

Canonical Portfolios: Optimal Asset and Signal Combination*

Nikan Firoozye[†]

Vincent Tan[‡]

Stefan Zohren[§]

January 2023

Abstract

We present a novel framework for analyzing the optimal asset and signal combination problem, which builds upon the dynamic portfolio selection problem of [Brandt and Santa-Clara \(2006\)](#) in two phases. First, we reformulate their original investment problem into a tractable vehicle, which admits a closed-form solution that scales to large dimensions by imposing a joint Gaussian structure on the asset returns and signals. Second, we recast the optimal portfolio of correlated assets and signals into a set of uncorrelated managed portfolios through the lens of Canonical Correlation Analysis of [Hotelling \(1936\)](#). The new investment environment of uncorrelated managed portfolios offers unique economic insights into the joint correlation structure of our optimal portfolio policy. We also operationalize our theoretical framework to bridge the gap between theory and practice and showcase the improved performance of our proposed method over natural competing benchmarks.

JEL Classification: G11, D81, C1.

Keywords: Canonical Correlation Analysis, Portfolio Selection, Statistical Arbitrage

*The authors would like to thank Mihai Cucuringu, Brian Healy, Anthony Ledford, Ana Maria Pires, Slavi Marinov, Attilio Meucci, and Yin Cheng Ng, as well as from seminar participants at University of Oxford and Man AHL for the helpful comments. This work was supported by the Oxford-Man Institute of Quantitative Finance.

[†]Department of Computer Science, University College London. E-mail: n.firoozye@ucl.ac.uk

[‡]Oxford-Man Institute of Quantitative Finance, University of Oxford. E-mail: vincent.tan@eng.ox.ac.uk

[§]Oxford-Man Institute of Quantitative Finance, University of Oxford. E-mail: stefan.zohren@eng.ox.ac.uk

1 Introduction

The investment decisions of portfolio managers are typically guided by signals that encompass their view on future returns.¹ In order to harness the predictive contents of these signals to drive those decisions, there is a standard approach championed by [Markowitz \(1952\)](#), which still to this day remains the workhorse of modern portfolio theory and forms the bedrock of how we think about diversification. Markowitz’s portfolio selection is traditionally performed in two stages. In the first stage, the investor draws inferences about the return-generating process either through plug-in estimation or subjective belief formation. In the second stage, the optimal portfolio weights are formed with estimates from the first stage.

In this paper, we borrow inspiration from [Brandt and Santa-Clara \(2006\)](#) and consider solving a dynamic portfolio selection problem with multiple return-predictive signals in a single stage. Our first contribution is to reformulate the authors’ original investment problem into a tractable one by modeling the joint distribution of the returns and signals as a Gaussian, which yields a solution that scales to large dimensions. The end result is a powerful framework that (1) integrates both return forecasting and asset allocation processes together, (2) exposes the multivariate contributions that are embedded in both processes, which may impact the final allocation of capital, and (3) is simple to implement. However, along with these benefits come concomitant challenges. Unlike the standard mean-variance framework that only requires one large-dimensional covariance as an input, our framework necessitates two additional inputs; namely, the covariance of the signals and the cross-covariances between both the returns and signals. These new inputs not only warrant estimation but their interactions—between and within variables—also introduce an additional layer of complexity that we have to distill.

Our second contribution is to provide a simplification of this problem by reorganizing the set of assets and accompanying signals into a set of uncorrelated managed portfolios. As was hinted in [Firoozye and Koshiyama \(2020, Section 4\)](#), this can be achieved by a powerful tool from multivariate analysis known as canonical correlation analysis (CCA) pioneered by [Hotelling \(1936\)](#). CCA is a generalization of principal component analysis (PCA) of [Hotelling \(1933\)](#) to two sets of random variables and a study of their joint association. This is pertinent to our case where we have multiple correlated assets and correlated signals that are mutually linked by a joint correlation objective. CCA allows us to express any portfolio returns derived from our framework in terms of its exposures to uncorrelated sources of returns by reweighting the original set of financial variables such that their correlation is maximally retained; we refer to these special weights as *canonical portfolios*.²

The third contribution is to render our theoretical framework applicable to practice, even in large

¹The terminology “signals” is also often referred to as predictors, characteristics, state variables, cross-sectional anomalies, etc. in the finance literature.

²The idea of recasting the asset universe into their orthogonal components through PCA can be traced back to [Partovi and Caputo \(2004\)](#) with further extensions and applications from [Meucci \(2009\)](#) and [Avellaneda and Lee \(2010\)](#), as well as from [Kozak et al. \(2018, 2020\)](#) in the context of risk pricing in a no near-arbitrage framework. Our work aims to broaden that horizon to *two* sets of variables by also considering the signal as an important input in the analysis of their joint association with returns.

dimensions. This is a challenging task given that all three matrix inputs have degrees of freedom that scale quadratically in the size of the problem. Thus, regularization of these large-dimensional objects is paramount to encourage stability in the out-of-sample results. Our prescription to this problem is by employing shrinkage to the covariances, which is commonly used in portfolio selection problems; see, for example, [Frost and Savarino \(1986\)](#) and [Ledoit and Wolf \(2004a,b, 2017\)](#). Additionally, we retain the leading few canonical portfolios that exhibit the greatest predictability while discarding noisy and unstable ones. We show through a series of backtest simulations that our proposed method responds positively to these innovations that we introduced and outperforms competitive benchmarks.

Our paper builds upon a series of seminal works from [Brandt \(1999\)](#), [Ferson and Siegel \(2001\)](#), [Brandt and Santa-Clara \(2006\)](#), and [Brandt et al. \(2009\)](#) that directly model the portfolio weights. Of relevance to our study is [Brandt and Santa-Clara \(2006\)](#), where the authors analyzed portfolio policies conditioning on a small number of state variables through augmentation of the asset space. However, we provide a counterpoint to the authors on two fronts. First, we adopt a tractable approach to their problem that allows us to reveal the correlation structure of our optimal portfolio through CCA. Second, since our approach does not expand the dimension of the asset space, our method is more computationally tractable in large dimensions, allowing it to accommodate large-scale cross-sectional applications such as the equity universe. The price that we have to pay for this tractability, however, is the need to confront three large-dimensional objects.

Our study also extends the work of [Firoozye and Koshiyama \(2019\)](#) on the use of total least squares (TLS) of [Golub and Van Loan \(1980\)](#) for optimally combining signals on a univariate return. Their original study focused on a novel objective function for return prediction. While most research analysts will seek to find a good forecast for future returns via methods such as ordinary least squares (OLS), or in the nonlinear context, via a large suite of machine learning-based methods, [Firoozye and Koshiyama \(2019, 2020\)](#) argued that if the goal is to maximize the Sharpe ratio of trading strategy, then this can be achieved through a linear combination of signals that maximizes the correlation between their combination and return. In the linear context, the solution to this problem comes out via TLS, which is an errors-in-variables formulation of regression. TLS has been well-studied primarily among numerical analysts and is used less formally on trading desks of investment banks and hedge funds, typically under the moniker of PCA regression. Given the close relation between TLS and CCA, our work extends this research towards a multivariate context, which is more relevant for problems in portfolio selection.

The remainder of the paper is organized as follows. [Section 2](#) gives the description of the dynamic portfolio selection problem. [Section 3](#) provides the financial interpretation of our optimizer with CCA and details our proposed estimation approach. [Section 4](#) describes the empirical methodology and presents the results of the out-of-sample backtest exercise with Fama-French equity sorted portfolios. [Section 5](#) concludes. [Appendix A–B](#) contain all the figures, tables, and additional mathematical derivations.

2 Setting the Stage

2.1 Notation

In this section, we introduce the notational convention for our analysis. Let the subscript i index the variables such that $i \in \{1, \dots, N\}$, where N denotes the dimension of the asset universe. The subscript t indexes the trading dates such that $t \in \{1, \dots, T\}$, where T denotes the number of observations. The notation $\text{Cov}(\cdot, \cdot)$ represents the covariance matrix between two random vectors, $\text{Tr}(\cdot)$ represents the trace of a square matrix, and $\text{Diag}(\cdot)$ represents a diagonal matrix of a given vector. Denote \mathbb{I}_N to be an identity matrix of dimensions $N \times N$. For any N -dimensional vector z , its ℓ_2 -norm is given by $\|z\|_2 := \sqrt{z_1^2 + \dots + z_N^2}$, and for any real matrix Z , its Frobenius norm is $\|Z\|_F := \sqrt{\text{Tr}(Z'Z)}$. Here, the symbol $(\cdot)'$ denotes the transpose operator of a vector or matrix, the symbol $:=$ is a definition sign, and the symbol $=$ is the equal sign.

Let $r_{t,i}$ be the return for a risky asset i sampled at trading date t , stacked into a vector $r_t := (r_{t,1}, \dots, r_{t,N})'$. Also, let $x_{t,i}$ be an M -dimensional vector return-predictive signal at date t for asset i , stacked into a vector $x_t := (x_{t,1}, \dots, x_{t,N})'$. This means every asset is accompanied by at least one signal and so M is necessarily a positive integer. For example, if there are two signals for one asset, then $M = 2$. The signals are understood to proxy the future expected returns conditional on the investor's information set at time t . We shall assume that both the asset returns and the signals exhibit stationarity and ergodicity.

Their multivariate distributions are assumed to be Gaussian with zero expectations and covariances $\text{Var}(r_{t+1}) := \text{Cov}(r_{t+1}, r_{t+1}) = \Sigma_r$ and $\text{Var}(x_t) := \text{Cov}(x_t, x_t) = \Sigma_x$, and a cross-covariance $\text{Cov}(r_{t+1}, x_t) = \Sigma_{rx} = \Sigma'_{xr}$. The column vector of stacked subsequent returns and signals $(r_{t+1}, x_t)'$ has a joint covariance matrix expressed in the following partitioned form:

$$\Sigma = \begin{pmatrix} \Sigma_r & \Sigma_{rx} \\ \Sigma'_{rx} & \Sigma_x \end{pmatrix}. \quad (1)$$

In our notation, we denote estimated quantities of their population analogs with a hat accent ($\hat{\cdot}$) on them. The sample covariances of the asset returns and signals and the sample cross-covariance between the returns and signals are denoted as $S_r := T^{-1} \sum_{t=1}^T r_{t+1} r'_{t+1}$, $S_x := T^{-1} \sum_{t=1}^T x_t x'_t$, and $S_{rx} := T^{-1} \sum_{t=1}^T r_{t+1} x'_t$, respectively.

Due to the symmetric positive semidefinite property of the covariance matrices, they admit a spectral decomposition. For example, $\Sigma_r = P_r \text{Diag}(d_{r,1}, \dots, d_{r,N}) P'_r$, where $P_r := [p_{r,1}, \dots, p_{r,N}]$ is an orthogonal matrix ($P'_r P_r = P_r P'_r = \mathbb{I}_N$) whose columns contain the eigenvectors and $\text{Diag}(d_{r,1}, \dots, d_{r,N})$ is a diagonal matrix containing the eigenvalues. The eigenvalues are assumed to be sorted in ascending order. The matrix factors $\Sigma_r^{1/2}$ can be obtained by retaining the eigenvectors of Σ_r , but reassembling them with the square roots of the eigenvalues given by $(\sqrt{d_{r,1}}, \dots, \sqrt{d_{r,N}})'$. An analogous notation applies for the decomposition of the covariance of signals Σ_x and its matrix of factors $\Sigma_x^{1/2}$.

2.2 Dynamic Strategies

We begin our analysis by considering the portfolio weights obtained at time t to be linear in the vector of signals, that is, $w_t := A'x_t$, where A is a $NM \times N$ matrix of coefficients.³ These weights are conditional on the strength of the signals but the matrix A is a static object where each column maps the signal vector into a portfolio weight invested in each basis asset. From this parameterization, the conditional portfolio selection problem that an investor faces at time t can be formulated in terms of a quadratic utility objective as follows

$$\max_A \mathbb{E}_t[x_t'Ar_{t+1}] - \frac{\gamma}{2} \text{Var}_t[x_t'Ar_{t+1}], \quad (2)$$

where γ is an investor's risk aversion parameter.⁴ This objective function underscores that the investor chooses to simultaneously allocate between the assets and signals in order to yield a trading strategy that optimizes the returns of the portfolio at time $t + 1$, which we denote as r_{t+1}^w .

Brandt and Santa-Clara (2006) showed that this problem formulation is equivalent to an unconditional portfolio selection problem since the matrix of coefficients is constant through time. Indeed, the matrix A that optimizes the investor's conditional problem at any given date is the same for all dates and hence, it also optimizes the investor's unconditional problem. Hence, we can drop the subscript t in (2) and consider the following revised problem instead

$$\max_A \mathbb{E}[x_t'Ar_{t+1}] - \frac{\gamma}{2} \text{Var}[x_t'Ar_{t+1}]. \quad (3)$$

Based on the assumption of joint Gaussianity between the returns and signals, we can express this objective in terms of the unconditional second-moments with the following proposition.

Proposition 1. The objective function to the investor's problem (3) is given by

$$\max_A \text{Tr}(A\Sigma_{rx}) - \frac{\gamma}{2} (\text{Tr}(\Sigma_x A \Sigma_r A') + \text{Tr}(\Sigma_{rx} A \Sigma_{rx} A)). \quad (4)$$

A closed-form solution to this optimization problem is available in Theorem 2 from Appendix B.

It is analytically convenient to assume that $\text{Tr}(\Sigma_{rx} A \Sigma_{rx} A) \approx 0$ in Equation (4) from Proposition 1.⁵ The following auxiliary problem is sufficient for us to study.

³Note that the assumption that the optimal portfolio weights are linear functions of the signals is not entirely restrictive. Indeed, each of the signals in x_t can be composed of non-linear functions of a more primitive set of underlying signals.

⁴Other reformulations of this problem include the maximization of mean return or minimization of risk. Regardless of the choice, all three problems result in the same mean-variance trade-off.

⁵This approximation allows us to express the optimal portfolio policy solely in terms of the unconditional moments without any recourse to canonical correlation analysis, which is required to solve problem (4). It is a reasonable approximation provided the squared expected returns of the dollar-neutral managed portfolios—whose returns are of the form x_t times the return of each asset—as well as their weighted sum contributions are small and can be approximated with zero. Notwithstanding this approximation, we have also empirically tested both approaches in Section 4.

Proposition 2. Consider the following objective function

$$\max_A \text{Tr}(A\Sigma_{rx}) - \frac{\gamma}{2} \text{Tr}(\Sigma_x A \Sigma_r A'). \quad (5)$$

The solution to this problem is given by

$$A = \frac{1}{\gamma} \Sigma_x^{-1} \Sigma'_{rx} \Sigma_r^{-1}, \quad (6)$$

The weight w_t allocated to each asset conditional on the signal at time t is then

$$w_t = \frac{1}{\gamma} \Sigma_r^{-1} \Sigma_{rx} \Sigma_x^{-1} x_t. \quad (7)$$

In absence of any further constraints on w_t , this is the frictionless portfolio policy that invests directly in each asset, and the size of the portfolio is determined by the level of the investor’s risk aversion. The dynamics of the portfolio is guided by the trajectory of the signal x_t at each period. We see that the portfolio takes into account the cross-sectional information from the asset returns, the signals as well as their cross-relationships. The cross-covariance matrix Σ_{rx} can be interpreted as a matrix consisting of managed portfolios that encapsulate the potential earnings opportunities that are available to an investor.⁶ Each managed portfolio, formed in the spirit of [Hansen and Richard \(1987\)](#), [Cochrane \(1996\)](#) and [Brandt and Santa-Clara \(2006\)](#) among others, invests in each asset a proportion determined by the size of the conditioning variable. For this reason itself, we term this matrix a “conditional portfolios” matrix.

Example 1 (Univariate Portfolio). Let us consider one return-accompanying signal and postulate that all covariances in Equation (7) are the identity matrix. In this case, the weight is proportional to the conditioning variable. This portfolio policy is interesting in its own right because it ignores the impact that the returns and signals can have on the covariances. Hence, it is essentially a univariate system in which the signals only forecast their associated asset. We shall henceforth refer to this conditional portfolio as the univariate factor (UNI).

Example 2 (Two-Assets Portfolio). Suppose that $N = 2$, $M = 1$, and $\gamma = 1$. If the asset returns and signals have unit variances, then the covariances are

$$\Sigma_r = \begin{pmatrix} 1 & \rho_r \\ \rho_r & 1 \end{pmatrix}, \quad \Sigma_x = \begin{pmatrix} 1 & \rho_x \\ \rho_x & 1 \end{pmatrix}, \quad \text{and} \quad \Sigma_{rx} = \begin{pmatrix} \xi_{11} & \xi_{12} \\ \xi_{21} & \xi_{22} \end{pmatrix}. \quad (8)$$

For simplicity, also suppose that the signals are uncorrelated, that is, $\rho_x = 0$. Inserting these objects

⁶In [Kelly et al. \(2022\)](#), the authors term the cross-covariance matrix as the “prediction matrix”, which is also another valid interpretation since the correlation between an asset return and some signal is one measure of signal-return predictability. Given that a cross-covariance matrix is not symmetric in general, the predictive strength of a signal i on asset j may be different from that of signal j on asset i .

into the optimal matrix A , we can write the weights invested in the two basis assets as

$$w_t = \left(\frac{\xi_{11} - \rho_r \xi_{21}}{1 - \rho_r^2} x_1 + \frac{\xi_{12} - \rho_r \xi_{22}}{1 - \rho_r^2} x_2, \frac{\xi_{21} - \rho_r \xi_{11}}{1 - \rho_r^2} x_1 + \frac{\xi_{22} - \rho_r \xi_{12}}{1 - \rho_r^2} x_2 \right). \quad (9)$$

Moreover, the expected portfolio return is

$$\text{Tr}(A\Sigma_{rx}) = \text{Tr}(\Sigma_r^{-1}\Sigma_{rx}\Sigma_x^{-1}\Sigma'_{rx}) = \frac{\xi_{11}^2 + \xi_{12}^2 + \xi_{21}^2 + \xi_{22}^2 - 2\rho_r(\xi_{11}\xi_{21} + \xi_{12}\xi_{22})}{1 - \rho_r^2}. \quad (10)$$

This expression is non-negative given $\rho_r \in [-1, 1]$ since $\xi_{11}^2 + \xi_{21}^2 \geq 2\rho_r \xi_{11}\xi_{21}$ and $\xi_{12}^2 + \xi_{22}^2 \geq 2\rho_r \xi_{12}\xi_{22}$ by the inequality of arithmetic and geometric means. Let us consider three cases. First, if $\rho_r = 0$, then the expected return is simply the squared expected returns of all conditional portfolios that exploit own- and cross-predictabilities. Second, if $\rho_r \in (-1, 0)$, then there is an additional positive contribution to the conditional portfolios that arises from diversifying into imperfectly negatively correlated assets. The position is leveraged according to a factor of $1/(1 - \rho_r^2)$, which increases as $|\rho_r|$ approaches one from below. Third, if $\rho_r \in (0, 1)$, then the expected portfolio return is now reduced since the gains to diversification are limited by the extent to which the assets are positively correlated. Notwithstanding the diminished returns, the optimal strategy is also leveraged with a factor $1/(1 - \rho_r^2)$ in order to increase the portfolio returns to meet its profit objective. A similar conclusion holds for the case $\rho_r = 0$ and $\rho_x \neq 0$.

