prediction matrix as $\Pi = E(R_{t+1}S_t') \in \mathbb{R}^{N\times N}$ where $R_{t+1} = (R_{i,t+1})_{i=1}^N \in \mathbb{R}^N$ is the vector of returns and $S_t = (S_{i,t})_{i=1}^N \in \mathbb{R}^N$ is the vector of signals. The diagonal part of the prediction matrix tracks the own-signal prediction effects, which are the focus of standard asset pricing. For example, if the signal $S_{i,t}$ represents each asset's own momentum, then $\Pi_{i,i} = E(R_{i,t+1}S_{i,t})$ is the expected profit of trading asset i based on its own momentum signal. In other words, we think of the signal $S_{i,t}$ as the portfolio holding and $R_{i,t+1}S_{i,t}$ as the corresponding return.

Importantly, the prediction matrix also tracks cross-predictability phenomena. Indeed, the off-diagonal part of the prediction matrix, $\Pi_{i,j} = E(R_{i,t+1}S_{j,t})$, shows how asset j's signal predicts asset i's return. Cross-predictability exists very generally in conditional asset pricing models, be they equilibrium in nature or purely statistical. Knowledge of the entire prediction matrix, as opposed to the typical focus on diagonal elements alone, is critical to devising optimal portfolios and understanding their risk-return tradeoff.

Our main contribution is to develop an extensive theoretical understanding of the prediction matrix and the asset pricing information it carries. The main tools of our analysis are singular value decompositions, analogous to using principal components analysis (PCA) to study variance-covariance matrices. The leading components (singular vectors) of Π are defined as those responsible for the lion's share of covariation between signals and future returns. This is where cross-predictability information becomes valuable. Like the diagonal elements, off-diagonal elements of Π are informative about the joint dynamics in signals and returns.

We refer to Π's singular vectors as "principal portfolios." They are a set of normalized portfolios ordered from those most predictable by S to those least predictable. The top principal portfolios are thus the most "timeable" portfolios, and as such they offer the highest unconditional expected returns for an investor that faces a leverage constraint (one who cannot hold infinitely large positions).

A key insight of our approach is that applying a singular value decomposition directly to Π conflates two different and opposing economic phenomena. We propose first splitting Π into a symmetric part (which is equal to its transpose and denoted Π^s) and an anti-symmetric part (which is equal to minus its transpose and denoted Π^a), and applying separate matrix decompositions to Π^s and Π^a. The symmetry separation of Π,

$$\Pi = \underbrace{\frac{1}{2}(\Pi + \Pi')}_{\Pi^s} + \underbrace{\frac{1}{2}(\Pi - \Pi')}_{\Pi^a}, \tag{1}$$

is a powerful device. With eigenvalue decompositions of each part, we can take a complicated collection of predictive associations in the Π matrix and decode them into a set of wellorganized facts about expected returns. These facts describe i) the nature of each predictive pattern represented in Π and ii) the strength of these patterns.

The nature of a predictive pattern is described by its classification as either symmetric or anti-symmetric, which, amazingly, translate into beta and alpha. In particular, we show that eigenvectors of the symmetric matrix Π^s are effective ways to achieve exposure (beta) to the factor based on the signal S, while eigenvectors of the anti-symmetric matrix Π^a are powerful factor-neutralized (alpha) strategies with respect to this factor. We refer to strategies arising from eigenvectors of the symmetric component as "principal exposure portfolios" (PEPs) and the strategies arising from the anti-symmetric part as "principal alpha portfolios" (PAPs). Once classified as "exposure" versus "alpha," prediction patterns (principal portfolios) are then ordered from strongest to weakest and based on the size of their associated eigenvalues. In particular, we prove that the unconditional average returns of PEPs and PAPs are exactly proportional to their respective eigenvalues.

This decomposition has a close connection to equilibrium asset pricing. When signals are betas to the pricing kernel and there is no-arbitrage, all PAPs must deliver zero expected excess returns (because they have zero factor exposure) and all PEPs must deliver non-negative average returns (because they have positive exposure to the pricing kernel). These insights are the groundwork for a new asset pricing test based on eigenvalues of the symmetric and anti-symmetric components of the prediction matrix. In rational asset pricing models, there should not be any alpha relative to the pricing kernel. When we pick signals that are supposed to be proportional to covariances with the pricing kernel (e.g., market betas), then the corresponding prediction matrix should have a zero anti-symmetric part—meaning that Π should be symmetric and there should be no alpha portfolios. Moreover, negative eigenvalues of the symmetric part of Π correspond to strategies with negative factor exposure and positive expected returns, another form of alpha. Since rational asset pricing also rules out both forms of alpha, we get the asset pricing test that Π should be symmetric and positive definite. In other words, when signals capture exposure to the pricing kernel, all PEPs should deliver non-negative returns and all PAPs should deliver zero returns.

We also develop theoretical underpinnings for practical empirical usage of the prediction matrix from the perspective of robust statistics and machine learning. Our main theoretical results are developed in population, where Π is known, and, with N assets, this requires estimating N^2 parameters. Such rich parameterization can lead to noisy estimates and overfit that deteriorate the out-of-sample performance of principal portfolios. In the literature and financial practice, signals are often analyzed or traded in the form $\sum_i S_{i,t}R_{i,t+1}$, which essentially restricts the signal-based analysis to testing a single parameter equal to average own-predictability, $\sum_i E(S_{i,t}R_{i,t+1})$. While this may benefit from some robustness, restricting the analysis to a one-parameter problem is harsh—it forfeits any and all useful information about heterogeneity in own-predictability or cross-predictability in the estimated Π matrix. Principal portfolios are well suited to balance the joint considerations of exploiting potentially rich information from throughout Π while controlling parameterization to reduce noise and overfit. We show that low-rank approximations of Π and its symmetry-based components Π^s and Π^a offer a means of balancing both considerations in a data-driven way in order achieve robust out-of-sample portfolio performance.

We implement the methodology empirically using several data sets. We conduct out-of-sample analyses that, at each time period t, estimates the prediction matrix from past signals and returns (i.e., only information that is available through time t). Estimating the prediction matrix is easy: $\hat{\Pi}_t = \frac{1}{120}\sum_{\tau=t-120}^{t-1} R_{\tau+1}S'_\tau$, where we use a backward looking window of 120 time periods in most of our analysis. Having estimated the prediction matrix, the singular value decompositions of Π as well as its symmetric and anti-symmetric parts immediately yield PPs, PEPs, and PAPs, and we track their out-of-sample performance.

As a simple illustration of our method, we first consider the empirical performance of principal portfolios using standard Fama-French portfolios as base assets and momentum (i.e., past returns) as the trading signal. We find that the leading principal portfolios deliver significant positive returns out of sample. As further evidence consistent with our framework, PEP portfolios have significant factor exposure while PAP have little to no factor exposure and significant alpha. These results are robust across a number of specifications and alternative samples (international equities; futures contracts on equity indices, bonds, commodities, and currencies).

We also consider a more comprehensive sample of more than 100 standard equity factors with more than 100 different trading signals (e.g., accruals, book-to-market, cash-to-assets, etc.). Our approach can only handle one signal at a time, so we conduct our principal portfolios analysis one signal at a time. We also aggregate the resulting principal portfolios for each of these signals into a combined strategy. We find that factor timing strategies based on principal portfolios perform well overall and across the majority of signals. In fact, principal portfolios have significant alpha relative to standard factor models and to the principal components-based factor timing approach of Haddad et al. (2020).

Our paper is related to several literatures. Our asset pricing test complements other such tests, including Gibbons et al. (1989) and Hansen and Jagannathan (1991) (see Cochrane (2009) for an overview). Our method to uncover new forms of predictability complements existing methods based on regressions (see Welch and Goyal (2008) and references therein), portfolio sorts (a recent example is Fama and French (2015)), and machine learning (Gu et al. (2018)). We consider factor timing based on a host of signals, complementing the work on factor timing based on value spreads (Asness et al. (2000), Cohen et al. (2003), Haddad et al. (2020)) and factor momentum (Arnott et al. (2019), Gupta and Kelly (2019)). Finally, we consider linear trading strategies, which have also been studied in the context of dynamic trading with transaction costs by Gârleanu and Pedersen (2013, 2016), Collin-Dufresne et al. (2015), Collin-Dufresne et al. (2019), and others. While this literature focuses on linear-quadratic programming, we instead consider eigenvalue methods.

In summary, we present a new way to uncover return predictability and test asset pricing models. We illustrate how the method works empirically with a wide range of encouraging results for out-of-sample principal portfolio performance across signals and samples.

2 Principal Portfolio Analysis

In this section, we lay out our principal portfolio analysis (PPA) framework. We describe the concept of linear strategies of predictive signals, show how linear strategies are intimately linked to the prediction matrix, derive optimal strategies, and introduce the notion of principal portfolios that implement optimal strategies.

Let us first introduce the setting and notation that we use throughout. The economy has N securities traded at discrete times. At each time t, each security i delivers a return in excess of the risk-free rate, $R_{i,t}$. All excess returns at time t are collected in a vector, $R_t = (R_{i,t})_{i=1}^N$ and their conditional variance-covariance matrix is $\Sigma_{R,t} = \mathrm{Var}_t(R_{t+1})$.

For each time and security, we have a "signal" or "characteristic" $S_{i,t}$, and all signals are collected in a vector, $S_t = (S_{i,t})_{i=1}^N$. We can think of these predictive characteristics as market betas, valuation ratios, momentum scores, or other observable signals that proxy for conditional expected returns.

2.1 Linear Trading Strategies

How can an investor best exploit predictive information in an asset characteristic S? To answer this question in a tractable way, we work in the context of general linear trading strategies based on S. Then, we derive an optimal linear strategy subject to leverage constraints and show the intimate connection between the optimal linear strategy and the prediction matrix Π.

A *linear strategy* based on S has portfolio weights of the form $w_t' = S_t'L$. We refer to $L \in \mathbb{R}^{N\times N}$ as the *position matrix* because each column of L translates signals into a portfolio position in each asset. For example, the first column $L_1 = (L_{i,1})_{i=1}^N$ of L translates all the signals into a position in asset 1, $S_t'L_1$. The return of a linear strategy is naturally the positions times the returns, that is,

$$R_{t+1}^{w_t} = w_t' R_{t+1} = \sum_j \underbrace{(S_t' L_j)}_{\text{position in } j} \underbrace{R_{j,t+1}}_{\text{return of } j} = S_t' L R_{t+1}\,. \tag{2}$$

We see that a linear strategy generally allows the position $S_t'L_j$ in any asset j to depend on the signals of *all* assets. Interestingly, these strategies can potentially exploit both predictability using each asset's own signal as well as cross-predictability using other signals.

The large majority of return prediction patterns in the empirical literature focus on strategies that are agnostic of cross-predictability. The literature's default portfolio construction based on a characteristic S builds a simple tradable factor of the form:

$$\widetilde{F}_{t+1} = \sum_j S_{j,t} R_{j,t+1} \tag{3}$$

We refer to $\widetilde{F}_{t+1}$ as the "simple factor" henceforth. We note that the simple factor is a linear strategy with identity position matrix ($L = \mathrm{Id}$):

$$\widetilde{F}_{t+1} = \sum_i S_{i,t} R_{i,t+1} = S_t' R_{t+1} = S_t' \mathrm{Id} R_{t+1}. \tag{4}$$

Hence, our framework nests the standard framework, and allows more general strategies.

The simplicity of the conventional strategy $\widetilde{F}$ makes it a helpful reference point for the strategies we advocate in this paper. It is a portfolio that relies only on own-signal predictions with no cross-prediction. Moreover, it imposes that own-signal predictions enter into the portfolio uniformly, with no regard for heterogeneity in predictive effects across assets. When a researcher reports that this type of simple factor has a positive average return, $E(\widetilde{F}_{t+1}) > 0$, it is the same as saying that the signal positively predicts own-asset returns on average.

2.2 The Prediction Matrix

A central part of our analysis makes use of what we call the **prediction matrix**:

$$\Pi = E(R_{t+1}S_t'). \tag{5}$$

Π encodes predictive information for how the signals predict all returns, based on assets' own signals as well as cross-predictability. A strategy that literally chooses an asset's position equal to its own signal $S_{i,t}$ earns a return of $R_{i,t+1}S_{i,t}$, and $\Pi_{i,i}$ is the expected value of this return. Likewise, a strategy that takes a position in asset i based on the signal of another asset j earns average returns of $\Pi_{i,j}$.

If $S_{j,t}$ predicts $R_{j,t+1}$ on average across securities, then this is the same as saying that the prediction matrix has a positive trace (tr, the sum of its diagonal elements):

$$E(\widetilde{F}) = E\left(\sum_j S_{j,t}R_{j,t+1}\right) = \mathrm{tr}(\Pi) \ > \ 0\,. \tag{6}$$

This notion of positive own-predictability on average across securities has emerged as the standard criterion by which predictive signals are measured in the empirical finance literature and is typically evaluated based on the sample average of the strategy in (3).

Average own-predictability not only abstracts from information in off-diagonal elements of Π, but also from heterogeneity in own-effects on the main diagonal. In short, strategies predicated on average own-predictability are highly constrained in the information they consider regarding the predictive content of S. Proposition 1 shows that the *entire* Π matrix is necessary (and sufficient!) for understanding the returns of more general linear strategies.

Proposition 1 (Return of Linear Strategies) *The expected excess return of a linear trading strategy* $w_t' = S_t' L$ *is*

$$E\left(R_{t+1}^{w_t}\right) = E\left(S_t' L R_{t+1}\right) = \mathrm{tr}(L\Pi). \tag{7}$$

An interesting linear strategy in its own right is to take positions in every asset based on the magnitude of its predictability by the signal S, whether that information comes from its own signal or from another asset's signal. This amounts to using Π itself as the position matrix ($L = \Pi'$) or using a positive multiple of Π:

Proposition 2 (Trading the Prediction Matrix) *Let* M *be an arbitrary positive semi-definite matrix. Then, the linear strategy with position matrix* $L = M\Pi'$ *has positive expected excess return:*

$$E(S_t' L R_{t+1}) \;=\; \mathrm{tr}(M\,\Pi'\Pi) \;=\; \mathrm{tr}((\Pi M^{1/2})'\,(\Pi M^{1/2})) \;\geq\; 0\,. \tag{8}$$

Moreover, the inequality is strict if and only if $M^{1/2}\Pi'$ *is not identically zero.*

We see that the prediction matrix plays two important roles: First, Π tells us the return of any linear strategy as seen in Proposition 1. Second, Π' is itself a return-generating linear strategy as seen in Proposition 2.

2.3 The Prediction Matrix vs. a Predictive Regression

We note that the prediction matrix is closely related to the following regression of the vector of returns on the vector of signals:

$$R_{t+1} = BS_t + \varepsilon_{t+1}. \tag{9}$$

Here, the regression coefficient is $B = \Pi E(S_t S_t')^{-1}$, which depends on the prediction matrix Π. Further, one could try to estimate the mean-variance efficient portfolio weights,

$$w_t^{\text{mean-variance}} = \text{Var}_t(\varepsilon_{t+1})^{-1} E_t(R_{t+1}) = \text{Var}_t(\varepsilon_{t+1})^{-1} \Pi E(S_t S_t')^{-1} S_t \tag{10}$$

but this would require the estimation of a large number of parameters and the inversion of two large estimated matrices. Instead, the factor literature has focused on the simple linear factor in (3) for tractability.

We seek to enhance the set of linear strategies without going all the way to a regression-based mean-variance approach. Our approach is focused on finding a linear portfolio L that "works well" across signals. To accomplish this goal, we consider a novel objective – defined next – with a tractable solution that lends itself to dimension reduction. We note that our focus on the prediction matrix means that our method is likely closer to the mean-variance solution (10) when signals and returns are properly scaled as discussed further in the empirical section.[1]

2.4 Objective Function

We consider the objective to maximize the expected return of a linear strategy subject to a portfolio constraint on the position matrix L:

$$\max_{L:\|L\|\leq 1} \; E\left(S_t' L R_{t+1}\right) . \tag{11}$$

We naturally need a portfolio constraint, since otherwise we can increase the expected return by simply increasing position sizes, e.g., the strategy $2L$ doubles the expected return of the

[1] Empirically, we use scale signals such that $S_{t,i} \in [-\frac{1}{2}, \frac{1}{2}]$, implying that $E(S_t S_t')$ becomes closer to being proportional to the identity matrix. Also, we consider test assets that are hedged (and scaled by volatility in the case of futures contracts with vastly different risk levels) such that $\text{Var}_t(\varepsilon_{t+1})$ is not too far from being proportional to the identity. These scalings mean that our optimal solution may approximate the regression based approach, but none of our theoretical results rely on these restrictions.

strategy L. To understand the specific constraint that we use in (11), note first that we are interested in a constraint that depends on the position matrix L, not the random portfolio holdings $w_t' = S_t' L$, since we are maximizing over position matrices.

Specifically, in (11) we maximize the expected return over the set of all position matrices with matrix norm of at most one. We define the standard matrix norm as $\|L\| = \sup\{\|Lx\| : x \in \mathbb{R}^m \text{ with } \|x\| = 1\}$, where $\|x\| \equiv (\sum_i x_i^2)^{1/2}$ is the standard Euclidean norm of a vector $x \in \mathbb{R}^N$ and we note that $\|L\| = \|L'\|$.

The economic meaning of this constraint is that we consider strategies with a bounded portfolio size. The linear strategy has portfolio weight $S_t' L$, which has a size of $\|L' S_t\| \le \|L'\| \, \|S_t\| \le \|S_t\|$ when $\|L\| \le 1$. So we consider linear strategies where the position size is always bounded by the position size of the simple strategy. Further, if S_t is normalized such that $\|S_t\| = 1$ for all signals, then the linear strategy has a position size that is similarly bounded, $\|L' S_t\| \le 1$.[2]

We can also interpret the objective function as a robust mean-variance problem. For example, when the return variance-covariance matrix is given by $\Sigma_{R,t} = \sigma^2 \mathrm{Id}$ for some $\sigma \in \mathbb{R}$, the objective function (11) is identical to the following:

$$\max_L E(S_t' L R_{t+1}) \text{ subject to } \max_{S:\mathrm{Var}_t(S' R_{t+1}) \le 1} \mathrm{Var}_t(S' L R_{t+1}) \le 1 \,. \tag{12}$$

In words, we maximize expected return subject to a risk constraint. This risk constraint is robust in the sense that we require that the variance is bounded *regardless* of the signal realization S. This robust objective where we maximize risk with respect to S, rather than considering the risk conditional on S, is natural for a linear strategy — since the position matrix L is constant over time and should "work" for all signals. To see the equivalence of

[2] Here we discuss "position size" in terms of the Euclidian norm, while the notional leverage of a position x is normally calculated as $\|x\|_1 = \sum_k |x_k|$. However, the portfolio constraint $\|L\| \le 1$ also implies a constraint on notional leverage. Indeed, since $\|x\|_1 \le \|x\| n^{1/2}$, notional leverage is bounded: $\|L' S_t\|_1 \le \|L' S_t\| n^{1/2} \le n^{1/2}$.

(11) and (12), note that

$$\max_{S:\mathrm{Var}_t(S'R_{t+1})\leq 1} \mathrm{Var}_t(S'LR_{t+1}) = \max_{S:S\neq 0} \frac{\mathrm{Var}_t(S'LR_{t+1})}{\mathrm{Var}_t(S'R_{t+1})} = \max_{S:S\neq 0} \frac{\sigma^2\|LS\|^2}{\sigma^2\|S\|^2} = \|L\|^2\,. \tag{13}$$

The risk constraint says that the risk of the linear strategy should be at most as high as that of the simple factor. Another way to get the same result is to require that the risk is limited when the signals are limited, $\max_{S:\|S\|\leq 1} \mathrm{Var}_t(S'LR_{t+1}) \leq \sigma^2$.

The assumption $\Sigma_{R,t} = \sigma^2\mathrm{Id}$ is appropriate if volatilities are similar in the cross section (or has been made similar, as we do our empirical study of futures) and if the correlation matrix is close to, or has been shrunk to, the identity—and such shrinkage can be useful in an optimization setting (Pedersen et al. (2020)). In any event, when we have general variance-covariance matrix $\Sigma_{R,t}$, then our portfolio constraint $\|L\| \leq 1$ still serves to control both risk, leverage, and the portfolio norm.[3] Further, we show how to solve a robust mean-variance problem for general $\Sigma_{R,t}$ in Appendix A. The appendix also shows how to solve the mean-variance problem with a risk penalty driven by risk aversion (instead of the risk constraint used here).

2.5 Optimal Linear Strategies

Given the objective (11), the optimal strategy is surprisingly elegant as shown in the following proposition.

[3]The portfolio constraint $\|L\| \leq 1$ implies a limit on the portfolio norm by definition, a leverage limit described in Footnote 2, and the following risk limit:

$$\max_{S:\|S\|\leq 1} \sqrt{\mathrm{Var}_t(S'LR_{t+1})} \;=\; \|\Sigma_{R,t}^{1/2}L'\| \leq \|\Sigma_{R,t}^{1/2}\|\|L\| \leq \|\bar{\Sigma}\|$$

when the variance-covariance matrix is bounded, $\Sigma_{R,t} \leq \bar{\Sigma}$.

Proposition 3 *The solution to (11) is given by $L = M\Pi'$ with $M = (\Pi'\Pi)^{-1/2}$, and*

$$\max_{L:\|L\|\le 1} E\left(S_t' L R_{t+1}\right) = \sum_{i=1}^{N} \bar{\lambda}_i \,,$$

where $\bar{\lambda}_1 \ge \cdots \ge \bar{\lambda}_N$ are the singular values of Π, i.e., the eigenvalues of $(\Pi'\Pi)^{1/2}$.

Proposition 3 shows that the prediction matrix Π is integral to optimal linear strategies based on the signal S_t. The solution is given in closed firm and of the form described in Proposition 2. Further, the solution depends on the singular values of Π, which in general depend on all its elements – not just the diagonal – so it has the potential to outperform the simple factor.

2.6 Principal Portfolios

We next decompose the optimal solution into a collection of linear strategies that we refer to as **principal portfolios** (PP) of the signal S. Principal portfolios are the building blocks that sum to form the optimal linear strategy in Proposition 3.

The construction of PPs uses the singular value decomposition of Π. Namely, let

$$\Pi = U \bar{\Lambda} V', \tag{14}$$

where $\bar{\Lambda} = \operatorname{diag}(\bar{\lambda}_1, \cdots, \bar{\lambda}_N)$ is the diagonal matrix of singular values, and U, V are orthogonal matrices with columns denoted u_k and v_k, respectively. Now, the optimal L from Proposition 3 can be rewritten as

$$(\Pi'\Pi)^{-1/2}\Pi' = V\bar{\Lambda}^{-1}V'V\bar{\Lambda}U' = VU' = \sum_{k=1}^{N} v_k u_k' \,.$$

We define the k^{th} principal portfolio as the linear strategy with position matrix $L_k = v_k\,(u_k)'$, which has a return of

$$PP_{t+1}^k = S_t' \underbrace{v_k\, u_k'}_{L_k} R_{t+1} = \underbrace{S_t' v_k}_{S_t^{v_k}} \underbrace{u_k' R_{t+1}}_{R_t^{u_k}}. \tag{15}$$

We see that there are two interpretations of a principal portfolio. First, it is a simple linear strategy with a position matrix L_k of rank 1. Second, it is a strategy that trades the portfolio u_k (with return $R_t^{u_k}$) based on the signal coming from the portfolio v_k (i.e., with signal $S_t^{v_k}$). This latter interpretation plays a key role when we discuss the beta components in the next section.

The construction of principal portfolios is actually very simple. All you need to do is use your favorite program to compute the singular value decomposition of Π (a standard feature of most computing programs), take the column vectors of U and V, and you are done.

Decomposing the optimal strategy into its principal portfolios is similar to decomposing the variance into the principal components. The difference is that principal component analysis decomposes the *variance*, but principal portfolio analysis decomposes the *expected return*. Just like the variance of each principal component equals its corresponding eigenvalue, the expected return of each principal portfolio is its singular value:

$$E(PP_{t+1}^k) \;=\; \mathrm{tr}(\Pi\, v_k u_k') \;=\; \mathrm{tr}(U\,\bar{\Lambda}\,V' v_k u_k') \;=\; \mathrm{tr}(U\,\bar{\Lambda}\,e_k u_k') \;=\; \mathrm{tr}(\bar{\lambda}_k u_k u_k') \;=\; \bar{\lambda}_k\,. \tag{16}$$

The following proposition summarizes the results of this section.

Proposition 4 *The expected return of each principal portfolio is given by its corresponding singular value,*

$$E(PP_{t+1}^i) = \bar{\lambda}_i, \tag{17}$$

and the sum of principal portfolios is the optimal linear strategy:

$$\max_{\|L\|\leq 1} E(S_t' L R_{t+1}) = E\left(\sum_{i=1}^{N} PP_{t+1}^i\right) = \sum_{i=1}^{N} \bar{\lambda}_i. \tag{18}$$

The following example provides some intuition for this result.

Example (Signals are Expected Returns). If signals are equal to conditional expected returns, $S_{i,t} = E_t(R_{i,t+1})$, one might question the usefulness of principal portfolios. But even in this simple setting principal portfolios are insightful about the optimal strategy. In this case, the prediction matrix reduces to the unconditional second moment of S_t, denoted Σ_S,

$$\Pi = E(R_{t+1} S_t') = E(E_t(R_{t+1}) S_t') = E(S_t S_t') = \Sigma_S. \tag{19}$$

Therefore, principal portfolios are given by the principal components of Σ_S. The matrix Σ_S describes the joint dynamics in conditional expected returns. Its leading principal component describes the portfolio of assets with the most variable expected return. In other words, the first principal component of Σ_S is the most "timeable" portfolio. It is the most attractive portfolio to trade for an investor facing a position size constraint and delivers the highest unconditional average profitability. The second principal component is the next most attractive, and so on. Singular values of Π relate to variability of expected returns, which explains why unconditional expected returns on principal portfolios are pinned down by the size of singular values in (17). And in this example, all principal portfolios have positive expected excess returns (assuming that Σ_S is non-degenerate), so the optimizing investor holds them all, as in (18). However, if the prediction matrix is estimated with error, it may be more robust to focus on the top PPs, as discussed in Section 5.

3 Optimal Alpha and Beta Strategies

We next derive the return of the optimal alpha and beta strategies, and show how these can be decomposed into principal portfolios, just as in the general solution in Propositions 3–4.

3.1 Alpha-Beta Symmetry Decomposition

To decompose the return into its alpha and beta components, we must first specify the factor that is used to compute the beta. In other words, how do we characterize the riskiness of linear strategies? To address this question, Lemma 1 introduces a factor having the special property that $S_{i,t}$ exactly describes asset i's conditional exposure to the factor.

Lemma 1 (Characteristics as Covariances) *Define the factor F_{t+1} as*

$$F_{t+1} = \left(\frac{1}{S_t' \left(\Sigma_{R,t}\right)^{-1} S_t} \left(\Sigma_{R,t}\right)^{-1} S_t \right)' R_{t+1}. \tag{20}$$

F_{t+1} is the unique tradable factor with the property that

$$S_{i,t} \;=\; \frac{\mathrm{Cov}_t(R_{i,t+1}, F_{t+1})}{\mathrm{Var}_t(F_{t+1})}. \tag{21}$$

This factor (referred to as the "latent factor" henceforth) is an economically important reference point.[4] It has a natural risk factor interpretation—it is the factor that unifies the expected return interpretation of $S_{i,t}$ and the risk exposure interpretation of $S_{i,t}$. No other factor based on S shares this property (including the literature's standard "simple factor," $\widetilde{F}$).

[4] Kelly et al. (2020a,b) propose a modeling approach and extensive empirical study of this point. Lemma 1 shows that we can always think of any signals as exposures to a factor, but it does not necessarily imply that the return predictability embodied by S is "rational" in the sense that the factor F covaries with risks that investors care about, namely the pricing kernel, and that certain eigenvalue bounds are satisfied, as discussed later.

