Principal Portfolios

Kelly, Bryan T.; Malamud, Semyon; Pedersen, Lasse Heje

Document Version
Final published version

Published in:
Journal of Finance

DOI:
10.1111/jofi.13199

Publication date:
2023

The Journal of FINANCE
The Journal of THE AMERICAN FINANCE ASSOCIATION

THE JOURNAL OF FINANCE • VOL. LXXVIII, NO. 1 • FEBRUARY 2023

Principal Portfolios

BRYAN KELLY, SEMYON MALAMUD, and LASSE HEJE PEDERSEN*

ABSTRACT

We propose a new asset pricing framework in which all securities' signals predict each individual return. While the literature focuses on securities' own-signal predictability, assuming equal strength across securities, our framework 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.

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). Alternatively, $S_{i,t}$ may be a predictor that is agnostic of equilibrium considerations, such as each asset's valuation ratio or its recent price momentum. Standard analyses—for example, evaluating characteristic-sorted portfolios or asset pricing tests in the spirit of Gibbons, Ross, and Shanken (1989)—focus on own-asset predictive signals; that is, the association between $S_{i,t}$ and the return only on asset i, $R_{i,t+1}$.

In this paper, we propose a new approach to analyzing asset prices through the lens of what we call the "prediction matrix." This approach shows (i) how

*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 Lando, 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. Malamud gratefully acknowledges support from the Swiss Finance Institute and the Swiss National Science Foundation and Pedersen gratefully acknowledges support from the Center for Financial Frictions (Grant no. DNRF102) and the Center for Big Data in Finance (Grant no. DNRF167).

Correspondence: Lasse Heje Pedersen, Department of Finance, Copenhagen Business School, Solbjerg Plads 3, A5, DK-2000 Frederiksberg, Denmark; e-mail: lhp001@gmail.com.

DOI: 10.1111/jofi.13199

to optimally invest in light of cross-predictability, where optimal refers to the return-maximizing strategy among a class of linear trading strategies; (ii) how to find optimal alpha and beta strategies; and (iii) 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 can 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 trade-off.

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 component 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" (PPs). They are a set of normalized portfolios ordered from those most predictable by S to those least predictable. The top PPs are thus the most "timeable" portfolios, and as such offer the highest unconditional expected returns for an investor that faces a leverage constraint (i.e., 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 by Π^s) and an antisymmetric part (which is equal to minus its transpose and denoted by 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 well-organized 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 antisymmetric, which, amazingly, translates 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 antisymmetric matrix Π^a are powerful factor-neutralized (alpha) strategies with respect to this factor.[1] We refer to strategies arising from eigenvectors of the symmetric component as "principal exposure portfolios" (PEPs) and the strategies arising from the antisymmetric part as "principal alpha portfolios." Once classified as "exposure" versus "alpha," prediction patterns are then ordered from strongest to weakest and based on the size of the principal portfolios' 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 nonnegative 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 antisymmetric components of the prediction matrix. If the pricing kernel is correctly specified, 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 antisymmetric 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. We thus 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 nonnegative 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 reduce the out-of-sample performance of PPs. 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,

[1] Of course, the decomposition into alpha and beta depends on the benchmark factor with which they are computed. For example, our principal alpha portfolios (PAPs) deliver alpha with respect to the factor generated by the signals S, but not necessarily with respect to other factors.

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. PPs 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 to 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 date t, estimate the prediction matrix from past signals and returns (i.e., 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 periods in most of our analyses. Having estimated the prediction matrix, the singular-value decompositions of Π as well as its symmetric and antisymmetric 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 PPs using standard Fama-French portfolios as base assets and momentum (i.e., past returns) as the trading signal. We find that the leading PPs deliver significant positive returns out of sample. As further evidence consistent with our framework, PEPs have significant factor exposure while PAPs 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 handle only one signal at a time, so we conduct our PPs analysis one signal at a time. We also aggregate the resulting PPs for each of these signals into a combined strategy. We find that factor-timing strategies based on PPs perform well overall and across the majority of signals. In fact, PPs have significant alpha relative to standard factor models and relative to the principal components-based factor-timing approach of Haddad, Kozak, and Santosh (2020).

Our paper is related to several literatures. Our asset pricing test complements other such tests, including Gibbons, Ross, and Shanken (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 (see Fama and French (2015) for a recent example), and machine learning (Gu, Kelly, and Xiu (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, Polk, and Vuolteenaho (2003), Haddad, Kozak, and Santosh (2020)) and factor momentum (Arnott et al. (2022), 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, Daniel, and Sağlam (2020), among 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 PP performance across signals and samples.

I. Principal Portfolio Analysis (PPA)

In this section, we lay out our PPA framework. We describe the concept of linear strategies of predictive signals, show how linear strategies are closely linked to the prediction matrix, derive optimal strategies, and introduce the notion of PPs that implement optimal strategies.

We begin by introducing 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 $\Sigma_{R,t} = \text{var}_t(R_{t+1})$.

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.

A. Linear Trading Strategies

How can an investor best exploit predictive information contained in asset characteristic S? To address this question in a tractable way, we work in the context of general linear trading strategies based on S. We then derive an optimal linear strategy subject to leverage constraints and show the close 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 of 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,

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

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 and 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. 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 for 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 the portfolio uniformly, abstracting from heterogeneity in predictive effects across assets. When a researcher reports that this type of simple factor has a positive average return, $E(\widetilde{F}_{t+1}) > 0$, they are effectively saying that the signal positively predicts own-asset returns on average.

B. The Prediction Matrix

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

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

where 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 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 for measuring predictive signals 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*

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

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 semidefinite matrix. Then the linear strategy with position matrix* $L = M\Pi'$ *has positive expected excess return:*

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

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.

C. The Prediction Matrix versus 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 $B = \Pi E(S_tS_t')^{-1}$, which depends on the prediction matrix Π. One could also try to estimate the mean-variance efficient portfolio weights,

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

but this would require the estimation of a large number of parameters and the inversion of two large estimated matrices. Instead, the factor literature focuses 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 focuses on finding a linear portfolio L that "works well" across signals. To do so, 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 implies 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 below.[2]

D. 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, for example, the strategy $2L$ doubles the expected return of the 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 $\|L' S_t\| \leq \|L'\| \, \|S_t\| \leq \|S_t\|$ when $\|L\| \leq 1$. So we consider linear strategies for which the position size is always bounded by the position size of the simple strategy. Furthermore, 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, $\|L' S_t\| \leq 1$.[3]

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

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

[2] Empirically, we scale signals such that $S_{t,i} \in [-\frac{1}{2}, \frac{1}{2}]$, implying that $E(S_t S_t')$ is 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 $\mathrm{var}_t(\varepsilon_{t+1})$ 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 relies on these restrictions.