Remark 1 (Conditional Portfolio Policies). We can gain further intuition into the role of conditioning that our problem posits. Consider the following conditional portfolio selection problem:

$$\max_{w_t} \mathbb{E}_t[w'_t r_{t+1}] - \frac{\gamma}{2} \text{Var}_t[w'_t r_{t+1}]. \quad (11)$$

Since the weights are known at time t , the problem can be solved to give

$$w_t = \frac{1}{\gamma} \text{Var}_t[r_{t+1}]^{-1} \mathbb{E}_t[r_{t+1}]. \quad (12)$$

Observe that this optimal portfolio is determined by the mean and variance of returns that incorporate all available information up to time t . Hence, restricting the weights to be linear functions of a set of signals also restricts the solution set, which necessarily leads to a suboptimal solution. However, specifying the conditional means and variances of the returns is a notoriously difficult task. Our weights restriction simplifies these dual tasks by converting a dynamic problem into an equivalent static problem for which a conditional portfolio policy is available solely in terms of the unconditional moments of the assets and signals. That is, our conditional policy (7) does not rely on how the conditional moments of returns depend on the signals or how the signals vary over time.

Remark 2 (Regression-Based Policies). Notice that the portfolio weights (7) can be seen as one that had been obtained from a two-stage process of forecasting and asset allocation. To see this, let

us consider the following return-generating process:

$$r_{t+1} = B'x_t + \varepsilon_{t+1}, \quad (13)$$

where B is a matrix of coefficients and ε_t is an idiosyncratic error that is independent and identically distributed (i.i.d.). A simple forecasting approach that captures this relationship between both variables would be to perform an ordinary least squares (OLS) regression. In particular, one would first perform N independent univariate OLS regressions of the returns to assets on the signals in order to obtain a vector of cross-sectional predictive signals, and then insert them into (12) to arrive at the vector of allocations:

$$w_t = \frac{1}{\gamma} \Sigma_\varepsilon^{-1} \Sigma_{rx} \Sigma_x^{-1} x_t, \quad (14)$$

where $\Sigma_\varepsilon = \text{Cov}(\varepsilon_t) = \Sigma_r - \Sigma_{rx} \Sigma_x^{-1} \Sigma_{rx}'$ is the covariance matrix of the residuals. In fact, this portfolio is also optimal under our assumption of joint Gaussianity between the asset returns and signals. By invoking the standard properties of the multivariate Gaussian distribution, the conditional distribution of r_{t+1} given x_t is also multivariate Gaussian distributed with means and variances given by

$$\mathbb{E}_t[r_{t+1}] = \Sigma_{rx} \Sigma_x^{-1} x_t, \quad (15)$$

$$\text{Var}_t[r_{t+1}] = \Sigma_r - \Sigma_{rx} \Sigma_x^{-1} \Sigma_{rx}. \quad (16)$$

Inserting these expressions into (12) gives portfolio (14). In contrast to our portfolio (7), this portfolio depends on the covariance structure of the unexplained parts of the returns. Another key difference here is that our portfolio policy respects symmetry; exchanging the roles of the assets and signals yields identical returns to the portfolio. This symmetry property may be warranted in situations where the signals have an influence on the asset returns but at the same time, the reverse association is also economically plausible. It also implies that our approach is more closely related to total least squares (TLS) than it is to OLS since TLS is an errors-in-variables regression model that optimizes for the correlation of both variables, which is a symmetric metric.

2.3 Fully Invested Constraint

In quantitative equity investing, it is commonplace to impose economic constraints on the portfolio weights. A prominent example would be the fully-invested portfolio, which enforces a budget constraint so that the weights add to one. This equality constraint can be included in our optimization problem and given that it is linear in the matrix of coefficients A , the portfolio weights can be fortunately solved in closed form.

Proposition 3. We can formulate the optimal asset and signal combination problem with a fully-

invested constraint as follows

$$\begin{aligned} & \max_A \text{Tr}(A\Sigma_{rx}) - \frac{\gamma}{2} \text{Tr}(\Sigma_x A \Sigma_r A') \\ & \text{subject to } \mathbf{1}' A' x_t = 1, \end{aligned} \tag{17}$$

where $\mathbf{1}$ denotes the conformable vector of ones of dimension N . The problem has the following analytical solution

$$w_t^{\text{FI}} = (1 - \kappa) \frac{\Sigma_r^{-1} \mathbf{1}}{\mathbf{1}' \Sigma_r^{-1} \mathbf{1}} + \kappa \frac{\Sigma_r^{-1} \Sigma_{rx} \Sigma_x^{-1} x_t}{\mathbf{1}' \Sigma_r^{-1} \Sigma_{rx} \Sigma_x^{-1} x_t}, \tag{18}$$

where $\kappa := \gamma^{-1} \mathbf{1}' \Sigma_r^{-1} \Sigma_{rx} \Sigma_x^{-1} x_t$ is a scalar value.

Proposition 3 shows that the portfolio weight is expressed as a convex linear combination of the global minimum variance portfolio and the optimal unconstrained portfolio policy that is renormalized to one. The latter portfolio resembles the so-called ‘tangency portfolio’, which is the highest Sharpe ratio of a portfolio of risky assets from the standard mean-variance framework. The fully-invested portfolio w_t^{FI} has a similar form to the one obtained from the standard Markowitz paradigm found in [Huang and Litzenberger \(1988, Sections 3.8 and 3.9\)](#).

Example 3. Suppose there exists a vector of coefficients θ that maps the vector of signals $x_{t,i}$ into a real-valued signal $y_{t,i}$ for asset i according to $y_{t,i} = \theta' x_{t,i}$. Furthermore, suppose that all the covariances in the fully-invested portfolio are identity matrices with respect to this newly defined N -dimensional signal vector y_t . Then our fully-invested portfolio (18) simplifies to

$$w_t^{\text{FI}} = \frac{1}{N} \mathbf{1} + \frac{1}{\gamma} \left(y_t - \frac{\mathbf{1}' y_t}{N} \right). \tag{19}$$

The resulting portfolio can now be seen as a combination of an equally-weighted portfolio and a long-short portfolio that sums to zero. This corresponds to the parametric portfolio policy of [Brandt et al. \(2009\)](#) but specialized in an equally-weighted deterministic benchmark portfolio, and a quadratic utility preference.

2.4 Relation to Existing Literature

At this juncture, it is important to highlight the differences between our work from those of the existing literature. We first draw our attention to the fact that the objective function (5) under our consideration is similar to that of [Brandt and Santa-Clara \(2006\)](#).⁷ If we assume joint Gaussianity of the returns and signals and the squared returns of conditional managed portfolios are relatively

⁷A subtle difference worth pointing out is that the authors model the portfolio policy to be *affine* in the state variables, and hence the allocations are also driven by the assets themselves. Here, we work in a slightly restricted setting where the portfolio policy is only driven by the conditional managed portfolios themselves.

small and can be approximately zero.⁸ Then we can express the optimal matrix of coefficients A from Equation (6) in vectorized form as

$$\text{vec}(A) = \frac{1}{\gamma} \text{vec}(\Sigma_x^{-1} \Sigma'_{rx} \Sigma_r^{-1}) \quad (20)$$

$$= \frac{1}{\gamma} (\Sigma_x^{-1} \otimes \Sigma_r^{-1}) \text{vec}(\Sigma'_{rx}) \quad (21)$$

$$= \frac{1}{\gamma} (\Sigma_x \otimes \Sigma_r)^{-1} \text{vec}(\Sigma'_{rx}) \quad (22)$$

$$\approx \frac{1}{\gamma} \mathbb{E}[(x_t x'_t) \otimes (r_{t+1} r'_{t+1})]^{-1} \mathbb{E}[x_t \otimes r_{t+1}], \quad (23)$$

where $\text{vec}(\cdot)$ is an operator that stacks the columns of matrix A into a vector, and \otimes is the Kronecker product of two matrices. The fourth (approximate) equality invokes our sparsity assumptions so that the second-moment matrix of the conditional product returns $x_t \otimes r_{t+1}$ is a Kronecker product of two covariances. This corresponds to the solution that one obtains by rewriting the portfolio returns as $x'_t A r_{t+1} = \text{vec}(A)(x_t \otimes r_{t+1})$, and optimizing over $\text{vec}(A)$ with respect to $x_t \otimes r_{t+1}$.

The authors considered this new optimization scheme to be an augmented asset space form of the mean-variance optimization. They had removed one step in the process of forecasting and allocation and combined them into one step, thus reducing noise in the allocation process. However, there are two issues with this approach. First, augmenting the asset space results in a severe expansion of the problem dimension. The unrestricted covariance matrix in (23) has $N^2 M(N^2 M + 1)/2$ parameters, while our Kronecker product factorization in (22) has only $N(N+1)/2 + NM(NM+1)/2$ parameters. Clearly, the former becomes less computationally tractable when the dimensions of both the asset and signal space are large. Second, there is less clarity on the structure of their optimal solution due to the complicated interactions in the covariance between the subsequent returns and signals. Notwithstanding these shortcomings, our efforts in recasting the optimal strategy (23) into (6) (or (46) in Theorem 2) help to amend their approach towards large-scale financial applications. Although this requires us to model the joint distribution of returns and signals, our solution lends itself nicely to the study of the joint correlation structure of the optimal matrix of coefficients A ; see Section 3.3.

Second, our work is related to the classic portfolio selection problem of Markowitz (1952) in its static form. If we assume independent and identically distributed (i.i.d.) returns with constant moments and time-invariant weights $w_t \equiv w$, the investment problem (11) simplifies to an unconditional one, which has the following well-known solution

$$w = \frac{1}{\gamma} \text{Var}[r_{t+1}]^{-1} \mathbb{E}[r_{t+1}]. \quad (24)$$

This portfolio policy is typically implemented through plug-in estimation. In this paper, we focus on the more realistic setting of non-i.i.d. returns, which motivates our model parameterization and use

⁸For example, if a portfolio managed on the basis of a momentum signal has an average return of 5%, then the square of its return is 0.25%.

of conditioning variables that affect the joint distribution of returns.

Finally, our work is also related to that of [Kelly et al. \(2022\)](#) on cross-predictability. In their work, the authors pursued a similar parameterization for the portfolio weights but optimize an objective function that is subject to a robust risk constraint that controls for the leverage of A . As a result, their optimal portfolio policy depends only on the cross-covariance matrix. When viewed in the context of our framework, the analysis is simplified and avoids the challenges of inverting two additional large-dimensional matrices. However, this departs from our paper on two fronts. First, we assume that the investors have quadratic utility preferences in a similar spirit to the standard mean-variance problem.⁹ Second, our application of canonical correlation analysis formally suggests that the cross-covariance between *decorrelated* asset returns and signals is the more appropriate object for investment analysis as opposed to the cross-covariance of the original asset returns and signals. This is because the latter contains non-trivial variations that are embedded in both financial variables that may obscure the inference.

3 Strategy Diversification

3.1 Canonical Correlation Analysis

The goal of canonical correlation analysis (CCA) is to perform dimension reduction on two different data sets that comprise a large number of interrelated variables, while simultaneously retaining as much of the correlation present in the two data sets. This is made possible through a transformation of the original set of variables into a new set of mutually orthogonal paired variables, which are ranked so that the largest few retain most of the correlation present in all of the original variables.

There are several standard approaches to solving the CCA problem as outlined in [Uurtio et al. \(2017\)](#). Indeed, this may involve solving a standard eigenvalue problem ([Hotelling, 1936](#)), or a generalized eigenvalue problem ([Bach and Jordan, 2002](#)), or through singular value decomposition (SVD) ([Healy, 1957](#)). We adopt the latter approach in this brief exposition of the classical subject by emphasizing a sequence of two change of basis operations to arrive at a reduced problem whose financial interpretation we provide in [Section 3.3](#).

Given that scale invariance is a property of correlation, the first step is to orthogonalize the random vectors r and x such that their covariances are the identity matrices. This can be achieved through the following linear transformations $\tilde{r} := \Sigma_r^{-1/2}r$ and $\tilde{x} := \Sigma_x^{-1/2}x$. The transformed objects are now uncorrelated and have a unit variance and their joint covariance matrix is given by

$$\begin{pmatrix} \mathbb{I}_N & \Sigma_r^{-1/2}\Sigma_{rx}\Sigma_x^{-1/2} \\ \Sigma_x^{-1/2}\Sigma_{rx}'\Sigma_r^{-1/2} & \mathbb{I}_{NM} \end{pmatrix}. \quad (25)$$

⁹[Levy and Markowitz \(1979\)](#) and [Markowitz \(1991\)](#) argue that the quadratic utility preferences serve as a reasonable approximation to other utility preferences in portfolio selection problems.

It is helpful to introduce the object

$$\Sigma_{\tilde{r}\tilde{x}} := \mathbb{E}[\tilde{r}\tilde{x}'] = \Sigma_r^{-1/2}\Sigma_{rx}\Sigma_x^{-1/2}, \quad (26)$$

as the adjusted cross-correlation matrix between the transformed variables \tilde{r} and \tilde{x} .¹⁰

The second step is to perform an SVD operation on the adjusted cross-covariance matrix $\Sigma_{\tilde{r}\tilde{x}}$. Let $((s_1, \dots, s_N); (u_1, \dots, u_N); (v_1, \dots, v_N))$ denote a system of singular values and singular vectors of the adjusted cross-covariance matrix $\Sigma_{\tilde{r}\tilde{x}}$. We assume that the singular values s_i are sorted in increasing order. The canonical correlations correspond to the singular values. The canonical variates of r and x are defined as $u'_1\tilde{r}, \dots, u'_N\tilde{r}$ and $v'_1\tilde{x}, \dots, v'_N\tilde{x}$, respectively. As the singular values, s_i are sorted in increasing order, the canonical variate with the largest correlation is given by the pair $(u'_N\tilde{r}, v'_N\tilde{x})$ while the canonical variate with the smallest correlation is $(u'_1\tilde{r}, v'_1\tilde{x})$.

Finally, since the solutions are expressed on a different basis than our original problem, we have to translate them back to our original basis. By inverting the change of variables we made, the i th canonical variates can be written as $(u'_i\Sigma_r^{-1/2}r, v'_i\Sigma_x^{-1/2}x)$. This pair can be seen as a linear combination of the original assets and signals with coefficients given by $q_{r,i} := \Sigma_r^{-1/2}u_i$ and $q_{x,i} := \Sigma_x^{-1/2}v_i$, which are the so-called canonical directions. The application of change of basis operations simplifies the covariance structure considerably.

In order to implement the CCA in practice, a common approach is to replace the population second-moments Σ_r, Σ_x , and Σ_{rx} , with their sample counterparts S_r, S_x , and S_{rx} computed from random samples r_2, \dots, r_{T+1} and x_1, \dots, x_T . Let $((\hat{s}_1, \dots, \hat{s}_N); (\hat{u}_1, \dots, \hat{u}_N); (\hat{v}_1, \dots, \hat{v}_N))$ be a system of singular values (sorted in increasing order) and singular vectors of the sample analog of (26) denoted by $S_{\tilde{r}\tilde{x}}$. Then the sample canonical variates of r and x are given by $\hat{u}'_1\tilde{r}, \dots, \hat{u}'_N\tilde{r}$ and $\hat{v}'_1\tilde{x}, \dots, \hat{v}'_N\tilde{x}$, respectively. However, one should be skeptical of sample canonical correlations as they are not reflective of the population canonical correlations. Indeed, the true canonical correlation of the i th canonical variate pair $(\hat{u}'_i\tilde{r}, \hat{v}'_i\tilde{x})$ is $\hat{v}'_i\Sigma'_{\tilde{r}\tilde{x}}\hat{u}_i$, as opposed to $\hat{s}_i = \hat{v}'_iS'_{\tilde{r}\tilde{x}}\hat{u}_i$. This is because the sample canonical correlations are known to be inconsistent estimates of their population counterparts when the dimensionality of the problem for both variables is large relative to the number of observations.

To bring this point home, let us consider a specific case when the two variable sets r and x are independent and each set consists of i.i.d. Gaussian random variables. Figure 1 shows the distribution of sample canonical correlations for various values of the limiting ratios $N/T \in \{0.003, 0.03, 0.3\}$ and $M/T \in \{0.006, 0.06, 0.6\}$. We obtained this figure by invoking the asymptotic formula derived by Wachter (1980, Theorem 3.1).¹¹ We observe that the sample canonical correlations are biased

¹⁰Note that the matrix (26) is different from the *unadjusted* cross-covariance matrix Σ_{rx} . The adjusted cross-covariance matrix is the central object in CCA as it internalizes the variabilities from the asset returns and signals that might potentially obscure the relationships between both of the variables. Moreover, since the optimal matrix of coefficients for our problem (5) is indeed the adjusted cross-covariance matrix under an appropriate change of basis (see Section 3.3), it makes sense for us to focus on this object for our analysis.