To interpret this result, it is again helpful to consider the example in which $S_t = E_t(R_{t+1})$. In this case, F_{t+1} is the conditional tangency portfolio, and is thus the tradable representation of the pricing kernel. As a result, the expected returns and the factor exposures are equal up to a constant positive scale factor. And, being the tangency portfolio, all assets have zero alpha versus this factor in the absence of arbitrage. Importantly, while this factor is useful for interpreting some of our results, none of our results rely on actually observing F—since we don't observe it. We don't observe F because it depends on the conditional variance-covariance matrix $\Sigma_{R,t}$, which can only be estimated with noise. Instead, we develop methods that can beat the simple factor $\widetilde{F}$ without relying on observing, much less inverting, $\Sigma_{R,t}$.

The risk factor interpretation of the latent factor F is helpful in characterizing the risk and return of signal-based linear strategies. To characterize risk and return of linear strategies, recall that any square matrix $B \in \mathbb{R}^{N \times N}$ is decomposable into its *symmetric part*, $B^s = \frac{1}{2}(B + B')$, and its *anti-symmetric part*, $B^a = \frac{1}{2}(B - B')$, where $B = B^s + B^a$. The symmetric part equals its own transpose while the anti-symmetric part equals minus its own transpose, and both parts have a number of interesting properties. For example, since $B^a = -(B^a)'$, the anti-symmetric part has zeros along the main diagonal.

Hence, any linear strategy can be seen as a sum of a symmetric and anti-symmetric part, $L = L^s + L^a$. This decomposition has a deep economic interpretation, as we show next.

Proposition 5 (Alpha-Beta Symmetry Decomposition) *The conditional latent factor exposure and expected return of the strategy* $R^{w_t}_{t+1} = S_t' L R_{t+1} = S_t' L^s R_{t+1} + S_t' L^a R_{t+1}$ *is*

$$\underbrace{\frac{\mathrm{Cov}_t(R^{w_t}_{t+1},\ F_{t+1})}{\mathrm{Var}_t(F_{t+1})}}_{\textit{factor beta}} = S_t' L^s S_t \tag{22}$$

$$E(R^{w_t}_{t+1}) = \mathrm{tr}(L^s \Pi^s) + \mathrm{tr}(L^a \Pi^a)\,. \tag{23}$$

This proposition shows that the risk (beta to the latent factor) of a linear strategy $S_t'L$ is purely determined by its symmetric part; while the expected return is determined by both the symmetric or anti-symmetric parts via their interaction with the respective components of the prediction matrix, Π^s and Π^a.

This proposition has wide-ranging implications. First, an anti-symmetric strategy is always factor neutral. Second, an anti-symmetric strategy can nevertheless deliver positive returns as long as $\Pi^a \neq 0$. In this case, an anti-symmetric strategy can deliver positive expected return with zero factor exposure, that is, pure alpha with respect to F! The fact that factor exposures depend only on the symmetric component, L^s, regardless of the symmetry of Π is a direct implication of Lemma 2.

Lemma 2 *For any symmetric matrix $B \in \mathbb{R}^{N\times N}$ and any anti-symmetric matrix $A \in \mathbb{R}^{N\times N}$, we have $\operatorname{tr}(BA) = \operatorname{tr}(AB) = 0$ and $x'Ax = 0$ for all vectors $x \in \mathbb{R}^N$.*

In other words, anti-symmetric matrices nullify certain matrix multiplications, which translates into factor-neutrality of trading strategies.

Proposition 5 also shows how symmetric strategies can deliver returns via the interaction with Π^s. Symmetric strategies have a beta to the factor given by $S_t'L^sS_t$, which can be positive or negative. A symmetric strategy has positive factor beta for all possible realizations of the signal vector S_t if and only if L is positive definite. So, as we analyze in more detail in the next section, eigenvalues are key to understanding both risk and return. Finally, a symmetric strategy that always has negative factor beta corresponds to a negative definite L.

As an example application of Proposition 5, consider the riskiness of the simple factor $\widetilde{F}$ in (3), which is a linear strategy with identity position matrix ($L = \mathrm{Id}$) as seen in Equation 4. Hence, this simple factor has expected return $\operatorname{tr}(L^s\Pi^s) = \operatorname{tr}(\Pi^s) = \operatorname{tr}(\Pi)$ and it always has a positive exposure to the latent factor, $\operatorname{Cov}_t(\widetilde{F}_{t+1}, F_{t+1}) = \operatorname{Var}_t(F_{t+1})S_t'S_t > 0$.

The optimal linear strategy in Proposition 3 and the corresponding principal portfolios do not distinguish whether expected returns originate from factor exposure or alpha. In the remainder of this section, we show that Π^s and Π^a lie at the heart of optimal symmetric and anti-symmetric trading strategies. We derive symmetric and anti-symmetric analogues of principal portfolios, and show that these are the building blocks to optimal symmetry-decomposed strategies with either pure factor exposure and no alpha, or pure alpha and no factor exposure.

Said simply, symmetry is beta, and anti-symmetry is alpha. We next derive the optimal beta and alpha, respectively.

3.2 Symmetric Strategies: Principal Exposure Portfolios

As shown in equation (4), the simple factor is a simple symmetric linear strategy that trades each asset based on its own signal. The idea that symmetric strategies trade based on their own signals holds more generally. In particular, any strategy that scales the portfolio position in proportion to the signal aggregated to the portfolio level—that is, any portfolio that trades on the portfolio's own signal—is a symmetric strategy.

To see this, consider a portfolio $w \in \mathbb{R}^N$. The portfolio w has excess return $R^w_{t+1} = \sum_i w_i R_{i,t+1}$. Aggregating the underlying signals based on these weights means that the portfolio-level own signal is $S^w_t = \sum_i w_i S_{i,t}$. Trading the portfolio based on its own signal means using its signal as portfolio weight, which generates a return of

$$S^w_t R^w_{t+1} = S'_t w w' R_{t+1}. \tag{24}$$

We see that trading the portfolio based on its own signal is a linear strategy with a symmetric, positive semi-definite position matrix $L = ww'$. It's expected return is therefore

$$E\left(S^w_t R^w_{t+1}\right) = E\left(w' S_t R'_{t+1} w\right) = w'\Pi w \ = w'\Pi^s w, \tag{25}$$

which shows, in a different way from (23), that the return depends only on the symmetric part of the prediction matrix (the last equality uses Lemma 2).

All symmetric linear strategies can be represented as combinations of portfolios traded based on their own signals. This is achieved through the eigendecomposition of any symmetric position matrix L based on its eigenvalues λ_k and orthonormal eigenvectors w_k:

$$L = \sum_{k=1}^{K} \lambda_k \, w_k \, (w_k)' \, . \tag{26}$$

Furthermore, the position matrix satisfies our portfolio constraint $\|L\| \leq 1$ if $|\lambda_k| \leq 1$ for all k.

This result provides intuition for why symmetric linear strategies have factor exposure. They trade portfolios based on the portfolio's own signal. In this sense, they do what the signal prescribes, which anchors their behavior to that of the factor F. For example, if the signal $S_{i,t}$ is each security's momentum, then a symmetric linear strategy consists of trading different portfolios based on their own momentum—in the same spirit as the factor.

We next consider *optimal* symmetric linear strategies. We know from (23) that an optimal symmetric strategy maximizes $\mathrm{tr}(L\Pi^s)$, so we can use Proposition 3 with Π replaced by Π^s. The solution can be written simply based on the eigenvalue-decomposition

$$\Pi^s \;=\; W\Lambda^s W' = \sum_{k=1}^{N} \lambda_k^s \, w_k^s \, (w_k^s)' \, , \tag{27}$$

where $W = (w_1^s, ..., w_N^s)$ is the matrix of eigenvectors corresponding to the eigenvalues $\lambda_1^s \geq ... \geq \lambda_N^s$. We see that the optimal symmetric strategy is:

$$(\Pi^s\Pi^s)^{-1/2}\Pi^s = W|\Lambda^s|^{-1}W' \; W\Lambda^s W' = W \, \mathrm{sign}(\Lambda^s) W' = \sum_{k=1}^{N} \mathrm{sign}(\lambda_k^s) \, w_k^s \, (w_k^s)' \, . \tag{28}$$

The optimal strategy decomposes into N components, which we call **principal exposure**

portfolios (PEPs). That is, the k^{th} PEP is a linear strategy with position matrix $w_k^s(w_k^s)'$ and a return of

$$PEP_{t+1}^k = S_t^{w_k^s} R_{t+1}^{w_k^s} = S_t' w_k^s (w_k^s)' R_{t+1}. \tag{29}$$

The next result characterizes the returns of PEPs:

Proposition 6 *The expected return of each PEP is equal to its corresponding eigenvalue*

$$E(PEP_{t+1}^k) = E\left(S_t^{w_k^s} R_{t+1}^{w_k^s}\right) = E\left(S_t' w_k^s (w_k^s)' R_{t+1}\right) = \lambda_k^s, \tag{30}$$

Going long PEPs with positive eigenvalues and short those with negative is the optimal symmetric linear strategy:

$$\max_{\|L\|\leq 1,\ L=L'} E(S_t' L R_{t+1}) = \sum_{k=1}^{N} \text{sign}(\lambda_k^s) E(PEP_{t+1}^k) = \sum_{k=1}^{N} |\lambda_k^s|. \tag{31}$$

The first result shows that returns of PEPs equal their eigenvalues. The second result shows that the collection of PEPs yield the symmetric linear strategy with the highest unconditional expected return, subject to leverage constraint $\|L\| \leq 1$. This optimal performance is achieved by trading PEPs while accounting for the direction of their predictability. The optimal strategy takes long positions of size 1 in all PEPs with positive expected returns (i.e., positive eigenvalues) and short positions of size -1 in PEPs with negative expected returns.

We next consider how the PEPs relate to the simple factor $\widetilde{F}$.

Proposition 7 (Beating the Factor) *The simple factor, $\widetilde{F}$, can be decomposed as*

$$\widetilde{F}_{t+1} = \sum_{i=1}^{N} S_{i,t} R_{i,t+1} = \sum_{k=1}^{N} S_t^{w_k^s} R_{t+1}^{w_k^s} = \sum_{k=1}^{N} PEP_{t+1}^k. \tag{32}$$

If all eigenvalues are non-negative, $\lambda_k^s \geq 0$, then $\widetilde{F}$ is the optimal symmetric strategy. Otherwise, $\widetilde{F}$ has a lower expected return than buying the subset of PEPs with positive eigenvalues, which is lower than that the optimal strategy from Proposition 6:

$$E\left(\widetilde{F}_{t+1}\right) = \sum_{k=1}^{N} \lambda_k^s \; \leq \; \sum_{k:\lambda_k^s>0} \lambda_k^s \leq \sum_{k=1}^{N} |\lambda_k^s| \,. \tag{33}$$

Interestingly, the simple factor actually equals the sum of all PEPs as seen in (32). In fact, $\widetilde{F}$ can be viewed as the sum of all possible returns of symmetric strategies, not just the PEPs. Namely, for any orthonormal basis of portfolios $B = \{b_k\}_{k=1}^N$, we have that $BB' = \mathrm{Id}$ and, hence,

$$\sum_{i=1}^{N} S_{i,t}\, R_{i,t+1} \;=\; S_t' R_{t+1} \;=\; S_t' B B' R_{t+1} \;=\; \sum_{k=1}^{N} S_t^{b_k}\, R_{t+1}^{b_k} \,. \tag{34}$$

That is, *trading the simple factor on stocks is equivalent to trading it on portfolios.*

The fact $\widetilde{F}$ equals the sum of PEPs together with Equation (30) imply that the expected excess return of the simple factor equals the sum of the eigenvalues, $E(\widetilde{F}_{t+1}) \;=\sum_{k=1}^N \lambda_k^s$. Therefore, when a researcher documents that a simple strategy $\widetilde{F}_{t+1}$ has significantly positive average returns, we learn that the sum of eigenvalues of Π^s is positive.

When all eigenvalues are non-negative, the simple factor is in fact optimal among all symmetric strategies. So, in this case, the simple strategy is not just simple — our analysis sheds new light on why it is a natural starting point.

When $E(\widetilde{F}_{t+1}) \;=\; \sum_{k=1}^N \lambda_k^s \;>\; 0$, eigenvalues can nevertheless be negative. Negative eigenvalues correspond to those surprising PEPs that are *negatively* predicted by their own signals. When there exist PEPs with negative eigenvalues, we can beat the simple factor by leaving these PEPs out, buying only the PEPs that "work". Trading all the PEPs with positive eigenvalues is the optimal strategy among all linear strategies that always have positive factor exposure (i.e., among strategies with positive semi-definite L).

If we are willing to have a factor exposure that may switch sign, we can achieve an ever higher return. Indeed, negative eigenvalues also describe useful prediction patterns, just in the opposite direction. Therefore, an investor can do even better by also shorting the PEPs with negative eigenvalues, as shown in equation (33).

The analogy between principal portfolio analysis and principal component analysis is remarkably close when we focus on the symmetric part of the prediction matrix as highlighted in Table 1. As seen in the table, PCA and PPA share five key properties. While PCA decomposes the variance into its components, PPA decomposes the expected excess return. Both have similar connections to eigenvalues, orthogonality, the trace, and optimality across orthonormal portfolios.

Example (Diagonal Prediction Matrix). Suppose there is no cross-predictability and signals are mean zero ($E(S_{j,t}) = 0$). Then $\Pi_{ij} = E(R_{i,t+1}S_{j,t}) = 0$ for all $i \neq j$. Hence, Π is symmetric, so there are no anti-symmetric (zero exposure) strategies within Π. Furthermore, the PEPs are simply the unit vectors, $w_k^s = \mathbf{1}_k$.[5] The optimal strategy is long assets with positive own-predictability and short those with negative own-predictability.

3.3 Anti-symmetric Strategies: Principal Alpha Portfolios

We now turn to anti-symmetric linear trading strategies. The most basic type of anti-symmetric matrix has the form $L = xy' - yx'$, which we call *rank-2 anti-symmetric strategies*. These building blocks are analogous to the rank-1 symmetric trading strategies, $L = ww'$, that are the basic building blocks of all symmetric trading strategies. Each rank-2 anti-symmetric strategy generates a return of

$$S_t'(x_j y_j' - y_j x_j')R_{t+1} \;=\; S_t^{x_j} R_{t+1}^{y_j} - S_t^{y_j} R_{t+1}^{x_j}. \tag{35}$$

[5] Here, $\mathbf{1}_k = (0, \ldots, 1, 0, \ldots, 0)'$, where 1 is in the k'th position.

The first part of this portfolio is the return to trading the portfolio y_j based on the signal coming from the portfolio x_j. In other words, a strong signal for x_j ($S_t' x_j$) recommends scaling up the position in y_j ($y_j' R_{t+1}$), and this generates a return of $S_t^{x_j} R_{t+1}^{y_j}$. The second part is similar but flips the roles x_j and y_j and shorts the associated strategy (due to the minus sign). Thus, anti-symmetric strategies are understandable as long-short strategies that trade two portfolios against each other based on the strength of each other's signal.

The next result shows that all anti-symmetric strategies can be represented as a sum of these basic building blocks.

Lemma 3 *Any anti-symmetric matrix A has an even number $2K$ of non-zero eigenvalues. The non-zero eigenvalues are purely imaginary and come in complex-conjugate pairs: $i\lambda_k$ and $-i\lambda_k$. The corresponding orthonormal eigenvectors are $z_k = \frac{1}{\sqrt{2}}(x_k + iy_k)$ and the complex conjugate $\bar{z}_k = \frac{1}{\sqrt{2}}(x_k - iy_k)$, where $x_k, y_k \in \mathbb{R}^N$ with $\|x_k\| = \|y_k\| = 1$, $x_k' y_k = 0$, and $x_k' x_l = x_k' y_l = y_k' y_l = 0$ for all $k \neq l$, $k, l \leq K \leq N/2$. The corresponding eigendecomposition is given by*

$$A = \sum_{k=1}^{K} \lambda_k (x_k y_k' - y_k x_k'). \tag{36}$$

In other words, this result shows how to break any anti-symmetric matrix into its basic building blocks of form $x_k y_k' - y_k x_k'$ for $k = 1, ..., K$.[6] But why does this result in zero conditional factor exposure, as guaranteed by Proposition 5? The next example helps develop intuition for the absence of factor risk in anti-symmetric strategies.

Example (Beta-neutral Strategy). Consider an economy of N assets that satisfies the CAPM, save for asset 1, which has a positive alpha. That is, $E_t(R_{i,t+1}) = \alpha \mathbf{1}_{i=1} + \beta_{i,t}\theta_t$, where[7] $\theta_t \geq 0$ is the market risk premium, $\beta_{i,t}$ is the conditional CAPM beta of stock i, and

[6] An anti-symmetric strategy satisfies the portfolio constraint, $\|A\| \leq 1$ as long as $|\lambda_k| \leq 1$ in (36).

[7][8]

$\alpha > 0$. Suppose further that signals are defined to be the conditional betas, $S_{i,t} = \beta_{i,t}$. A standard beta-neutral strategy to exploit this scenario takes a long position in asset 1 having size equal to 1 (i.e., the size is set equal to the factor's beta on itself). The conditional beta from the long position is equal to $\beta_{1,t}$, so beta-neutrality is achieved with a position of $-\beta_{1,t} = -S_{1,t}$ in the factor. This strategy is a rank-2 anti-symmetric strategy with $L = yx' - xy'$. The long position in asset 1 corresponds to $x = (1, 0, \cdots, 0)'$, and the short position in the factor corresponds to $y = (1, 1, ..., 1)'$. In other words, the beta-neutral strategy has zero symmetric component, non-zero anti-symmetric component, and positive expected return, rendering it an alpha strategy with expected return:

$$E(S_t' L R_{t+1}) = E(\beta_t'(yx' - xy')R_{t+1}) = E\left(\sum_i \beta_{i,t} R_{1,t+1} - \beta_{1,t} \sum_i R_{i,t+1}\right) = \alpha E\left(\sum_{i=2}^{N} \beta_{i,t}\right),$$

which is positive as long as betas are positive on average. This is not the only pure alpha strategy, as a long position in asset 1 can be hedged with any other asset or combination of assets. Below, we show how to construct optimal pure alpha strategies with respect to the factor F using the eigendecomposition of Π^a.

The example illustrates that the fundamental $yx' - xy'$ structure underlying all anti-symmetric strategies is closely related to the familiar approach to factor neutralization. To eliminate factor exposures, the position size in each must be equal to the factor exposure of the other, and with appropriately opposing signs.

Next, we derive *optimal* anti-symmetric strategies. The first step is to apply the eigen-decomposition in (36) to the anti-symmetric part of the transposed prediction matrix, $(\Pi^a)'$. By Lemma 3, the matrix $(\Pi^a)'$ has $2N^a$ non-zero and purely imaginary eigenvalues, $i\lambda_k^a$ and $-i\lambda_k^a$, for some $N^a \leq N/2$,. Their imaginary parts, $\lambda_k^a \in \mathbb{R}$, can be ordered as

$$\lambda_1^a \geq \cdots \geq \lambda_{N^a}^a \geq 0 \geq -\lambda_{N^a}^a \geq \cdots \geq -\lambda_1^a. \tag{37}$$

For each eigenvalue λ_j^a, we denote the corresponding real and imaginary parts of the eigenvectors by x_j and y_j, respectively.

We define the j^{th} **principal alpha portfolio (PAP)** as the linear strategy based on the j^{th} eigenvectors: $L_j = x_j y_j' - y_j x_j'$ for $j = 1, ..., N^a$. We note that, since $N^a \leq N/2$, there exist at most $N/2$ principal alpha strategies and they are orthonormal (Lemma 3). The PAP buys portfolio y_j based on the signal coming from the portfolio x_j and simultaneously shorts portfolio x_j based on the signal from y_j. Similar to the result for PEPs, we find that PAP expected returns are proportional to their eigenvalues and that the sum of PAPs is in fact the optimal anti-symmetric linear trading strategy.

Proposition 8 *A principal alpha strategy has expected return $E(PAP_{t+1}^j) = 2\lambda_j^a$ and zero factor exposure. The sum of PAPs is the optimal anti-symmetric linear strategy:*

$$\max_{\|L\|\leq 1,\ L=-L'} E(S_t' L R_{t+1}) = \sum_{k=1}^{N^a} E(PAP_{t+1}^k) = \sum_{k=1}^{N^a} 2\lambda_j^a. \tag{38}$$

The next example helps illustrate the properties of PEPs and PAPs.

Example (Constant Signals). Suppose that signals are constant over time, $S_t = S$.[9] In this case, the prediction matrix is especially simple, $\Pi = E(R_{t+1} S_t') = RS'$, where we use the short-hand notation $R := E(R_{t+1})$. We can now compute the PEPs and PAPs explicitly.

First, consider a case in which returns align with signals exactly, $R = S$. In this case, we have $\Pi = SS'$. This matrix is symmetric and has a rank of one. Hence, there is a single principal exposure portfolio with a non-zero eigenvalue, namely the eigenvector S, and no principal alpha portfolios. Therefore, this PEP is the only meaningful portfolio, and it is the same as the simple factor, S, with expected return $S'R = R'R > 0$.

[9]As a concrete example, consider sorting stocks into value (book-to-market) deciles, using the decile portfolios as the baseline assets, and using a value signal defined as the decile number of each asset as the predictive signal. This is in contrast to, for example, forming assets as value-sorted portfolios, but using portfolio momentum as the trading signal. In this case signals are far from constant over time, and this is what we do empirically.

Next, consider the case in which expected returns do *not* line up perfectly with the signal. Then $\Pi = RS'$ is no longer symmetric. The symmetric part is $\Pi^s = 0.5(RS' + SR')$, which has a rank of 2. Hence, Π^s has at most two non-zero eigenvalues, $\lambda_1^s = 0.5(R'S + \|R\|\,\|S\|) > 0 \geq \lambda_N^s = 0.5(R'S - \|R\|\,\|S\|)$ and the corresponding PEPs are[10]

$$w_1^s = c_1^s \left(\frac{R}{\|R\|} + \frac{S}{\|S\|} \right), \; w_N^s = c_N^s \left(\frac{R}{\|R\|} - \frac{S}{\|S\|} \right),$$

where c_1^s, c_N^s are constants chosen such that $\|w_1^s\| = \|w_N^s\| = 1$. We see that the first principal exposure portfolio bets on securities with high average returns and high signals, while the last PEP bets on securities with high average returns and low signals. The negative eigenvalue PEP isolates losses due to the erroneous component of S and exploits them with a short position.

In this example, the prediction matrix also has an anti-symmetric part. The strategy that trades this is $L = \Pi^{a\prime} = 0.5(SR' - RS')$. To derive the PAP, note that $\Pi^{a\prime}$ has at most two non-zero eigenvalues with purely imaginary parts $\lambda_1^a = 0.5(\|R\|\,\|S\| - R'S)^{1/2} \geq 0 \geq \lambda_N^a = -\lambda_1^a$ and the corresponding PAP is the linear strategy with position matrix $L = xy' - yx'$, where[11]

$$y = c^a \left(R\|S\|^2 - S(R'S) \right), \; x = S/\|S\|.$$

The short part of the portfolio (x) is exactly the factor hedge. It is in place to ensure that the constraint (zero factor exposure) is satisfied. The remaining part of the problem is to find the highest average return subject to the constraint. Since the factor uses all (and only) the information in S, the remaining information that the PAP has at its disposal comes from the unconditional mean of returns. Thus the long side of the PAP (y) is determined by the

[10] These eigenvalues and eigenvectors can be verified by checking that $\Pi^s w_k^s = \lambda_k^s w_k^s$ for $k = 1, N$.

[11] c^a is determined such that $\|y\| = 1$

information in R that is missed by S, hence the emergence in y of the difference between R and S.

Example (Betting Against Beta: PAP is the new BAB). Proceeding from the prior example, suppose that the erroneous signals S are chosen to be the expected returns in an asset pricing model, i.e. $S_j = \text{Cov}(-M_t, R_{j,t})$ where M denotes the model's pricing kernel. Suppose further that there is less dispersion in true expected returns than predicted by the model—for simplicity, suppose that $\mathbf{1}'S/N = 1$ and suppose that $R = \mathbf{1}$, where $\mathbf{1}$ is the vector with all coordinates equal to one. Then the alpha portfolio arising from the anti-symmetric part of the prediction matrix has portfolio weight $w' = S'(SR' - RS') = (S'S)\mathbf{1}' - NS'$. This strategy goes long the equal-weighted portfolio (given by $R = \mathbf{1}$), while shorting the beta-weighted portfolio, S. To keep the portfolio beta-neutral, the equal-weighted portfolio (which is lower beta) is scaled up relative to the beta-weighted portfolio,[12] $S'S > N$. Hence, this strategy resembles the betting-against-beta (BAB) strategy of Frazzini and Pedersen (2014). This strategy has expected excess return of $w'R = NS'S - N^2 > 0$ and a beta of $w'S = NS'S - NS'S = 0$.

3.4 Static and Dynamic Bets

In the preceding examples, signals are constant, which makes the math particularly tractable to illustrate intuitive aspects of principal portfolios. But constant signals imply that there are only static trading opportunities. In general, signals fluctuate over time, and principal portfolios use information about both static and dynamic trading opportunities. The prediction matrix can be written as a sum of its static and dynamic components:

$$\Pi = E(R_{t+1}S_t') = E(R_{t+1})E(S_t') + \text{Cov}(R_{t+1}, S_t'). \tag{39}$$

[12] This result follows from Cauchy-Schwarz, which yields that $N^2 = (\mathbf{1}'S)^2 \leq (\mathbf{1}'\mathbf{1})(S'S) = NS'S$, and the inequality is strict since we assume that betas vary across stocks.

Suppose that signals do not predict future returns in the sense that $\text{Cov}(S_{i,t}, R_{j,t+1}) = 0$ for all i, j. In this case, Π simplifies to the constant signal example, $\Pi = E(R)\, E(S')$, and we have up to two PEPs and one PAP with strictly positive expected return, but these are purely based on the signals' time series average. The first term on the right side of equation (39) thus embodies information in the prediction matrix regarding "static bets."

The second term summarizes information in the prediction matrix regarding "dynamic bets." To focus purely on dynamic bets, then we can demean signals in the time series, looking at $\tilde{S}_{i,t} = S_{i,t} - E(S_{i,t})$. This redacts static information from Π and concentrates only on dynamic opportunities:

$$E(R_{t+1}\tilde{S}_t') = \text{Cov}(R_{t+1}, \tilde{S}_t') = \text{Cov}(R_{t+1}, S_t'). \tag{40}$$

Our approach allows both static and dynamic bets since both may be useful. Static bets are useful if they pick up that certain assets generally have higher returns, and if it's possible to time one's portfolio positions, then dynamic bets are profitable. We find in our empirical analysis that many of the effects we see are driven by dynamic bets.

To summarize, as the above examples illustrate, there are potentially two ways to earn alpha relative to the factor. The first stems from the observation that if Π^s has any negative eigenvalues, then shorting the corresponding PEPs yields a positive expected return with a negative factor exposure, which is alpha with respect to the factor. The second is to identify anti-symmetric strategies with positive expected returns. Because an anti-symmetric strategy is guaranteed to have zero factor exposure, it is also alpha to the factor.

4 Asset Pricing Tests: Positivity Bounds

We next propose a test for whether our signal S is an exposure (i.e., beta) to the true pricing kernel. Said differently, we wish to test whether the factor F corresponding to S is proportional to the true pricing kernel, $F_{t+1} \propto -M_{t+1}$ (or M's projection on the tradable space; recall that Lemma 1 shows how F is related to S). For example, we can consider signals given by betas to the market return, R^m_{t+1}, which corresponds to testing that the pricing kernel is of the form $M_{t+1} = a_t - b_t R^m_{t+1}$ for $a_t, b_t \in \mathbb{R}$ (i.e., the CAPM). Or, we can consider signals based on exposure to consumption, corresponding to testing that the pricing kernel is of the form $M_{t+1} = \beta u'(c_{t+1})/u'(c_t)$ (consumption CAPM).