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

In words, we maximize expected returns subject to a risk constraint. This risk constraint is robust in the sense that we require variance to be bounded *regardless* of the signal realization S. This robust objective where we maximize risk with respect to S, rather than consider 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 (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 be 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 in our empirical study of futures) and if the correlation matrix is close to, or has been shrunk to, the identity matrix. Such shrinkage can be useful in an optimization setting (Pedersen, Babu, and Levine (2021)). In any event, when we have general variance-covariance matrix $\Sigma_{R,t}$, our portfolio constraint $\|L\| \leq 1$ still serves to control for risk, leverage, and the portfolio norm.[4] Furthermore, we show how to solve a robust mean-variance problem for general $\Sigma_{R,t}$ in Appendix Section I. In Internet Appendix Section I, we also show how to solve the mean-variance problem with a risk penalty driven by risk aversion (instead of the risk constraint used here).[5]

E. Optimal Linear Strategies

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

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

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

where $\bar{\lambda}_1 \geq \cdots \geq \bar{\lambda}_N \geq 0$ are the singular values of Π, that is, 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 form and of

[4] The portfolio constraint $\|L\| \leq 1$ implies a limit on the portfolio norm by definition, a leverage limit described in footnote , and the 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, $\Sigma_{R,t} \leq \bar{\Sigma}$.

[5] The Internet Appendix may be found in the online version of this article.

the form described in Proposition 2. Furthermore, the solution depends on the singular values of Π, which in general depend on all of its elements—not just the diagonal—so it has the potential to outperform the simple factor.

F. Principal Portfolios

We next decompose the optimal solution into a collection of linear strategies that we refer to as "principal portfolios" of the signal S. PPs 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. Specifically, let

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

where $\bar{\Lambda} = \text{diag}(\bar{\lambda}_1, \ldots, \bar{\lambda}_N)$ 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} PP 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 PP. 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 PPs is actually very simple. All one needs to do is use their favorite program to compute the singular-value decomposition of Π (a standard feature of most computing programs) and take the column vectors of U and V.

Decomposing the optimal strategy into its PPs is similar to decomposing the variance into the principal components. The difference is that PCA decomposes the *variance*, whereas PPA decomposes the *expected return*. Just like the variance of each principal component equals its corresponding eigenvalue, the expected return of each PP is its singular value:

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

The following proposition summarizes the results of this section.

PROPOSITION 4: *The expected return of each PP is given by its corresponding singular value,*

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

and the sum of PPs is the optimal linear strategy

$$\max_{\|L\|\le 1} E(S'_t L R_{t+1}) = E\left(\sum_{i=1}^{N} PP^i_{t+1}\right) = \sum_{i=1}^{N} \bar{\lambda}_i. \tag{18}$$

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 PPs. But even in this simple setting PPs are insightful about the optimal strategy. In this case, the prediction matrix reduces to the unconditional second moment of S_t, denoted by 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, PPs 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 PPs are pinned down by the size of singular values in (17). Moreover, all PPs have positive expected excess returns (assuming that Σ_S is nondegenerate), 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 IV.

II. 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 PPs, just as in the general solution in Propositions 3 and 4.

A. Alpha-Beta Symmetry Decomposition

To decompose the return into its alpha and beta components, we must first specify the factor 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*

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

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.[6] 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," $\widetilde{F}$.

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 thus is the tradable representation of the pricing kernel. Moreover, 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 relies on actually observing F—since we do not observe it. We do not 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 the risk and return of linear strategies, recall that any square matrix $B \in \mathbb{R}^{N\times N}$ can be decomposed into its *symmetric part*, $B^s = \frac{1}{2}(B + B')$, and its *antisymmetric part*, $B^a = \frac{1}{2}(B - B')$, where $B = B^s + B^a$. The symmetric part equals its own transpose while the antisymmetric part equals minus its own transpose. Both parts have a number of interesting properties. For example, since $B^a = -(B^a)'$, the antisymmetric part has zeros along the main diagonal.

Hence, any linear strategy can be seen as a sum of a symmetric and antisymmetric 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 $R^{w_t}_{t+1} = S_t' L R_{t+1} = S_t' L^s R_{t+1} +$*

[6] Kelly, Pruitt, and Su (2020a, 2020b) 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 we discuss later.

$S_t' L^a R_{t+1}$ *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 determined entirely by its symmetric part, while the expected return is determined by both the symmetric and antisymmetric 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 antisymmetric strategy is always factor neutral. Second, an antisymmetric strategy can nevertheless deliver positive returns if $\Pi^a \neq 0$. In this case, an antisymmetric strategy can deliver positive expected return with zero factor exposure, that is, pure alpha with respect to F. (Of course, while an antisymmetric strategy always has zero exposure to F, it could have exposure to other factors not considered here, as we analyze in Internet Appendix III.)

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 antisymmetric matrix* $A \in \mathbb{R}^{N\times N}$, *we have* $\mathrm{tr}(BA) = \mathrm{tr}(AB) = 0$ *and* $x'Ax = 0$ *for all vectors* $x \in \mathbb{R}^N$.

In other words, antisymmetric 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 $S_t' L^s S_t$, 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. Thus, 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). This simple factor has expected return $\mathrm{tr}(L^s\Pi^s) = \mathrm{tr}(\Pi^s) = \mathrm{tr}(\Pi)$, and it always has positive exposure to the latent factor, $\mathrm{cov}_t(\widetilde{F}_{t+1}, F_{t+1}) = \mathrm{var}_t(F_{t+1}) S_t' S_t > 0$.

The optimal linear strategy in Proposition 3 and the corresponding PPs 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 antisymmetric trading strategies. We derive symmetric and antisymmetric analogs of PPs, and show that they are the building blocks of optimal symmetry-decomposed strategies with either pure factor exposure and no alpha or pure alpha and no factor exposure.

Put simply, symmetry is beta and antisymmetry is alpha. We next derive the optimal beta and alpha.

B. Symmetric Strategies: PEPs

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 semidefinite 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 than (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 $\|L\| \le 1$ if $|\lambda_k| \le 1$ 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 $W = (w_1^s, \ldots, w_N^s)$ is the matrix of eigenvectors corresponding to the eigenvalues $\lambda_1^s \geq \ldots \geq \lambda_N^s$. 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 can be decomposed into N components, or PEPs. 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 eigenvalues is the optimal symmetric linear strategy:

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

The first result shows that returns of PEPs equal their eigenvalues. The second result shows that the collection of PEPs yields 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 nonnegative, $\lambda_k^s \geq 0$, 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 of 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 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. Indeed, for any orthonormal basis of portfolios $B = \{b_k\}_{k=1}^N$, we have that $BB' = \mathrm{Id}$ and hence

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

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

The fact that $\widetilde{F}$ equals the sum of PEPs together with equation (30) implies 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 shows that a simple strategy $\widetilde{F}_{t+1}$ has significantly positive average returns, we learn that the sum of the eigenvalues of Π^s is positive.