¹¹Wachter's result also holds under a less restrictive setting of i.i.d. data with finite second-moments; see, for example, Yang and Pan (2012, Theorem 1).

upwards away from zero and the severity of the bias worsens as N/T or M/T or both increases. This distributional shape is expected to hold when N , M and T are approximately large and it only depends on the ratio N/T and M/T ; this is true regardless of any particular realization of the adjusted sample cross-covariance matrix. We attempt to address these problems with regularization in Section 3.6.

3.2 Relation to PCA, PLS, and RRR

Having laid the groundwork for CCA, it is interesting to see how CCA is related to a host of linear dimension-reduction techniques commonly used to uncover underlying economic structures. We shall restrict our attention to the principal component analysis (PCA), partial least squares (PLS), and reduced rank regression (RRR) due to their tight connection with one another and extensive applications in the finance and econometrics literature.¹²

Starting with PCA developed by [Hotelling \(1933\)](#), the method seeks to break down the variation of the data into mutually orthogonal components whose variances can be ranked from the smallest to the largest. In the context of a paired dataset consisting of asset returns with their associated signals, PCA focuses on searching for directional vectors that have the maximum variation in each variable separately. In contrast to CCA, however, it does not take into account any relationship between the two financial variables. In this paper, we would like to focus on how financial variables are related, not how much they vary. We summarize the key properties distinguishing PCA and CCA in Table 1.

On the other hand, PLS was introduced by [Wold \(1975\)](#) as an econometric technique that aims to explain the relationship of a paired dataset by studying the (cross-)covariance. This can be understood by relaxing the variance constraints in the optimization problem of CCA, which results in a maximum (cross-)covariance problem. The solution(s) can be obtained via singular value decomposition of the cross-covariance matrix, Σ_{rx} , and the obtained managed portfolios closely mimic that of [Kelly et al. \(2022\)](#). This differs from CCA where the cross-covariance matrix is normalized with respect to the covariances from both the r_{t+1} and the x_t variables.

Finally, the RRR proposed by [Anderson and Rubin \(1949, 1950\)](#) and [Anderson \(1951\)](#) can be understood through the regression framework (13) with the following error minimization problem:

$$\min_B \mathbb{E} \left[\left\| \Gamma^{1/2} (r_{t+1} - B' x_t) \right\|_2^2 \right], \quad (27)$$

¹²For example, PCA has been used to extract common factors from a covariance matrix of individual stocks among others by [Roll and Ross \(1980\)](#), [Connor and Korajczyk \(1986, 1988\)](#), and [Lehmann and Modest \(1988\)](#). In the context of describing risk compensation through estimation of the stochastic discount factor, [Kozak et al. \(2018, 2020\)](#) invokes no near-arbitrage arguments for the use of PCA to identify factor risk exposures that price the cross-section of expected returns of characteristic-managed portfolios. For PLS, one of the earliest applications to finance can be ascribed to [Kelly and Pruitt \(2013\)](#), where the authors used a three-pass regression filter (which is a special case of PLS) to forecast market return and cash flow growth. The RRR, alongside CCA, features prominently in the estimation of the celebrated vector-error-correction model (VECM) of [Engle and Granger \(1987\)](#) as a means to detect cointegration, which are economic relationships that cannot be obtained through standard regression approaches; see [Johansen \(1988, 1991, 1995\)](#).

where Γ is a $N \times N$ weighting matrix, and the matrix B is subject to a rank constraint; see also, [Velu and Reinsel \(2013\)](#). Expressing the objective function in terms of the unconditional moments, we have

$$\mathbb{E}[\|\Gamma^{1/2}(r_{t+1} - B'x_t)\|_2^2] = \mathbb{E}[\|\Gamma^{1/2}r_{t+1}\|_2^2] - 2\mathbb{E}[x_t' B \Gamma r_{t+1}] + \mathbb{E}[x_t' B \Gamma B' x_t] \quad (28)$$

$$= N - 2\text{Tr}(B \Gamma \Sigma_{rx}) + \text{Tr}(B \Gamma B' \Sigma_x) \quad (29)$$

$$= \|\Gamma^{1/2} B \Sigma_x^{1/2} - \Gamma^{1/2} \Sigma_{rx} \Sigma_x^{-1/2}\|_F^2 + \text{constant}. \quad (30)$$

The solution to this rank-restricted minimization problem follows from the theorem of [Eckart and Young \(1936\)](#) and it relies on the singular value decomposition of $\Gamma^{1/2} \Sigma_{rx} \Sigma_x^{-1/2}$. The typical objective function of RRR occurs when $\Gamma = \mathbb{I}_N$. If additionally, $\Sigma_x = \mathbb{I}_M$, then RRR coincides with PLS. Finally, observe that if $\Gamma = \mathbb{I}_N$ and $r_{t+1} = x_t$, then the objective function coincides with that of PCA.

Interestingly, if $\Gamma = \Sigma_r^{-1}$, then the solution of RRR generalizes to CCA. This speaks to the difference between optimizing a regression-based objective and a correlation-based objective. Indeed, regression attempts to explain as much of the variation in the returns r_{t+1} as possible. However, if r_{t+1} is decorrelated and isotropic along all variance directions, then minimizing a regression loss function becomes equivalent to maximizing a correlation objective. Unlike conventional regression-based approaches, there is no notion of independent and dependent variables in CCA because the correlation metric is symmetric; that is, if we exchange the roles of asset returns and signals, CCA yields identical outcomes for the canonical directions and canonical correlations.

In summary, RRR, PLS, and CCA can all be viewed under the same framework, which is solving a cross-covariance maximization problem. Each of these linear dimensional reduction techniques is subject to different normalization schemes with CCA serving as the most general one that nests these methods.

3.3 Reformulation as Canonical Portfolios

The portfolio returns r_{t+1}^w at time $t+1$ obtained with (6) is not particularly intuitive as the expression involves large-dimensional matrix inversion and multiplication operations. However, we can make progress by performing a CCA in order to decompose the portfolio selection problem into one that we can provide financial interpretation.

We start by expressing the portfolio returns in terms of their transformed objects as

$$x_t' A r_{t+1} = \frac{1}{\gamma} x_t' \Sigma_x^{-1} \Sigma_{rx}' \Sigma_r^{-1} r_{t+1} = \frac{1}{\gamma} \tilde{x}_t' \Sigma_{\tilde{r}\tilde{x}} \tilde{r}_{t+1}, \quad (31)$$

Since the decorrelated objects \tilde{r}_{t+1} and \tilde{x}_t span the same universe as the original assets and signals, we shall term them “synthetic assets” and “synthetic signals”. These newly defined objects have identity covariances, and so this is the cross-sectional version of risk parity.

From expression (31), we can see that the portfolio returns depend intricately on the synthetic

asset returns and synthetic signals, which are coupled through their adjusted cross-covariances. The portfolio returns do not depend on whether we exchange the role of the asset returns or signals; what matters is how they are related to each other.

Our next step is to choose basis vectors such that $\Sigma_{\tilde{r}\tilde{x}}$ is diagonal. In particular, we perform an SVD operation such that all cross-relationships are eliminated in the new basis in order to arrive at the following simplified expression for the portfolio returns:

$$x_t' A r_{t+1} = \frac{1}{\gamma} \sum_{i=1}^N s_i (v_i' \tilde{x}_t) (u_i' \tilde{r}_{t+1}). \quad (32)$$

Thus, we see that a generic strategy return can be viewed either as a combination of the original assets and signals that have been optimally blended with the matrix of coefficients A , or as a combination of N uncorrelated managed long-short portfolios weighted by their canonical correlations. The N long-short managed portfolios span the same space of investment opportunities as the original assets and signals.

Economically, we interpret the change of basis as an operation that reorganizes the set of N synthetic assets and N synthetic signals into a set of N uncorrelated managed portfolios. Each of these managed portfolios is expressed as a certain weighted combination of the assets and signals themselves. These particular combinations are determined by the singular vectors or canonical directions; in either case, we shall refer to them as *canonical portfolios* henceforth to reinforce the idea that these vectors are essentially managed portfolios that are loading onto their respective asset and signal variables. The higher the i th canonical correlation, the higher the return of the i th canonical portfolio. Said differently, the canonical portfolios are ordered in such a way that the highest canonical correlation corresponds to the *most linearly predictable portfolio*, and the second highest canonical correlation, the *second most linearly predictable portfolio*, and so on. This notion of predictability distinguishes canonical portfolios from other concepts of orthogonal managed portfolios such as those constructed by PCA.

Circling back to Equation (32), the source of the portfolio returns can be seen to be distributed across the canonical portfolios of $\Sigma_{\tilde{r}\tilde{x}}$. This reinforces the economic notion of diversification: Having exposure to different uncorrelated canonical portfolios is akin to putting each of your eggs into different baskets. We see that the capital assigned to each canonical portfolio is proportional to its original returns r_{t+1} and signal x_t , and proportional to the correlation s_i . Put differently, we want to assign capital to strategies that have high correlations but are also orthogonal to each other.

3.4 Canonical Portfolio Analysis

With our CCA decomposition, we can also gain some insight into how the canonical portfolios impact the returns of a portfolio. We start by working out the expected returns and variance of these canonical portfolios with the next proposition.

Proposition 4. Let $\pi_{ti} := s_i (v_i' \tilde{x}_t) (u_i' \tilde{r}_{t+1})$ be the return that the i th canonical portfolio generates.

The expected value and variance return of each canonical portfolio are

$$\mathbb{E}[\pi_{ti}] = s_i^2, \quad \text{and} \quad \text{Var}[\pi_{ti}] = s_i^2(1 + s_i^2). \quad (33)$$

Furthermore, the expected value and variance of the optimal portfolio is

$$\mathbb{E}[x'_t A r_{t+1}] = \frac{1}{\gamma} \sum_{i=1}^N s_i^2, \quad \text{and} \quad \text{Var}[x'_t A r_{t+1}] = \frac{1}{\gamma} \sum_{i=1}^N s_i^2. \quad (34)$$

We see that the expected portfolio returns are positive, and hence, the optimal portfolio is a profit-generating strategy that has economic value. Moreover, the expected portfolio returns can be expressed as the sum of the contributions from the squared canonical correlations; the larger the expected return of the i th canonical portfolio, the larger its contribution towards the overall returns.¹³ This suggests that the squared canonical correlations can serve as a natural measure to rank the performance of the canonical portfolios.

Interestingly, the canonical correlations are also intimately connected to the Sharpe ratio of the returns associated with our portfolio policy.

Corollary 1. The Sharpe ratio of the i th canonical portfolio and the optimal portfolio is given by

$$\text{SR}_i := \frac{\mathbb{E}[\pi_{ti}]}{\sqrt{\text{Var}[\pi_{ti}]}} = \frac{s_i}{\sqrt{1 + s_i^2}}, \quad \text{and} \quad \text{SR} := \frac{\mathbb{E}[x'_t A r_{t+1}]}{\sqrt{\text{Var}[x'_t A r_{t+1}]}} = \sqrt{\sum_{i=1}^N s_i^2}. \quad (35)$$

From Corollary 1, we see the squared canonical correlations also influence the portfolio's Sharpe ratio. We end this section with two simple examples to give further intuition into these results.

Example 4 (Equal canonical correlations). If we suppose $s_i = s$ for all i , then Equation (35) simplifies to $\text{SR} = s\sqrt{N}$. In this special case, we arrive at the Fundamental Law of Active Management of Grinold (1989), where the so-called ‘‘information coefficient’’ is s and the effective number of statistically independent investments (also referred to as ‘breadth’) is N . We see that the Sharpe ratio improves with the average canonical correlation and the number of canonical portfolios, with the latter exhibiting diminishing gains to returns due to the square-root scaling in its exponent. In the general case where the canonical correlations are not identical across all canonical portfolios, the canonical correlations are thought to be the information coefficient over clusters of signals.

Example 5 (Bias of In-Sample Returns). Let $\hat{A} := \gamma^{-1} S_x^{-1} S'_{rx} S_r^{-1}$ is the optimal matrix of coefficients that replaces the population moments in A with sample moment-based estimates. For

¹³The formulas (33) and (34) have some close similarities to a result in Kelly et al. (2022, Proposition 4) for the expected returns. One notable difference, however, is that our singular values are raised to the power of two. This is implicit in the fact that the expected return of each canonical portfolio is scaled by a leverage factor of s_i . On the other hand, this multiplicative factor is absent in their expression since leverage is explicitly controlled for in their problem formulation.

simplicity, suppose that $S_r = \Sigma_r$ and $S_x = \Sigma_x$. Since the Frobenius norm is a convex function, by Jensen's inequality, the expected portfolio return satisfies

$$\mathbb{E}[\text{Tr}(\hat{A}S_{rx})] = \frac{1}{\gamma}\mathbb{E}[\text{Tr}(S'_{\hat{r}\hat{x}}S_{\hat{r}\hat{x}})] \geq \frac{1}{\gamma}\text{Tr}(\mathbb{E}[S_{\hat{r}\hat{x}}]\mathbb{E}[S'_{\hat{r}\hat{x}}]) = \frac{1}{\gamma}\mathbb{E}[\text{Tr}(S'_{\hat{r}\hat{x}}\Sigma_{\hat{r}\hat{x}})], \quad (36)$$

where we used the fact that $\mathbb{E}[S_{rx}] = \Sigma_{rx}$ in the last equality due to the unbiasedness of the sample cross-covariance. Hence,

$$\sum_{i=1}^N \mathbb{E}[\hat{s}_i^2] \geq \sum_{i=1}^N s_i^2 = \sum_{i=1}^N \mathbb{E}[\hat{s}_i s_i^\circ], \quad (37)$$

where $s_i^\circ := \hat{q}'_{r,i}\Sigma_{rx}\hat{q}_{x,i}$ is the out-of-sample 'canonical correlation' associated with the i th estimated canonical portfolio $\hat{q}_{r,i}$ and $\hat{q}_{x,i}$.¹⁴ Hence, on average, the in-sample returns are always optimistic but the out-of-sample evaluation disappoints. This is because both the in-sample singular values and singular vectors are estimated with a bias. Hence, in order to ensure that the in-sample and out-of-sample returns are more in sync, we have to shrink the in-sample singular values and align singular vectors closer to the truth.

3.5 Static and Dynamic Returns Decomposition

In our standing assumptions, we supposed that both the asset returns and signals are centered and have zero mean. However, when evaluating the out-of-sample performance of the portfolio, the original non-demeaned returns and signals tend to be used instead. Moreover, the means of both financial variables may be non-negligible and contain useful information for describing the cross-predictability of future returns. In general, we can write the second-moment matrix of conditional portfolios using the following decomposition:

$$\Sigma_{rx} = \mathbb{E}[r_{t+1}x'_t] = \mathbb{E}[r_{t+1}]\mathbb{E}[x_t]' + \text{Cov}(r_{t+1}, x_t). \quad (38)$$

The first component is a rank-one outer product matrix of the unconditional means in the asset returns and signals while the second component is the cross-covariance matrix of the demeaned returns and signals. In particular, the first component places emphasis on the cross-sectional differences driven by the unconditional levels since all time-series variabilities have been averaged out, while the second component captures time-series variations in the returns and signals that are expressed as the deviations from their unconditional means. [Kelly et al. \(2022\)](#) describe the first component as static bets and the second component as dynamic bets.

We can derive the expected returns of the optimal portfolio in this setting in the next proposition.

Proposition 5. If the expected value of the returns and signals are $\mathbb{E}[r_{t+1}] = \mu_r$ and $\mathbb{E}[x_t] = \mu_x$,

¹⁴In fact, a more precise relationship between the in-sample and out-of-sample portfolio returns has been established by [Benaych-Georges et al. \(2019, Proposition 2.9\)](#).

then the expected portfolio return utilizing the second-moment matrix (38) is given by

$$\mathbb{E}[x'_t A r_{t+1}] = \frac{1}{\gamma} (\mu'_r \Sigma_r^{-1} \mu_r) (\mu'_x \Sigma_x^{-1} \mu_x) + \frac{1}{\gamma} \sum_{i=1}^N s_i^2. \quad (39)$$

Proposition 5 shows that there is an additional non-negative contribution to the portfolio returns, which is due to the (squared) maximum Sharpe ratio achievable from the assets. This observation aligns nicely with our interpretation that the sum of the squared canonical correlations is related to the squared Sharpe ratio of optimal portfolio returns from Corollary 1. That is, the expected portfolio returns can be attributed to the Sharpe ratios from both static and dynamic bets.

We shall let our portfolio exploit both investment opportunities that arise from the second-moment matrix of managed portfolio returns. This is possible within our framework since the assumption of zero-mean returns and signals strictly applies to the variance of the portfolio returns, and so only the covariances of the returns and signals have to be centered. However, we shall apply CCA to the cross-covariance matrix to fulfill its modeling assumptions since otherwise, it can result in a generic top canonical portfolio that is mainly driven by the unconditional levels.¹⁵

3.6 Estimation

Our optimal portfolio policy requires the knowledge of the population covariances of the asset returns, the covariances of the signals, and cross-covariance between both variables. These objects are generally unknown to us, and so in order to render our framework to practice, we have to estimate them with real data. Given the large-dimensional nature of our problem, it is necessary to regularize the covariances of our portfolio to reduce the estimation errors. Unfortunately, simply maintaining the top canonical portfolio is not sufficient for empirical analysis. Indeed, if either the assets or the signals has rank T (which tends to be the case when $N \gg T$), then the canonical portfolios can take on any arbitrary values. Moreover, since the smallest amount of variability in each dataset gets rescaled to one, CCA can produce spurious outcomes.

The challenge of estimating the covariance matrix of financial covariances is well known amongst practitioners (Jobson and Korkie, 1980). A standard approach is to use the sample covariance matrix. However, when the dimensionality of the problem is large relative to the number of observations, estimation error of the sample covariance matrix can create issues for portfolio optimizers; they tend to place extreme bets on low-risk sample eigenvectors. In fact, this observation led Michaud (1989) to refer to mean-variance optimizers as “error maximization” schemes. There have been several approaches from practice to address this problem using methods from bootstrapping (Michaud and Michaud, 2008) to Bayesian estimators (Black and Litterman, 1992; Lai et al., 2011).