Specifically, suppose that our signal $S_{i,t}$ is proportional to pricing kernel exposure, $\mathrm{Cov}_t(R_{j,t+1}, -M_{t+1})$, where we only assume proportionality (rather than equality) since we may not know the equity premium in the CAPM or the risk aversion in CCAPM. Then, signals should be closely related to expected returns. Indeed, the definition of a pricing kernel is a process M with $E_t((1 + R^f_t + R_{j,t+1})M_{t+1}) = 1$ for all assets, where R^f_t is the risk-free rate, which implies[13]

$$E_t(R_{j,t+1}) = (1 + R^f_t)\mathrm{Cov}_t(R_{j,t+1}, -M_{t+1}) = \theta_t \, S_{j,t} \,, \tag{41}$$

where $\theta_t > 0$ is a factor of proportionality due to the risk-free rate and to our assumption that the signal S is proportional to (but not necessarily equal to) the covariance.

For example, if we are testing the CAPM, then the signal $S_{j,t}$ is typically the market beta, $\beta_{j,t} = \mathrm{Cov}_t(R_{j,t+1}, R^m_{t+1})/\mathrm{Var}_t(R^m_{t+1})$. In this case, the expected excess return is $E_t(R_{j,t+1}) = E_t(R^m_{t+1})\beta_{j,t}$, so here θ_t is the market risk premium, $E_t(R^m_{t+1})$. We would like to develop a test that does not require knowledge of θ_t because we may not know $E_t(R^m_{t+1})$ (or the coefficients a_t, b_t in $M_{t+1} = a_t - b_t R^m_{t+1}$).

The key insight is that, when the signal is proportional to the beta to the pricing kernel, the prediction matrix must be symmetric and positive definite—regardless of the factor of

[13]To see this result, note that the definition of a pricing kernel applied for the risk-free asset (which has 0 excess return) yields $(1 + R^f_t)E_t(M_{t+1}) = 1$, which implies that $E_t(R_{j,t+1}M_{t+1}) = 0$ for excess returns. Therefore, $E_t(R_{j,t+1}) = (1 + R^f_t)E_t(M_{t+1})E_t(R_{j,t+1}) = (1 + R^f_t)(E_t(R_{j,t+1}M_{t+1}) - \mathrm{Cov}_t(R_{j,t+1}, M_{t+1})) = (1 + R^f_t)\mathrm{Cov}_t(R_{j,t+1}, -M_{t+1})$.

proportionality, θ. To see that, note that any off-diagonal element of the prediction matrix is

$$\Pi_{j,i} \;=\; E(S_{i,t}R_{j,t+1}) \;=\; E(S_{i,t}E_t(R_{j,t+1})) \;=\; E(\theta_t S_{i,t}S_{j,t}) \;=\; \Pi_{i,j}\,. \tag{42}$$

which shows that Π is symmetric. Further, we see that the prediction matrix is positive semi-definite since, for any $w \in \mathbb{R}^N$:

$$w'\Pi w = w'E(\theta_t S_t S_t')w = E(\theta_t[w'S_t]^2) > 0 \tag{43}$$

This finding provides new asset pricing tests as summarized here:

Proposition 9 (Positivity of Prediction Matrix) *If there exists $\theta_t \in \mathbb{R}$ such that*

$$E(R_{i,t+1}|\theta_t, S_t) \;=\; \theta_t S_{i,t} \tag{44}$$

for all i, then the corresponding prediction matrix Π is symmetric. If $\theta_t \geq 0$, then Π is positive semi-definite, and, equivalently, all the corresponding PEPs have non-negative expected returns and all PAPs have zero expected returns. The premise (44) holds, for example, if there is no arbitrage so a pricing kernel exists, and the signal $S_{i,t}$ is proportional to exposures to the pricing kernel.

The intuition behind this result follows from our earlier portfolio theory: We know that negative eigenvalues of Π^s and a non-zero Π^a give rise to alpha strategies (Sections 3.2 and 3.3, respectively). Since alpha strategies cannot exist in a rational asset pricing model, all eigenvalues of Π^s must be positive and Π^a must be zero. In other words, Π must be symmetric and positive semi-definite.

These restrictions provide novel asset pricing tests. One benefit of this approach is that we do not need to know θ_t, we just need to observe signals and returns, and then consider the positivity of the corresponding prediction matrix. Another helpful feature is that the test is unconditional, i.e., it relies on an unconditional expected value, $\Pi = E(R_{t+1}S_t')$, even if the underlying asset pricing model is conditional. Hence, while some tests require an understanding of how the risk premium varies over time or make assumptions to get from a conditional CAPM to an unconditional test, we have a test of the conditional CAPM (and other conditional models) based on an unconditional moment condition. Further, this restriction also tests cross-asset effects.

These restrictions are straightforward to implement in practice. To test symmetry, one can simply calculate average PAP returns and test whether they are statistically different from zero. To test positive definiteness, we can test whether all of eigenvalues of the Π^s are non-negative or, equivalently, whether the PEP returns are non-negative. Appendix D presents Central Limit Theorems (CLTs) that justify this approach. In particular, Proposition 16 provides a CLT for the distribution of eigenvalues of Π^s, Proposition 17 derives the CLT for Π^a, and Propositions 18 and 19 present CLTs for the returns on trading symmetric and anti-symmetric linear strategies.

We implement these tests empirically in Appendix G.2. Among other things, we find that these tests are powerful and able to reject the 5-factor Fama-French model. We note that our method also works when signals are noisy, as seen in the next result.

Proposition 10 (Noisy Signals) *Suppose that $S_{i,t} = \kappa_t E_t[R_{i,t+1}] + \eta_{i,t}$, where $\kappa_t \in \mathbb{R}$ and $\eta_{i,t} = \beta_t E_t[R_{i,t+1}] + \gamma_{i,t}$ with $E[\gamma_{i,t} R_{j,t+1}] = 0$ for all i, j. Then Π is symmetric and, if $\kappa_t + \beta_t \geq 0$, then Π is also positive semi-definite.*

Finally, Appendix C presents several extensions of these asset pricing tests. First, while Proposition 9 shows that the standard asset pricing condition (44) implies symmetry and positive definiteness, Proposition 12 shows that the reverse is also true. Propositions 13, 14, and 15 show how our results are modified when the model is mis-specified (due to a common alpha in returns or missing factor exposures).

5 Robust Strategies: Shrinkage via Principal Portfolios

Our theoretical analysis up to now has taken place in population with the prediction matrix, $\Pi = E(R_{t+1}S_t')$, known. In reality, Π is unknown and must be estimated. Unfortunately, this is a highly parameterized framework; it requires estimating N^2 parameters. The standard tradable factor approach from the literature (3) essentially restricts the set of linear strategies to a single parameter problem—i.e., signals are typically assessed based only on their average own-predictability $\sum_i E(S_{i,t}R_{i,t+1})$. This approach can be viewed as a regularization device that exploits a signal while imposing many restrictions to minimize the number of parameters. But these restrictions may be unnecessarily severe. They sacrifice any and all useful information about heterogeneity in own-predictability (differences among diagonal elements of Π) or cross-predictability (off-diagonal elements).

Principal portfolios are ideally suited to balance two considerations: 1) exploiting potentially rich information from throughout the predictability matrix, and 2) controlling parameterization to reduce overfit and ensure robust out-of-sample portfolio performance. In this section, we develop robust principal portfolio trading strategies by shrinking the predictability matrix.

The analysis in Sections 2 and 3 shows that a singular value decomposition of Π (or of its symmetric and anti-symmetric parts) finds orthonormal portfolios and orders them from highest expected return to lowest. This eigendecomposition has another great benefit in that it lends itself naturally to a convenient form of regularization. In particular, if we reconstitute the Π matrix by retaining only the K largest singular values and zeroing out the rest, we obtain the matrix of rank K that is as close as possible to the original Π. This idea is familiar from principal components analysis, which finds low-rank approximations to a variance-covariance matrix by zeroing out all but its largest eigenvalues.

The following proposition operationalizes the idea of robust optimal trading strategies by constraining the parameter space to position matrices with $\text{rank}(L) \leq K$. Here, K is a tuning parameter that can be chosen empirically. To add further generality and another convenient tuning parameter, we introduce the Schatten p-norm for a matrix L (see, Horn and Johnson (1991)):

$$\|L\|_p = \left(\sum_{k=1}^{N} |\bar{\lambda}_k(L)|^p\right)^{1/p},$$

where $\bar{\lambda}_k(L)$ is the k-th singular value of L and $p \in [1, \infty]$. The limiting case $p = \infty$ corresponds to the standard matrix norm $\|L\| = \|L\|_\infty$, whereas $p = 2$ corresponds to the sum of squares of all elements $\|L\|_2 = (\sum_{k,l} L_{l,k}^2)^{1/2}$ (Frobenius norm). Interestingly, we show that different matrix norms correspond to different ways of weighting the principal portfolios. These insights are formalized in the following proposition, which generalizes all the optimization problems that we considered so far (Propositions 3, 6, and 8).

Proposition 11 (General Solution) *Optimal portfolios subject to* $\mathrm{rank}(L) = K$ *and* $||L||_p \leq 1$*, where* $p = [1, \infty]$ *and* q *is defined by* $1/p + 1/q = 1$*, satisfy:*

1. *The solution with no symmetry constraints depends on the top* K *singular values,* $\bar{\lambda}_k$*, of* Π*:*

$$\max_{\|L\|_p \leq 1,\ \mathrm{rank}(L) \leq K} E(S_t' L R_{t+1}) = \left(\sum_{k=1}^{K} \bar{\lambda}_k^q\right)^{1/q}. \tag{45}$$

The optimal L *is* $S_t' L R_{t+1} = c \sum_{k=1}^{K} \bar{\lambda}_k^{q-1} PP_{t+1}^k$*, where* $c = \left(\sum_{k=1}^{K} \bar{\lambda}_k^q\right)^{-1/p}$*.*

2. *The solution when restricting to symmetric strategies depends on the set* $\mathcal{K}$ *of the* K *largest absolute eigenvalues* $|\lambda_k^s|$ *of* Π^s*:*

$$\max_{\|L\|_p \leq 1,\ \mathrm{rank}(L) \leq K,\ L = L'} E(S_t' L R_{t+1}) = \left(\sum_{k \in \mathcal{K}} |\lambda_k^s|^q\right)^{1/q}. \tag{46}$$

The optimal L is $S_t' L R_{t+1} = c \sum_{\mathcal{K}} |\lambda_k^s|^{q-1} \operatorname{sign}(\lambda_k^s) PEP_{t+1}^k$, *where* $c = \left(\sum_{\mathcal{K}} |\lambda_k^s|^q\right)^{-1/p}$.

3. *The solution when restricting to anti-symmetric strategies depends on the eigenvalues λ_k^a of Π^a:*

$$\max_{\|L\|_p \leq 1,\ \operatorname{rank}(L) \leq 2K,\ L=-L'} E(S_t' L R_{t+1}) = \left(2 \sum_{k=1}^{K} (\lambda_k^a)^q\right)^{1/q}. \tag{47}$$

The optimal L is $S_t' L R_{t+1} = c \sum_{k=1}^{K} (\lambda_k^a)^{q-1} PAP_{t+1}^k$, *where* $c = \left(2 \sum_{k=1}^{K} (\lambda_k^a)^{q-1}\right)^{-1/p}$.

Proposition 11 shows that optimal low-dimensional trading strategies are the same as the general optimality results proven earlier, with the exception that the strategies use only the leading principal portfolios. This is true regardless of whether one considers general linear strategies (L), symmetric and hence factor-exposed strategies ($L = L'$), or anti-symmetric pure alpha strategies ($L = -L'$). By truncating the strategy at the top K principal portfolios, these robust strategies replace the lesser singular values with zeros.[14] The lesser components may be dominated by noise, and are therefore likely to have poor out-of-sample performance. Zeroing them out regularizes the optimal strategy to control overfit and its adverse out-of-sample impact. The number of principal portfolios included in a robust strategy, K, determines the extent of regularization. It serves as a hyperparameter that can be controlled by the researcher or tuned via cross-validation.

What are the implications of the more general norm $\|\cdot\|_p$ in this proposition, and what economic role does it play? Proposition 11 shows that the optimal strategy is a weighted sum of principal portfolios for any norm. This result shows that PPs are very general building blocks. The choice of norm simply affects how the principal portfolios are weighted, which also illustrates the connection between the tuning parameters p and K: The less important PPs can be "zeroed out" by the choice of K and down-weighted by the choice of p.

[14] Note that singular values of a symmetric or an anti-symmetric matrix coincide with the absolute values of its eigenvalues.

At that same time, the idea of constraining trading strategy leverage in the optimization problem has a natural economic motivation—risk and institutional frictions impose leverage considerations on every real-world investor. The way real-world investors try to manage their leverage concerns is dictated in part by the performance of strategies in their opportunity set. This raises an interesting practical implication of Proposition 11. The norm exponent p can be treated as a hyperparameter that can be tuned via cross-validation. An investor that tunes p along with K in essence chooses the form of leverage constraint that lends itself to robust out-of-sample trading performance.

Interestingly, when $p = 2$, part 1 of the proposition is similar to trading a version of the Π matrix that has been estimated via a reduced rank regression (see, e.g., Velu and Reinsel (1998)).[15] Further, when $p = 2$ and we do not impose a rank restriction (i.e., we let $K = N$), then the solution is $L = \Pi'/\|\Pi\|_2$. So, in this case, we uncover the prediction matrix itself as the optimal strategy. For $p = \infty$, i.e. $q = 1$, the solution selects components that are large in absolute value, in the spirit of lasso applied to singular values, and with no rank restriction we recover Proposition 4.

The results in Sections 2 through 4 lay out a theoretical basis for principal portfolios, and Proposition 11 prescribes a machine learning approach to implementing principal portfolios in practice. Data-driven choices for hyperparameters K and p can allow the researcher to select the level of principal portfolio model complexity best suited for constructing optimal out-of-sample strategies.

6 Empirical Results

We next present two empirical applications of our method.

[15] Reduced rank regression (RRR) seeks to minimize the mean squared error $E(\|R_{t+1} - L'S_t\|^2) = E(\|R_{t+1}\|^2) - E(S_t'LR_{t+1}) + E(S_t'LL'S_t)$ under a rank constraint on the matrix L. By direct calculation, this objective is equivalent to maximizing $\mathrm{tr}(L\Pi) - \mathrm{tr}(LL'\Sigma_S)$. Thus, reduced rank regression amounts to maximizing the expected return, $\mathrm{tr}(L\Pi)$, with a punishment term for signal variance. If $\Sigma_S = \mathrm{Id}$, the punishment term coincides with $\|L\|_2^2$, and hence RRR is a modification of the problem solved in Proposition 11 for $p = 2$.

6.1 Fama-French Portfolio Momentum

Our first application uses principal portfolios on the 25 Fama-French portfolios.[16] As this is one of the simplest and most well-studied data sets in finance, it is an ideal empirical setting for demonstrating properties of our method in a transparent way. These portfolios are constructed by double-sorting U.S. stocks by their size (measured by market capitalization) and valuation ratio (book-to-market) and we use daily data from July 1963 until the end of 2019.

In order to have a simple time-varying predictive signal for each portfolio, we use momentum. A portfolio's own lagged monthly return is a strong positive predictor of subsequent monthly returns in a wide range of equity portfolios around the world (Gupta and Kelly, 2019), as well as in other asset classes (Moskowitz et al., 2012). For each asset in each sample, we compute its cumulative return over the past 20 trading days (approximately one month), then standardize the signal each period by converting it to a cross-sectional rank and dividing by the number of assets and subtracting the mean (mapping the signal into the [-0.5,0.5] interval).[17] We use this to predict subsequent monthly (20-day cumulative) returns on each portfolio.[18]

We estimate the prediction matrix as the sample counterpart of the definition $\Pi = E(R_{t+1}S_t')$ using a rolling "training window." The training window is the past 120 time periods. In our base case, the training period consists of the past 120 non-overlapping

[16] https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

[17] All of our theoretical results apply to cross-sectionally demeaned signals. If we start with any signal S, we can work with the cross-sectionally demeaned signal: $\tilde{S}_{j,t} = S_{j,t} - \frac{1}{N}\sum_{k=1}^{N} S_{k,t}$. The corresponding simple factor $\tilde{F}$ is dollar neutral. The eigenvalues of the prediction matrix with respect to $\tilde{S}$ and S have the same signs, except for at most two eigenvalues (see Proposition 20 in the Appendix). Further, demeaning means that we only exploit cross-sectional predictability, not time series predictability, which essentially leads to the "loss" of one eigenvalue (Proposition 22 in the Appendix).

[18] We cross-sectionally demean returns to focus prediction on cross section differences in returns rather than time series fluctuations in the common market component of returns. This approach corresponds to choosing each test asset to be a long position in portfolio i, hedged by going short an equal-weighted average of all portfolios (clearly an implementable strategy). In a robustness analysis we show that our results are similar if we do not hedge out the market return in this way.

20-day time periods. The estimated prediction matrix at time period t is

$$\hat{\Pi}_t = \frac{1}{120} \sum_{\tau=t-120}^{t-1} R_{\tau+1} S'_\tau \,. \tag{48}$$

Based on this empirical prediction matrix, we compute its singular vectors to form principal portfolios and we compute and the eigenvectors of its symmetric and anti-symmetric parts, giving rise to the empirical principal exposure portfolios and principal alpha portfolios. We compare these to the simple factor $\widetilde{F}_t$ defined in (3). To limit the undue effects of illiquidity on our conclusions, we always add an extra 1-day buffer between the last day in the training sample and the first day in the forecast window.

We investigate whether empirical principal portfolios behave in accordance with our theoretical predictions. Figure 1.A shows the singular values of the prediction matrix, averaged over time. Recall that, according to the theory, these singular values correspond to the expected returns of the corresponding PPs. The realized (out-of-sample) next-month returns of the PPs are plotted in Figure 1.D, along with their confidence bands. We find that the realized returns roughly match the shape of the ex ante singular values, with the low-numbered PPs having large eigenvalues and high realized returns. However, while this relation would be perfect on an in-sample basis, we naturally see some degradation of realized returns relative to the eigenvalues when looking out-of-sample.

In a similar vein, Figures 1.B and 1.C show the eigenvalues of the symmetric and anti-symmetric parts of the prediction matrix, respectively. Figures 1.E and 1.F report the out-of-sample realized returns of the corresponding PEPs and PAPs, respectively. Again we see a close relation between the ex ante predicted returns, and the out-of-sample realized ones. In this sample, only the first two PPs and first two PEPs appear to have a significant out-of-sample return, and only the first PAP return is significant.

One might wonder what these portfolios look like? We explore this in the case of PEPs and PAPs. Figure 2.A plots the out-of-sample weights of the eigenvector w_1 underlying the first PEP. Interestingly, this eigenvector tends to be long value and short growth stocks, and simultaneously tends to be long larger stocks and short smaller ones. Recall that PEP1 trades w_1 based its own signal, that is, PEP1 is going long or short a size-value bet based on its own momentum. Said differently, when large-value has recently outperformed, then PEP1 buys large-value, and, otherwise, it buys small-growth. To illustrate this strategy further, Figure 2.B plots the momentum, $S'w_1$, of the eigenvector. Lastly, Figure 2.C show the overall portfolio weight, $S'w_1w_1'$, averaged over time. Similarly, Figure 2, Panels D–E illustrate the PAP1 trading strategy.

Figure 3 summarizes the risk-adjusted out-of-sample performance of the PPs, PEPs, and PAPs. For simplicity, we only report the return of the sum of the top three principal portfolios (among each version: PP, PEP, and PAP), and the combination of the top 3 PEPs plus top 3 PAPs.[19] In each case, we compare their performance to that of the simple factor, which is just the sum-product of signals and returns. When analyzing factor performance, we use the exact same signal construction for the factor and PPs and evaluate both over the same forecast horizons, so each group of bars is an apples-to-apples out-of-sample comparison. We see that the PEP has a similar Sharpe ratio (SR) to that of the simple factor, where SR is the average excess return divided by volatility. The PAP has a higher SR, and the combination of PEP and PAP is higher yet, more than double the SR of the simple factor. The PP strategy performs similarly to PAP, handily beating the simple factor. The best overall performance is achieved by the combination of PEPs and PAPs. Throughout, when we report Sharpe ratios and information ratios, we also report ± 2 standard error bars around each estimate based the approximate standard error formula Lo (2002).

Figure 3 also plots the out-of-sample information ratio (IR) and its confidence interval as a measure of the risk-adjusted return of the principal portfolios. Specifically, the IR is computed by regressing the return of the PP (or PEP, PAP, or their combination) on the

[19]When combining PEPs and PAPs, we rescale the PAP component to have the same volatility as the PEP component, then take a 50/50 combination.

simple factor ($\widetilde{F}$) and the five Fama-French factors (the market MKT, the size factor SMB, the value factor HML, the profitability factor RMW, and the investment factor CMA):

$$PP_t = \alpha + \beta^0 \widetilde{F}_t + \beta^1 MKT_t + \beta^2 SMB_t + \beta^3 HML_t + \beta^4 RMW_t + \beta^5 CMA_t + \varepsilon_t \tag{49}$$

The IR is the alpha divided by residual volatility, IR$= \alpha/\sigma(\varepsilon_t)$, which can be interpreted as the Sharpe ratio when all the factors on the right hand side are hedged out (i.e., the alpha expressed as a Sharpe ratio).

Table 2 reports the details of this regression. As seen from Table 2 (and the confidence intervals in Figure 6.B), the PEP does not have a significant alpha (or, equivalently, a significant IR), but the PAP is highly significant (t-statistic of 4.42) and so is the PP strategy and the combination of PEP and PAP. Interestingly, Table 2 also shows that PEP has a highly significant loading on the simple factor with a high R^2, while, in contrast, PAP has small and insignificant factor loadings and low R^2. These findings are consistent with the idea that PEP provides factor exposure while PAP provides uncorrelated alpha.

Extended Momentum Analysis. We report a variety of extensions and robustness tests in Appendix G. We find that the out-of-sample principal portfolios perform even better at shorter forecast horizons (Figure 6). Principal portfolios also work across several other data sets: 25 U.S. size and operating profitability portfolios, 25 U.S. size and investment portfolios, the international counterparts of the three sets of Fama-French portfolios (i.e., developed countries excluding the U.S.), and a sample of 52 futures contracts for commodities, equity indices, sovereign bonds, and currencies (Figure 7). We show robustness with respect to other momentum signals, namely those based on 40, 60, 90, 120, and 250 day past returns, following on the standard practice of considering momentum signals up to 1 year (Figure 8). Lastly, we consider subsample analysis by decade (Figure 9) and demonstrate robustness of our findings when we do not cross-sectionally demean signals and returns (Figure 10).

6.2 Factor Timing

Our second empirical analysis investigates the power of principal portfolios for factor timing using an extensive data set of "anomaly" portfolios. The majority of empirical asset pricing is focused on long-run average returns of common factors and their ability to explain differences in long-run average returns across stocks. However, several recent papers document the returns of common factors are predictable. As emphasized by Haddad et al. (2020), factor return predictability has implications for our understanding of the stochastic discount factor (SDF) due to its close link to conditional mean-variance efficient portfolios in the market. Furthermore, factor return predictability implies that a dynamic combination of factors outperforms static positions in these factors, increasing the unconditional Sharpe ratio and volatility of the SDF. In doing so, the evidence on factor predictability poses a quantitative challenge for leading theoretical asset pricing models, which tend to generate SDFs with only moderate Sharpe ratios and are too smooth.

While the SDF implications due to factor predictability are economically important, approaches for quantifying factor predictability have only recently emerged in the literature. The main focus has been on two types of predictors: factor momentum (Gupta and Kelly, 2019; Arnott et al., 2019) and factor valuation ratios (Cohen et al., 2003; Kelly and Pruitt, 2013; Haddad et al., 2020). The method of principal portfolios is well suited for quantifying predictability in factor portfolios. It allows for more general predictive associations than considered in the factor timing literature to date, such as heterogeneous factor predictability and cross-factor predictive effects while maintaining the robustness of low parameterization thanks to its built-in dimension reduction.

We analyze factor timing through the lens of principal portfolios and do so for a larger collection of factor predictors than studied in the prior literature. We begin from the large set of 153 US equity characteristics and associated factors from Jensen et al. (2021).[20] We discard

[20] Data and code available at `https://github.com/bkelly-lab/GlobalFactor`.

15 factors/signals whose sample begins later than 1963. For each signal, the factor portfolio is formed from a high-low tercile spread and is value-weighted in each tercile. We construct factor return predictors by aggregating each of the 138 stock-level signals into a factor-level characteristic by applying the same factor weighting scheme used to construct the factor portfolio return.[21] Thus, returns for each of the factor portfolios (e.g., the book-to-market factor) are accompanied by a set of 138 time-series predictors (e.g., accruals, 12-month momentum, cash-to-assets, book-to-market, etc., at the factor level). In our application of principal portfolios, the set of base asset returns corresponds to the 138 long-short anomaly factors, and the set of signals corresponds to each of the 138 factor-level predictors. Our final data set, which covers the 684 months from 1963-2019, is a balanced panel of 138 factor portfolios (a 684×138 array), each possessing 138 different predictor variables for each factor (a three-dimensional $684 \times 138 \times 138$ array).

We conduct our principal portfolios analysis one signal at a time. For example, we construct the set of principal portfolios among the 138 factors using just the accruals predictor for each factor. Then we construct principal portfolios based on just the book-to-market predictor of each factor. We proceed in this way, building one set of out-of-sample principal portfolio returns for each predictor. For each predictive signal, we estimate the factor prediction matrix in a rolling 120-month training sample ending at time t, then use the fitted parameters to construct out-of-sample principal portfolio returns at $t + 1$.

Figure 4 reports the performance of principal portfolios on average across the 138 different signals. We report average Sharpe ratios for portfolios corresponding to the first ten eigenvalues of the prediction matrix, and we overlay the average standard error bar. In Panel A, we see that the leading PP achieves an annualized Sharpe ratio of 0.6, and performance drops for the second and higher eigenvalue portfolios. Panel B shows the performance of PEP portfolios corresponding to the top five and bottom five eigenvalues of the symmetric prediction matrix.

[21]Following standard practice in the literature, stock-level characteristics are cross-sectionally ranked and mapped into the [-0.5,0.5] interval before they are aggregated up to form factor-level predictors.

The first and last PEPs earn annualized Sharpe ratios of 0.4 and -0.4, respectively. This indicates that own-factor predictability tends to be heterogenous on average across signals, with a given signal exhibiting positive own-factor predictability for some factors by negative own-factor predictability for others. Panel C shows that the overall performance of PPs is dominated by the leading PAP, which delivers an out-of-sample Sharpe ratio of 0.7 per annum.

In Panel D, we report the average performance of the simple factor. This factor restricts portfolio construction to homogenous own-factor predictions and produces an average annualized Sharpe ratio of 0.2. We introduce an additional benchmark portfolio in Panel D that weights factors based on their historical mean returns over an expanding sample. Consistent with the findings of Jensen et al. (2021), the performance of factors is remarkably stable over time. This suggests that "static bets" overweighting historically successful factors perform well out-of-sample. Indeed this simple strategy achieves a Sharpe ratio of 0.5, and it contributes in part to the performance of principal portfolios, which also detects attractive static bets.

We are interested in whether principal portfolios can successfully exploit return predictability to time factors. To this end, we are especially conservative in calculating information ratios of principal portfolios and control not only for the simple factor and the Fama-French five-factor model but also for the factor portfolio based on historical mean weights. Indeed, the historical mean weight factor explains the bulk of the performance of PEP strategies, resulting in small and insignificant information ratios. However, the leading PP and PAP portfolios continue to produce large and significant information ratios. The leading PAP portfolio achieves an information ratio of 0.8 on average across all signals.

Next, rather than reporting averages of principal portfolio strategies based on individual signals, we study the performance of strategies that combine principal portfolios across all signals. In particular, we report the performance of equal-weight averages of the leading

PP, PEP, and PAP across the set of 138 signals. These combined strategies are shown in Figure 5, where we report annualized Sharpe ratios and information ratios (controlling for the simple factor, the five Fama-French factors, and the historical mean weight factor), along with 95% confidence intervals for each. We find that the equal-weight combination of leading PPs earns a Sharpe ratio of 0.6, slightly higher than the average Sharpe ratio of the individual leading PPs (same to one decimal place). However, the information ratio of the combined strategy is 1.1 and is highly statistically significant. Next, information ratios for equal-weight combinations of PEPs (the first and last PEP, denoted PEP 1 and PEP N, respectively) are large and highly significant. The first PEP, which hones in on factors with positive symmetric predictability, earns an information ratio of 0.6 after controlling for benchmark factors. What is more intriguing is that the last PEP is able to reliably exploit heterogeneity in own-factor predictability. It identifies a combination of factors that are predicted by the signal with the opposite sign as the first PEP. An optimizing trader that takes a short position in the PEP N strategy earns an annualized information ratio of 1.0. The strongest effect of combining strategies across signals appears for the PAP combination, which achieves an out-of-sample information ratio of 1.3.