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

When $E(\widetilde{F}_{t+1}) = \sum_{k=1}^{N} \lambda_k^s > 0$, some 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 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 semidefinite L).

If we are willing to have a factor exposure that may switch sign, we can achieve even higher returns. 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).

Table I

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 of the 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 the variance (return) among all portfolios that are orthogonal to the first k portfolios.

	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)	$\pi_k' \Sigma_R \pi_l = 0$ i.e., $\mathrm{cov}(R^{\pi_k}_{t+1}, R^{\pi_l}_{t+1}) \ = \ 0$	$(w^s_k)' \Pi^s w^s_l = 0$ i.e., $E(S^{w^s_k}_t R^{w^s_l}_{t+1}) \ + \ E(S^{w^s_l}_t R^{w^s_k}_{t+1}) = 0$
(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)	$(\pi_k) = \arg\max_{\text{orthon.}\{x_k\}_{k=1}^K} \sum_k \mathrm{var}(R^{x_k}_{t+1})$	$(w^s_k) = \arg\max_{\text{orthon.}\{x_k\}_{k=1}^K} \sum_k E(S^{x_k}_t R^{x_k}_{t+1})$
(v)	$\pi_{k+1} = \arg\max_{x \perp \{\pi_1,\ldots,\pi_k\}} \mathrm{var}(R^x_{t+1})$	$w^s_{k+1} = \arg\max_{x \perp \{w^s_1,\ldots,w^s_k\}} E(S^x_t R^x_{t+1})$

The analogy between PPA and PCA is remarkably close when we focus on the symmetric part of the prediction matrix as highlighted in Table I. 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 have 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 antisymmetric (zero exposure) strategies within Π. Furthermore, the PEPs are simply the unit vectors $w^s_k = \mathbf{1}_k$.[7] The optimal strategy is long assets with positive own-predictability and short those with negative own-predictability.

C. Antisymmetric Strategies: PAPs

We now turn to antisymmetric linear trading strategies. The most basic type of antisymmetric matrix has the form $L = xy' - yx'$, which we call *rank-2 antisymmetric 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 antisymmetric strategy generates a return of

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

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

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

$(S_t'x_j)$ 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, antisymmetric strategies can be interpreted 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 antisymmetric strategies can be represented as a sum of these basic building blocks.

LEMMA 3: *Any antisymmetric matrix* A *has an even number* $2K$ *of nonzero eigenvalues. The nonzero 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 is* $\bar{z}_k = \frac{1}{\sqrt{2}}(x_k - iy_k)$, *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* $k \neq l$ *and* $k, l \leq K \leq N/2$. *The corresponding eigendecomposition is given by*

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

This lemma shows how to break any antisymmetric matrix into its basic building blocks of form $x_k y_k' - y_k x_k'$ for $k = 1, \ldots, K$.[8] 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 antisymmetric strategies.

Example (Beta-Neutral Strategy). Consider an economy of N assets that satisfies the capital asset pricing model (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[9] $\theta_t \geq 0$ is the market risk premium, $\beta_{i,t}$ is the conditional CAPM beta of stock i, and $\alpha > 0$. 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 with size equal to one (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 antisymmetric strategy with $L = yx' - xy'$. The long position in asset 1 corresponds to $x = (1, 0, \ldots, 0)'$, and the short position in the factor corresponds to $y = (1, 1, \ldots, 1)'$. In other words, the beta-neutral strategy has zero symmetric component, nonzero antisymmetric component, and positive expected return, rendering it an alpha strategy with expected return

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

[8] An antisymmetric strategy satisfies the portfolio constraint $\|A\| \leq 1$ as long as $|\lambda_k| \leq 1$ in (36).
[9] Here, $\mathbf{1}_{i=1}$ equals one if i=1 and zero otherwise.

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 antisymmetric 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, with appropriately opposing signs.

Next, we derive *optimal* antisymmetric strategies. The first step is to apply the eigendecomposition in (36) to the antisymmetric part of the transposed prediction matrix, $(\Pi^a)'$. By Lemma 3, the matrix $(\Pi^a)'$ has $2N^a$ nonzero 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} PAP as the linear strategy based on the j^{th} eigenvectors: $L_j = x_j y_j' - y_j x_j'$ for $j = 1, \ldots, N^a$. 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 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 the optimal antisymmetric 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 antisymmetric 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, $S_t = S$.[10] 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 the 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.

[10] 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.

Hence, there is a single PEP with a nonzero eigenvalue, namely, the eigenvector S, and no PAP. Therefore, this PEP is the only meaningful portfolio, and it is the same as the simple factor, S, with expected return $S'R = R'R > 0$.

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 nonzero 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[11]

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

where $c_1^s,\ c_N^s$ are constants chosen such that $\|w_1^s\| = \|w_N^s\| = 1$. We see that the first PEP 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 antisymmetric part. The strategy that trades this is $L = \Pi^{a'} = 0.5(SR' - RS')$. To derive the PAP, note that $\Pi^{a'}$ has at most two nonzero 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[12]

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

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

Example (Betting against Beta (BAB): PAP is the New BAB). Proceeding from the prior example, suppose that the signals S are chosen to be the expected returns in an asset pricing model, that is, $S_j = \mathrm{cov}(-M_t, R_{j,t})$, 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 antisymmetric 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

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

[12] Note that c^a is determined such that $\|y\| = 1$.

equal-weighted portfolio (which is lower beta) is scaled up relative to the beta-weighted portfolio,[13] $S'S > N$. Hence, this strategy resembles the 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 $w'S = NS'S - NS'S = 0$.

D. Static and Dynamic Bets

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

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

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 based only 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 on dynamic bets, 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; if it is 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 examples above 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 antisymmetric strategies with positive expected returns. Because an antisymmetric strategy is guaranteed to have zero factor exposure, it is also alpha to the factor.

[13] 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.

III. 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. Put 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). Alternatively, 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, $\text{cov}_t(R_{j,t+1}, -M_{t+1})$, where we only assume proportionality (rather than equality) since we may not know the equity premium in the CAPM or the risk aversion in the consumption CAPM. 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[14]

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

where $\theta_t > 0$ is a factor of proportionality due to the risk-free rate and due 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, $\beta_{j,t} = \text{cov}_t(R_{j,t+1}, R^m_{t+1})/\text{var}_t(R^m_{t+1})$. 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 proportionality, θ. To see this, 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. Furthermore, we see that the prediction matrix is positive semidefinite 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 next.