We obtain regularized covariance matrices by applying the linear shrinkage technology from Ledoit and Wolf (2004b). The covariances of returns will be estimated as $\hat{\Sigma}_r = (1 - \delta_r) S_r + \delta_r N^{-1} \text{Tr}(S_r) \mathbb{I}_N$, where the shrinkage intensity δ_r is determined based on an asymptotic formula. We also apply

¹⁵CCA is sometimes applied to non-demeaned variables. In such a case, this is more closely related to using a cosine similarity objective instead of a correlation objective.

the same linear shrinkage technology for the covariances of signals $\hat{\Sigma}_x$ with a shrinkage intensity parameter δ_x . The choice of linear shrinkage of the covariances has a nice interpretation in our context in that varying the shrinkage intensities allows us to interpolate between the maximum covariance (PLS) and maximum correlation (CCA) problems. Shrinking towards maximum covariance helps to break the singularities by considering managed canonical portfolios that have better out-of-sample properties. In the extreme regularization setting with both $\hat{\Sigma}_r$ and $\hat{\Sigma}_x$ being identity matrices, our estimated canonical portfolios mimics the principal portfolios approach of Kelly et al. (2022).

With both of these estimated covariances at hand, we proceed to build a regularized adjusted cross-covariance matrix by pre-conditioning the sample covariances S_{rx} on both sides with the estimated matrix factors $\hat{\Sigma}_r^{1/2}$ and $\hat{\Sigma}_x^{1/2}$. One can then choose to maintain the top few canonical correlations of the regularized adjusted sample cross-covariance matrix defined as

$$\hat{\Sigma}_r^{-1/2} S_{rx} \hat{\Sigma}_x^{-1/2}, \quad (40)$$

and set the bottom ones to zero.¹⁶ This thresholding operation has the effect of regularizing the problem as it reduces the effective number of parameters that we have to estimate by maintaining the top few canonical portfolios that are the most predictable.

4 Empirical Analysis

4.1 Data and Portfolio Construction Rules

For our empirical analysis, we download six datasets from Kenneth French’s data library, which are characteristic-sorted long-short stock portfolios.¹⁷ They are daily returns on portfolios of stocks sorted on the basis of size and book-to-market (FF), size and operating profitability (ME/OP), and size and investment (ME/INV), each of which is of universe size 25 and 100. In comparison to individual stocks, each return of a portfolio is an average return of a group of stocks sharing similar characteristics and so they are less subject to large variabilities due to idiosyncratic risks. Hence, they serve as useful test assets that may allow us to more easily harvest the predictability in stock returns.

Although some of these portfolio returns have been available since 1926, we conduct most of our analysis on the period from July 1963 to October 2022, for which most of the returns are available. For simplicity, we suppose that 21 consecutive days constitute one trading ‘month’ and 252 consecutive days as one trading ‘year’. We adopt a sequential updating scheme and rebalance the portfolio every ‘month’ on a rolling walk-forward basis. To obtain a well-defined investment universe for which we can estimate the portfolios, we use the following rule. For each rebalancing

¹⁶The regularized sample adjusted cross-covariance matrix (40) has a close similarity to that of Vinod (1976), where ridge regression was proposed as a means of regularizing the sample adjusted cross-covariance matrix. We resort to the class of linear shrinkage estimators due to its ability to also reduce the influence of large variance directions.

¹⁷The description of all portfolio construction can be found on Kenneth French’s website: https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html. At the time of writing, these datasets were based on the 10-2022 CRSP database.

date, we select test assets that have a complete return history over the most recent $T = 120$ months as well as a complete return ‘future’ over the subsequent trading month. This provides us with 578 monthly out-of-sample returns, which covers the out-of-sample investment period from 09-09-1974 through 10-21-2022.

4.2 Signal Construction

As a demonstration of our method to a dynamic portfolio selection problem between equity sorted portfolios, we start with a single conditioning variable as our signal. We consider the momentum signal as there is substantial empirical evidence that documents this anomaly in the returns of individual stocks (Jegadeesh, 1990; Jegadeesh and Titman, 1993), industries (Moskowitz and Grinblatt, 1999), and of size and value portfolios (Lewellen et al., 2010). This application is similar to Kelly et al. (2022) for reproducibility.

To construct a momentum signal, we compute for each asset the lagged one-month return defined as the simple average return over the previous 21 trading days. Similar to Asness et al. (2019), Freyberger et al. (2020), Kozak et al. (2020) and Kelly et al. (2022), we rank the momentum signals across the assets from 1 to N , dividing the ranks by the number of assets, and then centering the normalized ranks to map the signals into the range $[-0.5, 0.5]$. This provides us with a set of dollar-neutral signals that are insensitive to outliers for which we further divide by the sum of their absolute values. This keeps the gross exposure (that is, the sum of the absolute amount of long and short positions) fixed since otherwise doubling the number of assets at any time t will result in signals that are two times more aggressive even though the investment opportunities remain the same. We assume \$1 of capital is invested to \$1 of long and short positions.

We collect the individual momentum values of the N assets to yield a predictive signal x_t for the subsequent monthly returns. Additionally, we impose a one-day buffer between the constructed signals and the subsequent returns to limit the effects of illiquidity from driving our results and to bring our backtest simulations closer to being tradeable practice.¹⁸

4.3 Candidate Portfolios

Given the time series panels of asset returns and signals, we consider the following portfolios in our study:

- **CP2**: Our proposed portfolio construction methodology based on Equation (7).
- **MVO**: The mean-variance optimization portfolio of Markowitz (1952).
- **PP2**: The Principal Portfolios of Kelly et al. (2022).

¹⁸The challenge working with the equity portfolios from Kenneth French’s website is that they contain small illiquid stocks. Moreover, given that we are using daily returns, which are close-to-close returns, asynchronous trading at the end of the day may arise; see, for example, Lo and MacKinlay (1990). Consequently, this may induce some lead-lag relationships among the stocks, which can lead to autocorrelation in the portfolio returns and spurious correlation estimates due to the Epps (1979) effect. The latter problem can be particularly acute given our use of multivariate techniques in the estimation of portfolios. Therefore, we employ a one-day buffer along with monthly returns to mitigate the effects that asynchronous or infrequent trading can have on the portfolio returns.

- **UNI**: The univariate factor where the weights are the signals.

The competing portfolios have been chosen because they can be subsumed in our proposed method and hence, serve as natural benchmarks for us to determine where the contribution to any improved performance comes from. The suffix number attached to the portfolio labels CP and PP indicates the number of managed portfolios that we retain; for example, CP2 means that we choose to keep the leading two most predictable managed portfolios.¹⁹

The covariance of returns in CP2 and MVO is estimated with the linear shrinkage of [Ledoit and Wolf \(2004b\)](#). Additionally, we also apply linear shrinkage to the covariance of signals for CP2 but choose a high shrinkage intensity with value $\delta_x = 0.9$ without relying on the asymptotic formula from [Ledoit and Wolf \(2004b\)](#) for this purpose since it was developed for financial returns that are assumed to be independent and identically distributed.

Finally, we renormalize the estimated portfolios so that the sum of the absolute value of their weights equals one. This allows all the portfolios to be comparable in scale. It also implies that the gross exposure for all portfolios is one dollar by construction, that is, we apply one dollar of capital for one dollar of long and short positions. This is sensible for long-short equity hedge fund managers who face institutional constraints such as limits on gross exposure by their prime brokerage.²⁰

4.4 Evaluation Methodology

To evaluate the performance of the different portfolios, we report three main out-of-sample performance measures: the average cumulative out-of-sample returns, the standard deviation of the out-of-sample returns, and the Sharpe ratio defined as the ratio of the average returns to the standard deviation of returns. For ease of interpretability, all performance measures are annualized with 12 trading ‘months’. The Sharpe ratio is computed with respect to the actual returns (as opposed to returns in excess of the risk-free rate) since we believe it is more relevant in our context where the portfolios are formed solely on the basis of risky assets.

We also report three additional performance measures based on the out-of-sample returns in excess of a 6-factor benchmark; that is, the 5 factors from [Fama and French \(2015\)](#) augmented with out-of-sample returns from UNI. We compute the Jensen’s alpha, beta to the out-of-sample UNI returns, and information ratio from a 6-factor regression model. This is done for all portfolios except for UNI. The alpha and information ratio are annualized with 12 trading ‘months’. We also provide the t-statistics of the Sharpe ratio and information ratio, which are computed with approximate standard errors from [Lo \(2002\)](#).

Additionally, we report the following portfolio weight statistics averaged over the 578 trading months: turnover, proportional leverage (computed as the fraction of negative weights), the sum of negative weights, and the minimum and maximum weight. Note that these statistics are not

¹⁹We cross-sectionally demean the returns in the construction of the cross-covariance matrix for PP as suggested in [Kelly et al. \(2022\)](#) to focus on the cross-sectional differences. However, we choose to ignore otherwise for the other portfolios.

²⁰The insights are similar if we renormalize the portfolios to achieve a target level of volatility or return.

our primary focus since our proposed method is not optimized to account for these measures. Nevertheless, they are provided to give a better overview of the different methods.

4.5 Application of Canonical Portfolio Analysis

We can use the results of Proposition 4 to identify the sources of portfolio returns by estimating the canonical portfolios that contribute most to its profitability. To this end, we will consider the FF25, ME/OP25, and ME/INV25 test assets for this purpose. The FF25 is widely studied amongst academics and it allows us to check if our method produces the expected results. Moreover, we know from Lewellen et al. (2010) that the size and value portfolio returns have a strong factor structure explained mostly by the three-factor model Fama and French (1993). Kozak et al. (2018) finds that retaining the first three principal components extracted from FF25, closely reproduces the Fama and French three-factor model. Thus, this observation could potentially be exploited in our method.

The left panels of Figure 3 show the squared sample canonical correlations of the demeaned conditional portfolios matrix for the three test assets. It is also overlaid with the squared canonical correlations generated from randomly permuting the signals observations for each asset and repeating a similar exercise.²¹ The squared canonical correlations are averaged over the rebalancing dates and ordered from the smallest to the largest. From this in-sample analysis, we can see that the leading two squared sample canonical correlations have values larger than their pseudo-random generated counterparts and that the top one ‘sticks out’ and extends beyond the value of one.²² Moreover, the spacing between the sample canonical correlations is wider at the top end of the spectrum and more uniform at the bottom end of the spectrum.

We contrast these findings with the right panels of Figure 3, which shows the out-of-sample returns of each canonical portfolio computed as the product of the in-sample and out-of-sample canonical correlation, accompanied by their standard error. Not surprisingly, we find a performance deterioration in the out-of-sample performance due to the bias in the in-sample predictions. Nevertheless, there is some coherence between the in-sample predictions and the out-of-sample returns in that the top canonical portfolio possesses the most realized returns followed by the second one, while the bottom ones are close to zero.

Given the prominence of the leading canonical correlation, we plot its corresponding weights. The left panels of Figure 4 show the top canonical portfolios (that is, the canonical directions) averaged over the rebalancing dates. To ensure the signs of the canonical portfolios are consistent across time, we flip the sign of the canonical portfolios at any given rebalancing date if its cosine similarity with the canonical portfolios obtained from the previous rebalancing date is negative. While there is not any clear pattern of trades that we can immediately discern, it nonetheless differs from other research findings vis-à-vis PCA; for example, Kelly et al. (2022) finds that the top principal component of a

²¹Randomly shuffling the time series helps generate a null distribution. It is a useful heuristic to determine the importance of each canonical correlation relative to a random benchmark compared to a formal statistical test of significance from Yang and Pan (2015).

²²The values of (squared) regularized canonical correlations can exceed one unlike its sample-based analog, which is constrained to the interval (0,1).

symmetrized cross-covariance matrix goes long (short) on big (small) equity portfolios, and long (short) on value (growth) portfolios. The right panels of Figure 4 provide the final weights invested in each asset.

4.6 Empirical Results

Table 2 summarizes the performance of the various portfolio methods for the different test assets. Restricting our attention to the FF25 column, we see that CP2 has an average return, which is lower than that of PP2 (2.20% versus 3.76%) but has much lower volatility (2.18% versus 5.85%).²³ Altogether, this translates into Sharpe ratio of 1.01 which is a 50% improvement over the closest competitor, PP2, at 0.66. The alpha of the CP2 is 1.84% and has a low beta to the univariate factor return of 0.15. Adjusting the alpha by the idiosyncratic volatility of 2.04% gives an information ratio of 0.9 for CP2. A similar conclusion holds for the ME/OP25 and ME/INV25 test assets, although we observe a deterioration in the Sharpe ratio for all portfolios in larger-sized test asset possibly due to estimation errors. CP2 also underperforms PP2 in the ME/INV100 dataset.

Table 3 describes the distribution of the portfolio weights of the estimated portfolios. In the FF25 column, we see that CP2 has the lowest turnover. This is interesting given that we make no effort to control the trajectory of the weights. The average sum of negative weights in the CP2 is -0.496 and the average proportional leverage in CP2 is less than 0.5 indicating a slight tilt towards long positions. The weights are CP2 appear to be the most dispersed but do but are relatively not extreme. We draw a similar conclusion for the other test assets.

It would be interesting to investigate where the performance improvement of CP2 comes from: Is better at selecting assets that have historically performed well on average or at varying the positions in the assets dynamically? One way to discern between the two possible explanations is to decompose the portfolio returns into the following components:

$$\mathbb{E}[w_t' r_{t+1}] = \underbrace{\mathbb{E}[w_t]' \mathbb{E}[r_{t+1}]}_{\text{static}} + \underbrace{\text{Cov}(r_{t+1}, w_t)}_{\text{dynamic}}. \quad (41)$$

Intuitively, static bets refer to the investor’s ability to get the long-run allocations right. On the other hand, the covariance between the signal and subsequent returns refers to an investor’s ability to tactically ‘time’ the market, this is, to accurately identify movements in the assets and gain exposure to those assets accordingly.

Table 4 reports the results of this decomposition, where we estimate the static and dynamic components of returns according to Equation (41) with their corresponding sample analogs. Panels A and B show that both the static and dynamic components contribute positively to the overall performance of the portfolio. Panel C of Table 4 breaks down the share of total returns due to taking dynamic bets for CP2 and P2. It shows that on average the returns of CP2 and PP2 come from

²³Note that the magnitudes of the returns for all portfolios appear to be low in comparison to those published in hedge fund return indices. This is expected since the gross exposure of our portfolios are all constrained to one. In practice, one would typically apply a leverage factor greater than one in order to magnify the returns.

taking dynamic bets. Therefore, the performance of our proposed method does not come from its ability to long (short) more highly performing (underperforming) assets; it must emanate from its ability to tactically time the market.

Finally, we also isolate the portfolio returns due to the long and short legs of the trade. We write $r_t^w = l((r_t^w)_+ - (r_t^w)_-)$ where $(r_t^w)_+$ is the return on the long leg and $(r_t^w)_-$ is the return on the short leg, with weights of both trade legs normalized to sum to one. Here, l reflects the leverage of the long-short portfolio, but since our estimated portfolios have a unit gross exposure by construction, l is close to one. Panel D and E of Table 4 present the statistics of the long and short legs, respectively. For CP2, the average return on the long leg is 8.01% and the short leg is 5.71%. This indicates that the profits of our portfolio come from the long side of the trades.

4.7 Robustness Checks

In this section, we inspect whether the outperformance of our proposed portfolio construction methodology is robust to different revisions in the current empirical set-up. In particular, we will be interested in results based on (1) subsample period, (2) forecast horizon (3) shrinkage in the signal covariance, (4) momentum lookback window, and (5) ‘approximate’ versus ‘actual’ portfolio policy.

4.7.1 Sub-Period Analysis

In this section, we check if there are any peculiar subsample effects that may drive the performances of our proposed scheme. We divide the out-of-sample period into four roughly equally-sized subsamples of 144 months (that is, 12 trading years) each: (1) 1986-09-25 to 1986, (2) 1986 to 1998, (3) 1998 to 2010, and (4) 2010 to 2022. Then we perform the same procedure in each subsample. The results are provided in Table 5.

Generally, the performance of all portfolios appears stronger in the earlier periods of the sample but poorer in the recent decade. We see that the outperformance of CP2 over the competing portfolios is consistent over time for FF25 and ME/INV25. However, CP2 underperforms PP2 in (1) ME/INV100 for most of the subsamples, and (2) FF100, ME/OP25, and ME/OP100 in the earlier subsample.

4.7.2 Forecast Horizon

To make a forecast of subsequent returns, we have used a horizon length of 21 trading days, which roughly corresponds to one month. Given our use of non-overlapping observations between the signals and subsequent returns, this implies that the data are sampled on a monthly basis. which also corresponds to the frequency with which we rebalance our portfolios. We now change the forecast horizon from 21 days to 1, 5, and 10 days. Each of these forecast horizons covers an investment period from (1) 01-06-1965 to 10-21-2022, (2) 12-06-1966 to 10-21-2022, and (3) 06-12-1969 to 10-21-2022; this provides us with 14550 daily, 2814 weekly, and 1347 fortnightly out-of-sample returns, respectively.

Table 6 shows that the annualized Sharpe ratio of all portfolios tends to be better at shorter holding periods. Barring the potential side effects of illiquidity in daily returns, this suggests that the conditioning information becomes more relevant as the holding period decreases. This makes sense since the conditional portfolios become more reactive to changes in the states of the market. Overall, the ranking of the methods remains similar relative to Table 2 with the exception of the 1-day horizon, where CP2 is the best performer for all test assets.

4.7.3 Shrinkage Intensity

We examine the effect of shrinkage on the covariance of signals in our proposed method. Our default choice in the analysis was $\delta_x = 0.9$. The annualized Sharpe ratios for CP2 corresponding to shrinkage values of $\delta_x \in \{0, 0.1, 0.5, 0.8, 1\}$ are presented in Table 7 over different subsamples.