To help interpret the contribution of these strategies to the stochastic discount factor, we analyze the ex-post tangency portfolio using principal portfolios (along with standard asset pricing factors) as inputs. We consider a set of 11 assets as inputs to the tangency portfolio. First are the five Fama-French factors. Next are two benchmark factors defined as the equal-weight average of simple factors across signals and the factor constructed using historical mean weights of the 138 underlying anomaly factors. Lastly, we include the equal-weight combination of the first PP for all signals, the equal-weight combination of the first PEP (PEP 1), the combination of the last PEP (PEP N), and the combination of leading PAPs. To better interpret the weights, all assets input to the tangency portfolio calculation are scaled to have the same volatility.

As a frame of reference, the first column of Table 3 reports the ex-post tangency portfolio weights for the five Fama-French factors alone. We see that four of the five factors contribute significantly to the tangency portfolio. The ex-post tangency Sharpe ratio of the Fama-French model is 1.1. We interpret this ex-post tangency portfolio as a representation of the benchmark SDF from the asset pricing literature.

The second column of Table 3 supplements the Fama-French factors with the two additional benchmark factors and the principal portfolio strategies. Because the expected return on PEP N is negative, we switch its sign to align it with the remaining factors, which all have positive expected returns. When including principal portfolios, the market factor, SMB, and RMW remain significant components of the SDF, but HML and CMA become insignificant. The two additional benchmark factors are also insignificant. By far, we see that the favorite input to the tangency portfolio is the PAP strategy, which the Markowitz solution levers up with a weight of 2.1. At the same time, the tangency portfolio places a large negative weight on PEP 1 due to its correlation with other factors.

Principal portfolios generate an enormous gain in the mean-variance efficiency of the SDF. The tangency portfolio Sharpe ratio is 2.2, essentially double that of the Fama-French five-factor model. To aid interpretation in the face of correlated PP strategies, the third column estimates tangency portfolio weights with a non-negativity constraint. In this case, we see that all principal portfolios other than PAP drop out of the SDF. The market and PAP portfolios remain the largest contributors to the SDF, each receiving more than a quarter of tangency weight, followed by RMW and CMA, which also receive significantly positive weight. With a non-negativity constraint, the ex-post tangency Sharpe ratio is 1.5, representing a gain in efficiency of 40% versus the Fama-French model.

This evidence indicates that principal portfolios are a potent method for factor return prediction and, in turn, for constructing factor timing strategies. Next, we compare this result to a recent advance in factor timing methodology proposed by Haddad et al. (2020), or

HKS henceforth. HKS advocate a shrewd principal components reduction of the cross-section of anomaly factors. The leading principal components among anomaly factors amount to a few portfolios that account for the bulk of the covariance among anomaly factors. By focusing their factor timing analysis on these riskiest dimensions of the factor space, HKS leverage their economic prior to narrow the search for factor predictability to the most plausible subspace of anomaly returns. Indeed, HKS show that returns of the first few anomaly principal components portfolios are robustly predictable by the book-to-market ratios of those portfolios. While HKS focus on a single predictor (book-to-market ratio) in their factor timing analysis, we extend this by applying their procedure to all 138 predictors in our data set.

Our implementation of the HKS procedure is designed to produce a factor timing strategy that can be compared on an apples-to-apples basis with principal portfolio strategies. We estimate the factor return covariance matrix $\hat{\Sigma}_t \ = \mathrm{Cov}(R_{[t-120:t]})$ using a rolling 120-month training window and and compute the associated eigenvalue decomposition,

$$\hat{\Sigma}_t \ = \ Q_t D_t Q_t'.$$

Here, Q is the matrix of principal components of R_{t+1} that provide an orthogonal rotation of the factors ordered from highest to lowest variance, $R_{PC,t+1} = Q_t' R_{t+1}$. The predictors are rotated to align with the principal component returns using the same Q matrix, $S_{PC,t} = Q_t' S_t$. Following HKS, we retain only the five most volatile components, motivated by their prior of "absence of near-arbitrage." Next, in the same training window, we estimate predictive regressions of the form

$$R_{PC,j,t+1} = a_{j,t} + b_{j,t} S_{PC,j,t} + e_{j,t+1}, \quad j = 1, ..., 5$$

and construct fitted values $\hat{E}_t(R_{PC,j,t+1}) = \hat{a}_{j,t} + \hat{b}_{j,t} S_{PC,j,t}$, Finally, using the vector of fitted predictions $\hat{E}_t(R_{PC,t})$ and the training sample return covariance matrix $\hat{\Sigma}_t$, we construct the out-of-sample factor timing portfolio return at $t+1$ as the tangency portfolio of leading components:

$$HKS_{t+1} = \hat{E}_t(R_{PC,t+1})' \hat{\Sigma}_t^{-1} R_{PC,t+1}.$$

Note that in the construction of principal portfolios, we do not directly use information on the covariance matrix of returns. For an apples-to-apples comparison with principal portfolios, we also report a version of the HKS timing strategy that only uses predictive information and excludes $\hat{\Sigma}$:

$$HKS_{t+1}^{\text{No Cov.}} = \hat{E}_t(R_{PC,t+1})' R_{PC,t+1}.$$

We report results for the HKS methodology on the right side of Figure 5. As in the case of principal portfolios, we construct an equal-weight average of HKS factor timing portfolios across the 138 possible predictors. As with principal portfolios, we calculate the annualized out-of-sample Sharpe ratios for both the HKS and $HKS^{\text{No Cov.}}$ portfolios, and information ratios versus the benchmarks described earlier. The HKS portfolio Sharpe ratio is nearly identical to that for the PAP portfolio (0.7 in both cases). However, the information ratio of HKS drops to 0.4, compared to 1.3 for PAP. The $HKS^{\text{No Cov.}}$ strategy suffers compared to the main HKS portfolio, with a Sharpe ratio of zero and a negative information ratio. This suggests that the performance of the HKS strategy derives from covariance timing and not expected return timing. Lastly, if we add HKS and $HKS^{\text{No Cov.}}$ to the ex-post tangency portfolio studied in Table 3, we find that they receive weights of 0.00 and 0.02 in the unconstrained portfolio, respectively, and are statistically insignificant (untabulated). In the constrained non-negative version, they receive tangency weights of 0.04 and 0.00, respectively, and are again insignificant.

It turns out that the outperformance of principal portfolios relative to the factor timing approach of Haddad et al. (2020) is predicted by theory. A key observation is that trading strategies investigated in Haddad et al. (2020) fit into the class of symmetric linear strategies, thus they cannot capture any anti-symmetric components of Π. As we show above, it is precisely the PAPs that drive success of principal portfolios strategies. We discuss this and related theoretical results in Appendix E.

7 Conclusion: The Power of Principal Portfolio Analysis

We present a new method for analyzing return predictability. Our main contribution is a new theoretical understanding of the prediction matrix, Π, founded on the decomposition of this matrix into principal portfolios. We classify predictive patterns in Π as either symmetric or anti-symmetric and derive theoretical results that translate these patterns into into beta and alpha. These results, in turn, give rise to a novel test of asset pricing models, for which we derive a complete distribution theory.

Our analysis provides theoretical guidance on how to optimally invest based on return-predictive signals, even when this predictability involves complex phenomena such as cross-asset predictability or violations of equilibrium asset pricing restrictions. We demonstrate the practical impact of this guidance in an extensive empirical analysis. We find that the leading principal portfolios based on a wide range (over 100) stock return prediction signals deliver large and significant risk-adjusted average returns out-of-sample. Our empirical principal portfolios significantly expand the mean-variance frontier relative to benchmarks in the literature, including the Fama-French five-factor model and the factor timing strategies of Haddad et al. (2020).

References

Arnott, Robert D, Mark Clements, Vitali Kalesnik, and Juhani T Linnainmaa, "Factor momentum," *Available at SSRN 3116974*, 2019.

Asness, Clifford S, Jacques A Friedman, Robert J Krail, and John M Liew, "Style timing: Value versus growth," *The Journal of Portfolio Management*, 2000, *26* (3), 50–60.

Bhatia, Rajendra, *Matrix Analysis*, Springer, 1997.

Britten-Jones, Mark, "The sampling error in estimates of mean-variance efficient portfolio weights," *The Journal of Finance*, 1999, *54* (2), 655–671.

Cochrane, John H, *Asset pricing: Revised edition*, Princeton university press, 2009.

Cohen, Randolph B, Christopher Polk, and Tuomo Vuolteenaho, "The value spread," *The Journal of Finance*, 2003, *58* (2), 609–641.

Collin-Dufresne, Pierre, Kent D Daniel, Ciamac C Moallemi, and Mehmet Sağlam, "Dynamic asset allocation with predictable returns and transaction costs," *Available at SSRN 2618910*, 2015.

—, **Kent Daniel, and Mehmet Sağlam**, "Liquidity regimes and optimal dynamic asset allocation," *Journal of Financial Economics*, 2019.

Fama, Eugene F and Kenneth R French, "A five-factor asset pricing model," *Journal of financial economics*, 2015, *116* (1), 1–22.

Fan, Ky, "On a theorem of Weyl concerning eigenvalues of linear transformations.," *Proceeding of the National Academy of Sciences USA*, 1950, *36*, 31–35.

— **and A. J. Hoffman**, "Some metric inequalities in the space of matrices," *Proceedings of the American Mathematical Society*, 1955, *6*, 111–116.

Frazzini, Andrea and Lasse Heje Pedersen, "Betting against beta," *Journal of Financial Economics*, 2014, *111* (1), 1 – 25.

Gârleanu, Nicolae and Lasse Heje Pedersen, "Dynamic Trading with Predictable Returns and Transaction Costs," *The Journal of Finance*, 2013, *68* (6), 2309–2340.

___ **and** ___ , "Dynamic Portfolio Choice with Frictions," *Journal of Economic Theory*, 2016, *165*, 487–516.

Gibbons, Michael R, Stephen A Ross, and Jay Shanken, "A test of the efficiency of a given portfolio," *Econometrica: Journal of the Econometric Society*, 1989, pp. 1121–1152.

Gu, Shihao, Bryan Kelly, and Dacheng Xiu, "Empirical asset pricing via machine learning," Technical Report, National Bureau of Economic Research 2018.

Gupta, Tarun and Bryan Kelly, "Factor momentum everywhere," *The Journal of Portfolio Management*, 2019, *45* (3), 13–36.

Haddad, Valentin, Serhiy Kozak, and Shrihari Santosh, "Factor Timing," *The Review of Financial Studies*, 02 2020, *33* (5), 1980–2018.

Hansen, Lars Peter and Ravi Jagannathan, "Implications of security market data for models of dynamic economies," *Journal of political economy*, 1991, *99* (2), 225–262.

Horn, Toger A. and Charles R. Johnson, *Topics in Matrix Analysis*, Cambridge, 1st Edition, 1991.

Jensen, Theis Ingerslev, Bryan T Kelly, and Lasse Heje Pedersen, "Is There A Replication Crisis In Finance?," Technical Report, National Bureau of Economic Research 2021.

Kelly, Bryan and Seth Pruitt, "Market Expectations in the Cross-Section of Present Values," *The Journal of Finance*, 2013, *68* (5), 1721–1756.

___ **,** ___ **, and Yinan Su**, "Characteristics are Covariances: A Unified Model of Risk and Return," *Journal of Financial Economics*, 2020.

___ **,** ___ **, and** ___ , "Instrumented Principal Component Analysis," *Working paper*, 2020.

Lo, Andrew W, "The statistics of Sharpe ratios," *Financial analysts journal*, 2002, *58* (4), 36–52.

Marshall, Albert W. and Ingram Olkin, *Inequalities: Theory of Majorization and Applications*, Academic, 1979.

Moskowitz, Tobias J, Yao Hua Ooi, and Lasse Heje Pedersen, "Time series momentum," *Journal of financial economics*, 2012, *104* (2), 228–250.

Pedersen, Lasse Heje, Abhilash Babu, and Ari Levine, "Enhanced Portfolio Optimization," *Working paper, AQR and Copenhagen Business School*, 2020.

Tao, Terrence, *Topics in Random Matrix Theory*, Springer, 2012.

Velu, Raja and Gregory C. Reinsel, *Multivariate Reduced-Rank Regression. Theory and Applications*, Springer, 1998.

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

Table 1: Analogy between PCA and PPA

This table shows five analogies between principal component analysis (PCA) and principal portfolio analysis (PPA) for the symmetric part of the prediction matrix. For PCA (PPA): (i) the variance (expected excess returns) of each component equals its eigenvalue; (ii) different components $k \neq l$ are orthogonal; (iii) the sum variances (returns) of individual securities equals that of the components, and also equals the trace of the variance-covariance matrix (prediction matrix); (iv) the top K components maximize variance (return) for orthonormal portfolios; and (v) component $k+1$ maximizes variance (return) among all portfolios that are orthogonal to the first k ones.

	Principal Component Analysis	Principal Portfolio Analysis (Symmetric Part)
(i)	$\mathrm{Var}(R^{\pi_k}_{t+1}) \;=\; \lambda_k(\Sigma_R)$	$E(S^{w^s_k}_t R^{w^s_k}_{t+1}) \;=\; \lambda_k(\Pi^s)$
(ii)	$\mathrm{Cov}(R^{\pi_k}_{t+1}, R^{\pi_l}_{t+1}) \;=\; 0$	$E(S^{w^s_k}_t R^{w^s_l}_{t+1}) \;+\; E(S^{w^s_l}_t R^{w^s_k}_{t+1}) = 0$
(iii)	$\sum_k \mathrm{Var}(R_{k,t+1}) = \sum_k \mathrm{Var}(R^{\pi_k}_{t+1}) = \mathrm{tr}(\Sigma_R)$	$\sum_k E(S_{k,t}R_{k,t+1}) = \sum_k E(S^{w^s_k}_t R^{w^s_k}_{t+1}) = \mathrm{tr}(\Pi^s)$
(iv)	$(\pi_k) = \arg\max_{\mathrm{orthon.}\{x_k\}_{k=1}^K} \sum_k \mathrm{Var}(R^{x_k}_{t+1})$	$(w^s_k) = \arg\max_{\mathrm{orthon.}\{x_k\}_{k=1}^K} \sum_k E(S^{x_k}_t R^{x_k}_{t+1})$
(v)	$\pi_{k+1} = \arg\max_{x\perp\{\pi_1,\cdots,\pi_k\}} \mathrm{Var}(R^x_{t+1})$	$w^s_{k+1} = \arg\max_{x\perp\{w^s_1,\cdots,w^s_k\}} E(S^x_t R^x_{t+1})$

Table 2: Principal Portfolio Factor Exposures

This table reports regressions of out-of-sample principal portfolio returns on the simple factor and the five Fama-French factors. Portfolios are constructed from the Fama-French 25 size and value portfolios based on a 20-day momentum signal. The table reports regressions for the simple factor itself (the own-signal strategy, $\widetilde{F}$, computed as the sum-product of each asset's own signal and return, denoted "Factor"), the equal-weight average of the top three principal portfolios ("PP 1-3"), the equal-weight average of the top three principal exposure portfolios ("PEP 1-3"), the equal-weight average of the top three principal alpha portfolios ("PAP 1-3"), and the equal weighted average of the top three PEP's and PAP's combined ("PEP and PAP 1-3"). In each regression, the left-hand-side portfolio is scaled to have the same full-sample volatility as the excess market return. Results are shown for a 20-day forecast horizons, and each forecast is made on an out-of-sample basis using a rolling training sample of the 120 most recent non-overlapping return observations. Sample period is 1963-2019.

Portfolio	Factor	Mkt-Rf	SMB	HML	RMW	CMA	Alpha	R^2
Factor		-0.2	0.13	-0.28	-0.26	0.36	9.35	0.08
t-statistic		-4.59	1.9	-3.43	-3.03	2.79	4.12	
PP 1-3	0.82	0.03	0.02	0.15	-0.09	-0.02	4.69	0.67
t-statistic	32.69	1.09	0.53	3.05	-1.63	-0.29	3.38	
PEP 1-3	0.94	0.01	-0.02	0.06	-0.13	-0.01	0.89	0.89
t-statistic	67.00	0.88	-0.76	1.95	-4.4	-0.16	1.14	
PAP 1-3	-0.08	0.08	0.19	0.06	0.28	0.06	10.41	0.04
t-statistic	-1.94	1.71	2.65	0.72	3.1	0.42	4.42	
PEP and PAP 1-3	0.65	0.07	0.13	0.09	0.11	0.04	8.51	0.41
t-statistic	19.53	1.93	2.31	1.32	1.58	0.35	4.62	

Table 3: Ex Post Tangency Portfolios

Ex post tangency portfolio weights using as inputs the five Fama-French factors, the equal-weight average of simple factors across 138 signals, a factor portfolio with weights proportional to the historical means of the 138 underlying anomaly factors, and strategies formed from combinations of various principal portfolios. PP denotes the equal-weight average of leading PPs across 138 signals, PEP 1 is the equal-weight combination of the first PEP, PEP N is the equal-weight combination of the last PEP (multiplied by -1 to make it a positive expected return strategy), and the equal-weight combination of leading PAPs. To aid interpretation, all portfolios that contribute to the tangency portfolio are scaled to have the same volatility. * denotes that the portfolio weight is statistically significant at the 1% significance level based on the test of Britten-Jones (1999). The first column considers only Fama-French factors as a benchmark and the last column restricts tangency weights to be positive. The last row reports the annualized tangency portfolio Sharpe ratio. Each portfolio input is calculated on a rolling out-of-sample basis, then the ex post tangency analysis is conducted from the full time series of out-of-sample strategies.

Portfolio	FF5	FF5 + PP	Non-negative FF5 + PP
Mkt-Rf	0.29*	0.30*	0.27*
SMB	0.14*	0.19*	0.07
HML	-0.03	-0.08	0.04
RMW	0.27*	0.27*	0.21*
CMA	0.32*	0.10	0.16*
Simple Factor		-0.26	0.00
Hist. Mean Wts.		-0.08	0.00
PP		-0.86	0.00
PEP 1		-1.03*	0.00
$-1\times$PEP N		0.41	0.00
PAP		2.05*	0.26*
Sharpe Ratio	1.09	2.15	1.52

Figure 1: Prediction Matrix Eigenvalues

Panels A, B, and C show estimated eigenvalues of the prediction matrix and its symmetric and anti-symmetric components, respectively, averaged over training samples. Panels D, E, and F show average out-of-sample returns and ± 2 standard error confidence bands for corresponding principal portfolios, principal exposure portfolios, and principal alpha portfolios, respectively. Estimates are based on predictions of 20-day returns of the Fama-French 25 size and value portfolios based on a 20-day momentum signal. Each training sample consists of 120 non-overlapping 20-day return observations. Sample period is 1963-2019.

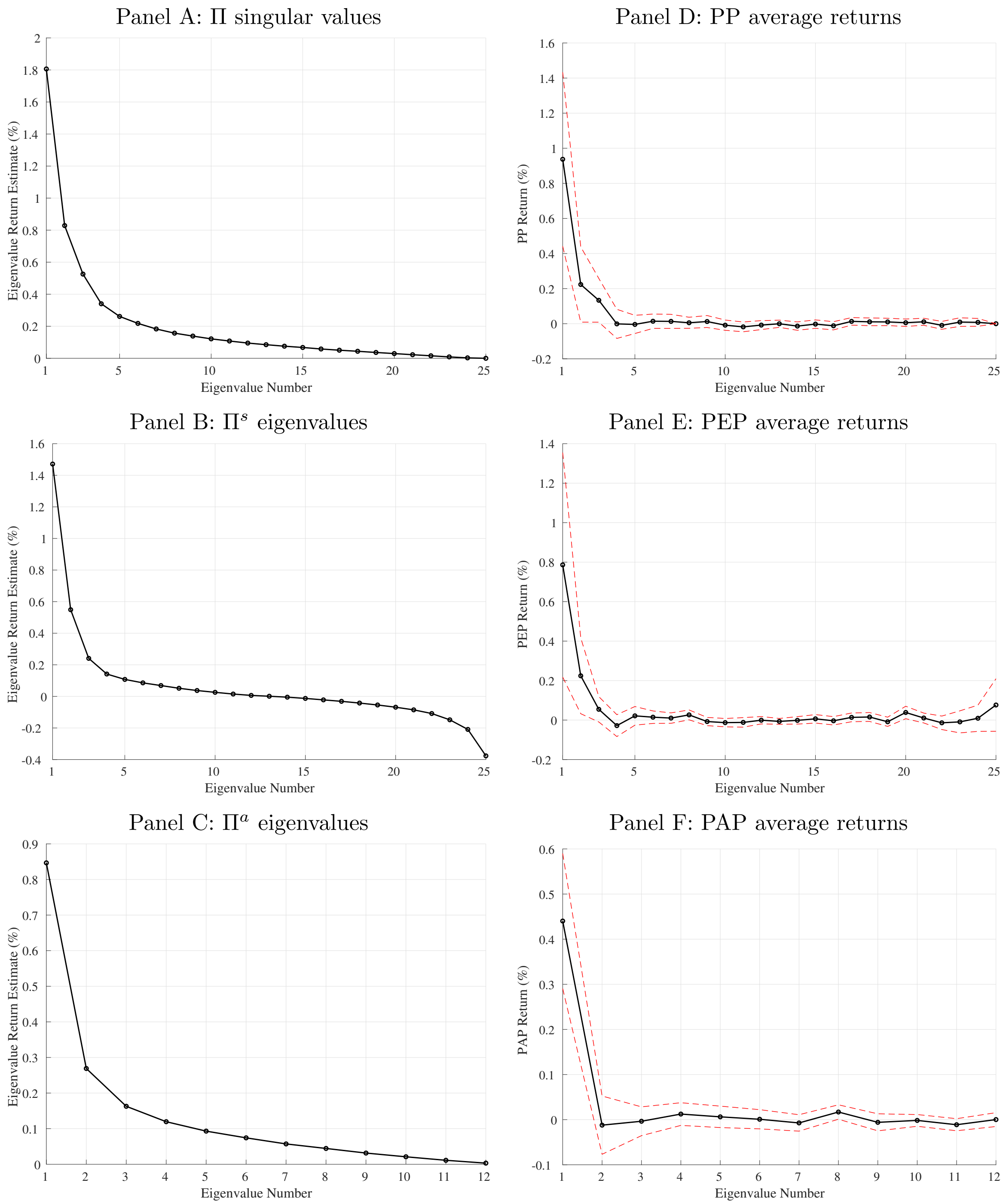


Figure 2: Portfolio Weights for Leading Principal Portfolios

Weights of the first principal exposure portfolio (Panel A) and first principal alpha portfolio (Panel B) on the 25 size and value portfolios, averaged over training samples. Portfolios are constructed based on a 20-day momentum signal and for a 20-day forecast horizon/holding period. Portfolios and estimates are made on an out-of-sample basis using a rolling training sample of the 120 most recent non-overlapping return observations. Sample period is 1963-2019.

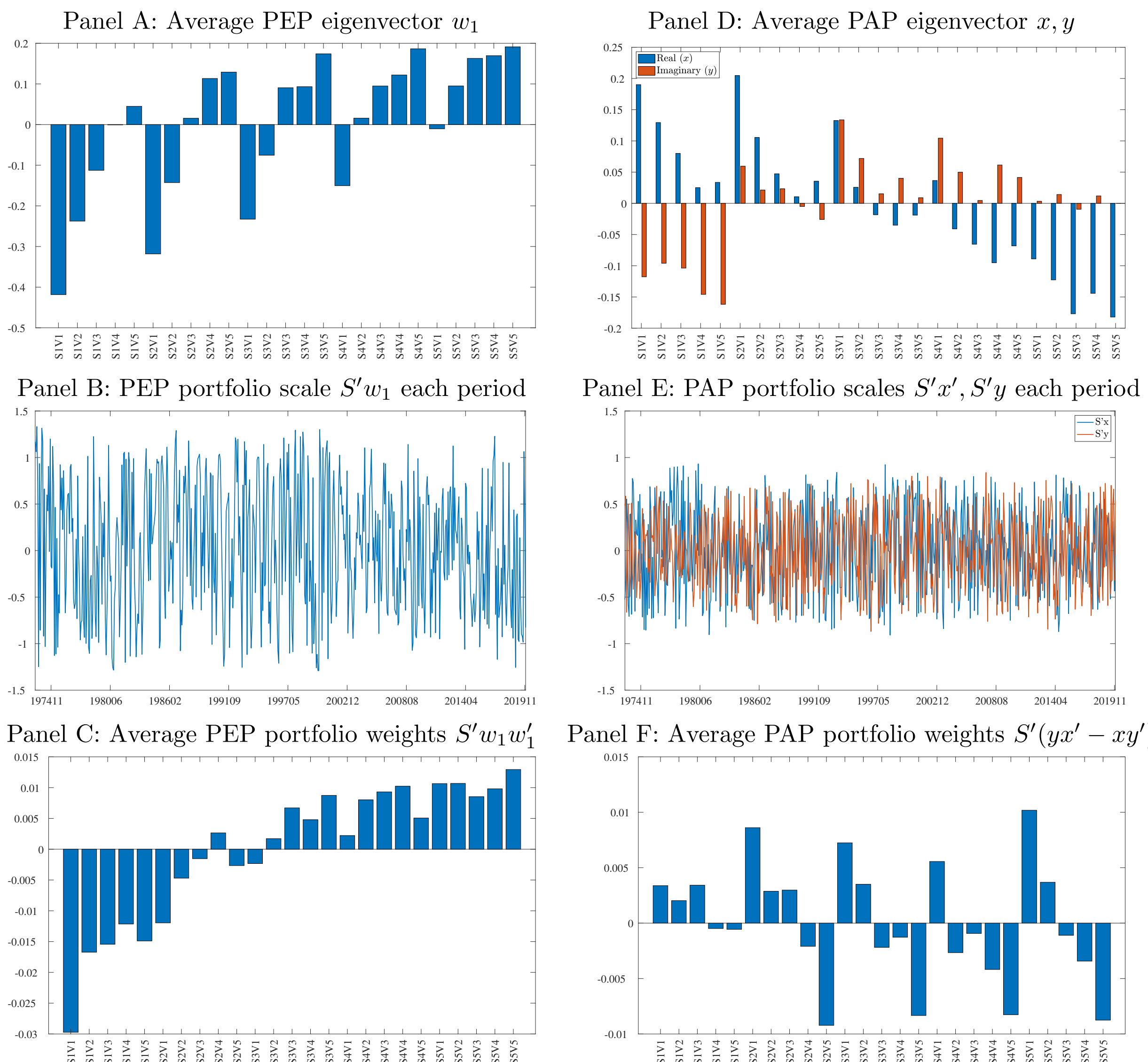


Figure 3: Principal Portfolio Performance Ratios

Out-of-sample performance of principal portfolios in terms of annualized Sharpe ratio (left set of bars) and annualized information ratio versus the own-predictor strategy and the Fama-French 5-factor model (right set of bars) along with ± 2 standard error band around each estimate. Portfolios are constructed from the Fama-French 25 size and value portfolios based on a 20-day momentum signal with a forecast horizons (and, equivalently, holding period) of 20 days. The figure reports performance of the simple factor ("Factor," that is, the standard own-signal strategy, included as a benchmark), the equal-weight average of the top three principal portfolios ("PP 1-3"), the equal-weight average of the top three principal exposure portfolios ("PEP 1-3"), the equal-weight average of the top three principal alpha portfolios ("PAP 1-3"), and the equal weighted average of the top three PEP's and PAP's combined ("PEP and PAP 1-3"). Each forecast is made on an out-of-sample basis using a rolling training sample of the 120 most recent non-overlapping return observations. Sample period is 1963-2019.