[14] To see this result, note that the definition of a pricing kernel applied for the risk-free asset (which has zero 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, $E_t(R_{j,t+1}) = (1 + R^f_t)E_t(M_{t+1})E_t(R_{j,t+1}) = (1 + R^f_t)(E_t(R_{j,t+1}M_{t+1}) - \text{cov}_t(R_{j,t+1}, M_{t+1})) = (1 + R^f_t)\text{cov}_t(R_{j,t+1}, -M_{t+1})$.

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 semidefinite, and, equivalently, all corresponding PEPs have nonnegative expected returns and all PAPs have zero expected returns.*[15]

The intuition behind this result follows from our earlier portfolio theory. We know that negative eigenvalues of Π^s and a nonzero Π^a give rise to alpha strategies (Sections II.B and II.C, respectively). Since alpha strategies cannot exist in an arbitrage-free asset pricing model with a correctly specified pricing kernel, all eigenvalues of Π^s must be positive and Π^a must be zero. In other words, Π must be symmetric and positive semidefinite.

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, that is, 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. Furthermore, 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 the eigenvalues of Π^s are nonnegative or, equivalently, whether the PEP returns are nonnegative. Internet Appendix Section IV presents Central Limit Theorems (CLTs) that justify this approach. In particular, Proposition IA.5 provides a CLT for the distribution of eigenvalues of Π^s, Proposition IA.6 derives the CLT for Π^a, and Propositions IA.7 and IA.8 present CLTs for the returns on trading symmetric and antisymmetric linear strategies. We note, however, that these CLTs rely on strong assumptions that may not hold in practice, so generalizing these results and benchmarking them to other asset pricing tests are important tasks for future research.

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

[15] 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 as shown in (41).

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 semidefinite.*

Internet Appendix Section III 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 IA.1 shows that the reverse is also true. Propositions IA.2, IA.3, and IA.4 show how our results change when the model is misspecified due to a common alpha in returns or missing factor exposures.

IV. Robust Strategies: Shrinkage via PPs

Our theoretical analysis thus far takes place in population with a known prediction matrix. 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, with signals 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 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 (differences among off-diagonal elements).

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

The analysis in Sections I and II shows that a singular-value decomposition of Π (or of its symmetric and antisymmetric parts) finds orthonormal portfolios and orders them from highest expected return to lowest. This eigendecomposition has another advantage 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 PCA, 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 PPs. These insights are formalized in the following proposition, which generalizes all of the optimization problems that we consider above (Propositions 3, 6, and 8).

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

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

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

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* $c = (\sum_{k=1}^{K} \bar{\lambda}_k^q)^{-1/p}$.

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

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

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

(iii) The solution when restricting attention to antisymmetric strategies depends on the eigenvalues λ_k^a *of* Π^a*:*

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

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* $c = (2 \sum_{k=1}^{K} (\lambda_k^a)^q)^{-1/p}$.

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 PPs. This is true regardless of whether one considers general linear strategies (L), symmetric and hence factor-exposed strategies ($L = L'$), or antisymmetric pure alpha strategies

($L = -L'$). By truncating the strategy at the top K PPs, these robust strategies replace the lesser singular values with zeros.[16] The lesser components may be dominated by noise and therefore are 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 PPs 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 PPs for any norm. This result shows that PPs are very general building blocks. The choice of norm simply affects how the PPs 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 downweighted by the choice of p.

At the 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 manage their leverage concerns is dictated in part by the performance of the 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 effect chooses the form of leverage constraint that lends itself to robust out-of-sample trading performance.

Interestingly, when $p = 2$, part (i) of the proposition is similar to trading a version of the Π matrix that has been estimated via a reduced-rank regression (RRR) (see, for example, Velu and Reinsel (1998)).[17] Furthermore, when $p = 2$ and we do not impose a rank restriction (that is, 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 $p = \infty$, that is, $q = 1$, 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 I through III lay out a theoretical basis for PPs, and Proposition 11 prescribes a machine learning approach to implementing PPs in practice. Data-driven choices for hyperparameters K and p can allow a researcher to select the level of PP model complexity best suited for constructing optimal out-of-sample strategies.

[16] Note that singular values of a symmetric or an antisymmetric matrix coincide with the absolute values of its eigenvalues.

[17] RRR seeks to minimize the mean squared error $E(\|R_{t+1} - L'S_t\|^2) = E(\|R_{t+1}\|^2) - 2E(S_t'LR_{t+1}) + E(S_t'LL'S_t)$ under a rank constraint on the matrix L. By direct calculation, this objective is equivalent to maximizing $2\mathrm{tr}(L\Pi) - \mathrm{tr}(LL'\Sigma_S)$. Thus, RRR 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 $p = 2$.

V. Empirical Results

We next present two empirical applications of our method.

A. Fama-French Portfolio Momentum

Our first application uses PPs on the 25 Fama-French portfolios.[18] 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 (as measured by market capitalization) and valuation ratio (book-to-market), and we use daily data from July 1963 through the end of 2019.

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, Ooi, and Pedersen (2012)). For each asset in each sample, we compute its cumulative return over the past 20 trading days (approximately one month). We 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).[19] We use this rank to predict subsequent monthly (20-day cumulative) returns on each portfolio.[20]

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 nonoverlapping 20-day periods. The estimated prediction matrix at 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 PPs and we compute the eigenvectors of its symmetric and antisymmetric

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

[19] 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 IA.9 in the Internet Appendix). Furthermore, demeaning implies that we only exploit cross-sectional predictability, not time-series predictability, which essentially leads to the "loss" of one eigenvalue (Proposition IA.11).

[20] 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.

parts, giving rise to the empirical PEPs and PAPs. 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 one-day buffer between the last day in the training sample and the first day in the forecast window.

We investigate whether empirical PPs behave in accordance with our theoretical predictions. Figure 1, Panel 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, Panel 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 a degradation of realized returns relative to the eigenvalues when looking out-of-sample, as reflected in the different y-axes in the left and right panels of Figure 1.