We see that there is no specific level of shrinkage that provides consistent outperformance for all test assets. This indicates that the shrinkage intensity δ_x is time-varying in nature. Shrinkage values greater than or equal to 0.5 appears to work well across different subsamples for larger-sized equity portfolios up until the recent decade, where there is some benefit of using more sample information from the signal correlations in datasets ME/OP100 and ME/INV100.

4.7.4 Momentum Lookback Window

The 21-day momentum signal that we used as a base case, which is perceived to be a relatively ‘fast’ signal in that it reacts quickly to changes in market conditions but is also one that is susceptible to noise. We now consider using momentum signals computed with a longer lookback window of sizes 42 days, 63 days, 84 days, 126 days, and 252 days. The remaining details remain similar.

Table 8 demonstrates that CP2 generally outperforms the other portfolios in terms of annualized Sharpe ratio for different momentum signals and different test assets. Although we observe some performance degradation for all portfolios as we increase the lookback size from 21 days to 126 days, the ranking of the methods remains similar to Table 2. There is, however, a significant performance gain in CP2 not seen in the other portfolios as we extend the momentum lookback size from 126 days to 252 days. This is comforting since momentum signals with longer lookback horizons tend to have lower turnover and allow portfolio managers to target higher overall dollar capacity to absorb large trades; for example, [Novy-Marx and Velikov \(2016\)](#) finds that a fund’s capacity decreases as turnover increases.

4.7.5 Approximate Versus Actual Solution

We have worked with the ‘approximated’ mean-variance problem from Proposition 2 throughout this paper since it helped us to simplify the analysis. Given that we also have a closed-form expression to the ‘full’ mean-variance problem from Theorem 2 in Proposition 1, we now check if it offers any practical benefits over our approximate solution. The only difference between both solutions is in terms of how the canonical correlations enter into the reconstruction of the optimal portfolio policy.

The full solution essentially nonlinearly adjusts the canonical correlations, while no adjustment takes place in the approximated solution. Figure 2 shows that both solutions are similar for small canonical correlation values but for large canonical correlation values, the conservative behavior of the optimal strategy is reinforced by downweighting its influence. This makes sense because we are taking into account more terms that affect the risk profile of the strategy.

From Table 9, we see that our approximated formula generally performs better than the full formula across different test assets. One reason for this observation is that the nonlinear adjustment of the canonical correlations may be too conservative relative to the unadjusted one. For example, a canonical correlation value of one gets reduced by half through the nonlinear adjustment. This holds mechanically irrespective of the data. Consequently, this behavior may inadvertently under-leverage the leading two canonical portfolios, which are the most profitable streams.

4.8 Extension: Two-Signal Case

Our canonical portfolios modeling framework is flexible enough to accommodate multiple signals. In this empirical exercise, we expand our signal vector to include two momentum signals of different lookback windows; one with a 21-day lookback window, and another with a 252-days lookback window. This expanded $2N$ -dimensional signal vector will serve as input for both CP and PP. On the other hand, we assume that both MVO and UNI take in an equal-weighted average of the two momentum signals as inputs since these methods do not have an ‘optimal’ way of blending different signals together in a single stage. The rest of the empirical setup remains unchanged.

From Table 10, we see that the Sharpe ratio of all portfolios, with the exception of MVO, generally improves in comparison to the one signal case. More importantly, the outperformance of CP2 over competing methods continues to hold up for different test assets.

Turning our attention to Table 11, we see that the turnover is reduced for all portfolios compared to the base case. This observation can be attributed to the inclusion of the signal with a long lookback window, which tends to have a lower turnover than one with a short lookback window. This is appealing since we can expect the performance after factoring in transaction costs to be better than using a single conditioning variable. We also observe lower proportional leverage throughout the portfolios indicating a tilt toward long positions. The weights also appear to be less dispersed than in the base case.

5 Conclusion

In this paper, we provide a novel framework for portfolio managers and academics alike to conceptualize the optimal asset and signal combination problem with canonical correlation analysis (CCA). Our contribution can be summarized as follows. First, we recast the original investment problem of Brandt and Santa-Clara (2006) into a tractable one that scales to large dimensions. Our solution can be seen to generalize the classic Markowitz (1952) mean-variance optimization scheme, such that it is able to ingest multiple return-predictive signals, and account for cross-predictability and uncertainty

in the signals. All of these properties are achieved by solving the portfolio selection problem in a single stage and assuming Gaussianity on the joint distribution of the assets and signals. Second, we lift the veil of complexity from our large-dimensional problem through a novel application of CCA. In particular, CCA breaks down the correlations of all of the asset returns and signals into independent long–short managed portfolios, which we term as *canonical portfolios*. Each of the canonical portfolios can be ranked from the one with the smallest correlation to the one with the highest; the canonical portfolios with the highest predictable returns get scaled up the most.

Having established the theoretical results of our method, we bring it to the empirical test. We ran backtest simulations on Fama-French sorted equity portfolios with a momentum signal, and we have found that our proposed method generally outperforms natural benchmarks. The performance of our method further improves when the analysis is extended to two momentum signals of different lookback windows. These results are made possible by introducing regularization techniques to overcome the instabilities in the estimation and exploiting the most predictable dimensions of the data.

In terms of future work, our proposed modeling framework is not set in stone and is flexible enough to accommodate further improvements. Some research ideas include introducing nonlinearities into the modeling process, relaxing the Gaussianity assumption on the joint return distribution, as well as designing signals and portfolios with maximal mutual information are promising avenues. Additionally, one may consider incorporating constraints on the portfolio weights such as transaction costs which will be pursued in a forthcoming work. Another interesting avenue to explore would be to apply these ideas to develop a test for asset pricing models or to form portfolios from different asset classes such as individual stocks portfolios, fixed-income portfolios, currency portfolios, and so forth. Last but not least, recasting the portfolio selection problem into a CCA framework enables researchers to leverage the insights and techniques from the rich literature of CCA that has been expanded by developments in machine learning.

References

- Anderson, T. W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *The Annals of Mathematical Statistics*, pages 327–351.
- Anderson, T. W. and Rubin, H. (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *The Annals of Mathematical Statistics*, 20(1):46–63.
- Anderson, T. W. and Rubin, H. (1950). The asymptotic properties of estimates of the parameters of a single equation in a complete system of stochastic equations. *The Annals of Mathematical Statistics*, pages 570–582.
- Asness, C. S., Frazzini, A., and Pedersen, L. H. (2019). Quality minus junk. *Review of Accounting Studies*, 24(1):34–112.
- Avellaneda, M. and Lee, J.-H. (2010). Statistical arbitrage in the US equities market. *Quantitative Finance*, 10(7):761–782.
- Bach, F. R. and Jordan, M. I. (2002). Kernel independent component analysis. *Journal of Machine Learning Research*, 3:1–48.
- Benaych-Georges, F., Bouchaud, J.-P., and Potters, M. (2019). Optimal cleaning for singular values of cross-covariance matrices. *arXiv preprint arXiv:1901.05543*.
- Black, F. and Litterman, R. (1992). Global portfolio optimization. *Financial Analysts Journal*, 48(5):28–43.
- Brandt, M. W. (1999). Estimating portfolio and consumption choice: A conditional Euler equations approach. *The Journal of Finance*, 54(5):1609–1645.
- Brandt, M. W. and Santa-Clara, P. (2006). Dynamic portfolio selection by augmenting the asset space. *The Journal of Finance*, 61(5):2187–2217.
- Brandt, M. W., Santa-Clara, P., and Valkanov, R. (2009). Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns. *The Review of Financial Studies*, 22(9):3411–3447.
- Cochrane, J. H. (1996). A cross-sectional test of an investment-based asset pricing model. *Journal of Political Economy*, 104(3):572–621.
- Connor, G. and Korajczyk, R. A. (1986). Performance measurement with the arbitrage pricing theory: A new framework for analysis. *Journal of financial economics*, 15(3):373–394.
- Connor, G. and Korajczyk, R. A. (1988). Risk and return in an equilibrium APT: Application of a new test methodology. *Journal of Financial Economics*, 21(2):255–289.

- Eckart, C. and Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3):211–218.
- Engle, R. F. and Granger, C. W. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica*, 55(2):251–276.
- Epps, T. W. (1979). Comovements in stock prices in the very short run. *Journal of the American Statistical Association*, 74(366a):291–298.
- Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56.
- Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1–22.
- Ferson, W. E. and Siegel, A. F. (2001). The efficient use of conditioning information in portfolios. *The Journal of Finance*, 56(3):967–982.
- Firoozye, N. and Koshiyama, A. (2019). Avoiding backtesting overfitting by covariance-penalties: An empirical investigation of the ordinary and total least squares cases. *The Journal of Financial Data Science*, 1(4):63–83.
- Firoozye, N. and Koshiyama, A. (2020). Optimal dynamic strategies on Gaussian returns. *Journal of Investment Strategies*, 9(1):23–53.
- Freyberger, J., Neuhierl, A., and Weber, M. (2020). Dissecting characteristics nonparametrically. *The Review of Financial Studies*, 33(5):2326–2377.
- Frost, P. A. and Savarino, J. E. (1986). An empirical Bayes approach to efficient portfolio selection. *Journal of Financial and Quantitative Analysis*, 21(3):293–305.
- Golub, G. H. and Van Loan, C. F. (1980). An analysis of the total least squares problem. *SIAM Journal on Numerical Analysis*, 17(6):883–893.
- Grinold, R. C. (1989). The fundamental law of active management. *The Journal of Portfolio Management*, 15(3):30–37.
- Haldane, J. (1942). Moments of the distributions of powers and products of normal variates. *Biometrika*, 32(3/4):226–242.
- Hansen, L. and Richard, S. F. (1987). The role of conditioning information in deducing testable. *Econometrica*, 55(3):587–613.
- Healy, M. (1957). A rotation method for computing canonical correlations. *Mathematics of Computation*, 11(58):83–86.

- Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6):417.
- Hotelling, H. (1936). Relations between two sets of variates. *Biometrika*, 28(3/4):321–377.
- Huang, C.-f. and Litzenberger, R. H. (1988). *Foundations for Financial Economics*. North-Holland.
- Isserlis, L. (1918). On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika*, 12(1/2):134–139.
- Jegadeesh, N. (1990). Evidence of predictable behavior of security returns. *The Journal of Finance*, 45(3):881–898.
- Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, 48(1):65–91.
- Jobson, J. D. and Korkie, B. (1980). Estimation for Markowitz efficient portfolios. *Journal of the American Statistical Association*, 75(371):544–554.
- Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, 12(2-3):231–254.
- Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica*, 59(6):1551–1580.
- Johansen, S. (1995). *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*. Oxford: Oxford University Press.
- Kelly, B., Malamud, S., and Pedersen, L. H. (2022). Principal portfolios. *The Journal of Finance*.
- Kelly, B. and Pruitt, S. (2013). Market expectations in the cross-section of present values. *The Journal of Finance*, 68(5):1721–1756.
- Kozak, S., Nagel, S., and Santosh, S. (2018). Interpreting factor models. *The Journal of Finance*, 73(3):1183–1223.
- Kozak, S., Nagel, S., and Santosh, S. (2020). Shrinking the cross-section. *Journal of Financial Economics*, 135(2):271–292.
- Lai, T. L., Xing, H., and Chen, Z. (2011). Mean–variance portfolio optimization when means and covariances are unknown. *The Annals of Applied Statistics*, 5(2A):798–823.
- Ledoit, O. and Wolf, M. (2004a). Honey, I shrunk the sample covariance matrix. *The Journal of Portfolio Management*, 30(4):110–119.
- Ledoit, O. and Wolf, M. (2004b). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2):365–411.

- Ledoit, O. and Wolf, M. (2017). Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets Goldilocks. *The Review of Financial Studies*, 30(12):4349–4388.
- Lehmann, B. N. and Modest, D. M. (1988). The empirical foundations of the arbitrage pricing theory. *Journal of Financial Economics*, 21(2):213–254.
- Levy, H. and Markowitz, H. M. (1979). Approximating expected utility by a function of mean and variance. *The American Economic Review*, pages 308–317.
- Lewellen, J., Nagel, S., and Shanken, J. (2010). A skeptical appraisal of asset pricing tests. *Journal of Financial Economics*, 96(2):175–194.
- Lo, A. W. (2002). The statistics of Sharpe ratios. *Financial Analysts Journal*, 58(4):36–52.
- Lo, A. W. and MacKinlay, A. C. (1990). An econometric analysis of nonsynchronous trading. *Journal of Econometrics*, 45(1-2):181–211.
- Lütkepohl, H. (1997). Handbook of matrices. *Computational Statistics and Data Analysis*, 2(25):243.
- Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7:77–91.
- Markowitz, H. M. (1991). Foundations of portfolio theory. *The Journal of Finance*, 46(2):469–477.
- Meucci, A. (2009). Managing diversification. *Risk*, pages 74–79.
- Michaud, R. (1989). The Markowitz optimization enigma: Is optimized optimal? *Financial Analysts Journal*, 45:31–42.
- Michaud, R. O. and Michaud, R. O. (2008). *Efficient Asset Management: A Practical Guide to Stock Portfolio Optimization and Asset Allocation*. Oxford University Press.
- Moskowitz, T. J. and Grinblatt, M. (1999). Do industries explain momentum? *The Journal of Finance*, 54(4):1249–1290.
- Novy-Marx, R. and Velikov, M. (2016). A taxonomy of anomalies and their trading costs. *The Review of Financial Studies*, 29(1):104–147.
- Partovi, M. H. and Caputo, M. (2004). Principal portfolios: Recasting the efficient frontier. *Economics Bulletin*, 7(3):1–10.
- Roll, R. and Ross, S. A. (1980). An empirical investigation of the arbitrage pricing theory. *The Journal of Finance*, 35(5):1073–1103.
- Uurtio, V., Monteiro, J. M., Kandola, J., Shawe-Taylor, J., Fernandez-Reyes, D., and Rousu, J. (2017). A tutorial on canonical correlation methods. *ACM Computing Surveys (CSUR)*, 50(6):1–33.
- Velu, R. and Reinsel, G. C. (2013). *Multivariate reduced-rank regression: Theory and applications*, volume 136. Springer Science & Business Media.

- Vinod, H. D. (1976). Canonical ridge and econometrics of joint production. *Journal of Econometrics*, 4(2):147–166.
- Wachter, K. W. (1980). The limiting empirical measure of multiple discriminant ratios. *The Annals of Statistics*, pages 937–957.
- Wick, G. (1950). The evaluation of the collision matrix. *Physical Review*, 80(2):268–272.
- Wold, H. (1975). Soft modelling by latent variables: The non-linear iterative partial least squares (NIPALS) approach. *Journal of Applied Probability*, 12(S1):117–142.
- Yang, Y. and Pan, G. (2012). The convergence of the empirical distribution of canonical correlation coefficients. *Electronic Journal of Probability*, 17:1–13.
- Yang, Y. and Pan, G. (2015). Independence test for high dimensional data based on regularized canonical correlation coefficients. *The Annals of Statistics*, 43(2):467–500.

A Figures and Tables

Figure 1: Wachter Law

This figure displays the limiting spectral density of the sample canonical correlations of Wachter (1980) for different ratios N/T and M/T . The population cross-covariance matrix is the zero matrix, and so all the population canonical correlation coefficients are equal to zero.

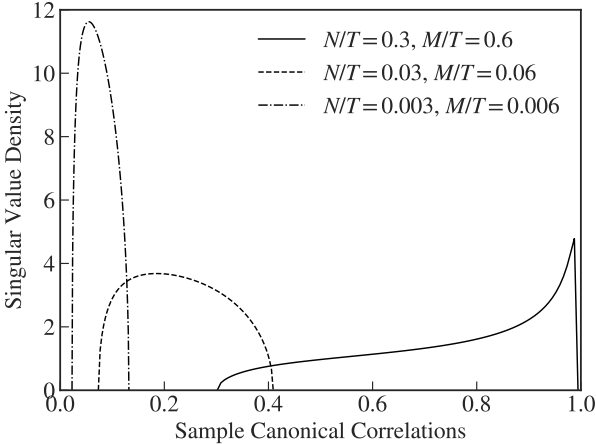


Figure 2: Optimal singular value adjustment versus linear approximation

This figure displays the optimal adjustment of the i th singular value versus an unadjusted i th singular value.