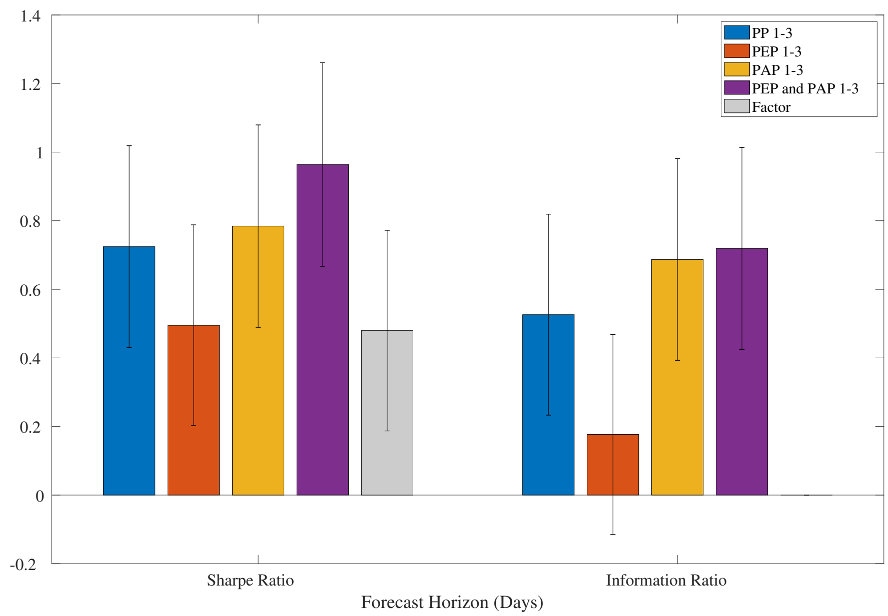


Figure 4: Average Performance of Factor Timing Strategies Based on Individual Signals

Out-of-sample performance of principal portfolios averaged across 138 signals. We report annualized Sharpe ratio and annualized information ratio versus the own-predictor strategy, the factor historical mean weight strategy, and the Fama-French 5-factor model. Portfolios are constructed using 138 anomaly factors as base assets, and principal portfolios are constructed for each of 138 different factor return predictors. The figure reports performance of the first 10 PPs (Panel A), the first five and last five PEPs (Panel B), and the first 10 PAPs (Panel C) averaged across signals, along with the average ± 2 standard error band. Panel D reports the average performance of simple factor strategies and the performance of a portfolio that weights factors based on their historical mean return (both of which are used as controls for the information ratios of principal portfolios). Each forecast is made on an out-of-sample basis using a rolling training sample of 120 monthly return observations. Sample period is 1963-2019.

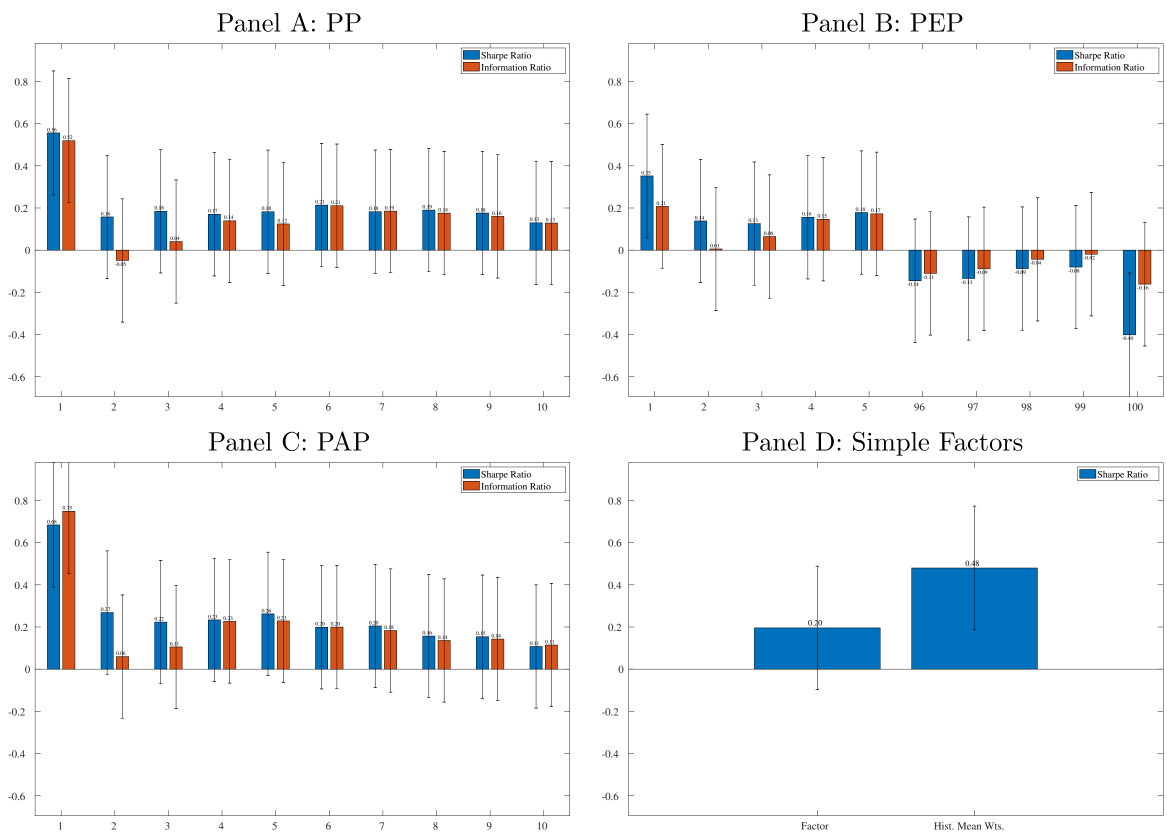
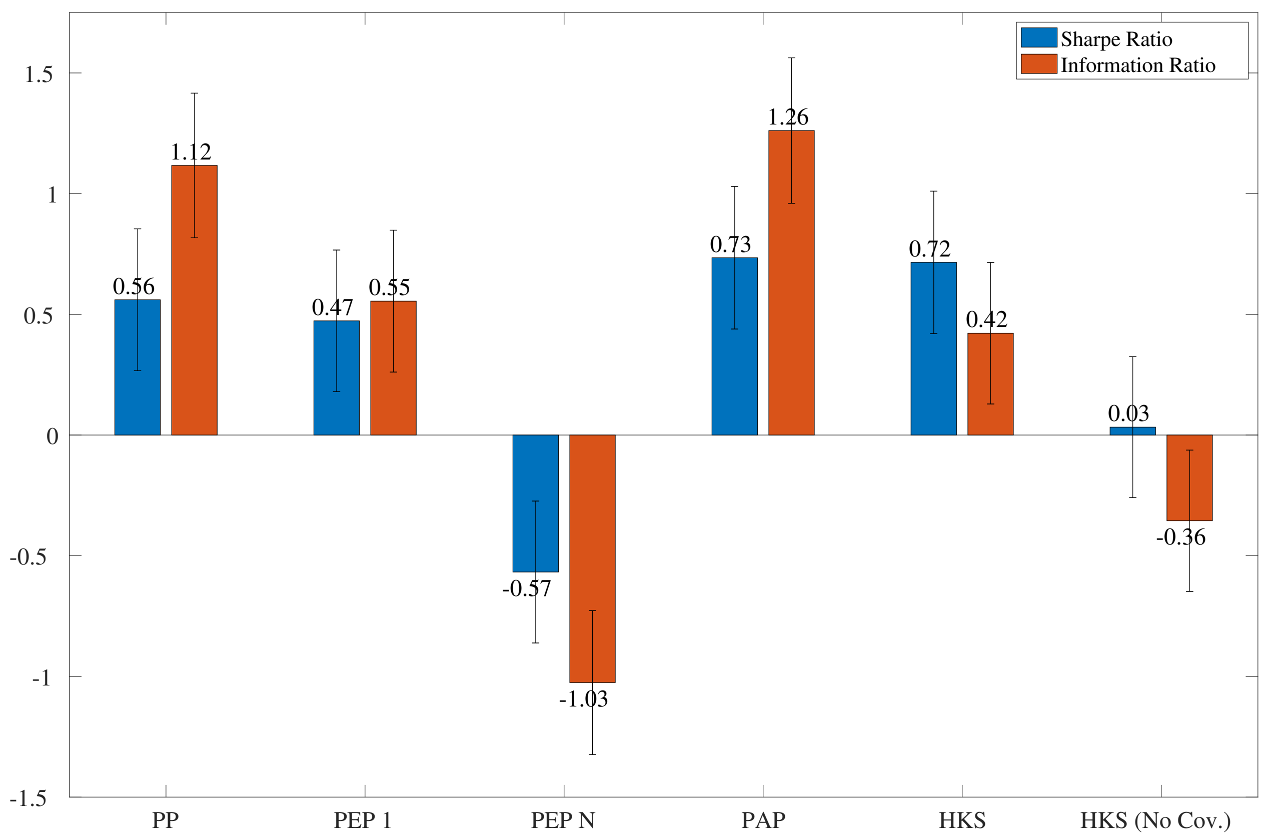


Figure 5: Performance of Factor Timing Strategies Combined Across Signals

Out-of-sample performance of principal portfolio strategies combined across across 138 signals. We report annualized Sharpe ratio and annualized information ratio versus the own-predictor strategy, the factor historical mean weight strategy, and the Fama-French 5-factor model, along with the average ± 2 standard error bands. Using 138 anomaly factors as base assets, principal portfolios are constructed for each of 138 different factor return predictors, then combined across signals. PP denotes the equal-weight average of the first PP across 138 signals, PEP 1 is the equal-weight combination of the first PEP, PEP N is the equal-weight combination of the last PEP, and PAP is the equal-weight combination of the first PAP across signals. We also report equal-weight combinations of HKS factor timing strategies with and without an inverse covariance matrix adjustement. Each portfolio is constructed on an out-of-sample basis using a rolling training sample of 120 monthly return observations. Sample period is 1963-2019.



Internet Appendix

A On Mean-Variance Optimization of Linear Strategies

A.1 Robust Mean-Variance Framework of Linear Strategies

Consider the following robust mean-variance objective function:

$$\max_L E(S_t' L R_{t+1}) \text{ subject to } \max_{S:\|S\|\leq 1} \mathrm{Var}_t(S' L R_{t+1}) \leq 1 \tag{50}$$

This objective is a robust in the sense that we require that the variance is bounded regardless of the signal S. The variance term can be written as

$$\max_{S:\|S\|\leq 1} \mathrm{Var}_t(S' L R_{t+1}) = \max_{S:\|S\|\leq 1} S' L \Sigma_{R,t} L' S = \max_{S:\|S\|\leq 1} \|\Sigma_{R,t}^{1/2} L' S\|^2 = \|\Sigma_{R,t}^{1/2} L'\|^2 \tag{51}$$

So, if the assets are normalized and uncorrelated such that $\Sigma_{R,t} = \mathrm{Id}$, then the robust variance constraint is the same as our matrix constraint $\|L\| \leq 1$.

We also get a similar solution when $\Sigma_{R,t} = \sigma^2 \mathrm{Id}$ for some $\sigma \in \mathbb{R}$, since, in this case, the portfolio constraint simply becomes $\|L\|\sigma \leq 1$. Hence, we just scale the position accordingly (i.e., use $L = L^*/\sigma$, where L^* is the standard solution with $\sigma = 1$). We note that, when $\Sigma_{R,t} = \sigma^2 \mathrm{Id}$, an alternative constraint is to require that the risk of the linear strategy cannot be greater than the risk of the simple factor as we do in (12), which is, again, equivalent to $\|L\| \leq 1$.

We can also easily solve a version of the robust mean-variance problem for general $\Sigma_{R,t}$. Solving (12) directly is not convenient since it would lead us to impose $\|\Sigma_{R,t}^{1/2} L'\| \leq 1$, which is not consistent with choosing a constant position matrix L. (Recall that the idea of linear strategies is to have a constant L, but rich portfolio dynamics, $S_t' L$, driven by the signals.) Instead, we consider the transformed "synthetic assets" with returns

$$\tilde{R}_{t+1} = \Sigma_{R,t}^{-1/2} R_{t+1}\,, \tag{52}$$

signals given by

$$\tilde{S}_t = \Sigma_{R,t}^{-1/2} S_t\,, \tag{53}$$

and the corresponding prediction matrix is $\tilde{\Pi} = E(\tilde{R}_{t+1} \tilde{S}_t') = E(\Sigma_{R,t}^{-1/2} \Pi \Sigma_{R,t}^{-1/2})$. For example, if the original assets are uncorrelated (i.e., a diagonal $\Sigma_{R,t}$), then creating these synthetic assets simply means scaling the assets to have the same constant volatility (as we do in the empirical analysis of futures).

Then we consider the robust mean-variance problem for the synthetic assets:

$$\max_{\tilde{L}} \; E\left(\tilde{S}_t'\tilde{L}\tilde{R}_{t+1}\right) \;\text{ subject to }\; \max_{\tilde{S}_t} \mathrm{Var}_t(\tilde{S}_t'\tilde{L}\tilde{R}_{t+1}) = \|\tilde{L}\| \leq 1. \tag{54}$$

So we see that this is our standard problem, expressed in terms of the synthetic assets. In other words, all our results apply for the robust mean-variance problem of the synthetic assets. For example, the optimal strategy for the synthetic assets is $\tilde{L} = (\tilde{\Pi}'\tilde{\Pi})^{-1/2}\tilde{\Pi}'$, using Proposition 3. Of course, this solution can be translated back to the original assets by noting that

$$L_t = \Sigma_{R,t}^{-1/2}\tilde{L}\Sigma_{R,t}^{-1/2}, \tag{55}$$

which holds since we must have that $\tilde{S}_t'\tilde{L}\tilde{R}_{t+1} = S_t'{\Sigma_R}^{-1/2}\tilde{L}{\Sigma_R}^{-1/2}R_{t+1} = \; S_t'L_tR_{t+1}$.

A.2 Risk Aversion instead of Risk Constraint

We can also consider a robust mean-variance problem in which the investor has risk aversion γ rather than a risk constraint:

$$\max_{L}\left(E(S_t'LR_{t+1}) - \frac{\gamma}{2}\left[\max_{S:\|S\|\leq 1} \mathrm{Var}_t(S'LR_{t+1})\right]\right) . \tag{56}$$

We can rewrite the objective using (51) as

$$\max_{L}\left(E(S_t'LR_{t+1}) - \frac{\gamma}{2}\|\Sigma_{R,t}^{1/2}L\|^2\right) \tag{57}$$

Let us first solve this portfolio problem when $\Sigma_{R,t} = \mathrm{Id}$. To find the solution, we start by finding the solution for each level of volatility:

$$L(c) = \arg\max_{L:\|L\|=c} \; E\left(S_t'LR_{t+1}\right) = c(\Pi'\Pi)^{-1/2}\Pi'. \tag{58}$$

where the last equality uses Proposition 3. Now we can solve the objective function (61) by maximizing over all possible volatilities, c:

$$\max_c \left(E\left(S_t' L(c) R_{t+1}\right) - \frac{\gamma}{2} c^2 \right) \;=\; \max_c \left(c \sum_{i=1}^N \bar{\lambda}_i - \frac{\gamma}{2} c^2 \right) \;=\; \frac{(\sum_{i=1}^N \bar{\lambda}_i)^2}{2\gamma}\,, \tag{59}$$

where the optimum is achieved by $c = \frac{\sum_{i=1}^N \bar{\lambda}_i}{\gamma}$, implying that the optimal strategy is

$$L = \frac{\sum_{i=1}^N \bar{\lambda}_i}{\gamma} (\Pi'\Pi)^{-1/2}\Pi'. \tag{60}$$

We see that the optimal strategy is the same as in Proposition 3, except for a scaling factor. The scaling factor naturally decreases in γ, reflecting that a more risk averse investor takes a smaller position. Similarly, the scaling factor increases in the sum of the singular values, since higher singular values imply stronger predictability, leading to a larger position.

Finally, consider the problem with a general variance-covariance matrix $\Sigma_{R,t}$. Rather than solving the objective function (61) (which could lead to a time-varying L), we consider the similar objective function for the synthetic assets:

$$\max_{\tilde{L}} \left(E(\tilde{S}_t' \tilde{L} \tilde{R}_{t+1}) - \frac{\gamma}{2} \left[\max_{\tilde{S}:\|\tilde{S}\|\le 1} \mathrm{Var}_t(\tilde{S}' \tilde{L} \tilde{R}_{t+1}) \right] \right) = \max_{\tilde{L}} \left(E(\tilde{S}_t' \tilde{L} \tilde{R}_{t+1}) - \frac{\gamma}{2} \|\tilde{L}\|^2 \right) \tag{61}$$

Since the synthetic assets have a variance-covariance matrix equal to the identity, their optimal solution is

$$\tilde{L} = \frac{\sum_{i=1}^N \bar{\lambda}_i(\tilde{\Pi})}{\gamma} (\tilde{\Pi}'\tilde{\Pi})^{-1/2}\tilde{\Pi}'\,, \tag{62}$$

which can be translated back to the original assets using (55).

A.3 Relation to the Standard Mean-Variance Framework

We next consider a standard mean-variance objective function, that is, we don't use the worst-case variance, but instead assume that conditional return variance is constant over time, $\Sigma_{R,t} = \Sigma_R$. In this case, we naturally recover the standard Markowitz solution. To see this, we use the notation $\Sigma_S = E(S_t S_t')$ as before and calculate:

$$\begin{aligned}
\max_L E\left(E_t(S_t' L R_{t+1}) - \frac{\gamma}{2}\mathrm{Var}_t(S_t' L R_{t+1})\right) &= \max_L E\left(S_t' L R_{t+1} - \frac{\gamma}{2} S_t' L \Sigma_R L' S_t\right) \\
&= \max_L \left(\mathrm{tr}(L\Pi) - \frac{\gamma}{2}\,\mathrm{tr}(L\Sigma_R L \Sigma_S)\right) \\
&= \max_{\tilde{L}} \left(\mathrm{tr}(\tilde{L}\tilde{\Pi}) - \frac{\gamma}{2}\,\mathrm{tr}(\tilde{L}'\tilde{L})\right) \\
&= \max_{\tilde{L}} \left(\mathrm{tr}(\tilde{L}\tilde{\Pi}) - \frac{\gamma}{2}\|\tilde{L}\|_2^2\right),
\end{aligned} \tag{63}$$

where we use the change of variable $\tilde{L} = \Sigma_S^{1/2} L {\Sigma_R}^{1/2}$ and $\tilde{\Pi} = {\Sigma_R}^{-1/2}\Pi\Sigma_S^{-1/2}$. So we see that this problem has the same form as our normal objective function, except that we have another matrix norm, namely the Frobenius 2-norm, $\|\cdot\|_2$. Since we solve the problem for all p-norms in Proposition 11 (and in the proof of Propostion 3), we know that solution, which is very simple: $\tilde{L} \ = c\tilde{\Pi}'$, where c is a constant that depends on the risk aversion γ. So the solution to the mean-variance problem using the original variables is

$$L = \Sigma_S^{-1/2}\tilde{L}{\Sigma_R}^{-1/2} = c\Sigma_S^{-1/2}\tilde{\Pi}'{\Sigma_R}^{-1/2} = c\Sigma_S^{-1}\Pi{\Sigma_R}^{-1} \tag{64}$$

In other words, the optimal portfolio is

$$w_t = L' S_t = c{\Sigma_R}^{-1}\Pi\Sigma_S^{-1}S_t = \ c\ {\Sigma_R}^{-1}E(R_{t+1}|S_t) \tag{65}$$

which is the standard Markowitz tangency portfolio (scaled by c depending on risk aversion). The last equality assumes that the conditional expected return can be computed using the multivariate regression of R_{t+1} on S_t, that is, $R_{t+1} = AS_t + \varepsilon_{t+1}$, and uses that the regression coefficient is $A = \Pi\Sigma_S^{-1}$.

So while our framework can nest the standard Markowitz solution, we seek to add robustness in several ways. First, we introduce the worst-case variance (captured by the operator matrix norm). Second, we avoid having to invert two matrices. Indeed, the Markowitz would first run a regression, requiring the matrix inversion Σ_S^{-1} and then perform a portfolio optimization, requiring the matrix inversion, ${\Sigma_R}^{-1}$, which is known to be unstable in practice. Instead, we formulate a simpler objective function, leading to a solution that is simply a sum of singular vectors.

B Proofs of Propositions

Proof of Proposition 1. Using the identity $\mathrm{tr}(AB) = \mathrm{tr}(BA)$ for any two square matrices A, B, we get

$$\mathrm{tr}(E(S_t' L R_{t+1})) \;=\; \mathrm{tr}(E(L R_{t+1} S_t')) \;=\; \mathrm{tr}(L\, E(R_{t+1} S_t')) \;=\; \mathrm{tr}(L\Pi) \;=\; \mathrm{tr}(\Pi L)\,.$$

$\square$

Proof of Proposition 2. The proof follows directly from Proposition 1 and the fact that $\mathrm{tr}(X'X) \geq 0$ for any matrix X. $\square$

Proof of Proposition 3. We provide a proof in the case of a general Schatten p-norm considered in Proposition 11, where $p = [1, \infty]$ and q is defined by $1/p + 1/q = 1$. First, the trace of any square matrix $A = UDV'$ is less than the sum of its singular values (d_k):

$$|\,\mathrm{tr}(A)| \;=|\,\mathrm{tr}(UDV')| = |\,\mathrm{tr}(V'UD)| = |\sum_k d_k (V'U)_{k,k}| \leq \sum_k d_k = \|A\|_1\,. \tag{66}$$

since $|(V'U)_{k,k}| = |V'_{\text{column } k} U_{\text{column } k}| \leq \|V_{\text{column } k}\|\;\|U_{\text{column } k}\| = 1$. Combining this inequality with Hölder's inequality for Schatten norms (see, e.g., Bhatia (1997), Corollary IV.2.6; or Tao (2012), p. 55, Exercise 1.3.9), we get:

$$|\,\mathrm{tr}(L\Pi)| \;\leq \|\Pi L\|_1 \leq \; \|\Pi\|_q\, \|L\|_p\,, \tag{67}$$

Finally, equality is achieved if L is proportional to $(\Pi'\Pi)^{q/2-1}\Pi'$. Thus,

$$\arg \max_{\|L\|_p \leq 1} \mathrm{tr}(L\Pi) \;=\; (\Pi'\Pi)^{q/2-1}\Pi' / \|(\Pi'\Pi)^{q/2-1}\Pi'\|_p\,. \tag{68}$$

$\square$

Proof of Proposition 4. Follows from the calculations in the main text. $\square$

Proof of Lemma 1. Suppose that there exists a tradable factor $F_{t+1} \;=\; x_t' R_{t+1}$ such that $S_{i,t} \;=\; \frac{\mathrm{Cov}_t(R_{i,t+1}, F_{t+1})}{\mathrm{Var}_t(F_{t+1})}$. We have

$$\mathrm{Cov}_t(R_{i,t+1}, F_{t+1}) \;=\; \mathrm{Cov}_t(R_{i,t+1}, x_t' R_{t+1}) \;=\; (\Sigma_t^R x_t)_i$$

and, hence, $S_t \;=\; \Sigma_t^R x_t\,/y$, where we have defined $y \;=\; \mathrm{Var}_t(F_{t+1})$. Furthermore,

$$\mathrm{Var}_t(F_{t+1}) \;=\; x_t'\Sigma_t^R x_t$$

Thus, we get $x_t \;=\; y(\Sigma_t^R)^{-1}S_t$, and we get a fixed point equation for y :

$$y \;=\; \mathrm{Var}_t(F_{t+1}) \;=\; x_t'\Sigma_t^R x_t \;=\; y^2 S_t'(\Sigma_t^R)^{-1}S_t \;\;\Leftrightarrow\;\; y \;=\; 1/S_t'(\Sigma_t^R)^{-1}S_t\,.$$

Reverting the arguments, we see that the converse is also true: the just computed portfolio x_t does satisfy $S_{i,t} \;=\; \frac{\mathrm{Cov}_t(R_{i,t+1},F_{t+1})}{\mathrm{Var}_t(F_{t+1})}$. □

Proof of Proposition 5. By Lemma 2, we have

$$\begin{aligned} &\mathrm{tr}(L\Pi) \;=\; \mathrm{tr}((L^s+L^a)(\Pi^s+\Pi^a)) \\ &=\; \mathrm{tr}(L^s\Pi^s) \;+\; \mathrm{tr}(L^s\Pi^a) \;+\; \mathrm{tr}(L^a\Pi^s) \;+\; \mathrm{tr}(\text{Ł}^a\Pi^a) \;=\; \mathrm{tr}(L^s\Pi^s) \;+\; \mathrm{tr}(L^a\Pi^a)\,. \end{aligned} \tag{69}$$

Finally,

$$\begin{aligned} \mathrm{Cov}_t(R^{w_t}_{t+1},F_{t+1}) &= \mathrm{Cov}_t(w_t'R_{t+1},\;F_{t+1}) = w_t'\mathrm{Cov}_t(R_{t+1},F_{t+1}) \\ &= \mathrm{Var}_t(F_{t+1})\,w_t'S_t \;=\; \mathrm{Var}_t(F_{t+1})\,S_t'LS_t \;=\; \mathrm{Var}_t(F_{t+1})\,S_t'L^sS_t\,, \end{aligned} \tag{70}$$

where the third identity uses the definition of F from Lemma 1 and the fact that $\mathrm{Var}_t(F_{t+1}) = 1/(S_t'\,(\Sigma_{R,t})^{-1}S_t)$, and the last identity follows because, by Lemma 2,

$$S_t'LS_t \;=\; S_t'L^sS_t \;+\; S_t'L^aS_t \;=\; S_t'L^sS_t\,. \tag{71}$$

□

Proof of Lemma 2. Since the trace of a matrix equals the trace of its transpose, we have

$$\mathrm{tr}(AB) = \mathrm{tr}((AB)') = \mathrm{tr}(B'A') = -\,\mathrm{tr}(BA) = -\,\mathrm{tr}(AB)$$

which shows that $\mathrm{tr}(AB) = \mathrm{tr}(BA) = 0$. Similarly, $x'Ax = (x'Ax)' = x'A'x = -x'Ax$, showing that $x'Ax = 0$. □

Proof of Proposition 6. To see the first result, note that the return equation (25) combined with the eigendecomposition (27) yield

$$E\left(S_t^{w_k^s} R_{t+1}^{w_k^s}\right) \;=\; (w_k^s)'\Pi^s w_k^s = (w_k^s)' \sum_{j=1}^{K} \lambda_j^s w_j^s (w_j^s)' w_k^s = \lambda_k^s\,. \tag{72}$$

The last claim follows directly from (68) for Π^s because, for $L = L'$, by Proposition 5 we have that

$$\max_{\|L\|_p \le 1,\; L=L'} \mathrm{tr}(L\Pi) \;=\; \max_{\|L\|_p \le 1,\; L=L'} \mathrm{tr}(L\Pi^s)$$

is attained by the symmetric matrix $c((\Pi^s)'\Pi^s)^{q/2-1}(\Pi^s)' \;=\; (|\Pi^s|)^{q-2}\Pi^s$. For $q = 1$, we get $|\Pi^s|^{-1}\Pi^s \;=\; \mathrm{sign}(\Pi^s)$. Here, we have used the standard functional calculus for symmetric matrices (Horn and Johnson (1991)): for any function $f(x)$ (such as $|x|$ or $\mathrm{sign}(x)$) we define $f(L) \;=\; W\,\mathrm{diag}(f(\lambda(L)))W'$ where $L = W\,\mathrm{diag}(\lambda(L))W'$ is the eigen-decomposition of a symmetric matrix L. $\square$

Proof of Proposition 7. Next, since $W = (w_1^s, ..., w_N^s)$ forms an orthonormal basis of $\mathbb{R}^N$, we have $WW' = Id$ so

$$\widetilde{F}_{t+1} = S_t' R_{t+1} = S_t' W W' R_{t+1} = (W' S_t) \cdot (W' R_{t+1}) = \sum_{k=1}^{N} S_t^{w_k^s} R_{t+1}^{w_k^s} \tag{73}$$

Hence, the result follows from the fact that including negative eigenvalues lowers the expected return relative to the other options considered. $\square$

Proof of Lemma 3. Equip $\mathbb{C}^N$ with the standard inner product

$$x \hat{\cdot} y \;=\; \sum_i x_i \bar{y}_i \tag{74}$$

and recall that the Hermitian adjoint of a matrix B is defined as $B^* \;=\; \bar{B}'$, where $\bar{B}$ is the complex adjoint of B. Furthermore, for any matrix B, we have

$$x \hat{\cdot} (By) \;=\; (B^* x) \hat{\cdot} y\,. \tag{75}$$

Let now A be a real anti-symmetric matrix. Consider the matrix iA. The first observation is that iA is a Hermitian matrix. Indeed, $\overline{iA}' \;=\; -iA' \;=\; iA$. Thus, iA has real eigenvalues and a basis of complex eigenvectors $\{w_k\}_{k=1}^N$. Let $\lambda \in \mathbb{R}$ be an eigenvalue of iA :

$$iAw \;=\; \lambda w \tag{76}$$

Then, taking a complex conjugate of this identity, we get

$$-iA\bar{w} \;=\; \lambda\,\bar{w} \tag{77}$$

and hence $\bar{w}$ is an eigenvector of iA with the eigenvalue $-\lambda$. Hence, all non-zero eigenvalues come in pairs.