In a similar vein, Figure 1, Panels B and C show the eigenvalues of the symmetric and antisymmetric parts of the prediction matrix, respectively. Figure 1, Panels E and 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 returns. 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, Panel 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 long or short a size-value bet based on its own momentum. Put differently, when large value has recently outperformed, PEP1 buys large value, otherwise it buys small growth. To illustrate this strategy further, Figure 2, Panel B plots the momentum, $S'w_1$, of the eigenvector. Figure 2, Panel C shows the overall portfolio weight, $S'w_1w_1'$, averaged over time. Similarly, Figure 2, Panels B, D, and 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 PPs (among each version: PP, PEP, and PAP), and the combination of the top three PEPs plus top 3 PAPs.[21] 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

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

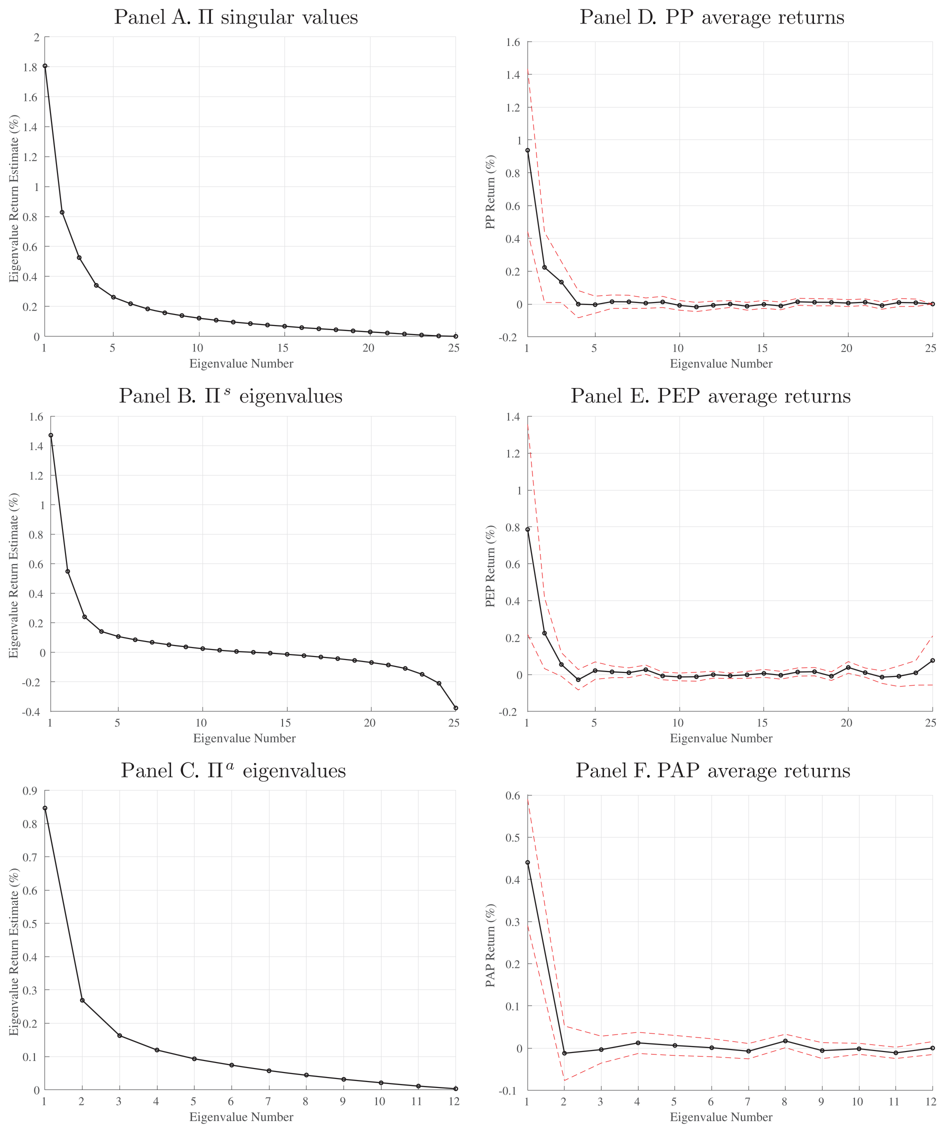

Figure 1. Prediction matrix eigenvalues. Panels A, B, and C show estimated eigenvalues of the prediction matrix and its symmetric and antisymmetric components, respectively, averaged over training samples. Panels D, E, and F show average out-of-sample returns and ±2 standard error confidence bands for corresponding principal portfolios (PPs), principal exposure portfolios (PEPs), and principal alpha portfolios (PAPs), respectively. Estimates correspond to 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 nonoverlapping 20-day return observations. The sample period is 1963 to 2019. (Color figure can be viewed at wileyonlinelibrary.com)

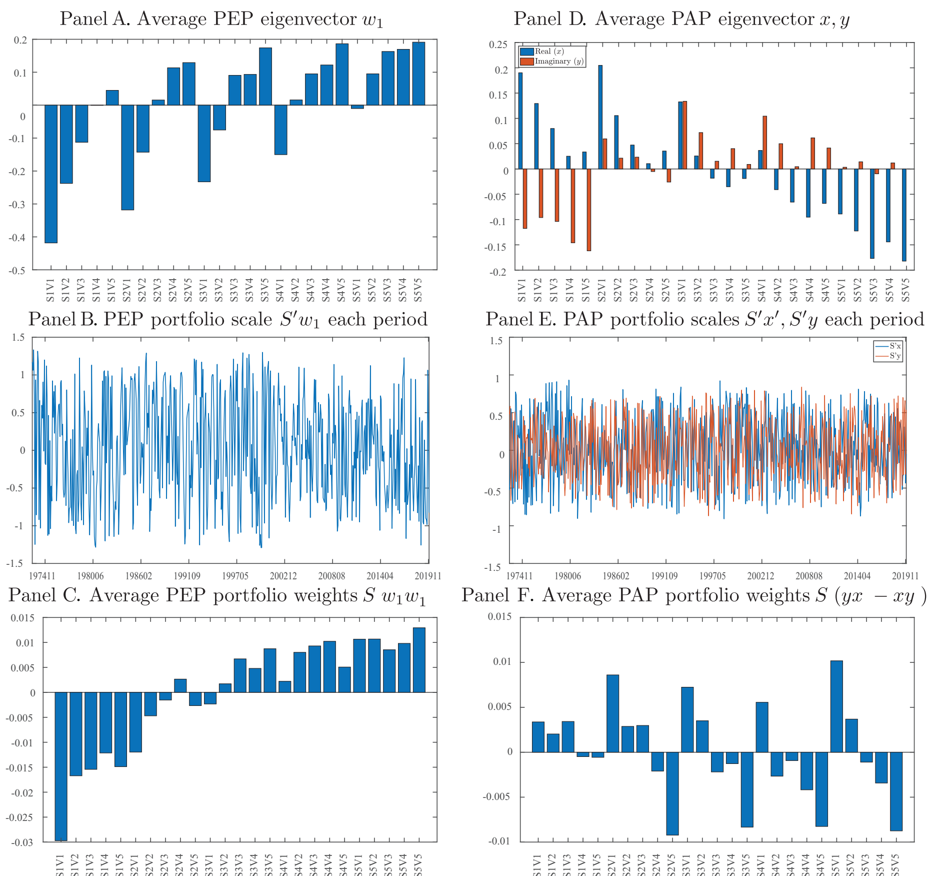

Figure 2. Portfolio weights for leading principal portfolios. Panels A and D plot the out-of-sample eigenvector underlying the first principal exposure portfolio (PEP) and first principal alpha portfolio (PAP) on the 25 size and value portfolios, averaged over training samples. Panels B and E plot the scale (in this case, interpreted as portfolio momentum), of the first PEP and PAP. Panels C and F show the overall portfolio weights averaged over time. Portfolios are constructed based on a 20-day momentum signal 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 nonoverlapping return observations. The sample period is 1963 to 2019. (Color figure can be viewed at wileyonlinelibrary.com)

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, at 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 SRs and information ratio (IR)s, we also report ±2 standard error bars around each estimate based the approximate standard error formula Lo (2002).