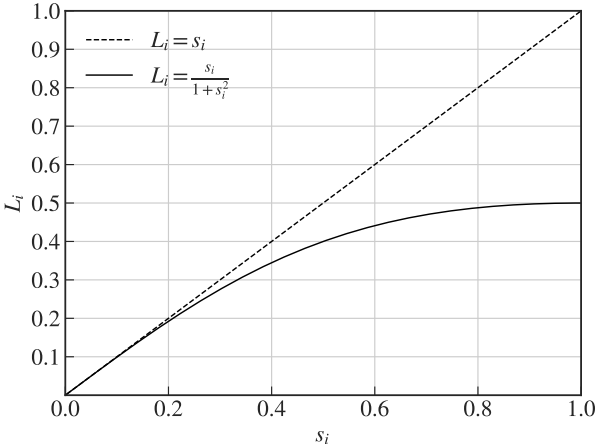
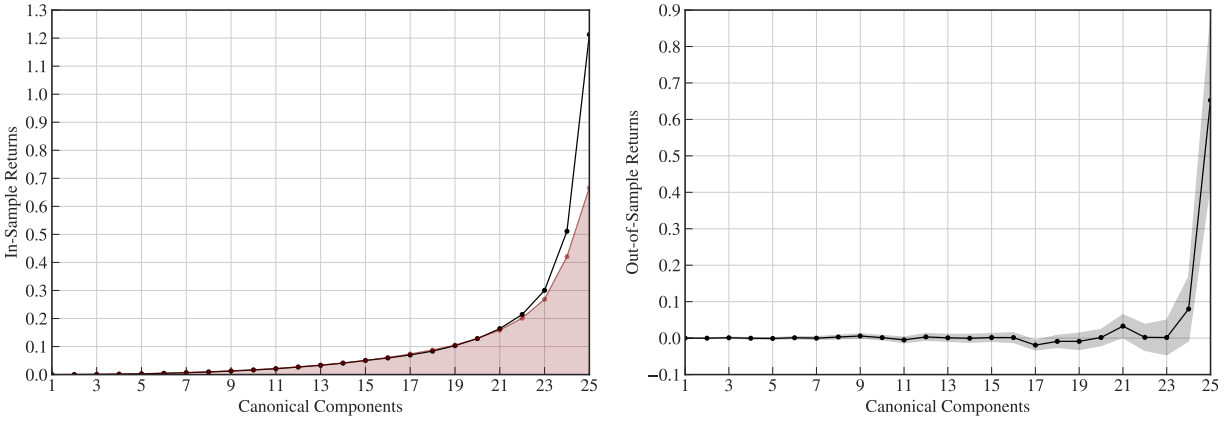


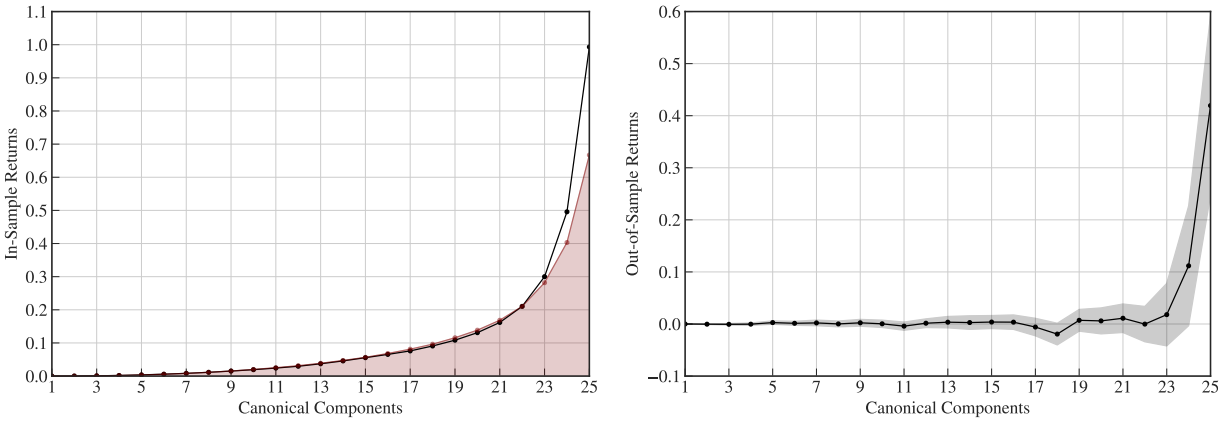
Figure 3: In-Sample and Out-of-Sample Canonical Portfolio Returns

This figure displays the in-sample and out-of-sample returns for each canonical portfolio. The left panel shows the squared in-sample regularized canonical correlations coefficients (black dots) overlaid with their pseudo-random generated counterparts (maroon dots). The right panel shows the out-of-sample returns of each canonical portfolio overlaid with a ± 2 standard error band. Each dot corresponds to values that are averaged over the rebalancing dates. The estimates are obtained by using a 21-day momentum signal to predict subsequent monthly returns and 120 non-overlapping monthly returns. The test assets are the FF25 (top panel), ME/OP25 (middle panel), and MEINV25 (bottom panel). The out-of-sample period is covered from 09-09-1974 until 10-21-2022.

(a) 25 Size and Book-to-Market



(b) 25 Size and Operating Profitability



(c) 25 Size and Investment

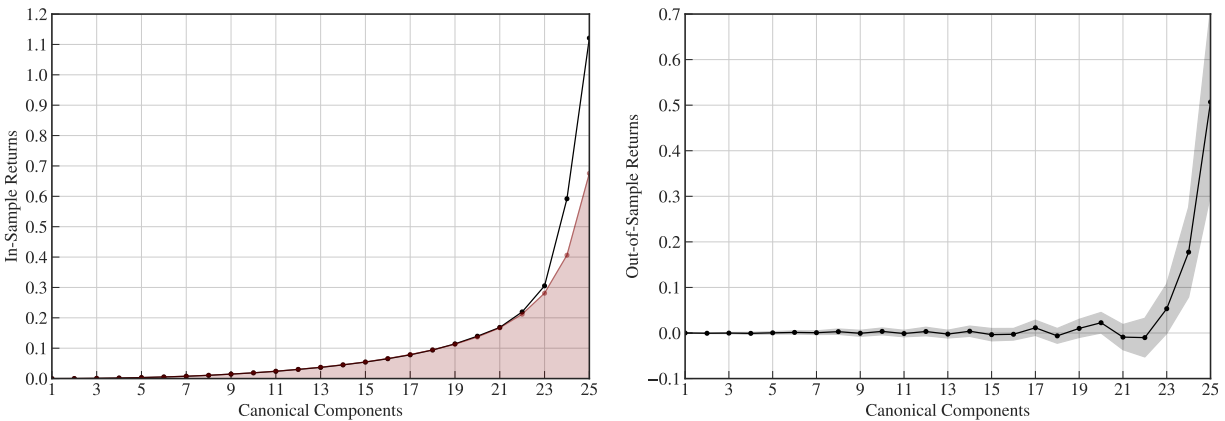
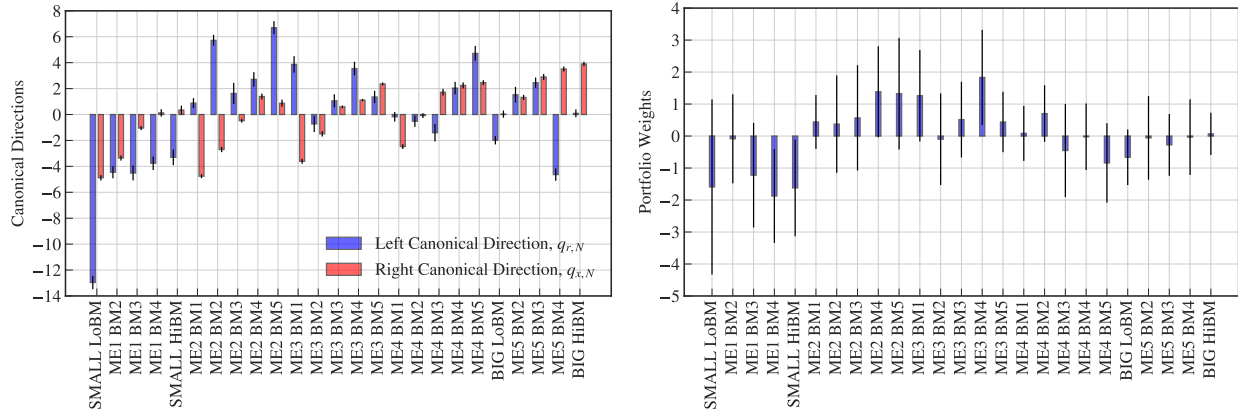


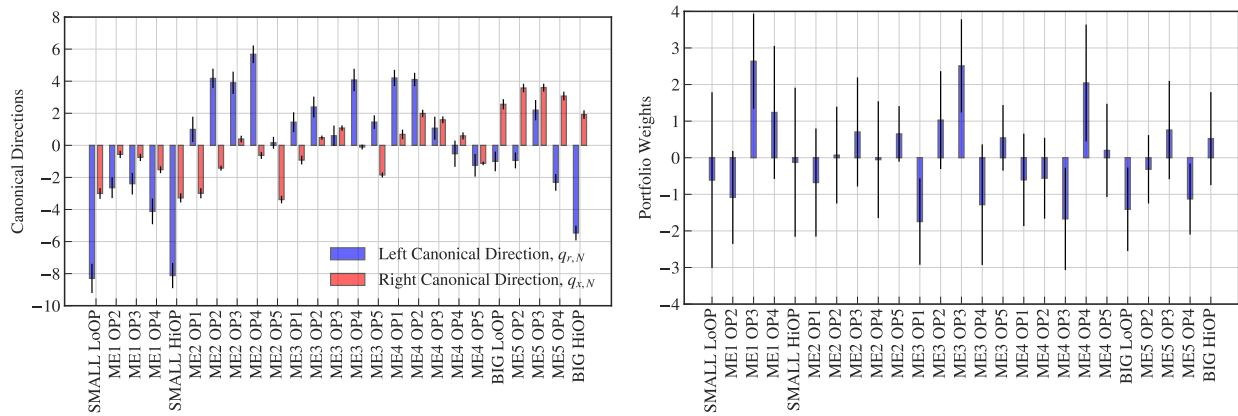
Figure 4: Weights for the Top Canonical Portfolio

This figure displays the weights of the leading canonical portfolio. The left panel shows the canonical ‘weights’ on the returns (blue) and signals (red). The right panel shows the final portfolio weights invested in each obtained from combining the returns and signals together through their canonical weights and scaling them up with the corresponding canonical correlation. The height of each bar corresponds to values that are averaged over the rebalancing dates, and they are overlaid with a ± 2 standard error band. The estimates are obtained by using a 21-day momentum signal to predict subsequent monthly returns and 120 non-overlapping monthly returns. The test assets are the FF25 (top panel), ME/OP25 (middle panel), and MEINV25 (bottom panel). The out-of-sample period is covered from 09-09-1974 until 10-21-2022.

(a) 25 Size and Book-to-Market



(b) 25 Size and Operating Profitability



(c) 25 Size and Investment

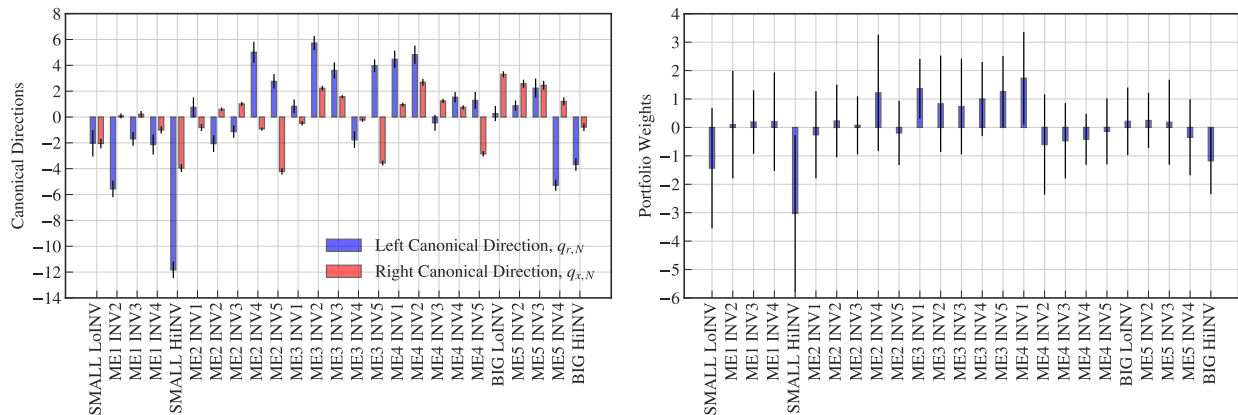


Table 1: Summary of PCA and CCA

This table summarizes the key distinguishing features of PCA and CCA applied to a paired dataset of asset returns r_{t+1} and signals x_t . For CCA (PCA), it displays (1) the correlation (variance) maximization objective, (2) the equivalent error minimization objective, (3) the i th canonical (principal) direction, and (4) the i th correlation (variance) of the corresponding canonical (principal) direction.

PCA	(1) Maximize Variance	$\max_{\substack{q_r, q_x \\ \ q_r\ _2 = \ q_x\ _2 = 1}} \text{Var}(q_r' r_{t+1}) + \text{Var}(q_x' x_t)$
	(2) Minimize Error	$\min_{\ q_r\ _2 = \ q_x\ _2 = 1} \mathbb{E}[\ r_{t+1} - q_r q_r' r_{t+1}\ _2^2] + \mathbb{E}[\ x_t - q_x q_x' x_t\ _2^2]$
	(3) Principal Direction	$q_{r,i}, q_{x,i} = \underset{\substack{\ q_r\ _2 = \ q_x\ _2 = 1 \\ q_r', q_{r,j} = 0, j=1, \dots, i-1 \\ q_x', q_{x,j} = 0, j=1, \dots, i-1}}{\text{argmax}} \text{Var}(q_r' r_{t+1}) + \text{Var}(q_x' x_t)$
	(4) Principal Variance	$d_{r,i} = \text{Var}(q_{r,i}' r_{t+1}), \quad d_{x,i} = \text{Var}(q_{x,i}' x_t)$
CCA	(1) Maximize Correlation	$\max_{\text{Var}(q_r' r_{t+1}) = \text{Var}(q_x' x_t) = 1} \text{Cov}(q_r' r_{t+1}, q_x' x_t)$
	(2) Minimize Error	$\min_{\text{Var}(q_r' r_{t+1}) = \text{Var}(q_x' x_t) = 1} \mathbb{E}[\ q_r' r_{t+1} - q_x' x_t\ _2^2]$
	(3) Canonical Direction	$q_{r,i}, q_{x,i} = \underset{\substack{q_r' \Sigma_r q_r = q_x' \Sigma_x q_x = 1 \\ q_r', q_{r,j} = 0, j=1, \dots, i-1 \\ q_x', q_{x,j} = 0, j=1, \dots, i-1}}{\text{argmax}} \text{Cov}(q_r' r_{t+1}, q_x' x_t)$
	(4) Canonical Correlation	$s_i = \frac{\text{Cov}(q_{r,i}' r_{t+1}, q_{x,i}' x_t)}{\sqrt{\text{Var}(q_{r,i}' r_{t+1})} \sqrt{\text{Var}(q_{x,i}' x_t)}}$

Table 2: Performance Summary

This table shows the out-of-sample performances (in percentages) of various portfolio selection methods using a 21-day momentum signal. The test assets are the 25 and 100 equity portfolios independently sorted on size and three other factors (book-to-market, operating profitability, and investment). Panels A-G display the average returns, the Sharpe ratio of returns, the alpha with respect to the Fama-French 5-factor regression augmented with the univariate factor (UNI), the beta of the simple factor, the idiosyncratic volatility, and the information ratio. The numbers in parenthesis are the t-statistics for the portfolio's Sharpe ratio and information ratio using the standard error of [Lo \(2002\)](#). The covariances are estimated on a rolling walk-forward basis with $T = 120$ non-overlapping monthly returns. All measures are computed with 579 monthly portfolio weights from 09-09-1974 until 10-21-2022.

	FF25	FF100	ME/OP 25	ME/OP 100	ME/INV 25	ME/INV 100
Panel A. Average Returns (%)						
CP2	2.196	1.334	1.741	0.953	1.809	0.782
PP2	3.762	3.462	3.806	3.490	3.377	3.432
MVO	0.381	0.188	0.358	-0.097	0.365	0.097
UNI	2.268	1.787	2.261	1.477	2.141	1.473
Panel B. Standard Deviation of Returns (%)						
CP2	2.175	1.366	2.048	1.260	2.057	1.248
PP2	5.853	5.690	4.912	4.931	4.642	4.433
MVO	1.857	1.366	1.685	1.301	1.595	1.261
UNI	5.280	4.814	4.173	3.663	4.384	3.970
Panel C. Sharpe Ratio						
CP2	1.010 (6.872)	0.977 (6.655)	0.854 (5.837)	0.759 (5.208)	0.883 (6.028)	0.631 (4.341)
PP2	0.662 (4.554)	0.628 (4.325)	0.786 (5.389)	0.722 (4.954)	0.740 (5.077)	0.785 (5.377)
MVO	0.214 (1.485)	0.145 (1.004)	0.220 (1.528)	-0.068 (-0.473)	0.236 (1.639)	0.083 (0.577)
UNI	0.452 (3.125)	0.393 (2.719)	0.557 (3.843)	0.419 (2.899)	0.506 (3.493)	0.389 (2.690)
Panel D. Alpha (%)						
CP2	1.844	1.189	1.116	0.823	1.340	0.683
PP2	2.252	2.240	2.183	2.347	2.069	2.565
MVO	-0.210	-0.097	-0.342	-0.391	-0.208	-0.214
Panel E. Beta to Simple Factor						
CP2	0.146	0.072	0.259	0.086	0.172	0.054
PP2	0.684	0.695	0.703	0.734	0.583	0.583
MVO	0.256	0.168	0.300	0.201	0.260	0.188
Panel F. Idiosyncratic Volatility (%)						
CP2	2.041	1.321	1.749	1.219	1.910	1.228
PP2	4.583	4.583	3.951	4.157	3.889	3.778
MVO	1.281	1.102	1.136	1.077	1.119	1.018
Panel G. Information Ratio						
CP2	0.903 (6.170)	0.900 (6.149)	0.638 (4.393)	0.675 (4.642)	0.701 (4.819)	0.556 (3.835)
PP2	0.491 (3.396)	0.489 (3.379)	0.553 (3.810)	0.565 (3.892)	0.532 (3.671)	0.679 (4.668)
MVO	-0.164 (-1.137)	-0.088 (-0.609)	-0.301 (-2.087)	-0.363 (-2.512)	-0.186 (-1.292)	-0.210 (-1.456)

Table 3: Weight Statistics

This table shows the average monthly weight statistics of various portfolio selection methods estimated using a 21-day momentum signal. The test assets are the 25 and 100 equity portfolios independently sorted on size and three other factors (book-to-market, operating profitability, and investment). Panels A-F display the turnover, the proportional leverage, the sum of negative weights in the portfolio, and the minimum and maximum portfolio weights. The covariances are estimated on a rolling walk-forward basis with $T = 120$ non-overlapping monthly returns. All measures are computed with 579 monthly portfolio weights from 09-09-1974 until 10-21-2022.