Furthermore, $\det(A) = \det(A') = \det(-A) = (-1)^N \det(A)$ so, if N is odd, A is degenerate and has a zero eigenvalue, whereas all non-zero eigenvalues come in pairs. Let us take all nonnegative eigenvalues $\lambda_1 \geq \cdots \geq \lambda_K$ of iA and let w_k be the respective complex eigenvectors. By the above, $-\lambda_k$ is also an eigenvalue, and the respective eigenvectors are $\bar{w}_k$. By the spectral decomposition theorem, we have

$$iA \;=\; \sum_{k=1}^{K}(\lambda_k w_k \bar{w}_k' \;-\; \lambda_k \bar{w}_k w_k') \tag{78}$$

where we have used that the orthogonal projection onto w_k is $w_k\bar{w}_k'$, where $\bar{w}_k$ is the complex conjugate vector. Now, we have

$$\begin{aligned}\lambda_k w_k \bar{w}_k' \;-\; \lambda_k \bar{w}_k w_k' &= 0.5\lambda_k(\,w_{k,1} + iw_{k,2})(w_{k,1} - iw_{k,2})' \\ &\quad - 0.5\lambda_k(w_{k,1} - iw_{k,2})(w_{k,1} + iw_{k,2})' \\ &= i\lambda_k(w_{k,2}w_{k,1}' - w_{k,1}w_{k,2}')\end{aligned} \tag{79}$$

and the claim follows.

For any Hermitian matrix (and, hence, also for iA), eigenvectors for different eigenvalues are always orthogonal. Thus, $w_{k,1} \pm iw_{k,2}$ must be orthogonal to $w_{j,1} \pm iw_{j,2}$ and hence $w_{k,1}, w_{k,2}$ are orthogonal to $w_{j,1},\ w_{j,2}$. Furthermore, $w_{k,1} \pm iw_{k,2}$ correspond to different eigenvalues $\pm\lambda_k$ and hence they also must be orthogonal:

$$0 \;=\; (w_{k,1} + iw_{k,2}) \cdot (w_{k,1} + iw_{k,2}) \;=\; \|w_{k,1}\|^2 - \|w_{k,2}\|^2 + 2iw_{k,1} \cdot w_{k,2} \tag{80}$$

and hence $\|w_{k,1}\| = \|w_{k,2}\|$ and $w_{k,1} \cdot w_{k,2} = 0$. Thus, the two vectors are $w_{k,1},\ w_{k,2}$ are also orthogonal.

Note that

$$iA(w_{k,1} + iw_{k,2}) \;=\; \lambda_k(w_{k,1} + iw_{k,2}) \tag{81}$$

is equivalent to $Aw_{k,1} \;=\; \lambda_k w_{k,2}$ and $Aw_{k,2} = -\lambda_k w_{k,1}$ implying that, in the basis $\{(w_{k,1}, w_{k,2})\}_{k=1}^K$ the matrix A is block-diagonal, composed of diagonal blocks

$$\begin{pmatrix} 0 & -\lambda_1 \\ \lambda_1 & 0 \end{pmatrix} \tag{82}$$

$\square$

Proof of Proposition 8. By definition, $(\Pi^a)'(x_j + iy_j) = i\lambda_j^a(x_j + iy_j)$, that is $(\Pi^a)'x_j = -\lambda_j^a y_j,\ (\Pi^a)'y_j = \lambda_j^a x_j$. The expected return is

$$\begin{aligned} E(R_{t+1}^{w_{j,t}^a}) =& E(S_t'(x_j y_j' \,-y_j x_j')R_{t+1}) = \;\operatorname{tr}((x_j y_j' \,-y_j x_j')\Pi^a) \;=\; -\operatorname{tr}((\Pi^a)'(x_j y_j' \,-y_j x_j')) \\ &= \lambda_j^a \operatorname{tr}(y_j' y_j \;+\; x_j' x_j) \;=\; 2\lambda_j^a\,. \end{aligned}$$

The last statement follows from (68) for Π^a because, for $L = -L'$, by Proposition 5 we have that

$$\max_{\|L\|_p\le 1,\ L=-L'} \operatorname{tr}(L\Pi) \;=\; \max_{\|L\|_p\le 1,\ L=-L'} \operatorname{tr}(L\Pi^a)$$

is attained by the anti-symmetric matrix $c((\Pi^a)'\Pi^a)^{q/2-1}(\Pi^a)' \;=\; (|i\Pi^a|)^{q-2}(\Pi^a)'$. For $q = 1$, we get $|i\Pi^a|^{-1}(\Pi^a)' \;=\; \operatorname{sign}(i\Pi^a)$. Here, we have used the standard functional calculus for Hermitian matrices (Horn and Johnson (1991)): for any function $f(x)$ (such as $|x|$ or $\operatorname{sign}(x)$) we define $f(L) \;=\; W\operatorname{diag}(f(\lambda(L)))W'$ where $L = W\operatorname{diag}(\lambda(L))W'$ is the eigen-decomposition of a Hermitian matrix L. $\square$

Proof of Proposition 9. Follows from calculations in the body of the paper. $\square$

Proof of Proposition 10. Direct calculation shows that

$$\Pi \;=\; (E[(\kappa_t + \beta_t)E_t[R_{i,t+1}]E_t[R_{j,t+1}]])_{i,j=1}^N$$

which yields the result. $\square$

Proof of Proposition 11. Let $X = L\Pi$. Then, by a result of Fan and Hoffman (1955), we have $\lambda_i(X^s) \;\le\; \bar{\lambda}_i(X)$. Furthermore, since $\operatorname{rank}(L) \le K$, we also have $\operatorname{rank}(X) \le K$ and

$\mathrm{rank}(X'X) = \mathrm{rank}(X) \leq K$, and hence there are at most K non-zero singular values of X. Thus,

$$\mathrm{tr}(L\Pi) \;=\; \mathrm{tr}(X) \;=\; \mathrm{tr}(X^s) \;=\; \sum_{i=1}^{N} \lambda_i(X^s) \;\leq\; \sum_{i=1}^{K} \bar{\lambda}_i(X)\,.$$

Second, by known result about singular values of products of matrices (see, for example, Marshall and Olkin (1979), p. 248), we have

$$\sum_{i=1}^{K} \bar{\lambda}_i(L\Pi) \;\leq\; \sum_{i=1}^{K} \bar{\lambda}_i(L)\,\bar{\lambda}_i(\Pi)$$

Third, by the Hölder inequality

$$\sum_{i=1}^{K} \bar{\lambda}_i(L)\,\bar{\lambda}_i(\Pi) \;\leq\; \left(\sum_{i=1}^{K} \bar{\lambda}_i(L)^p\right)^{1/p} \left(\sum_{i=1}^{K} \bar{\lambda}_i(\Pi)^q\right)^{1/q} \;=\; \|L\|_p \left(\sum_{i=1}^{K} \bar{\lambda}_i(\Pi)^q\right)^{1/q}\,.$$

Thus,

$$\max_{\|L\|_p \leq 1,\ \mathrm{rank}(L)\leq K} \mathrm{tr}(L\Pi) \;\leq\; \left(\sum_{i=1}^{K} \bar{\lambda}_i(\Pi)^q\right)^{1/q}\,.$$

Thus, it remains to verify that the equality holds with $L = c\sum_{k=1}^{K} \bar{\lambda}_k^{q-1} v_k(u_k)'$, where $c = \left(\sum_{k=1}^{K} \bar{\lambda}_k^q\right)^{-1/p}$. This follows directly from the identity $\mathrm{tr}(Lv_k(u_k)')$, established in (17)

Items 2 and 3 of the Proposition are in fact special cases of item 1. Indeed, for item 2, we have by Proposition 5 that

$$\begin{aligned}
&\max_{\|L\|_p \leq 1,\ L=L'\,\mathrm{rank}(L)\leq K} \mathrm{tr}(L\Pi) \;=\; \max_{\|L\|_p \leq 1,\ L=L',\ \mathrm{rank}(L)\leq K} \mathrm{tr}(L\Pi^s) \\
&\leq \max_{\|L\|_p \leq 1,\ \mathrm{rank}(L)\leq K} \mathrm{tr}(L\Pi^s) \;=\; \left(\sum_{i=1}^{K} \bar{\lambda}_i(\Pi^s)^q\right)^{1/q}\,.
\end{aligned} \tag{83}$$

Furthermore, $\bar{\lambda}_i(\Pi^s)$ is the i-th largest absolute eigenvalue of Π^s and the equality is achieved with $L = c\sum_{\mathcal{K}} |\lambda_k^s|^{q-1}\,\mathrm{sign}(\lambda_k^s) w_k^s (w_k^s)'$

Similarly, for the anti-symmetric part, we have

$$
\begin{aligned}
\max_{\|L\|_p\le 1,\ L=-L'\,\mathrm{rank}(L)\le K} \mathrm{tr}(L\Pi) &= \max_{\|L\|_p\le 1,\ L=-L',\ \mathrm{rank}(L)\le K} \mathrm{tr}(L\Pi^a) \\
\le \max_{\|L\|_p\le 1,\ \mathrm{rank}(L)\le K} \mathrm{tr}(L\Pi^a) &= \left(\sum_{i=1}^{K} \bar{\lambda}_i(\Pi^a)^q\right)^{1/q} .
\end{aligned}
\tag{84}
$$

Top $2K$ singular value of Π^a are just λ_i^a counted twice, and the equality is achieved with $L = c\sum_{k=1}^{K}(\lambda_k^a)^{q-1}(w_k^a(\bar{w^a}_k)' - \bar{w^a}_k(w_k^a)')$, where $w_k^a = x_k + iy_k$ are the *complex* eigenvectors of Π^a and come in complex conjugate pairs according to Lemma 3. $\square$

C Additional Asset Pricing Tests (Extensions of Section 4)

Proposition 12 *For any adapted stochastic process κ_t, let $\Pi(\kappa) = (E(\kappa_t R_{i,t+1} S_{j,t})_{i,j=1}^N) = E(\kappa_t R_{t+1} S_t')$. The following is true:*

- *We have $E_t(R_{i,t+1}) = \theta_t S_{i,t}$ with $\theta_t \ge 0$ if and only if $\Pi(\kappa)$ is symmetric and positive semi-definite for any adapted stochastic process $\kappa_t \ge 0$.*
- *We have $E_t(R_{i,t+1}) = \theta_t S_{i,t}$ with some θ_t if and only if $\Pi(\kappa)$ is symmetric for any adapted stochastic process κ_t.*

Proof of Proposition 12. The "only if" part follows directly from calculations in the body of the paper. To prove the converse, we note first that, by the law of iterated expectations, $\Pi(\kappa)$ is symmetric if and only if

$$E(\kappa_t(E_t(R_{i,t+1})S_{j,t} - E_t(R_{j,t+1})S_{i,t})] \;=\; 0\,.$$

Since κ_t is arbitrary nonnegative, and any random variable can be written down as a difference of two non-negatives, we get that

$$E_t(R_{i,t+1})S_{j,t} - E_t(R_{j,t+1})S_{i,t} \;=\; 0$$

This means that the vectors $E_t(R_{t+1})$ and S_t are co-linear and hence $E_t(R_{t+1}) = \theta_t S_t$ for some θ_t. Hence,

$$\Pi(\kappa) \;=\; E(\kappa_t \theta_t S_t S_t')\,.$$

Suppose for simplicity that $S_t \neq 0$ with probability one. Then, there exists a vector π such that $\pi' S_t \neq 0$ with probability one. Then, $\pi'\Pi(\kappa)\pi \;=\; E(\kappa_t\theta_t(\pi' S_t)^2) > 0$ for all $\kappa_t \geq 0$, which implies that $\theta_t \geq 0$. $\square$

Proposition 13 (Alpha and negative eigenvalues) *Suppose that*

$$R_{j,t+1} \;=\; \alpha_t \;+\; S_{j,t}F_{t+1} \;+\; \varepsilon_{j,t+1},\; E_t(\varepsilon_{j,t+1}) \;=\; 0$$

as in Frazzini and Pedersen (2014). Then, Π^s has at most one negative eigenvalue, while Π^a has at most one positive eigenvalue.

Furthermore, let $Q \;=\; (E(S_{j,t}\alpha_t))_{j=1}^N$, and define $\pi = Q/\|Q\| - \mathbf{1}_N N^{-1/2}$.

- *$\Pi^a \;=\; 0.5(\mathbf{1}Q' - Q\mathbf{1}')$ is non-zero if and only if Q is not proportional to $\mathbf{1}_N$.*
- *If*

$$E((\pi' S_t)^2\, E_t(F_{t+1})) \;<\; (1 - N^{-1/2}Q'\mathbf{1}/\|Q\|)^2\,,$$

then Π^s has exactly one negative eigenvalue.[22]

Proof of Proposition 13. The proof of the first claim follows directly from Weyl inequalities (99) in complete analogy with the proof of Proposition 20. The second claim follows because

$$\pi'\Pi^s\pi \;=\; E((\pi' S_t)^2\, E_t(F_{t+1})) \;-\; (1 - N^{-1/2}Q'\mathbf{1}/\|Q\|)^2$$

and the inequality $\pi'\Pi^s\pi$ implies that Π^s has a negative eigenvalue. $\square$

Proposition 14 (Omitted factors and negative eigenvalues) *Suppose that*

$$R_{j,t+1} \;=\; S_{j,t}\,F_{t+1} \;+\; \beta_{j,t}\,F^1_{t+1} \;+\; \varepsilon_{j,t+1},\; E_t(\varepsilon_{j,t+1}) \;=\; 0\,.$$

Suppose that $\mathrm{Cov}(S_{i,t}F^1_{t+1},\; \beta_{j,t}\,) \;=\; 0$ for all i, j. Then, Π^s has at most one negative eigenvalue, while Π^a has at most one positive eigenvalue. Furthermore, let $\bar{S} \;=\; (E(S_{j,t}F^1_{t+1}))_{j=1}^N$, $\bar{\beta} \;=\; (E(\beta_{j,t}))_{j=1}^N$, and define $\pi = \bar{S}/\|S\| - \bar{\beta}/\|\beta\|$. Then,

- *$\Pi^a \;=\; 0.5(\bar{\beta}\bar{S}' - \bar{S}\bar{\beta}')$ is non-zero if and only if $\bar{S}$ is not proportional to $\bar{\beta}$.*

[22] This condition is only sufficient, but simple. It is possible to derive a necessary and sufficient condition, but it is more involved.

- *If*

$$E((\pi' S_t)^2 \, E_t(F_{t+1})) \ < \ (1 - \bar{S}'\bar{\beta})^2 \, ,$$

then Π^s *has exactly one negative eigenvalue.*[23]

- *the result extends to the case when returns have a multi-factor structure, in which case* Π^s *has at most* K *negative eigenvalues if there are* K *omitted factors under the condition* $\mathrm{Cov}(S_{i,t}F^k_{t+1}, \ \beta_{j,k,t}) \ = \ 0$ *for all* i, j *and all factors* $k = 1, \cdots, K$*. Similarly,*

$$\Pi^a \ = \ 0.5 \sum_k (\bar{\beta}_k \bar{S}'_k - \bar{S}_k \bar{\beta}'_k)$$

where $\bar{S}_k \ = \ (E(S_{j,t}F^k_{t+1}))^N_{j=1}, \ \bar{\beta}_k \ = \ (E(\beta_{j,k,t}))^N_{j=1},$

Proof of Proposition 14. The proof is completely analogous to that of Proposition 13.

□

Proposition 15 (Sign-changing risk premium or omitted factors) *Suppose that returns have a* $(K+1)$*-factor structure with some latent factor exposures* $\beta_{i,k,t}$ *such that*

$$E_t[R_{i,t+1}] \ = \ \theta_t \, S_{i,t} \ + \ \sum_{k=1}^{K} \gamma_{k,t}\beta_{i,k,t} \, .$$

Suppose also that

$$E[\kappa_t \, S_{j,t} \sum_{k=1}^{K} \gamma_{k,t}\beta_{i,k,t}] \ = \ 0$$

for all $i \neq j$*. That is, there is no co-movement between factor exposures. Then,*

- $\Pi(\kappa)$ *is symmetric*

- $\theta_t \geq 0$ *if and only if* $\Pi(\kappa)$ *is positive semi-definite for any* $\kappa_t \geq 0$*.*

[23] This condition is only sufficient. It is possible to derive a necessary and sufficient condition, but it is more cumbersome.

D Central Limit Theorems for Asset Pricing Tests

We next develop central limit theorems (CLTs) for empirically testing the asset pricing bounds of Proposition 9. The starting point is to estimate of the prediction matrix, Π, by its sample counterpart:

$$\hat{\Pi}_{i,j}(T) \;=\; \frac{1}{T}\sum_{t=1}^{T} R_{i,t+1}S_{j,t}\,. \tag{85}$$

Based on this estimate, the question is whether we can reject that the estimated $\hat{\Pi}$ satisfies the conditions in Proposition 9.

Proposition 9 rules out two types of alpha when signals are betas: one coming from negative eigenvalues of $\hat{\Pi}^s$ and another coming from a non-zero $\hat{\Pi}^a$. Hence, we test (i) whether the negative eigenvalues of $\hat{\Pi}^s$ are statistically significant or are just sampling noise; and (ii) whether $\hat{\Pi}^a$ is significantly different from zero. We use the following distributional assumption.

Assumption 1 *Excess returns can be written as*

$$R_{i,t+1} \;=\; \theta_t\, S_{i,t} \;+\; \varepsilon_{i,t+1}\,, \tag{86}$$

where ε_{t+1} are i.i.d. over time with zero mean and covariance matrix Σ^ε; $\theta_t \geq 0$ are i.i.d. over time with finite first and second moments; signals $S_t \sim N(\bar{S},\ \Sigma^S)$ are i.i.d. over time; S_t, θ_t, and ε_{t+1} are jointly independent; and the eigenvalues of Π^s are pairwise disjoint, $\lambda_1^s > \ \cdots \ > \ \lambda_N^s$.

This assumption has several parts. First, (86) is simply a way of stating that we are working under the null hypothesis (44) from Proposition 9. Second, we make some distributional assumptions, and, lastly, we assume that none of the eigenvalues are the same, which holds generically (so this is a minimal loss of generality). The following proposition shows how to implement the first test.

Proposition 16 (Eigenvalue CLT) *Under Assumption 1, in the limit as $T \to \infty$, the estimated eigenvalues $\hat{\Lambda}^s \;=\; (\hat{\lambda}_k^s)_{k=1}^N$ of $\hat{\Pi}^s$ converge almost surely to the true eigenvalues, $\Lambda^s = (\lambda_k^s)_{k=1}^N$, and $\sqrt{T}(\hat{\Lambda}^s - \Lambda^s)$ converges in distribution to $N(0, \Sigma^\Lambda)$, where Σ^Λ has elements*

$$\Sigma^{\Lambda}_{k,l} = \lambda^s_k \lambda^s_l \frac{\mathrm{Var}(\theta)}{(E(\theta))^2} - 2E(\theta^2)(\bar{S}'w^s_k)^2(\bar{S}'w^s_l)^2 + \delta_{k,l}\lambda^s_k \left(2\frac{E(\theta^2)}{(E(\theta))^2}\lambda^s_k + \frac{(w^s_k)'\Sigma_\varepsilon w^s_k}{E(\theta)} \right) . \quad (87)$$

It is interesting that we can compute an asymptotic distribution of the ranked eigenvalues. The first two terms of (87) show the covariance between eigenvalues, while variances are driven by the last term since the kronecker delta, $\delta_{k,l}$, is 0 when $k \neq l$ and otherwise 1.[24]

This proposition shows how to test Proposition 9. Suppose for example that the estimated prediction matrix $\hat{\Pi}^s$ has 3 negative eigenvalues collected in a vector $\lambda_{neg} = (\lambda^s_k)^N_{k=N-2}$. Then we can use (87) to compute their 3-by-3 variance-covariance matrix, $\Sigma_{neg} = (\Sigma^{\Lambda}_{k,l})_{k,l\geq N-2}$, and the test whether these are in fact different from zero via $\lambda'_{neg}\Sigma^{-1}_{neg}\lambda_{neg}$, a Chi-squared test with 3 degrees of freedom.

Turning to the second test, we need to consider whether $\hat{\Pi}^a$ is significantly different from zero. Since $\hat{\Pi}^a$ is antisymmetric by construction, we know that the diagonal is zero and the lower triangle equals minus the upper triangle—so we only need to test that the upper triangle is zero. Therefore, we focus on $Z_T = \sqrt{T}\,\mathrm{vec}\left(\{\hat{\Pi}^a_{T,i,j}\}_{i=1,\ldots,N,j>i}\right) \in \mathbb{R}^{\frac{N(N-1)}{2}}$, which is the upper triangle "stacked" into a vector (the vec-operator means stacking) and scaled by sample size, T, to ensure a non-trivial asymptotic distribution.

Proposition 17 (Antisymmetric CLT) *Let $\bar{\Sigma}^S_{i,j} = E(S_{i,t}S_{j,t})$. The vector Z_T is asymptotically normal with zero mean and asymptotic covariance between $Z_{(i,j)}$ and $Z_{(k,l)}$ given by*

$$\Sigma^Z_{(i,j),(k,l)} = \frac{1}{4}\left(\bar{\Sigma}^S_{i,k}\Sigma^\varepsilon_{j,l} - \bar{\Sigma}^S_{j,k}\Sigma^\varepsilon_{i,l} - \bar{\Sigma}^S_{i,l}\Sigma^\varepsilon_{j,k} + \bar{\Sigma}^S_{j,l}\Sigma^\varepsilon_{i,k}\right) .$$

Hence, $Z'_T(\Sigma^Z)^{-1}Z_T$ is asymptotically χ^2-distributed with $\frac{N(N-1)}{2}$ degrees of freedom. Further, Z'_TZ_T has an asymptotic distribution given by $\sum_{k=1}^{N(N-1)/2}\lambda^Z_k Y_k$, where Y_k are i.i.d. χ^2_1 variables and $(\lambda^Z_k)_{k=1,\cdots,\frac{N(N-1)}{2}}$ are the eigenvalues of Σ^Z.

This proposition shows two ways to test whether the antisymmetric part, captured in Z, is zero. One test is a standard Chi-squared test. The second test is based on a slightly more complicated distribution, a weighted sum of Chi-squares, but this test has the

[24]The second term in (87) can make the eigenvalues' covariances negative, which intuitively is related to the fact that their sum is fixed (as the trace of the prediction matrix). Clearly, variances must be positive, which follows from $\lambda^s_k = (w^s_k)'\Pi^s w^s_k = E(\theta)((w^s_k)'\Sigma^S w^s_k + (w^s_k)'\bar{S}\bar{S}'w^s_k) \geq E(\theta)((w^s_k)'\bar{S}\bar{S}'w^s_k) = E(\theta)(\bar{S}'w^s_k)^2$, which implies that the third term is greater than the second, $\frac{E(\theta^2)}{(E(\theta))^2}(\lambda^s_k)^2 \geq E(\theta^2)(\bar{S}'w^s_k)^4$.

potential benefit that we need not invert the asymptotic covariance matrix, just compute its eigenvalues.

Proposition 18 *Fix a portfolio w. The finite sample estimate*

$$\sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T} S_t^w\, R_{t+1}^w \;-\; w^\top \Pi^s w\right) \;\sim\; N(0,\; \sigma_s^2(\pi))$$

where

$$\sigma_s^2(\pi) \;=\; (3(w'\bar{\Sigma}^S w)^2 \;-\; 2(\bar{S}'w)^4)\, E(\theta^2) \;+\; (w'\bar{\Sigma}w)\, w'\Sigma^\varepsilon w - \;(E(\theta)w'\bar{\Sigma}^S w)^2\,.$$

Proposition 19 *Fix a pair of portfolios x, y. The finite sample estimate of the mean return on the basic alpha portfolio*

$$\sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T}(S_t^x\, R_{t+1}^y - S_t^y\, R_{t+1}^x)\right) \;\sim\; N(0,\; \sigma_a^2(x,\, y))$$

where

$$\sigma_a^2(x,\, y) \;=\; x'\bar{\Sigma}^S x\, y'\Sigma^\varepsilon y \;+\; y'\bar{\Sigma}^S y\, x'\Sigma^\varepsilon x \;-\; 2x'\bar{\Sigma}^S y\, x'\Sigma^\varepsilon y\,.$$

D.1 Proof of CLTs

Proof of Proposition 16. The proof will be based on several technical lemmas

Lemma 4 *Consider a sequence of i.i.d. random vectors $X_t \in \mathbb{R}^d$ with finite first two moments and a covariance matrix Σ^X, and let*

$$\bar{X}_T \;=\; \frac{1}{T}\sum_{t=1}^{T} X_t\,.$$

Let $F:\ \mathbb{R}^d \;\to\; \mathbb{R}^K$ be a continuous function that is continuously differentiable in a small neighbourhood of $E(X)$, and let $A \;=\; DF(E(X))$ be the Jacobian of F evaluated at $E(X)$. Then, the following central limit theorem holds for F :

$$\sqrt{T}(F(\bar{X}_T) - F(E(X))) \;\to\; N(0, DF\,\Sigma^X\, DF')$$

in distribution.

Proof of Lemma 4. Standard results imply that it suffices to prove convergence of characteristic functions

$$E[e^{\alpha'\sqrt{T}(F(\bar{X}_T)-F(E(X)))}]$$

to that of the Gaussian distribution. Using the Taylor approximation, we get

$$\sqrt{T}(F(\bar{X}_T)-F(E(X))) \;=\; DF\sqrt{T}(\bar{X}_T-E(X)) \;+\; \sqrt{T}\,o(\|\bar{X}_T-E(X)\|)$$

By the central limit theorem, $\sqrt{T}(\bar{X}_T - E(X)) \;\to\; N(0,\Sigma^X)$ in probability and hence the error term $Y_T = \sqrt{T}\,o(\|\bar{X}_T - E(X)\|)$ converges to zero in probability, whereas $Z_T = DF\sqrt{T}(\bar{X}_T - E(X))$ converges to $N(0, DF\,\Sigma^X DF')$ in distribution. Pick $\varepsilon, \delta > 0$. Then, $Prob(|Y_T| > \varepsilon) < \delta$ for all large T. Then,

$$E[e^{i\alpha'\sqrt{T}(F(\bar{X}_T)-F(E(X)))}] \;=\; E[e^{i\alpha'(Z_T+Y_T)}]$$

and we know that $e^{i\alpha' Z_T}$ converges to the Gaussian characteristic function. Now,

$$\begin{aligned}
&|E[e^{i\alpha'(Z_T+Y_T)}] - E[e^{i\alpha'(Z_T)}] \;=\; |E[e^{i\alpha'(Z_T+Y_T)} - e^{i\alpha' Z_T}]| \\
&= \; |E[e^{i\alpha'(Z_T+Y_T)} - e^{i\alpha' Z_T}]| \;\le\; E[|e^{i\alpha' Y_T} - 1|] \\
&\le \; E[|e^{i\alpha' Y_T} - 1|(\mathbf{1}_{|Y|>\varepsilon} + \mathbf{1}_{|Y|\le\varepsilon})] \;\le C|\alpha|\varepsilon \;+\; 2\,Prob(|Y_T| > \varepsilon)
\end{aligned} \tag{88}$$

where we have used that $|e^{i\alpha'(Z_T+Y_T)}| = |e^{i\alpha' Y_T}| = |e^{i\alpha' Y_T}| = 1$ and where C is a large enough constant such that $|e^{i\alpha x} - 1| \le C|\alpha||x|$. Thus making ε, δ sufficiently small we see that the characteristic function comes arbitrarily close to that of Z_T, and the claim follows. $\square$

Lemma 5 *Let $\Lambda(A)$ be the function mapping a symmetric $N \times N$ matrix into the vector of its eigenvalues, arranged in the decreasing order. Let A be the matrix with disjoint eigenvalues $\lambda_1(A) > \cdots > \lambda_N(A)$, and let w_k be the corresponding eigenvectors. Then, Λ is continuously differentiable (in fact, real analytic) in a small neighbourhood of A and*

$$\lambda_i(A+B) \;=\; \lambda_i(A) \;+\; w_i' B w_i \;+\; O(\|B\|^2)\,.$$

Proof. Standard results imply that $\lambda_i(X)$ is a real analytic function in a small neighbourhood of A. Hence, it suffices to show that the gradient of λ_i is given by the map $B \to w_i' B w_i$.