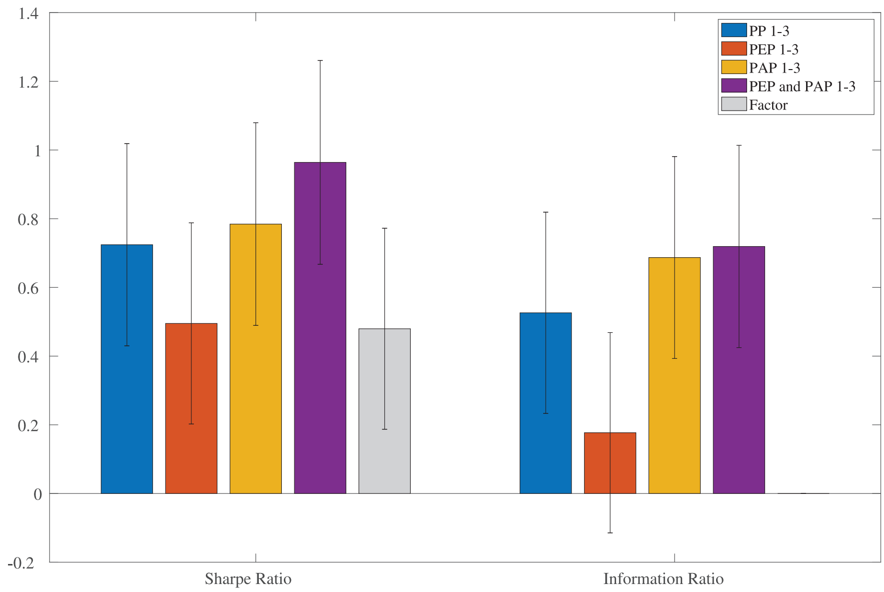

Figure 3. Principal portfolio performance ratios. This figure shows out-of-sample performance of principal portfolios in terms of the annualized Sharpe ratio (left set of bars) and the annualized IR versus the own-predictor strategy and the Fama-French five-factor model (right set of bars) along with ± 2 standard error bands 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 horizon (and, equivalently, holding period) of 20 days. The figure depicts performance for the simple factor ("Factor," that is, the standard own-signal strategy, included as a benchmark), the equal-weighted average of the top three principal portfolios ("PP 1-3"), the equal-weighted average of the top three principal exposure portfolios ("PEP 1-3"), the equal-weighted average of the top three principal alpha portfolios ("PAP 1-3"), and the equal-weighted average of the top three PEPs and PAPs 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 nonoverlapping return observations. The sample period is 1963 to 2019. (Color figure can be viewed at wileyonlinelibrary.com)

Figure 3 also plots the out-of-sample IR and its confidence interval as a measure of the risk-adjusted return of the PPs. Specifically, the IR is computed by regressing the return of the PP (or PEP, PAP, or their combination) on the 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 SR when all of the factors on the right-hand side are hedged out (i.e., the alpha expressed as an SR).

Table II

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-weighted average of the top three principal portfolios ("PP 1-3"), the equal-weighted average of the top three principal exposure portfolios ("PEP 1-3"), the equal-weighted average of the top three principal alpha portfolios ("PAP 1-3"), and the equal-weighted average of the top three PEPs and PAPs 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 horizon, and each forecast is made on an out-of-sample basis using a rolling training sample of the 120 most recent nonoverlapping return observations. The sample period is 1963 to 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 II reports results of this regression. As can be seen from Table II (and the confidence intervals in Figure 3), 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 II also shows that PEP has a highly significant loading on the simple factor with a high R^2, consistent with PEP providing exposure to this factor. In contrast, PAP has a small and insignificant factor loading on the simple factor, consistent with PAP providing factor-neutral alpha. (Some of the loadings on Fama-French factors are significant, but the overall R^2 is low.) In summary, these findings are consistent with the idea that PEP provides factor exposure while PAP provides nearly uncorrelated alpha.

Extended Momentum Analysis. We report a variety of extensions and robustness tests in Internet Appendix VII. We find that the out-of-sample PPs perform even better at shorter forecast horizons (Figure IA.1). PPs 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 United States), and a sample of 52 futures contracts for commodities, equity indices, sovereign bonds, and currencies (Figure IA.2). We show robustness with respect to other momentum signals, namely, those based on 40-, 60-, 90-, 120-, and 250-day past returns, following the standard practice of considering momentum signals up to one year (Figure IA.3). Finally, we consider subsample analysis by decade (Figure IA.4) and demonstrate robustness of our findings when we do not cross-sectionally demean signals and returns (Figure IA.5).

B. Factor-Timing

Our second empirical analysis investigates the power of PPs 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 that the returns of common factors are predictable. As emphasized by Haddad, Kozak, and Santosh (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 SR 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 SR 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. (2022)) and factor valuation ratios (Cohen, Polk, and Vuolteenaho (2003), Kelly and Pruitt (2013), and Haddad, Kozak, and Santosh (2020)). The method of PPs 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 PPs and do so for a larger collection of factor predictors than studied in prior literature. We begin with the large set of 153 U.S. equity characteristics and associated factors from Jensen, Kelly, and Pedersen (2022).[22] We discard 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.[23] Thus, returns for each of the factor

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

[23] 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 to form factor-level predictors.

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 PPs, 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 to 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 PPA one signal at a time. For example, we construct the set of PPs among the 138 factors using the accruals predictor for each factor. We then construct PPs based only on the book-to-market predictor of each factor. We proceed in this way, building one set of out-of-sample PP 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 and then use the fitted parameters to construct out-of-sample PP returns at $t + 1$.

Figure 4 reports the performance of PPs on average across the 138 different signals. We report average SRs for portfolios corresponding to the first 10 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 SR 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. The first and last PEPs earn annualized SRs of 0.4 and −0.4, respectively. This indicates that own-factor predictability tends to be heterogeneous on average across signals, with a given signal exhibiting positive own-factor predictability for some factors and 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 SR of 0.7 per annum.