	FF25	FF100	ME/OP 25	ME/OP 100	ME/INV 25	ME/INV 100
Panel A. Turnover						
CP2	1.158	1.202	1.179	1.225	1.165	1.217
PP2	1.149	1.148	1.167	1.154	1.173	1.162
MVO	1.377	1.424	1.378	1.416	1.381	1.428
UNI	1.252	1.282	1.261	1.295	1.263	1.298
Panel B. Proportional Leverage						
CP2	0.493	0.497	0.499	0.494	0.498	0.495
PP2	0.496	0.497	0.493	0.497	0.500	0.502
MVO	0.499	0.499	0.504	0.497	0.502	0.500
UNI	0.489	0.500	0.489	0.500	0.489	0.500
Panel C. Sum of Negative Weights						
CP2	-0.496	-0.497	-0.498	-0.498	-0.497	-0.498
PP2	-0.500	-0.500	-0.500	-0.500	-0.500	-0.500
MVO	-0.500	-0.500	-0.501	-0.501	-0.500	-0.500
UNI	-0.500	-0.500	-0.500	-0.500	-0.500	-0.500
Panel D. Minimum Weight						
CP2	-0.106	-0.033	-0.106	-0.033	-0.103	-0.032
PP2	-0.102	-0.034	-0.105	-0.037	-0.104	-0.033
MVO	-0.097	-0.030	-0.094	-0.030	-0.094	-0.029
UNI	-0.077	-0.020	-0.077	-0.020	-0.077	-0.020
Panel E. Maximum Weight						
CP2	0.104	0.033	0.103	0.032	0.103	0.032
PP2	0.099	0.033	0.100	0.036	0.103	0.033
MVO	0.098	0.030	0.095	0.029	0.096	0.030
UNI	0.077	0.020	0.077	0.020	0.077	0.020
Panel F. Standard Deviation of Weights						
CP2	0.051	0.013	0.051	0.013	0.051	0.013
PP2	0.051	0.013	0.052	0.013	0.052	0.013
MVO	0.050	0.012	0.050	0.012	0.050	0.012
UNI	0.047	0.012	0.047	0.012	0.047	0.012

Table 4: Portfolio Returns Decomposition

This table shows the average returns (in percentages) of various portfolio selection methods due to static/dynamic trades and long/short trades. Panels A-C display the results for the dynamic and static trades as well as the share of returns due to dynamic trades. Panels D-E display the results for the long and short trades. The signal is a 21-day momentum. The test assets are the 25 and 100 equity portfolios independently sorted on size and three other factors (book-to-market, operating profitability, and investment). The covariances are estimated on a rolling walk-forward basis with $T = 120$ non-overlapping monthly returns. All measures are computed with 579 monthly portfolio weights from 09-09-1974 until 10-21-2022.

	FF25	FF100	ME/OP 25	ME/OP 100	ME/INV 25	ME/INV 100
Panel A. Static Returns (%)						
CP2	0.184	0.204	0.182	0.159	0.178	0.138
PP2	0.043	0.032	0.209	0.165	0.081	0.026
MVO	0.118	0.159	0.084	0.103	0.142	0.153
UNI	0.168	0.170	0.190	0.166	0.157	0.158
Panel B. Dynamic Returns (%)						
CP2	2.008	1.127	1.561	0.791	1.640	0.657
PP2	3.718	3.429	3.614	3.349	3.302	3.426
MVO	0.263	0.029	0.268	-0.191	0.217	-0.067
UNI	2.097	1.614	2.059	1.303	1.988	1.316
Panel C. Share of Dynamic Returns (%)						
CP2	91.466	84.510	89.383	83.179	90.064	82.603
PP2	98.802	99.069	94.401	95.229	97.558	99.237
MVO	68.917	15.496	75.939	217.015	60.439	-77.498
UNI	92.467	90.326	91.399	88.562	92.652	89.249
Panel D. Long Leg Returns (%)						
CP2	8.012	7.621	7.666	7.505	8.021	7.725
PP2	8.893	8.532	8.664	8.506	8.716	8.924
MVO	7.506	7.306	7.248	7.197	7.570	7.457
UNI	8.057	7.916	7.940	7.822	8.255	8.015
Panel E. Short Leg Returns (%)						
CP2	5.714	6.297	5.785	6.522	6.011	6.848
PP2	4.764	4.747	4.543	4.697	5.081	5.240
MVO	6.939	6.979	6.637	7.152	7.001	7.224
UNI	5.476	5.861	5.442	6.163	5.866	6.347

Table 5: Subsample Analysis

This table shows the annualized out-of-sample Sharpe ratio of various portfolio selection methods over different subsamples. Panels A-D display the results for Sample 1 (1974 through 1986), Sample 2 (1986 through 1998), Sample 3 (1998 through 2010), and Sample 4 (2010 through 2022). The signal is a 21-day momentum. The test assets are the 25 and 100 equity portfolios independently sorted on size and three other factors (book-to-market, operating profitability, and investment). The numbers in parenthesis are the t-statistics for the portfolio's Sharpe ratio and information ratio using the standard error of Lo (2002). The covariances are estimated on a rolling walk-forward basis with $T = 120$ non-overlapping monthly returns.

	FF25	FF100	ME/OP 25	ME/OP 100	ME/INV 25	ME/INV 100
Panel A. Sharpe Ratio, Sample Period 1974 to 1986						
CP2	1.254 (4.208)	0.842 (2.874)	1.094 (3.685)	0.401 (1.378)	1.324 (4.411)	0.632 (2.163)
PP2	1.107 (3.741)	1.182 (3.979)	1.264 (4.224)	1.086 (3.660)	1.184 (3.973)	1.222 (4.093)
MVO	0.016 (0.056)	-0.152 (-0.527)	0.358 (1.232)	-0.168 (-0.580)	0.412 (1.419)	0.205 (0.707)
UNI	0.935 (3.181)	0.762 (2.607)	1.130 (3.802)	0.814 (2.773)	1.020 (3.446)	0.838 (2.850)
Panel B. Sharpe Ratio, Sample Period 1986 to 1998						
CP2	1.549 (5.117)	1.577 (5.199)	1.232 (4.124)	1.146 (3.854)	1.224 (4.100)	1.069 (3.604)
PP2	1.051 (3.560)	1.185 (3.989)	1.269 (4.239)	1.159 (3.892)	1.222 (4.094)	1.216 (4.074)
MVO	0.589 (2.025)	0.283 (0.977)	0.344 (1.185)	-0.125 (-0.431)	0.470 (1.616)	0.039 (0.134)
UNI	0.806 (2.755)	0.720 (2.466)	0.652 (2.230)	0.405 (1.393)	0.623 (2.134)	0.380 (1.307)
Panel C. Sharpe Ratio, Sample Period 1998 to 2010						
CP2	0.883 (3.010)	1.263 (4.237)	0.909 (3.086)	1.126 (3.788)	1.002 (3.390)	0.894 (3.034)
PP2	0.408 (1.407)	0.389 (1.342)	0.807 (2.747)	1.011 (3.418)	0.693 (2.369)	0.790 (2.692)
MVO	0.283 (0.977)	0.351 (1.214)	0.201 (0.693)	-0.109 (-0.378)	0.161 (0.557)	0.121 (0.416)
UNI	0.330 (1.140)	0.329 (1.137)	0.467 (1.606)	0.431 (1.483)	0.399 (1.373)	0.334 (1.151)
Panel D. Sharpe Ratio, Sample Period 2010 to 2022						
CP2	0.448 (1.539)	0.389 (1.339)	0.238 (0.819)	0.322 (1.105)	0.117 (0.401)	-0.174 (-0.599)
PP2	0.350 (1.205)	0.149 (0.515)	0.010 (0.034)	-0.213 (-0.731)	-0.136 (-0.468)	-0.054 (-0.185)
MVO	-0.062 (-0.212)	-0.006 (-0.021)	0.054 (0.186)	0.038 (0.132)	-0.041 (-0.140)	-0.160 (-0.552)
UNI	-0.011 (-0.037)	-0.031 (-0.108)	0.147 (0.505)	0.120 (0.412)	0.104 (0.358)	0.095 (0.327)

Table 6: Forecast Horizon

This table shows the annualized out-of-sample Sharpe ratio of various portfolio selection methods for different rebalancing frequencies. Panels A-C display the results for 1 day, 5 days, and 10 days. The signal is a 21-day momentum. The test assets are the 25 and 100 equity portfolios independently sorted on size and three other factors (book-to-market, operating profitability, and investment). The numbers in parenthesis are the t-statistics for the portfolio's Sharpe ratio and information ratio using the standard error of Lo (2002). The covariances are estimated on a rolling walk-forward basis with $T = 120$ non-overlapping monthly returns. Panel A is based on 14550 daily out-of-sample returns from 01-06-1965 until 10-21-2022, Panel B is based on 2814 weekly out-of-sample returns from 12-06-1966 until 10-21-2022, Panel C is based on 1347 fortnightly out-of-sample returns from 06-12-1969 until 10-21-2022.

	FF25	FF100	ME/OP 25	ME/OP 100	ME/INV 25	ME/INV 100
Panel A. Sharpe Ratio, Horizon: 1 day						
CP2	2.104 (15.920)	1.301 (9.869)	1.853 (14.031)	0.993 (7.535)	1.811 (13.717)	1.091 (8.280)
PP2	1.226 (9.299)	1.070 (8.119)	1.155 (8.767)	0.796 (6.042)	1.336 (10.133)	0.957 (7.266)
MVO	1.089 (8.264)	-0.381 (-2.896)	0.786 (5.971)	-0.727 (-5.518)	0.816 (6.195)	-0.685 (-5.201)
UNI	1.147 (8.702)	0.660 (5.016)	0.981 (7.444)	0.320 (2.434)	1.012 (7.679)	0.364 (2.762)
Panel B. Sharpe Ratio, Horizon: 5 days						
CP2	1.527 (11.279)	1.187 (8.810)	1.372 (10.156)	0.848 (6.313)	1.383 (10.233)	1.074 (7.980)
PP2	1.002 (7.448)	1.033 (7.676)	1.040 (7.728)	0.934 (6.946)	1.207 (8.953)	1.078 (8.004)
MVO	0.555 (4.140)	-0.147 (-1.096)	0.415 (3.097)	-0.449 (-3.351)	0.486 (3.627)	-0.179 (-1.336)
UNI	0.971 (7.222)	0.681 (5.075)	0.861 (6.410)	0.453 (3.378)	0.853 (6.350)	0.465 (3.469)
Panel C. Sharpe Ratio, Horizon: 10 days						
CP2	1.302 (9.356)	1.025 (7.416)	1.382 (9.917)	1.033 (7.469)	1.287 (9.255)	0.989 (7.160)
PP2	0.925 (6.702)	0.952 (6.892)	0.950 (6.882)	0.827 (6.005)	0.864 (6.266)	0.914 (6.627)
MVO	0.422 (3.078)	0.154 (1.123)	0.365 (2.666)	-0.154 (-1.125)	0.459 (3.351)	0.028 (0.202)
UNI	0.789 (5.727)	0.686 (4.993)	0.861 (6.245)	0.626 (4.555)	0.778 (5.650)	0.569 (4.142)

Table 7: Shrinkage Intensity

This table shows the annualized out-of-sample Sharpe Ratio of Canonical Portfolio with the leading two components for different shrinkage intensities in the estimation of the covariance of the signals. Panels A-E display the results for the full sample period, Sample 1 (1974 through 1986), Sample 2 (1986 through 1998), Sample 3 (1998 through 2010), and Sample 4 (2010 through 2022). The rows within each panel correspond to results with shrinkage parameter $\delta_x \in \{0, 0.2, 0.5, 1\}$. The signal is a 21-day momentum. The test assets are the 25 and 100 equity portfolios independently sorted on size and three other factors (book-to-market, operating profitability, and investment). The numbers in parenthesis are the t-statistics for the portfolio's Sharpe ratio and information ratio using the standard error of Lo (2002). The covariances are estimated on a rolling walk-forward basis with $T = 120$ non-overlapping monthly returns. All measures are computed with 579 monthly portfolio weights from 09-09-1974 until 10-21-2022.

	FF25	FF100	ME/OP 25	ME/OP 100	ME/INV 25	ME/INV 100
Panel A. Sharpe Ratio, Full Sample Period						
CP2 ($\delta_x = 1$)	0.986 (6.707)	0.939 (6.400)	0.842 (5.752)	0.634 (4.359)	0.849 (5.798)	0.588 (4.047)
CP2 ($\delta_x = 0.8$)	0.994 (6.764)	0.995 (6.766)	0.848 (5.791)	0.851 (5.817)	0.875 (5.970)	0.661 (4.545)
CP2 ($\delta_x = 0.5$)	0.973 (6.624)	0.849 (5.805)	0.836 (5.712)	0.914 (6.232)	0.869 (5.935)	0.728 (4.994)
CP2 ($\delta_x = 0.2$)	0.981 (6.679)	0.491 (3.393)	0.721 (4.949)	0.548 (3.778)	0.898 (6.125)	0.536 (3.694)
Panel B. Sharpe Ratio, Subsample Period 1974 to 1986						
CP2 ($\delta_x = 1$)	1.178 (3.969)	0.966 (3.284)	1.054 (3.558)	0.396 (1.364)	1.222 (4.092)	0.431 (1.482)
CP2 ($\delta_x = 0.8$)	1.315 (4.401)	0.757 (2.593)	1.103 (3.714)	0.495 (1.701)	1.384 (4.598)	0.752 (2.566)
CP2 ($\delta_x = 0.5$)	1.276 (4.278)	0.485 (1.672)	1.205 (4.040)	0.479 (1.645)	1.513 (4.990)	0.852 (2.896)
CP2 ($\delta_x = 0.2$)	1.280 (4.291)	0.059 (0.203)	0.937 (3.179)	-0.036 (-0.123)	1.580 (5.192)	0.486 (1.670)
Panel C. Sharpe Ratio, Subsample Period 1986 to 1998						
CP2 ($\delta_x = 1$)	1.465 (4.863)	1.406 (4.680)	1.134 (3.814)	0.728 (2.485)	1.159 (3.892)	1.043 (3.521)
CP2 ($\delta_x = 0.8$)	1.599 (5.267)	1.676 (5.493)	1.303 (4.348)	1.358 (4.518)	1.243 (4.158)	1.081 (3.644)
CP2 ($\delta_x = 0.5$)	1.703 (5.571)	1.548 (5.113)	1.288 (4.300)	1.489 (4.917)	1.061 (3.580)	1.147 (3.854)
CP2 ($\delta_x = 0.2$)	1.939 (6.245)	1.244 (4.176)	1.200 (4.023)	1.036 (3.500)	0.986 (3.336)	0.783 (2.669)
Panel D. Sharpe Ratio, Subsample Period 1998 to 2010						
CP2 ($\delta_x = 1$)	0.883 (3.009)	1.156 (3.896)	0.958 (3.244)	1.209 (4.052)	1.065 (3.594)	1.067 (3.599)
CP2 ($\delta_x = 0.8$)	0.852 (2.907)	1.295 (4.338)	0.869 (2.954)	1.025 (3.463)	0.912 (3.096)	0.798 (2.720)
CP2 ($\delta_x = 0.5$)	0.677 (2.322)	1.132 (3.819)	0.715 (2.441)	0.749 (2.556)	0.724 (2.473)	0.641 (2.193)
CP2 ($\delta_x = 0.2$)	0.489 (1.687)	0.448 (1.547)	0.769 (2.621)	0.285 (0.980)	0.812 (2.765)	0.591 (2.025)
Panel E. Sharpe Ratio, Subsample Period 2010 to 2022						
CP2 ($\delta_x = 1$)	0.487 (1.671)	0.376 (1.295)	0.228 (0.782)	0.121 (0.416)	0.087 (0.298)	-0.308 (-1.059)
CP2 ($\delta_x = 0.8$)	0.360 (1.240)	0.392 (1.348)	0.233 (0.802)	0.516 (1.764)	0.113 (0.390)	-0.023 (-0.078)
CP2 ($\delta_x = 0.5$)	0.437 (1.501)	0.272 (0.937)	0.376 (1.288)	0.984 (3.318)	0.330 (1.133)	0.338 (1.160)
CP2 ($\delta_x = 0.2$)	0.524 (1.799)	0.221 (0.762)	0.143 (0.490)	0.912 (3.083)	0.421 (1.442)	0.261 (0.898)

Table 8: Momentum Lookback Window

This table shows the annualized out-of-sample Sharpe Ratio of various portfolio selection methods based on momentum signals of different lookback window sizes. Panels A-E display the results for 42 days, 63 days, 84 days, 126 days, and 252 days. The test assets are the 25 and 100 equity portfolios independently sorted on size and three other factors (book-to-market, operating profitability, and investment). The numbers in parenthesis are the t-statistics for the portfolio's Sharpe ratio and information ratio using the standard error of Lo (2002). The covariances are estimated on a rolling walk-forward basis with $T = 120$ non-overlapping monthly returns. All measures are computed with 579 monthly portfolio weights from 09-09-1974 until 10-21-2022.