Indeed, differentiating the identity

$$w_k(t)'(A+tB)w_k(t) \;=\; \lambda_i(A+tB)$$

with respect to t, we get

$$\frac{d}{dt}w_k(t)'(A+tB)w_k(t) \;+\; w_k(t)'Bw_k(t)+w_k(t)'(A+tB)\frac{d}{dt}w_k(t) \;=\; \frac{d}{dt}\lambda_i(A+tB)\,. \quad (89)$$

But $w_k(t)'(A+tB)=\lambda_i w_k'(t)$ and $(A+tB)w_k(t)=\lambda_i w_k(t)$. Furthermore, differentiating the identity $w_k'(t)w_k(t)=1$, we get that $\frac{d}{dt}w_k(t)'w_k(t)=0$. Substituting all this into (89), we get

$$w_k(t)'Bw_k(t) \;=\; \frac{d}{dt}\lambda_i(A+tB)\,,$$

which is what had to be proved. $\square$

Lemma 6 *Let $Y_{i,j} \;=\; S_iR_j$. Then, the covariance matrix of X is given by*

$$\Sigma^X_{i,j,k,l} \;=\; \mathrm{Cov}(X_{i,j},\; X_{k,l}) \;=\; \bar\Sigma_{i,k}\Sigma^\varepsilon_{j,l}+\mathrm{Var}(\theta)\bar\Sigma^S_{i,j}\bar\Sigma^S_{k,l}+E(\theta)^2\left(\bar\Sigma^S_{i,k}\bar\Sigma^S_{j,l}\;+\;\bar\Sigma^S_{i,k}\bar\Sigma^S_{j,l}-2\bar S_i\bar S_j\bar S_k\bar S_l\right)$$

Proof of Lemma 6. We have

$$\begin{aligned}E(X_{i,j}X_{k,l}) \;&=\; E(S_{i,t}R_{j,t+1}S_{k,t}R_{l,t+1}) \;=\; E(S_{i,t}S_{k,t}E_t(R_{j,t+1}R_{l,t+1}))\\ &=\; E(S_{i,t}S_{k,t}(\theta_t^2S_{j,t}S_{l,t}+E_t(\varepsilon_{j,t+1}\varepsilon_{l,t+1}))) \;=\; E(\theta^2)E(S_{i,t}S_{k,t}S_{j,t}S_{l,t}) \;+\; E(S_{i,t}S_{k,t})E(\varepsilon_{j,t+1}\varepsilon_{l,t+1})\,,\end{aligned} \quad (90)$$

whereas

$$E(X_{i,j})E(X_{k,l}) \;=\; E(\theta)^2E[S_iS_j]E[S_kS_l]\,.$$

By assumption, S is Gaussian, and therefore

$$
\begin{aligned}
&E(S_{i_1}S_{i_2}S_{j_1}S_{j_2})\\
&= E((S_{i_1}-\bar S_{i_1}+\bar S_{i_1})(S_{i_2}-\bar S_{i_2}+\bar S_{i_2})(S_{j_1}-\bar S_{j_1}+\bar S_{j_1})(S_{j_2}-\bar S_{j_2}+\bar S_{j_2}))\\
&= E((S_{i_1}-\bar S_{i_1})(S_{i_2}-\bar S_{i_2})(S_{j_1}-\bar S_{j_1})(S_{j_2}-\bar S_{j_2}))\\
&+ \bar S_{i_1}\bar S_{i_2}\Sigma^S_{j_1,j_2}+\bar S_{i_1}\bar S_{j_1}\Sigma^S_{i_2,j_2}+\bar S_{i_1}\bar S_{j_2}\Sigma^S_{i_2,j_1}+\bar S_{i_2}\bar S_{j_1}\Sigma^S_{i_1,j_2}+\bar S_{i_2}\bar S_{j_2}\Sigma^S_{i_1,j_1}+\bar S_{j_1}\bar S_{j_2}\Sigma^S_{i_1,i_2}+\bar S_{i_1}\bar S_{i_2}\bar S_{j_1}\bar S_{j_2}\\
&= \Sigma^S_{i_1,i_2}\Sigma^S_{j_1,j_2}\ +\ \Sigma^S_{i_1,j_2}\Sigma^S_{i_2,j_1}\ +\ \Sigma^S_{i_1,j_1}\Sigma^S_{i_2,j_2}\\
&+\bar S_{i_1}\bar S_{i_2}\Sigma^S_{j_1,j_2}+\bar S_{i_1}\bar S_{j_1}\Sigma^S_{i_2,j_2}+\bar S_{i_1}\bar S_{j_2}\Sigma^S_{i_2,j_1}+\bar S_{i_2}\bar S_{j_1}\Sigma^S_{i_1,j_2}+\bar S_{i_2}\bar S_{j_2}\Sigma^S_{i_1,j_1}+\bar S_{j_1}\bar S_{j_2}\Sigma^S_{i_1,i_2}+\bar S_{i_1}\bar S_{i_2}\bar S_{j_1}\bar S_{j_2}\\
&= \bar\Sigma^S_{i_1,i_2}\bar\Sigma^S_{j_1,j_2}\ +\ \bar\Sigma^S_{i_1,j_2}\bar\Sigma^S_{i_2,j_1}\ +\ \bar\Sigma^S_{i_1,j_1}\bar\Sigma^S_{i_2,j_2}-2\bar S_{i_1}\bar S_{i_2}\bar S_{j_1}\bar S_{j_2}\,.
\end{aligned}
\tag{91}
$$

and, hence,

$$
\begin{aligned}
\mathrm{Cov}(X_{i,j},\ X_{k,l})\ &=\ E(S_{i,t}S_{k,t})E(\varepsilon_{j,t+1}\varepsilon_{l,t+1})\\
+\ E(\theta^2)&\left(\bar\Sigma^S_{i,j}\bar\Sigma^S_{k,l}\ +\ \bar\Sigma^S_{i,k}\bar\Sigma^S_{j,l}\ +\ \bar\Sigma^S_{i,k}\bar\Sigma^S_{j,l}-2\bar S_i\bar S_j\bar S_k\bar S_l\right)-E(\theta)^2\,E(S_{i,t}S_{j,t})E(S_{k,t}S_{l,t})\\
=\ \bar\Sigma_{i,k}\Sigma^\varepsilon_{j,l}&+\mathrm{Var}(\theta)\bar\Sigma^S_{i,j}\bar\Sigma^S_{k,l}+E(\theta)^2\left(\bar\Sigma^S_{i,k}\bar\Sigma^S_{j,l}\ +\ \bar\Sigma^S_{i,k}\bar\Sigma^S_{j,l}-2\bar S_i\bar S_j\bar S_k\bar S_l\right)
\end{aligned}
\tag{92}
$$

□

Now we are ready to prove the result. Note that $\Pi=\Pi^s=E(\theta)\bar\Sigma^S$. Since, by assumption, all eigenvalues of Π^s are pairwise disjoint, Lemma 5 implies that Lemma 4 is applicable with F being the map from the matrix to its vector of eigenvalues. Thus, eigenvalues will be asymptotically normal, and we can use the central limit theorem to compute the covariance of eigenvalues: By Lemma 5

$$
\sqrt{T}(\hat\lambda^s_k-\lambda^s_k)\ =\ \sqrt{T}(\lambda_k(\hat\Pi^s)-\lambda_k(\Pi^s))\ \approx\ \sqrt{T}(w^s_k)'(\hat\Pi^s-\Pi^s)w^s_k
$$

Let

$$
d\Pi_{i,j}\ =\ \sqrt{T}(\hat\Pi_{i,j}-\Pi_{i,j})\,.
$$

Then, $E(d\Pi) = 0$ and, by Lemma 4, the covariance matrix is asymptotically given by

$$\begin{aligned}
&\mathrm{Cov}(\sqrt{T}(\hat{\lambda}_k^s - \lambda_k^s), \sqrt{T}(\hat{\lambda}_k^s - \lambda_k^s)) \\
&= \mathrm{Cov}((w_k^s)' d\Pi^s w_k^s,\ (w_l^s)' d\Pi^s w_l^s) \ = \ E((w_k^s)' d\Pi^s w_k^s (w_l^s)' d\Pi^s w_l^s) \\
&= 0.25 E(\sum_{i_1,i_2} \sum_{j_1,j_2} (d\Pi_{i_1,i_2} + d\Pi_{i_2,i_1})(d\Pi_{j_1,j_2} + d\Pi_{j_2,j_1}) w_{k,i_1} w_{k,i_2} w_{l,j_1} w_{l,j_2}) \\
&= 0.25 \sum_{i_1,i_2} \sum_{j_1,j_2} \left(\Sigma^X_{i_1,i_2,j_1,j_2} + \Sigma^X_{i_2,i_1,j_1,j_2} + \Sigma^X_{i_1,i_2,j_2,j_1} + \Sigma^X_{i_2,i_1,j_2,j_1} \right) w_{k,i_1} w_{k,i_2} w_{l,j_1} w_{l,j_2}\,,
\end{aligned} \tag{93}$$

where Σ^X is given in Lemma 6. We have

$$\Sigma^X_{i_1,i_2,j_1,j_2} \ = \ \bar{\Sigma}_{i_1,j_1} \Sigma^\varepsilon_{i_2,j_2} + \mathrm{Var}(\theta) \bar{\Sigma}^S_{i_1,i_2} \bar{\Sigma}^S_{j_1,j_2} + E(\theta)^2 \left(\bar{\Sigma}^S_{i_1,j_1} \bar{\Sigma}^S_{i_2,j_2} + \bar{\Sigma}^S_{i_1,j_2} \bar{\Sigma}^S_{i_2,j_1} - 2 \bar{S}_{i_1} \bar{S}_{i_2} \bar{S}_{j_1} \bar{S}_{j_2} \right).$$

Thus, by direct calculation,

$$\begin{aligned}
&0.25 \sum_{i_1,i_2} \sum_{j_1,j_2} \left(\Sigma^X_{i_1,i_2,j_1,j_2} + \Sigma^X_{i_2,i_1,j_1,j_2} + \Sigma^X_{i_1,i_2,j_2,j_1} + \Sigma^X_{i_2,i_1,j_2,j_1} \right) w_{k,i_1} w_{k,i_2} w_{l,j_1} w_{l,j_2} \\
&= (w_k^s)' \bar{\Sigma}^S w_l^s \, (w_k^s)' \bar{\Sigma}^\varepsilon w_l^s + \mathrm{Var}(\theta) (w_k^s)' \bar{\Sigma}^S w_k^s (w_l^s)' \bar{\Sigma}^S w_l^s \\
&+ 2E(\theta)^2 (((w_k^s)' \bar{\Sigma}^S w_l^s)^2 - (\bar{S}' w_k^s)^2 (\bar{S}' w_l^s)^2)\,.
\end{aligned} \tag{94}$$

Since eigenvectors are orthogonal, we have $(w_k^s)' \bar{\Sigma}_S w_l^s = \lambda_k^S \delta_{k,l}$ so that the last expression is given by

$$\delta_{k,l} \lambda_k^S (w_k^s)' \bar{\Sigma}^\varepsilon w_k^s + \mathrm{Var}(\theta) \lambda_k^S \lambda_l^S + 2E(\theta)^2 (\delta_{k,l} (\lambda_k^S)^2 - (\bar{S}' w_k^s)^2 (\bar{S}' w_l^s)^2)\,,$$

and formula (87) follows because $\lambda_k^S = \lambda_k^s / E(\theta)$. □

Proof of Proposition 17. We have $E[S_{i,t} \varepsilon_{j,t+1} - S_{j,t} \varepsilon_{i,t+1}] = 0$ and hence

$$\begin{aligned}
&\mathrm{Cov}((\Pi_{i,j} - \Pi_{j,i}),\, (\Pi_{k,l} - \Pi_{l,k})) \ = \ E((\Pi_{i,j} - \Pi_{j,i})(\Pi_{k,l} - \Pi_{l,k})) \\
&= \frac{1}{T^2} E\left(\left(\sum_{t=1}^T (S_{i,t} \varepsilon_{j,t+1} - S_{j,t} \varepsilon_{i,t+1}) \right) \left(\sum_{t=1}^T (S_{k,t} \varepsilon_{l,t+1} - S_{l,t} \varepsilon_{k,t+1}) \right) \right) \\
&= \frac{1}{T^2} T E((S_i \varepsilon_j - S_j \varepsilon_i)(S_k \varepsilon_l - S_l \varepsilon_k)) \ = \ \frac{1}{T} \left(\bar{\Sigma}^S_{i,k} \Sigma^\varepsilon_{j,l} - \bar{\Sigma}^S_{j,k} \Sigma^\varepsilon_{i,l} - \bar{\Sigma}^S_{i,l} \Sigma^\varepsilon_{j,k} + \bar{\Sigma}^S_{j,l} \Sigma^\varepsilon_{i,k} \right)
\end{aligned} \tag{95}$$

The claim follows now directly from the Central Limit Theorem for i.i.d. random variables. $\square$

E Theoretical Relation to Haddad et al. (2020)

Here, we put the analysis of Haddad et al. (2020) into the general framework of principal portfolios and show how PPs can be used to (1) understand the origins and structure of predictability in Haddad et al. (2020) and (2) significantly improve the performance of factor timing strategies from Haddad et al. (2020).

Let $\tilde{R} = R_{PC}$ and $\tilde{S} = S_{PC}$. Our first observation is that the prediction matrix $\Pi = (\tilde{R}'_{t+1}\tilde{S}_t)$ computed in sample is dominated by noise: The results of Haddad et al. (2020) imply that naïvely exploiting the full Π matrix can lead to poor predictive performance. As we know from formula (32),

$$E\left(\sum_{i=1}^{50} R_{t+1}^{q_i} S_t^{q_i}\right) \;=\; E\left(\sum_{i=1} \tilde{R}_{i,t+1}\tilde{S}_{i,t}\right) \;=\; \sum_i \Pi_{i,i} \;=\; \mathrm{tr}(\Pi) \;=\; \mathrm{tr}(\Pi^s)$$

where $\{q_i\}$ is the orthonormal basis of PCs. The results of Haddad et al. (2020) imply that even an estimate of the trace of the prediction matrix is noisy and does not work well out-of-sample. Furthermore, any attempt of a dimension reduction depends on the structure of eigenvalues of the Π matrix. Another key observation is that all trading strategies investigated in Haddad et al. (2020) fall into the class of symmetric linear strategies. For example, their factor timing strategy satisfies

$$\tilde{R}_{t+1}^{\pi_t} \;=\; \pi_t'\tilde{R}_{t+1} \;=\; (\Sigma_{\tilde{R}}^{-1}(a + \mathrm{diag}(b)S_t))'\tilde{R}_{t+1} \;=\; (\Sigma_{\tilde{R}}^{-1}a)'\tilde{R}_{t+1} \;+\; \tilde{S}_t' L \tilde{R}_{t+1}\,,$$

where the matrix

$$L \;=\; \Sigma_{\tilde{R}}^{-1}\,\mathrm{diag}(b)$$

is symmetric (because $\Sigma_{\tilde{R}}$, the in-sample covariance of in-sample PCs, is diagonal). Thus, the dynamic pat of the factor timing strategy attempts to exploit a low rank linear symmetric strategy. Furthermore, empirical estimates of the b coefficients in Haddad et al. (2020) are all nonnegative, implying that L is in fact positive semi-definite. The following lemma complements Proposition 6 and provides an exact upper bound on the expected return of the factor timing strategy.

Lemma 7 *Let $\lambda_1^s \geq \cdots \geq \lambda_N^s$ be the eigenvalues of the matrix Π^s and $\bar{\lambda}_1 \geq \cdots \geq \bar{\lambda}_N$ the singular values of Π. Let also $(q_i)_{i=1}^N$ be an arbitrary orthonormal basis of $\mathbb{R}^N$ and $M \leq N$. Let L be an arbitrary $M \times M$ symmetric positive semi-definite[25] symmetric matrix with eigenvalues $\lambda_1(L) \geq \cdots \geq \lambda_M(L) \geq 0$. Let also $\hat{S}_t = (\tilde{S}_t^{q_i})_{i=1}^M$ and $\hat{R}_{t+1} = (\tilde{R}_{t+1}^{q_i})_{i=1}^M$. Let $\pi_t = L'\hat{S}_t$.*

$$E\left(\hat{R}^{\pi_t}\right) \;\leq\; \sum_{i=1}^{M} \mathbf{1}_{\lambda_i^s>0} E(PEP_i)\lambda_i(L) \;\leq\; \sum_{i=1}^{M} \lambda_i(L) E(PP_i) \tag{96}$$

The first upper bound it attained if and only if $\lambda_i^s \geq 0$ for all $i \leq M$ and q_i is the respective eigenvector of Π^s. The second upper bound is attained if and only if in addition Π is symmetric and $\lambda_M^s \geq 0$.

The result of Lemma 7 implies that no matter which rotation (using principal components or otherwise) we might use to exploit differences in predictability, we are limited by the eigenvalue structure of the Π^s matrix. Furthermore, if Π^s is not symmetric (i.e., $\Pi^a \neq 0$), then there will be a non-zero "gap" in the achievable excess return because the L for the factor timing strategy is symmetric and hence cannot capture the anti-symmetric part of predictability due to the identity (23). Of course, it may happen that the risk-return tradeoff for some arbitrary factor timing strategy appears better than principal portfolios. If so, it must be the case that the superior performance originates from better diversification, and not from better return prediction.

A key insight of the principal portfolio analysis is that eigenvectors for the largest eigenvalues of the different parts of the Π matrix capture a robust part of predictability and hence will tend to perform better than alternative prediction techniques out-of-sample.

F Cross-Sectionally Demeaning the Signals or the Returns

We start with a signal S, which is not cross-sectionally demeaned, and use the notation "$\sim$" (tilde) to indicate demeaning:

$$\tilde{S}_{j,t} \;=\; S_{j,t} \;-\; \frac{1}{N}\sum_{k=1}^{N} S_{k,t}$$

[25] An analogous result holds when L is not positive semi-definite, in which case the upper bound is expressed in terms of the largest absolute values of the eigenvalues of Π^s.

Similarly, the prediction matrix based on demeaned signals is

$$\tilde{\Pi} \;=\; (R_{i,t+1}\,\tilde{S}_{j,t})_{i,j=1}^N\,.$$

and $\tilde{\lambda}_k^s$ and $\tilde{\lambda}_k^a$ are the eigenvalues of $\tilde{\Pi}^s$ and $\tilde{\Pi}^a$, respectively.

Proposition 20 *The demeaned eigenvalues $\tilde{\lambda}_k^s$ are interlacing with the non-demeaned ones λ_k^s in the sense that*

$$\lambda_{k+1}^s \;\le \tilde{\lambda}_k^s \;\le\; \lambda_{k-1}^s \tag{97}$$

for all $k = 2, \ldots, K-1$, $\lambda_2^s \;\le\; \tilde{\lambda}_1^s$, and $\tilde{\lambda}_N^s \;\le \lambda_{N-1}^s$ and similarly for the antisymmetric eigenvalues. Hence, if Π^s has N^p positive eigenvalues, then $\tilde{\Pi}^s$ has between N^p-1 and N^p+1 positive eigenvalues. Furthermore, the total performance of the cross-sectional factor

$$E\left(\sum_{i=1}^N \tilde{S}_{i,t}\,R_{i,t+1}\right) \;=\; \sum_{i=1}^N \tilde{\lambda}_i^s$$

satisfies

$$\sum_{i=2}^N \lambda_i^s \;\le\; E\left(\sum_{i=1}^N \tilde{S}_{i,t}\,R_{i,t+1}\right) \;\le\; \sum_{i=1}^{N-1} \lambda_i^s\,. \tag{98}$$

In particular, if all eigenvalues λ_i^s are positive, then cross-sectional factor performs worse than the time series factor.

Proof. We use Weyl inequalities (Horn and Johnson (1991)): for any two symmetric or Hermitian matrices $A,\ B$,

$$\lambda_j(A) \;+\; \lambda_k(B) \;\le\; \lambda_i(A+B) \;\le\; \lambda_r(A) \;+\; \lambda_s(B) \tag{99}$$

whenever $j+k-N \ge i \ge r+s-1$.

Then, we note that, by direct calculation,

$$\tilde{\Pi}^s \;=\; \Pi^s \;+\; X\,,$$

where $X \,=\, 0.5(\pi\mathbf{1}' \;+\; \mathbf{1}\pi')$, and where the vector $\pi = (\pi_i) \;=\; -(E(R_{i,t+1}\frac{1}{N}\sum_{k=1}^N S_{k,t}))$. The matrix X has rank two and at most two non-zero eigenvalues that always have opposite signs: $\lambda_1(X) \;\ge\; 0 \;\ge\; \lambda_N(X)$. Thus, by the Weyl inequalities,

$$\lambda_{i+1}(\Pi^s) \;\le\; \lambda_N(X) \;+\; \lambda_{i+1}(\Pi^s) \;\le\; \lambda_i(\tilde{\Pi}^s) \;\le\; \lambda_{i-1}(\Pi^s) \;+\; \lambda_2(X) \;\le\; \lambda_{i-1}(\Pi^s)\,.$$

The proof for the antisymmetric part is analogous.

To prove the last inequality, define the orthogonal projection $P = Id - \frac{1}{N}\mathbf{1}_{N\times N}$ Then, $\tilde{S}_t = PS_t$ and hence, by direct calculation, $\tilde{\Pi} = \Pi P$. Furthermore, since signals are demeaned,

$$\sum_{i=1}^{N} \tilde{S}_{i,t}\, R_{i,t+1} \;=\; \sum_{i=1}^{N} \tilde{S}_{i,t}\, \tilde{R}_{i,t+1}$$

where

$$\tilde{R}_{t+1} \;=\; PR_{t+1}$$

and hence

$$\begin{aligned} E\left(\sum_{i=1}^{N} \tilde{S}_{i,t}\, R_{i,t+1}\right) \;&=\; E\left(\sum_{i=1}^{N} \tilde{S}_{i,t}\, \tilde{R}_{i,t+1}\right) \;=\; \mathrm{tr}(E(\tilde{R}_{t+1}\tilde{S}_t')) \\ =\; \mathrm{tr}(P\Pi P) \;&=\; \mathrm{tr}(P\Pi^s P)\,. \end{aligned} \tag{100}$$

The eigenvalues of $P\Pi^s P$ coincide with the $N-1$ eigenvalues $\{\hat{\lambda}_k^s\}_{k=1}^{N-1}$ of $P\Pi^s P\lceil_{P\mathbb{R}^N}$ restricted onto the subspace $P\mathbb{R}^N$, plus a zero eigenvalue. By the interlacing inequalities Horn and Johnson (1991), we have $\lambda_{k+1}^s \le \hat{\lambda}_k^s \le \lambda_k^s$, and therefore

$$\mathrm{tr}(P\Pi^s P) \;=\; \sum_{k=1}^{N-1} \hat{\lambda}_k^s$$

satisfies the required inequalities. $\square$

The same arguments imply that the following is true.

Proposition 21 *Let*

$$\tilde{R}_{i,t+1} \;=\; R_{i,t+1} \;-\; \frac{1}{N}\sum_{k=1}^{N} R_{k,t+1}$$

and

$$\hat{\Pi} \;=\; (\tilde{R}_{i,t+1} S_{j,t})_{i,j=1}^{N}$$

Then, the eigenvalues of $\hat{\Pi}^a$ *and* $\hat{\Pi}^s$ *satisfy the same inequalities as those of* $\tilde{\Pi}$ *(see Proposition 20).*

F.1 Cross-Sectionally Demeaning Portfolio Signals

Given any orthonormal tuple of portfolios $\{\pi_k\}_{k=1}^{K}$, we define the corresponding demeaned signals as

$$\tilde{S}_t^{\pi_k} \;=\; S_t^{\pi_k} \;-\; \frac{1}{K}\sum_{i=1}^{K} S_t^{\pi_i}\,.$$

Interestingly, we "lose" one eigenvalue when using demeaned factors rather than non-demeaned ones as seen in the following proposition where $\lambda_1^s \geq \cdots \geq \lambda_N^s$ are still the eigenvalues of Π^s (i.e., based on the original, non-demeaned signals).

Proposition 22 *The expected excess return of demeaned portfolios based on any orthonormal tuple of portfolios* $\{\pi_k\}_{k=1}^{K}$ *satisfies*

$$\sum_{i=1}^{K-1} \lambda_k^s \;\geq\; E\Big(\sum_{k=1}^{K} \tilde{S}_t^{\pi_k} R_{t+1}^{\pi_k}\Big) \;\geq\; \sum_{i=N-K+2}^{N} \lambda_k^s$$

and the bounds are exact.

Proof of Proposition 22. Let

$$\tilde{\pi}_k \;=\; \pi_k \;-\; \frac{1}{K}\sum_{i=1}^{K} \pi_i\,.$$

Then,

$$\tilde{S}_t^{\pi_k} \;=\; S_t^{\tilde{\pi}_k}$$

and, since $\sum_k S_t^{\tilde{\pi}_k} = 0$, we have

$$\sum_k S_t^{\tilde{\pi}_k} R_{t+1}^{\pi_k} \;=\; \sum_k S_t^{\tilde{\pi}_k} R_{t+1}^{\tilde{\pi}_k}\,.$$

Define the matrix

$$X \;=\; (\tilde{\pi}_1, \cdots, \tilde{\pi}_K),$$

with the columns given by $\tilde{\pi}_i$. Then,

$$E(\sum_k S_t^{\tilde{\pi}_k} R_{t+1}^{\tilde{\pi}_k}) \;=\; \mathrm{tr}(X'\Pi^s X) \;=\; \mathrm{tr}(\Pi^s X X')$$

Since, by assumption, π_k are orthonormal, we have

$$\tilde{\pi}_k'\tilde{\pi}_l \;=\; (\pi_k \;-\; \frac{1}{K}\sum_{i=1}^K \pi_i)'(\pi_l \;-\; \frac{1}{K}\sum_{i=1}^K \pi_i) \;=\; \delta_{k,l} - 1/K$$

and hence the matrix $P = X'X \in \mathbb{R}^{K\times K}$ has rank $K-1$ and eigenvalues 1 (of multiplicity $K-1$) and 0 and is therefore an orthogonal projection. Thus, we can write $X = UP$ where $U \in \mathbb{R}^{N\times K}$ is an orthogonal matrix satisfying $U'U = Id_K$. Thus,

$$\mathrm{tr}(X'\Pi^s X) \;=\; \mathrm{tr}(PU'\Pi^s UP)\,.$$

Let V be the orthogonal matrix such that $\tilde{P} = V'PV$ is the projection onto the span of the first $K-1$ standard basis vectors of $\mathbb{R}^N$. Then,

$$\mathrm{tr}(PU'\Pi^s UP) \;=\; \mathrm{tr}(PU'\Pi^s UPVV') \;=\; \mathrm{tr}(V'PVV'U'\Pi^s UVV'PV) \;=\; \mathrm{tr}(\tilde{P}\tilde{U}'\Pi^s\tilde{U}\tilde{P})$$

where $\tilde{U} = UV$. Then, $\tilde{U}$ is an arbitrary orthogonal matrix with columns $u_1, \cdots, u_k$, and

$$\mathrm{tr}(\tilde{P}\tilde{U}'\Pi^s\tilde{U}\tilde{P}) \;=\; \sum_{i=1}^{K-1} u_k'\Pi^s u_k$$

and the claim follows from the Ky Fan inequality (Fan (1950)). $\square$

G Additional Empirical Results

G.1 Robustness of Momentum Results

This appendix reports extension and robustness analysis to complement the base case portfolio momentum analysis of Section 6.1. We study portfolio momentum via principal portfolios in several other data sets, including 25 U.S. size and operating profitability portfolios, 25 U.S. size and investment portfolios, the international counterparts of the three sets of Fama-French portfolios (i.e., developed countries excluding the U.S.), and a sample of 52 futures contracts starting in 1985. Futures data includes contracts for 21 commodities, 17 equity indices, 5 sovereign bonds, and 9 currencies.