In Panel D, we report the average performance of the simple factor. This factor restricts portfolio construction to homogeneous own-factor predictions and produces an average annualized SR 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, Kelly, and Pedersen (2022), 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 SR of 0.5, and it contributes in part to the performance of PPs, which also detects attractive static bets.

We are interested in whether PPs can successfully exploit return predictability to time factors. To this end, we are especially conservative in calculating IRs of PPs 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 IRs. However, the leading PP and PAP portfolios continue to produce large and significant IRs. The leading PAP portfolio achieves an IR of 0.8 on average across all signals.

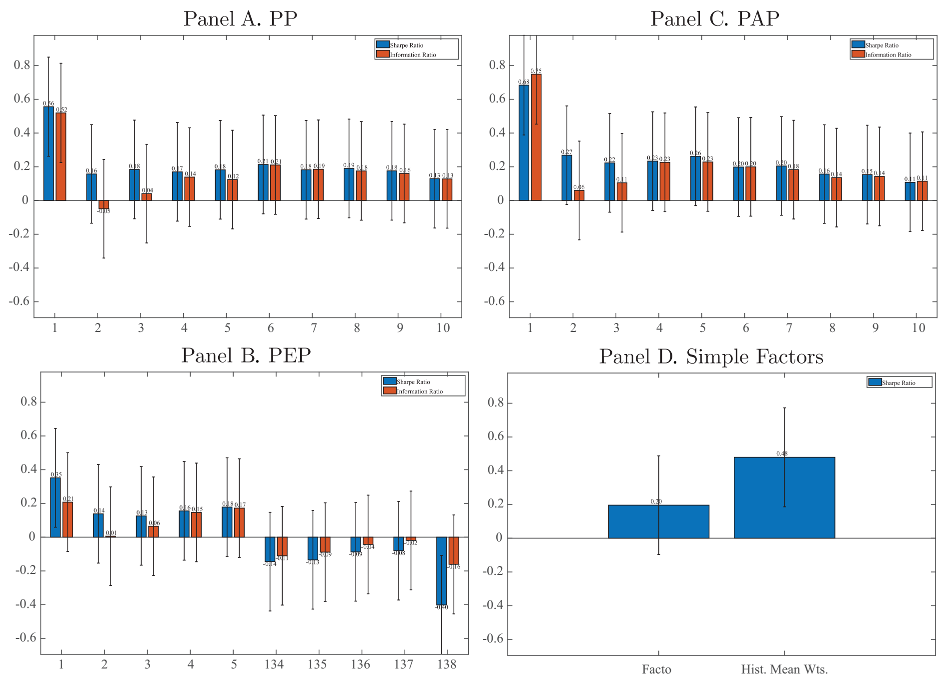

Figure 4. Average performance of factor-timing strategies based on individual signals. This figure shows out-of-sample performance of principal portfolios averaged across 138 signals. We show the annualized Sharpe ratio and the annualized IR versus the own-predictor strategy, the factor historical mean weight strategy, and the Fama-French five-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 depicts performance for the first 10 principal portfolios (PPs, Panel A), the first five and last five principal exposure portfolios (PEPs, Panel B), and the first 10 principal alpha portfolios (PAPs, Panel C) averaged across signals, along with the average ±2 standard error bands. Panel D shows 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 IRs of principal portfolios). Each forecast is made on an out-of-sample basis using a rolling training sample of 120 monthly return observations. The sample period is 1963 to 2019. (Color figure can be viewed at wileyonlinelibrary.com)

Next, rather than reporting averages of PP strategies based on individual signals, we study the performance of strategies that combine PPs across all signals. In particular, we report the performance of equal-weighted 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 SRs and IRs (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-weighted combination of leading PPs earns an SR of 0.6. The IR of the combined strategy is 1.1 and is highly statistically significant. Next, IRs for equal-weighted combinations of PEPs (the first and last PEP, denoted PEP 1 and PEP N, respectively) are large and highly significant. The first PEP,

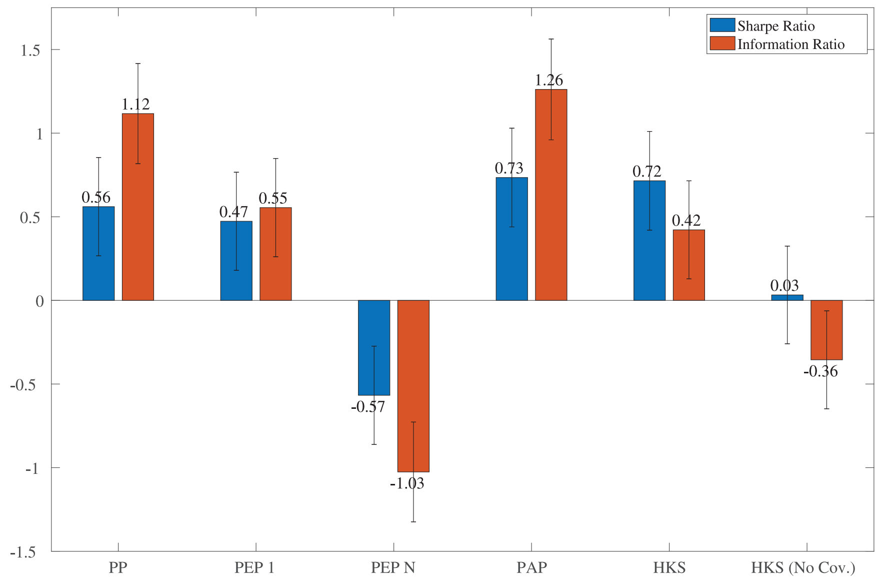

Figure 5. Performance of factor-timing strategies combined across signals. This figure shows out-of-sample performance of principal portfolio strategies combined across 138 signals. We show the annualized Sharpe ratio and the annualized IR versus the own-predictor strategy, the factor historical mean weight strategy, and the Fama-French five-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 and combined across signals. PP denotes the equal-weighted average of the first principal portfolio across 138 signals, PEP 1 is the equal-weighted combination of the first principal exposure portfolio, PEP N is the equal-weighted combination of the last principal exposure portfolio, and PAP is the equal-weighted combination of the first principal alpha portfolio across signals. We also report equal-weighted combinations of HKS factor-timing strategies with and without an inverse covariance matrix adjustment. Each portfolio is constructed on an out-of-sample basis using a rolling training sample of 120 monthly return observations. The sample period is 1963 to 2019. (Color figure can be viewed at wileyonlinelibrary.com)

which hones in on factors with positive symmetric predictability, earns an IR 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 IR of 1.0. The strongest effect of combining strategies across signals appears for the PAP combination, which achieves an out-of-sample IR of 1.3.

To help interpret the contribution of these strategies to the SDF, we analyze the ex post tangency portfolio using PPs (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-weighted average of simple factors across signals and

Table III

Ex Post Tangency Portfolios

This table reports ex post tangency portfolio weights using as inputs the five Fama-French factors, the equal-weighted 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-weighted average of leading principal portfolios across 138 signals, PEP 1 is the equal-weighted combination of the first PEP, PEP N is the equal-weighted combination of the last PEP (multiplied by −1 to make it a positive expected return strategy), and PAP is the equal-weighted combination of leading PAPs. To aid interpretation, all portfolios that contribute to the tangency portfolio are scaled to have the same volatility. * indicates that the portfolio weight is statistically significant at the 1% 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 SR. Each portfolio input is calculated on a rolling out-of-sample basis, and the ex post tangency analysis is then conducted from the full time series of out-of-sample strategies.

Portfolio	FF5	FF5 + PP	Nonnegative 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 × PEP N		0.41	0.00
PAP		2.05*	0.26*
Sharpe Ratio	1.09	2.15	1.52

the factor constructed using historical mean weights of the 138 underlying anomaly factors. Last, we include the equal-weighted combination of the first PP for all signals, the equal-weighted 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 III 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 SR 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 III supplements the Fama-French factors with the two additional benchmark factors and the PP 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 PPs, 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.

PPs generate an enormous gain in the mean-variance efficiency of the SDF. The tangency portfolio SR 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 nonnegativity constraint. In this case, we see that all PPs 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 nonnegativity constraint, the ex post tangency SR is 1.5, representing a gain in efficiency of 40% versus the Fama-French model.

This evidence indicates that PPs 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, Kozak, and Santosh (2020), 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 component 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 PP strategies. We estimate the factor return covariance matrix $\hat{\Sigma}_t = \text{cov}(R_{[t-120:t]})$ using a rolling 120-month training window and compute the associated eigenvalue decomposition,

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

Here, Q_t 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_t 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

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

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 PPs, we do not directly use information on the covariance matrix of returns. For an apples-to-apples comparison with PPs, 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 PPs, we construct an equal-weighted average of HKS factor-timing portfolios across the 138 possible predictors. As with PPs, we calculate the annualized out-of-sample SRs for both the *HKS* and $HKS^{\text{No Cov.}}$ portfolios, and IRs versus the benchmarks described earlier. The *HKS* portfolio SR is nearly identical to that for the PAP portfolio (0.7 in both cases). However, the IR 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 an SR of zero and a negative IR. This suggests that the performance of the *HKS* strategy derives from covariance timing and not expected return timing.

It turns out that the outperformance of PPs relative to the factor-timing approach of HKS is predicted by theory. A key observation is that the trading strategies investigated in HKS fit into the class of symmetric linear strategies, and thus they cannot capture any antisymmetric components of Π. As we show above, it is the PAPs that drive success of PPs strategies. We discuss this and related theoretical results in Internet Appendix Section V.

VI. Conclusion: The Power of PPA

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 PPs. We classify predictive patterns in Π as either symmetric or antisymmetric and derive theoretical results that translate these patterns into beta and alpha. These results give rise, in turn, 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 PPs based on a wide range of over 100 stock return prediction signals deliver large and significant risk-adjusted average returns out-of-sample. Our empirical PPs 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 HKS.

Initial submission: September 24, 2020; Accepted: August 4, 2021
Editors: Stefan Nagel, Philip Bond, Amit Seru, and Wei Xiong

REFERENCES

Arnott, Robert D., Mark Clements, Vitali Kalesnik, and Juhani T. Linnainmaa, 2022, Factor momentum, *Review of Financial Studies* (forthcoming).

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

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

Cochrane, John H., 2009, *Asset Pricing: Revised Edition* (Princeton Eniversity Press, Princeton, NJ).

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

Collin-Dufresne, Pierre, Kent D. Daniel, Ciamac C. Moallemi, and Mehmet Sağlam, 2015, Dynamic asset allocation with predictable returns and transaction costs, Working paper.

Collin-Dufresne, Pierre, Kent D. Daniel, and Mehmet Sağlam, 2020, Liquidity regimes and optimal dynamic asset allocation, *Journal of Financial Economics* 136, 379–406.

Fama, Eugene F., and Kenneth R. French, 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1–22.

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

Gârleanu, Nicolae, and Lasse Heje Pedersen, 2013, Dynamic trading with predictable returns and transaction costs, *Journal of Finance* 68, 2309–2340.

Gârleanu, Nicolae, and Lasse Heje Pedersen, 2016, Dynamic portfolio choice with frictions, *Journal of Economic Theory* 165, 487–516.

Gibbons, Michael R., Stephen A. Ross, and Jay Shanken, 1989, A test of the efficiency of a given portfolio, *Econometrica: Journal of the Econometric Society* 57, 1121–1152.

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

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

Haddad, Valentin, Serhiy Kozak, and Shrihari Santosh, 2020, Factor timing, *Review of Financial Studies* 33, 1980–2018.

Hansen, Lars Peter, and Ravi Jagannathan, 1991, Implications of security market data for models of dynamic economies, *Journal of Political Economy* 99, 225–262.

Horn, Toger A., and Charles R. Johnson, 1991, *Topics in Matrix Analysis* (Cambridge University Press, Cambridge).

Jensen, Theis Ingerslev, Bryan T. Kelly, and Lasse Heje Pedersen, 2022, Is there a replication crisis in finance? *Journal of Finance* (forthcoming).

Kelly, Bryan, and Seth Pruitt, 2013, Market expectations in the cross-section of present values, *Journal of Finance* 68, 1721–1756.

Kelly, Bryan, Seth Pruitt, and Yinan Su, 2020a, Characteristics are covariances: A unified model of risk and return, *Journal of Financial Economics* 134, 501–524.

Kelly, Bryan, Seth Pruitt, and Yinan Su, 2020b, Instrumented principal component analysis, Working paper.

Lo, Andrew W., 2002, The statistics of Sharpe ratios, *Financial Analysts Journal* 58, 36–52.

Moskowitz, Tobias J., Yao Hua Ooi, and Lasse Heje Pedersen, 2012, Time series momentum, *Journal of Financial Economics* 104, 228–250.
Pedersen, Lasse Heje, Abhilash Babu, and Ari Levine, 2021, Enhanced portfolio optimization, *Financial Analysts Journal* 77, 124–151.
Velu, Raja, and Gregory C. Reinsel, 1998, Multivariate Reduced-Rank Regression: Theory and Applications (Springer New York, NY).
Welch, Ivo, and Amit Goyal, 2008, A comprehensive look at the empirical performance of equity premium prediction, *Review of Financial Studies* 21, 1455–1508.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix S1: Internet Appendix.
Replication Code.