While the base case signal is each asset's 20-day momentum (approximately one month), we also consider momentum signals for robustness, namely based on 40, 60, 90, 120, and 250 day past returns. In each case, we standardize the signal each period by converting it to a cross-sectional rank and mapping it into the [-0.5,0.5] interval.

Turning to returns, the base-case measure of returns is each asset's 20-day return (again, one month). We also consider other forecast horizons, namely 1-day, 5-day, and 10-day returns, cross-sectionally demeaning returns to focus prediction on cross section differences in returns. Similarly to the base case, each of the other forecast horizons corresponds to a sample of non-overlapping time periods t of the same length. As advocated by Moskowitz et al. (2012), in the case of futures contracts we time-series de-volatize returns (both in the forecast target and in the momentum signal construction) using trailing 20-day return volatility, which helps avoid a situation in which results are unduly driven by the large cross-sectional differences in volatility of raw futures contract returns.[26]

Figure 6 shows that principal portfolios perform better at shorter forecast horizons, especially the PP and PEP strategies. Indeed, at the shorter forecast horizons, even PEP earns a higher SR than the simple factor, and the risk-adjusted return as measured by the IR becomes highly significant at 1-, 5-, and 10-day forecast horizons.

Figure 7.A and B show the Sharpe ratios and information ratios for seven other data sets: three samples of U.S. stocks (the base case from before, plus two other sets of Fama-French portfolios), three sets of international stocks (i.e., global stocks outside the U.S. sorted into similar portfolios), and a set of 52 futures contracts (consisting of equity index futures, bond futures, commodity futures, and currency forwards). In support of the model's predictive

[26] This adjustment has tiny effects in our equity asset analysis so, in the interest of simplicity, we do not de-volatilize equity returns.

power, we see that all of the SRs and IRs are positive, and several, but not all, are statistically significant. In further support of theoretical predictions, Figure 7.C shows that the ex ante eigenvalues are highly correlated to the ex post realized returns in each sample.

Finally, we analyze the robustness of our method across momentum horizons and subsamples. Figure 8 shows the performance of the leading PPs, PEPs, and PAPs for different look-back periods in the specification of the momentum signal. Panel A shows that the PEP performs similarly to the simple factor for all momentum horizons. However, PP, PAP, and the PEP/PAP combination deliver higher SR across all horizons. Turning to the information ratios in Panel B, we see the these are more statistically significant for the short- and medium-term momentum periods, and less significant for the longer-term momentum horizons.

Figure 9 reports the performance of the base-case strategy for each decade in the sample. We see that the performance tends to be positive across decades — so the strong overall performance is not being driven by a single decade — but, naturally, the statistical significance in each decade is reduced due to the short time window.

Lastly, Figure 10 shows that base case results are qualitatively similar when we omit cross-sectional demeaning of signals and returns from our data construction and instead allow for timing effects to enter into principal portfolio estimates.

G.2 Testing the Fama-French 5-Factor Model

In this subsection, we illustrate how one could use results in section 4 and Appendix C–D for asset pricing tests. Our test universe consists of the 25 Fama-French portfolios sorted by value and book-to-market, with monthly returns $R_{i,t+1},\ i = 1, \cdots, 25$. We look at monthly data for the period July 1963-January 2020. Our signals are constructed as follows. First, we compute $\beta_{i,t}(k),\ k = 1, \cdots, 5$ computed as rolling 24-month betas of FF portfolio returns with respect to the Mkt-Rf, SMB, HML, RMW, and CMA, also from Kenneth French website. We then run a panel regression

$$R_{i,t+1} - r_t^F \ = \ \sum_{k=1}^{5} \gamma_k \beta_{i,t}(k) \ + \ \varepsilon_{i,t+1}$$

over full sample and define

$$S_{i,t} \ = \ \sum_{k=1}^{5} \gamma_k \beta_{i,t}(k)\,.$$

Intuitively, we expect that $S_{i,t}$ should be the "true" conditional expected return on $R_{i,t+1}$ and, hence, the Π matrix (estimated over the whole sample) should be symmetric and positive definite. In order to be able to apply Propositions 17and 16, we need an estimate of $E(\theta)$, $\mathrm{Var}(\theta)$, and Σ^ε. We

- Estimate θ_t by running a cross-sectional regression every month in the data sample, and then compute its sample mean and variance.[27]

- Estimate the "true" $\Pi^s = E(\theta)\bar{\Sigma}^S$

- Compute residuals $\varepsilon_{i,t+1} = R_{i,t+1} - \theta_t S_{i,t}$ and their sample covariance matrix Σ^ε

For Propositions 18 and 19, we use, respectively, w as the principal exposure portfolio for the lowest negative eigenvalue of $\hat{\Pi}^s$, and x, y as the parts of the principal alpha portfolio for the largest positive eigenvalue of $\hat{\Pi}^a$.[28]

We compute the respective p-values:

- For Proposition 16, the chi-squared statistic for all negative eigenvalues is $< 10^(-20)$, so the null is rejected. The distribution of eigenvalues of Π^s is almost symmetric around zero if we exclude the largest positive eigenvalue. In particular, 12 out of 25 eigenvalues are negative, and t

- For Proposition 17, the matrix is not invertible. The generalized chi-squared statistic gives a p-value of 0.3. So we cannot reject the hypothesis: The test is just too noisy and hence is not very helpful.

- For Proposition 18, p-value is 0.002 for the lowest (most negative) eigenvalue and 0.006 for the second lowest, so we reject the null that these eigenvalues are positive.

- For Proposition 19, p-value is 10^{-16} for the largest anti-symmetric eigenvalue, and 10^{-8} for the second largest, so we reject the null.

Finally, we note that since $S_{i,t}$ was estimated so that $S_{i,t} \approx E_t[R_{i,t+1}]$, one might also argue that $\theta_t = 1$. We consider this case as well, and the corresponding test statistics slightly improve because $\mathrm{Var}(\theta) = 0$.

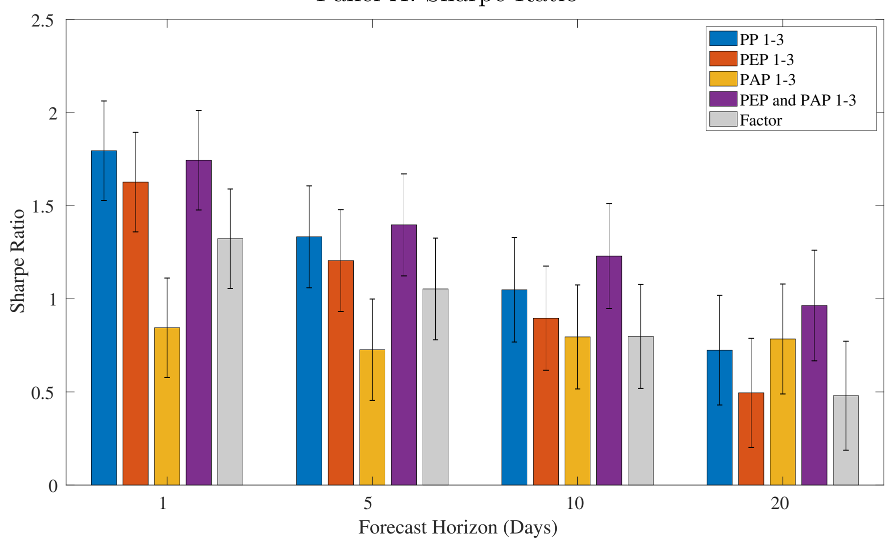
[27] Our variance estimate is huge, massively inflating standard errors and hence making the null even harder to reject.

[28] Note however that, asymptotically, the realized variance of the respective portfolio returns in Propositions 18 and 19 coincides with the variance predicted by the CLT. Since these returns are asymptotically Gaussian, we can just use the respective Sharpe ratios as the statistics.

Figure 6: Principal Portfolio Performance by Forecast Horizon

Out-of-sample performance of principal portfolios in terms of annualized Sharpe ratio (Panel A) and annualized information ratio versus the own-predictor strategy and the Fama-French 5-factor model (Panel B) along with ±2 standard error band around each estimate. Portfolios are constructed from the Fama-French 25 size and value portfolios based on a 20-day momentum signal. The figure reports performance of the simple factor ("Factor," that is, the standard own-signal strategy, included as a benchmark), the equal-weight average of the top three principal portfolios ("PP 1-3"), the equal-weight average of the top three principal exposure portfolios ("PEP 1-3"), the equal-weight average of the top three principal alpha portfolios ("PAP 1-3"), and the equal weighted average of the top three PEP's and PAP's combined ("PEP and PAP 1-3"). Results are shown for forecast horizons (and, equivalently, holding periods) of 1, 5, 10, and 20 days. Each forecast is made on an out-of-sample basis using a rolling training sample of the 120 most recent non-overlapping return observations. Sample period is 1963-2019.

Panel A: Sharpe Ratio



Panel B: Information Ratio

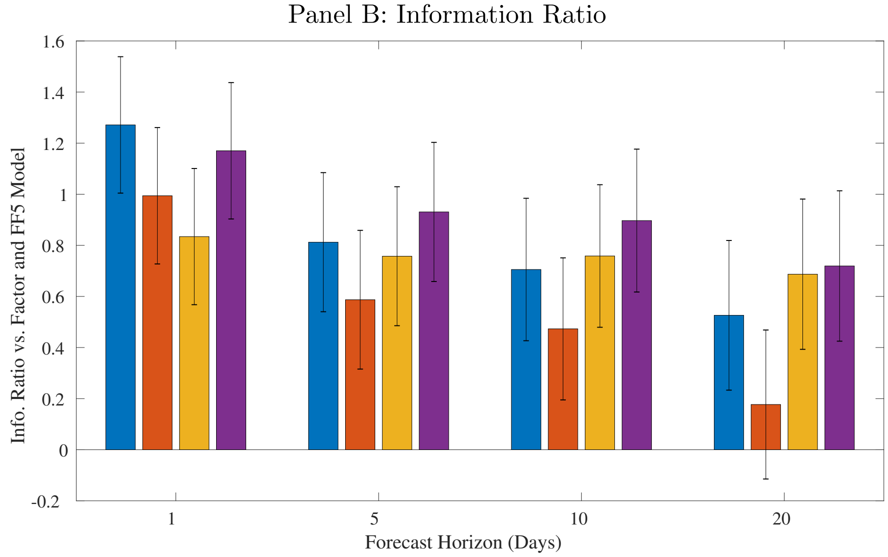
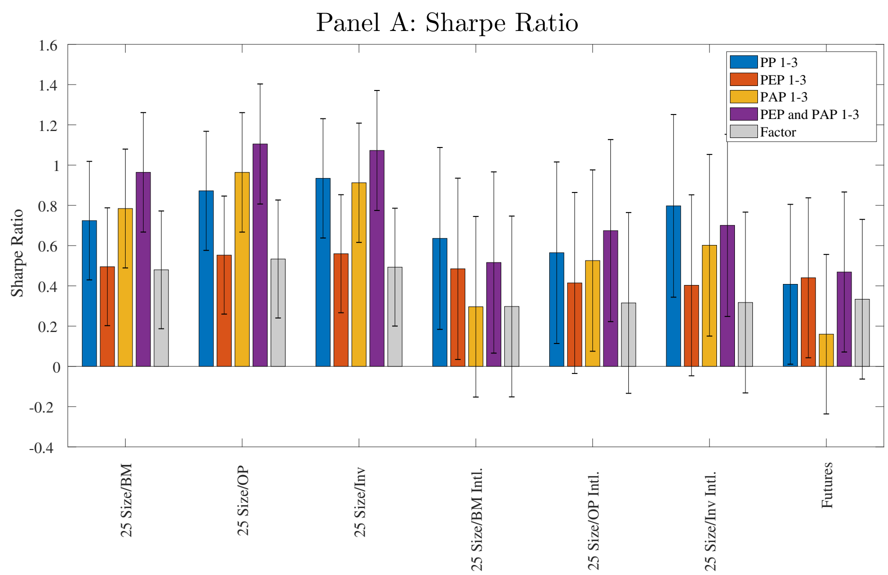
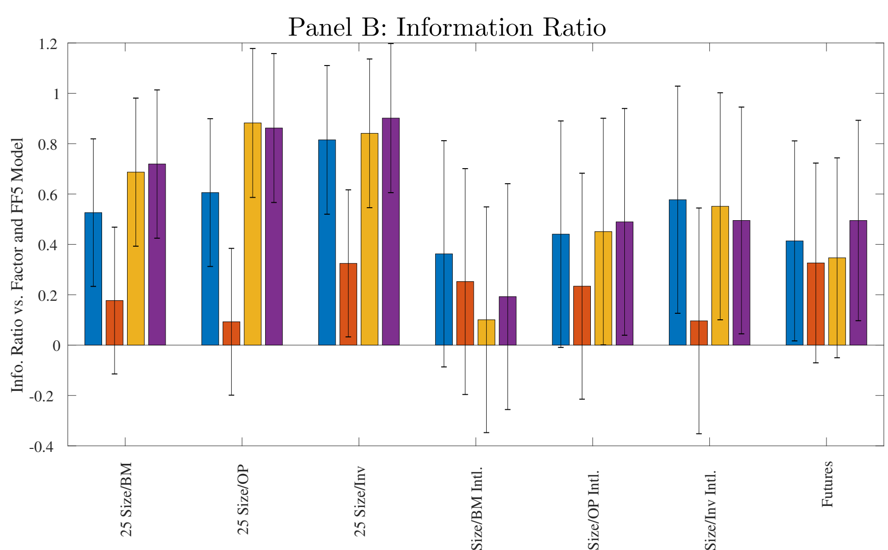


Figure 7: Principal Portfolio Performance in Other Asset Universes

Out-of-sample performance of principal portfolios in terms of annualized Sharpe ratio (Panel A) and annualized information ratio versus the own-predictor strategy and the Fama-French 5-factor model (Panel B) along with ±2 standard error band around each estimate. Portfolios are constructed based on a 20-day momentum signal from either the 25 U.S. size and value portfolios, 25 U.S. size and operating profitability portfolios, 25 U.S. size and investment portfolios, their international counterparts (developed countries excluding the U.S.), or 52 futures contracts. The figure reports performance of the simple factor ("Factor," that is, the standard own-signal strategy, included as a benchmark), the equal-weight average of the top three principal portfolios ("PP 1-3"), the equal-weight average of the top three principal exposure portfolios ("PEP 1-3"), the equal-weight average of the top three principal alpha portfolios ("PAP 1-3"), and the equal weighted average of the top three PEP's and PAP's combined ("PEP and PAP 1-3"). Panel C shows the correlation between out-of-sample average portfolio returns and eigenvalues of the prediction matrix. Blue bars show the correlation between PP's and singular values from the total prediction matrix, red bars show the correlation between PEP's and eigenvalues from the symmetric component, and yellow bars show the correlation between PAP's and eigenvalues from the anti-symmetric component. Results are shown for a 20-day forecast horizon/holding period. Each forecast is made on an out-of-sample basis using a rolling training sample of the 120 most recent non-overlapping return observations. Sample period is 1963-2019 for U.S. equity portfolios, 1990-2019 for international equity portfolios, and 1985-2019 for futures contracts.





Panel C: Correlation between average portfolio returns and eigenvalues

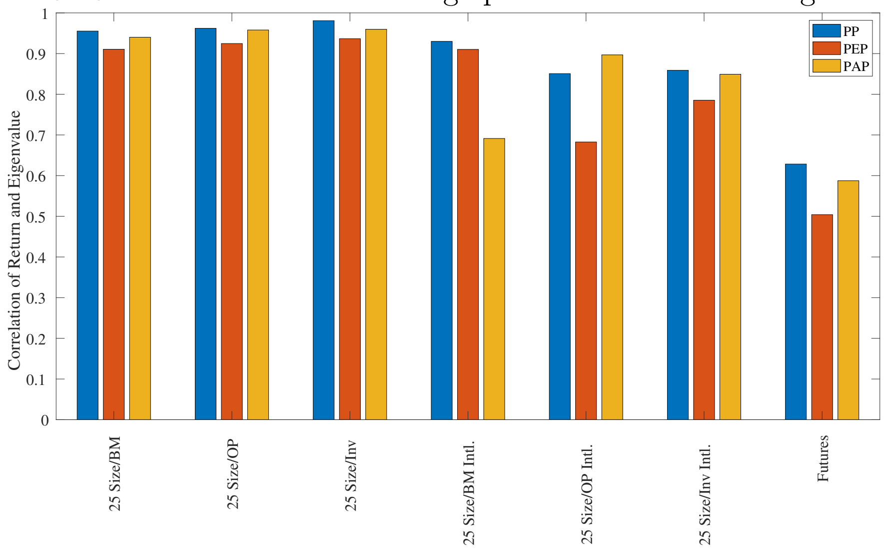


Figure 8: Principal Portfolio Performance by Momentum Lookback Window

Out-of-sample performance of principal portfolios in terms of annualized Sharpe ratio (Panel A) and annualized information ratio versus the own-predictor strategy and the Fama-French 5-factor model (Panel B) along with ±2 standard error band around each estimate. Portfolios are constructed based on a 20, 40, 60, 120 or 250-day momentum signal from 25 size and value portfolios. The figure reports performance of the simple factor ("Factor"), the equal-weight average of the top three principal portfolios ("PP 1-3"), the equal-weight average of the top three principal exposure portfolios ("PEP 1-3"), the equal-weight average of the top three principal alpha portfolios ("PAP 1-3"), and the equal weighted average of the top three PEP's and PAP's combined ("PEP and PAP 1-3"). Each forecast is made on an out-of-sample basis using a rolling training sample of the 120 most recent non-overlapping return observations. Sample period is 1963-2019.

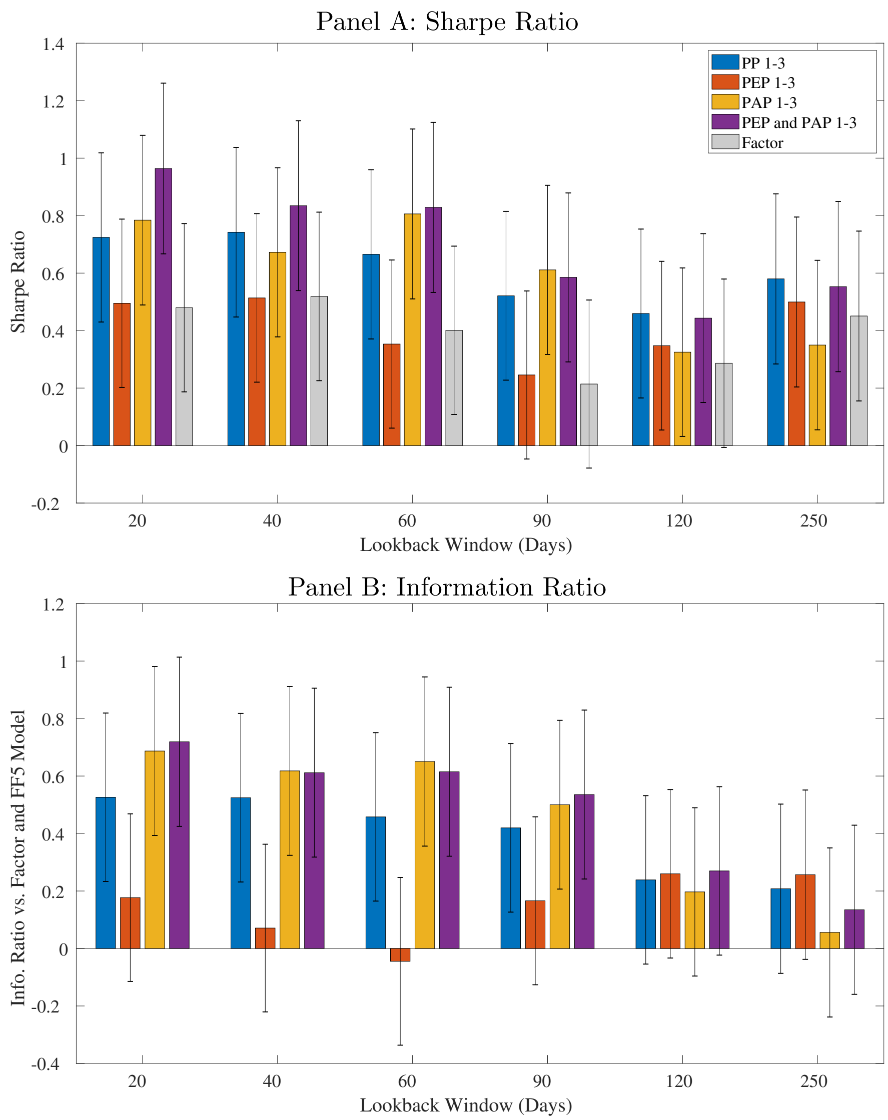
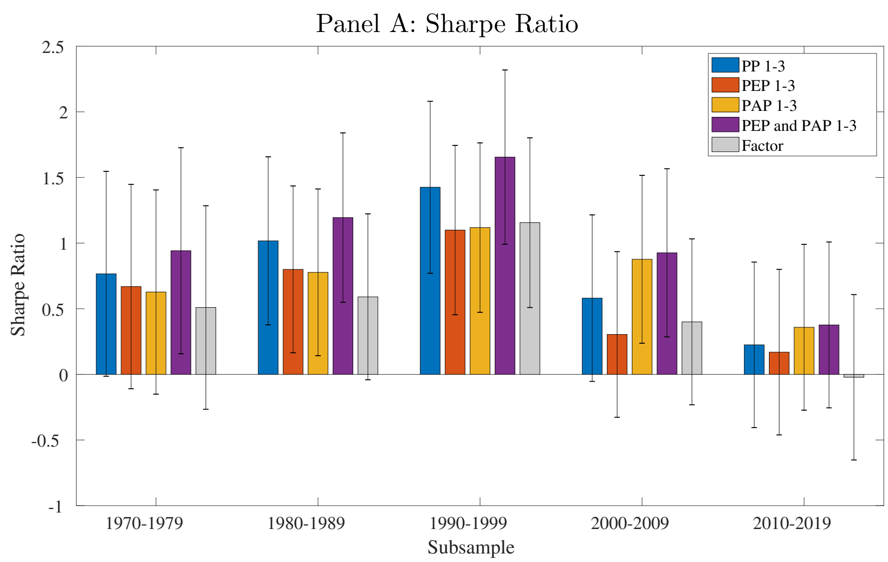


Figure 9: Principal Portfolio Performance in Subsamples

Out-of-sample performance of principal portfolios by decade in terms of annualized Sharpe ratio (Panel A) and annualized information ratio versus the own-predictor strategy and the Fama-French 5-factor model (Panel B) along with ±2 standard error band around each estimate. Portfolios are constructed based on a 20-day momentum signal from 25 size and value portfolios. The figure reports performance of the simple factor ("Factor"), the equal-weight average of the top three principal portfolios ("PP 1-3"), the equal-weight average of the top three principal exposure portfolios ("PEP 1-3"), the equal-weight average of the top three principal alpha portfolios ("PAP 1-3"), and the equal weighted average of the top three PEP's and PAP's combined ("PEP and PAP 1-3"). Results are shown for a 20-day forecast horizon/holding period. Each forecast is made on an out-of-sample basis using a rolling training sample of the 120 most recent non-overlapping return observations.



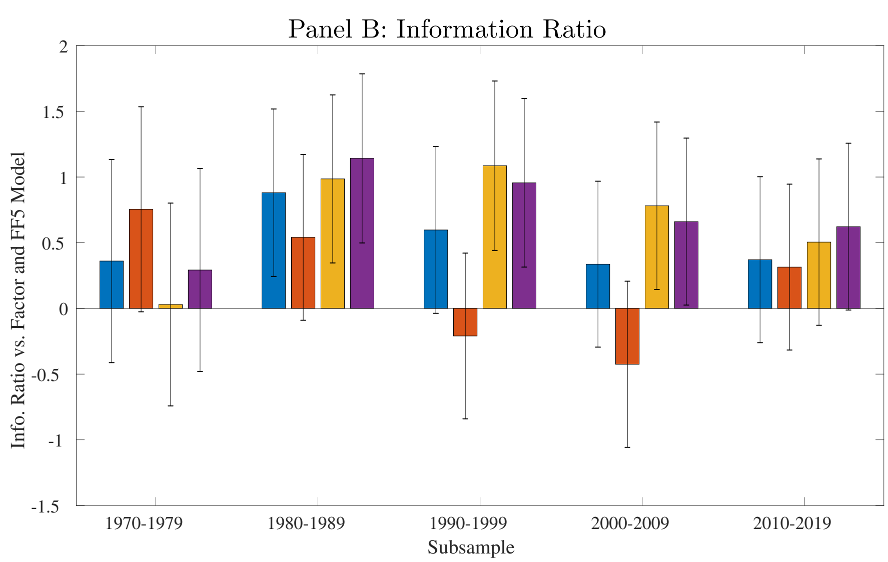
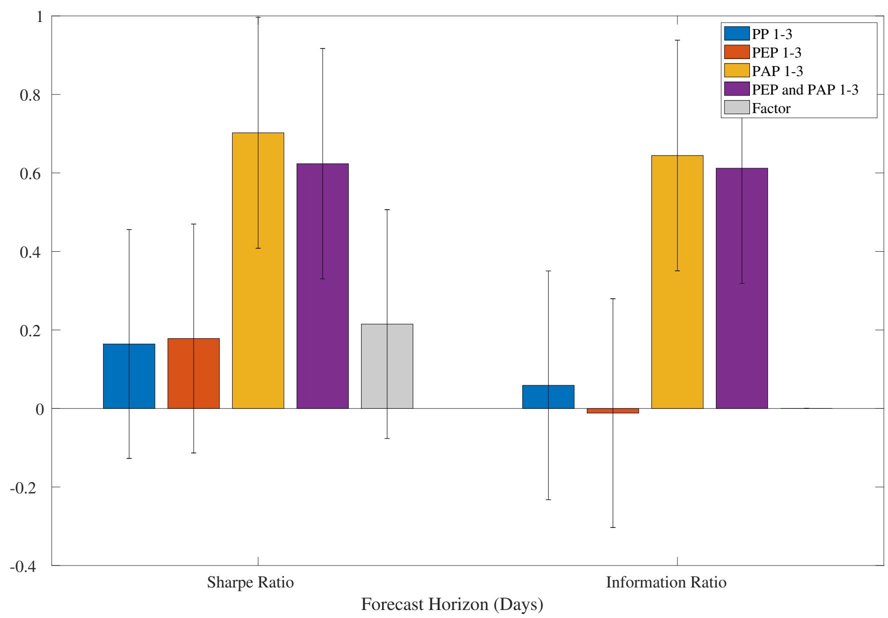


Figure 10: Principal Portfolio Performance Ratios With No Centering

This figure repeats the analysis of Figure 3 but with uncentered signals and returns.



Swiss Finance Institute

Swiss Finance Institute (SFI) is the national center for fundamental research, doctoral training, knowledge exchange, and continuing education in the fields of banking and finance. SFI's mission is to grow knowledge capital for the Swiss financial marketplace. Created in 2006 as a public–private partnership, SFI is a common initiative of the Swiss finance industry, leading Swiss universities, and the Swiss Confederation.

swiss:finance:institute

c/o University of Geneva, Bd. Du Pont d'Arve 42, CH-1211 Geneva 4
T +41 22 379 84 71, rps@sfi.ch, www.sfi.ch