	FF25	FF100	ME/OP 25	ME/OP 100	ME/INV 25	ME/INV 100
Panel A. Sharpe Ratio, Lookback Window Size: 42						
CP2	0.887 (6.054)	1.006 (6.837)	0.863 (5.895)	0.740 (5.074)	0.909 (6.198)	0.707 (4.853)
PP2	0.649 (4.465)	0.624 (4.298)	0.668 (4.587)	0.545 (3.759)	0.780 (5.342)	0.688 (4.728)
MVO	0.238 (1.647)	0.199 (1.380)	0.208 (1.444)	-0.051 (-0.355)	0.212 (1.468)	0.110 (0.761)
UNI	0.438 (3.031)	0.366 (2.532)	0.502 (3.462)	0.398 (2.748)	0.432 (2.982)	0.327 (2.265)
Panel B. Sharpe Ratio, Lookback Window Size: 63						
CP2	0.665 (4.575)	0.982 (6.681)	0.871 (5.948)	0.787 (5.389)	0.901 (6.143)	0.909 (6.199)
PP2	0.612 (4.216)	0.575 (3.962)	0.716 (4.913)	0.663 (4.555)	0.674 (4.627)	0.652 (4.479)
MVO	0.328 (2.273)	0.288 (1.995)	0.185 (1.284)	-0.195 (-1.354)	0.253 (1.753)	0.177 (1.229)
UNI	0.420 (2.901)	0.366 (2.536)	0.479 (3.308)	0.336 (2.325)	0.393 (2.716)	0.286 (1.979)
Panel C. Sharpe Ratio, Lookback Window Size: 84						
CP2	0.625 (4.306)	0.908 (6.196)	0.674 (4.628)	0.543 (3.743)	0.616 (4.237)	0.854 (5.834)
PP2	0.508 (3.504)	0.453 (3.129)	0.409 (2.826)	0.482 (3.329)	0.465 (3.211)	0.502 (3.460)
MVO	0.264 (1.829)	0.305 (2.110)	0.161 (1.117)	-0.097 (-0.675)	0.214 (1.480)	0.313 (2.169)
UNI	0.277 (1.917)	0.273 (1.891)	0.340 (2.352)	0.259 (1.796)	0.248 (1.715)	0.242 (1.678)
Panel D. Sharpe Ratio, Lookback Window Size: 126						
CP2	0.670 (4.605)	0.885 (6.042)	0.545 (3.755)	0.486 (3.352)	0.607 (4.179)	0.769 (5.270)
PP2	0.459 (3.169)	0.428 (2.957)	0.442 (3.055)	0.561 (3.865)	0.359 (2.486)	0.275 (1.907)
MVO	0.314 (2.172)	0.379 (2.620)	0.117 (0.808)	-0.107 (-0.743)	0.337 (2.331)	0.257 (1.780)
UNI	0.355 (2.455)	0.311 (2.154)	0.346 (2.390)	0.276 (1.910)	0.295 (2.040)	0.254 (1.760)
Panel E. Sharpe Ratio, Lookback Window Size: 252						
CP2	1.008 (6.851)	1.200 (8.089)	0.805 (5.511)	0.722 (4.953)	0.949 (6.460)	1.129 (7.629)
PP2	0.585 (4.035)	0.511 (3.526)	0.534 (3.683)	0.530 (3.654)	0.511 (3.521)	0.330 (2.282)
MVO	0.501 (3.458)	0.578 (3.986)	0.362 (2.502)	0.043 (0.300)	0.469 (3.240)	0.391 (2.701)
UNI	0.437 (3.023)	0.386 (2.672)	0.524 (3.612)	0.449 (3.103)	0.441 (3.042)	0.383 (2.650)

Table 9: Approximate versus Actual Solution

This table shows the out-of-sample performances (in percentages) of the ‘approximated’ versus ‘actual’ portfolio policy using a 21-day momentum signal. Panels A-G display the results for the average returns, the Sharpe ratio of returns, the alpha with respect to the Fama-French 5-factor regression augmented with the univariate factor (UNI), the beta of the simple factor, the idiosyncratic volatility, and the information ratio. The test assets are the 25 and 100 equity portfolios independently sorted on size and three other factors (book-to-market, operating profitability, and investment). The numbers in parenthesis are the t-statistics for the portfolio’s Sharpe ratio and information ratio using the standard error of Lo (2002). The covariances are estimated on a rolling walk-forward basis with $T = 120$ non-overlapping monthly returns. All measures are computed with 579 monthly portfolio weights from 09-09-1974 until 10-21-2022.

	FF25	FF100	ME/OP 25	ME/OP 100	ME/INV 25	ME/INV 100
Panel A. Average Returns (%)						
CP2 (Approx)	2.199	1.341	1.731	0.957	1.797	0.777
CP2 (Full)	2.253	1.175	1.578	0.822	1.675	0.596
Panel B. Standard Deviation of Returns (%)						
CP2 (Approx)	2.177	1.366	2.049	1.260	2.057	1.248
CP2 (Full)	2.210	1.356	2.119	1.229	2.061	1.309
Panel C. Sharpe Ratio						
CP2 (Approx)	1.011 (6.872)	0.982 (6.685)	0.848 (5.797)	0.763 (5.226)	0.877 (5.985)	0.626 (4.308)
CP2 (Full)	1.020 (6.933)	0.868 (5.934)	0.750 (5.140)	0.672 (4.620)	0.817 (5.588)	0.460 (3.177)
Panel D. Alpha (%)						
CP2 (Approx)	1.849	1.197	1.109	0.826	1.334	0.679
CP2 (Full)	1.908	1.043	0.942	0.739	1.224	0.536
Panel E. Beta to Simple Factor						
CP2 (Approx)	0.146	0.071	0.259	0.086	0.172	0.053
CP2 (Full)	0.144	0.067	0.267	0.057	0.165	0.046
Panel F. Idiosyncratic Volatility (%)						
CP2 (Approx)	2.043	1.321	1.749	1.220	1.911	1.229
CP2 (Full)	2.083	1.315	1.807	1.208	1.921	1.297
Panel G. Information Ratio						
CP2 (Approx)	0.905 (6.176)	0.906 (6.183)	0.634 (4.360)	0.677 (4.651)	0.698 (4.792)	0.553 (3.809)
CP2 (Full)	0.916 (6.247)	0.793 (5.434)	0.522 (3.596)	0.612 (4.210)	0.637 (4.379)	0.414 (2.858)

Table 10: Performance Summary for Two Signals Case

This table shows the out-of-sample performances (in percentages) of various portfolio selection methods using two momentum signals with a lookback window of 21 days and 252 days. Panels A-G display the results for the average returns, the Sharpe ratio of returns, the alpha with respect to the Fama-French 5-factor regression augmented with the univariate factor (UNI), the beta of the simple factor, the idiosyncratic volatility, and the information ratio. The test assets are 25 and 100 equity portfolios independently sorted on size and three other factors (book-to-market, operating profitability, and investment). The numbers in parenthesis are the t-statistics for the portfolio's Sharpe ratio and information ratio using the standard error of Lo (2002). The covariances are estimated on a rolling walk-forward basis with $T = 120$ non-overlapping monthly returns. All measures are computed with 579 monthly portfolio weights from 09-09-1974 until 10-21-2022.

	FF25	FF100	ME/OP 25	ME/OP 100	ME/INV 25	ME/INV 100
Panel A. Average Returns (%)						
CP2	2.925	1.807	2.308	1.027	2.274	1.551
PP2	4.160	4.143	3.779	4.082	2.893	2.522
MVO	0.761	0.581	0.595	-0.032	0.746	0.386
UNI	2.882	2.316	2.743	2.047	2.431	1.838
Panel B. Standard Deviation of Returns (%)						
CP2	2.343	1.384	2.153	1.312	2.072	1.293
PP2	6.042	6.053	5.288	5.488	4.707	4.760
MVO	1.893	1.431	1.807	1.332	1.642	1.347
UNI	5.489	5.097	4.543	4.095	4.619	4.280
Panel C. Sharpe Ratio						
CP2	1.244 (8.365)	1.302 (8.735)	1.072 (7.266)	0.786 (5.387)	1.096 (7.426)	1.197 (8.073)
PP2	0.706 (4.849)	0.702 (4.823)	0.729 (5.005)	0.758 (5.197)	0.630 (4.338)	0.548 (3.777)
MVO	0.410 (2.836)	0.412 (2.851)	0.337 (2.336)	-0.017 (-0.120)	0.461 (3.184)	0.293 (2.028)
UNI	0.546 (3.766)	0.476 (3.285)	0.619 (4.263)	0.516 (3.561)	0.544 (3.751)	0.448 (3.093)
Panel D. Alpha (%)						
CP2	2.287	1.576	1.508	0.781	1.641	1.300
PP2	2.048	2.097	1.691	2.263	1.318	1.205
MVO	0.025	0.185	-0.195	-0.391	0.117	-0.003
Panel E. Beta to Simple Factor						
CP2	0.206	0.100	0.277	0.106	0.202	0.115
PP2	0.735	0.866	0.783	0.922	0.600	0.716
MVO	0.254	0.174	0.291	0.171	0.257	0.198
Panel F. Idiosyncratic Volatility (%)						
CP2	2.040	1.285	1.745	1.240	1.856	1.198
PP2	4.494	4.155	3.909	3.986	3.858	3.673
MVO	1.288	1.124	1.235	1.142	1.136	1.058
Panel G. Information Ratio						
CP2	1.121 (7.584)	1.226 (8.254)	0.864 (5.906)	0.630 (4.336)	0.884 (6.037)	1.084 (7.349)
PP2	0.456 (3.149)	0.505 (3.485)	0.433 (2.990)	0.568 (3.914)	0.342 (2.364)	0.328 (2.272)
MVO	0.019 (0.133)	0.164 (1.139)	-0.158 (-1.095)	-0.343 (-2.373)	0.103 (0.715)	-0.003 (-0.019)

Table 11: Weight Statistics for Two Signals Case

This table shows the average monthly weight statistics of various portfolio selection methods estimated using a 21-day momentum signal. Panels A-F display the results for the turnover, the proportional leverage, the sum of negative weights in the portfolio, and the minimum and maximum portfolio weights. The test assets are the 25 and 100 equity portfolios independently sorted on size and three other factors (book-to-market, operating profitability, and investment). The covariances are estimated on a rolling walk-forward basis with $T = 120$ non-overlapping monthly returns. All measures are computed with 579 monthly portfolio weights from 09-09-1974 until 10-21-2022.

	FF25	FF100	ME/OP 25	ME/OP 100	ME/INV 25	ME/INV 100
Panel A. Turnover						
CP2	0.749	0.437	0.806	0.542	0.762	0.466
PP2	0.806	0.763	0.878	0.794	0.891	0.856
MVO	1.064	1.075	1.065	1.082	1.060	1.079
UNI	0.959	0.978	0.970	0.990	0.969	0.989
Panel B. Proportional Leverage						
CP2	0.477	0.473	0.480	0.488	0.470	0.486
PP2	0.478	0.483	0.472	0.473	0.477	0.484
MVO	0.506	0.499	0.509	0.503	0.497	0.497
UNI	0.485	0.498	0.488	0.497	0.485	0.498
Panel C. Sum of Negative Weights						
CP2	-0.483	-0.480	-0.486	-0.482	-0.481	-0.479
PP2	-0.500	-0.500	-0.500	-0.500	-0.500	-0.500
MVO	-0.499	-0.499	-0.500	-0.501	-0.498	-0.498
UNI	-0.500	-0.500	-0.500	-0.500	-0.500	-0.500
Panel D. Minimum Weight						
CP2	-0.121	-0.035	-0.113	-0.036	-0.114	-0.035
PP2	-0.104	-0.034	-0.108	-0.037	-0.104	-0.033
MVO	-0.099	-0.031	-0.096	-0.030	-0.097	-0.031
UNI	-0.090	-0.025	-0.089	-0.025	-0.089	-0.024
Panel E. Maximum Weight						
CP2	0.109	0.032	0.102	0.032	0.102	0.032
PP2	0.092	0.031	0.096	0.033	0.096	0.030
MVO	0.102	0.031	0.099	0.031	0.097	0.030
UNI	0.088	0.025	0.087	0.025	0.089	0.025
Panel F. Standard Deviation of Weights						
CP2	0.053	0.013	0.052	0.013	0.052	0.013
PP2	0.051	0.013	0.052	0.013	0.051	0.013
MVO	0.051	0.013	0.050	0.012	0.050	0.012
UNI	0.049	0.012	0.049	0.012	0.049	0.012

B Proofs for all the Propositions

Proof of Proposition 1. We provide the proof of this proposition in two parts.

Part 1. We drop the time subscript t for brevity. From the cyclic property of the trace operator and linearity of the expectation, the expected value of the portfolio returns is

$$\mathbb{E}[x'Ar] = \mathbb{E}[\text{Tr}(Arx')] = \text{Tr}(A\Sigma_{rx}). \quad (42)$$

In order to analyze the variance of the portfolio returns, we appeal to the following one-dimensional from Isserlis (1918) or Wick (1950), which expresses the higher moments of centered multivariate Gaussian vectors in terms of its second-moments.

Theorem 1. Let $z_1, z_2, z_3,$ and z_4 be jointly Gaussian random variables with mean zero. Then we have the following results:

$$\begin{aligned} \mathbb{E}[z_1] &= 0 \\ \mathbb{E}[z_1 z_2] &= \text{Cov}(z_1, z_2) \\ \mathbb{E}[z_1 z_2 z_3] &= 0 \\ \mathbb{E}[z_1 z_2 z_3 z_4] &= \mathbb{E}[z_1 z_2] \mathbb{E}[z_3 z_4] + \mathbb{E}[z_1 z_3] \mathbb{E}[z_2 z_4] + \mathbb{E}[z_1 z_4] \mathbb{E}[z_2 z_3] \end{aligned} \quad (43)$$

The last equation of (43) takes all partitions of size two of the four variables, which gives us three separate terms. By recasting fourth-order terms using the covariance terms, we can express the variance as

$$\begin{aligned} \text{Var}[x'Ar] &= \mathbb{E}[(x'Ar)^2] - \mathbb{E}[x'Ar]^2 \\ &= \sum_{i,j,k,l} \mathbb{E}[A_{ij}A_{kl}x_i r_j x_k r_l] - \sum_{i,j,k,l} \mathbb{E}[A_{ij}x_i r_j] \mathbb{E}[A_{kl}x_k r_l] \\ &= \sum_{i,j,k,l} A_{ij}A_{kl} \mathbb{E}[x_i r_j] \mathbb{E}[x_k r_l] + \sum_{i,j,k,l} A_{ij}A_{kl} \mathbb{E}[x_i x_k] \mathbb{E}[r_j r_l] \\ &\quad + \sum_{i,j,k,l} A_{ij}A_{kl} \mathbb{E}[x_i r_l] \mathbb{E}[x_k r_j] - \sum_{i,j,k,l} A_{ij}A_{kl} \mathbb{E}[x_i r_j] \mathbb{E}[x_k r_l] \\ &= \sum_{i,j,k,l} A_{ij}A_{kl} \mathbb{E}[x_i x_k] \mathbb{E}[r_j r_l] + \sum_{i,j,k,l} A_{ij}A_{kl} \mathbb{E}[x_i r_l] \mathbb{E}[x_k r_j] \\ &= \sum_{i,j,k,l} \mathbb{E}[x_k x_i] A_{ij} \mathbb{E}[r_j r_l] A_{kl} + \sum_{i,j,k,l} \mathbb{E}[r_l x_i] A_{ij} \mathbb{E}[r_j x_k] A_{kl}. \end{aligned} \quad (44)$$

where A_{ij} is the (i, j) -th entry of A . Reverting back to matrix notation, we have

$$\text{Var}[x'Ar] = \text{Tr}(\Sigma_x A \Sigma_r A') + \text{Tr}(\Sigma_{rx} A \Sigma_{rx} A). \quad (45)$$

Part 2. In order to solve the full mean-variance objective function (4), we use the tools from CCA that we have laid the ground in Section 3.1. To assist in our derivation, we expand our

notation and let Q_r and Q_x be matrices whose columns contain the so-called i th canonical directions $q_{r,i} := \Sigma_r^{-1/2} u_i$ and $q_{x,i} := \Sigma_x^{-1/2} v_i$, respectively, and $D := \text{Diag}(s_1, s_2, \dots, s_N)$ be a diagonal matrix containing the canonical correlations. We now state the main theorem:

Theorem 2. Suppose r_{t+1} and x_t are N -dimensional jointly Gaussian random variables. Then the objective function (2) is maximized at

$$A = Q_x \text{Diag}(L_1, \dots, L_N) Q_r', \quad (46)$$

where diagonal matrix consist of the following elements

$$L_i = \frac{1}{\gamma} \frac{s_i}{1 + s_i^2}, \quad (47)$$

for $i = 1, \dots, N$.

Without loss of generality, let us reparameterize the matrix of coefficients in terms of the canonical directions as $A = Q_x L Q_r'$ where L is a variable matrix that we now have to optimize on. The expected value of the portfolio returns at time t can be written as

$$\mathbb{E}[x_t' A r_{t+1}] = \text{Tr}(A \Sigma_{rx}) \quad (48)$$

$$= \text{Tr}(LD), \quad (49)$$

and the variance as

$$\text{Var}[x_t' A r_{t+1}] = \text{Tr}(\Sigma_x A \Sigma_r A') + \text{Tr}(\Sigma_{rx} A \Sigma_{rx} A) \quad (50)$$

$$= \text{Tr}(LL') + \text{Tr}(DLDL). \quad (51)$$

Putting all together, the objective function (4) can be written as:

$$\max_L \text{Tr}(LD) - \frac{\gamma}{2} (\text{Tr}(LL') - \text{Tr}(DLDL)). \quad (52)$$

Using the rules for matrix derivatives from Lütkepohl (1997), we take first-order conditions with respect to the matrix L to get

$$D - \gamma(L + DL'D) = 0. \quad (53)$$

Since the matrix D is diagonal and thus symmetric, we necessarily have the following relationship

$$\frac{1}{\gamma} D = L + DL'D = L' + DLD. \quad (54)$$

Let $L^a := L - L'$, which is an anti-symmetric matrix (that is, $L^{a'} = -L^a$) and has diagonal elements

zero. Then we have

$$L^a = DL^aD. \quad (55)$$

If we restrict our attention to the off-diagonal elements of (55), we see that $L_{ij}^a = s_i s_j L_{ij}^a$ for all $i \neq j$, so $(1 - s_i s_j)L_{ij}^a = 0$. Therefore, either $s_i s_j = 1$ for all $i \neq j$ or $L_{ij}^a = 0$. Consequently, L^a is identically zero and hence, L must be a symmetric matrix.

The benefit of variable matrix L being symmetric is that the condition (53) can now be written as

$$L + DLD = \frac{1}{\gamma}D. \quad (56)$$

If we focus on off-diagonal elements, we see that $(1 + s_i s_j)L_{ij} = 0$ for all $i \neq j$. But we also know that the canonical correlations satisfy the following ordering $0 \leq s_1 \leq \dots \leq s_N \leq 1$, and so it is impossible that $s_i s_j = -1$. Thus, $L_{ij} = 0$ for all $i \neq j$, and so L must be a diagonal matrix. Thus, optimizing over the elements $L_{ii} \equiv L_i$ for $i = 1, \dots, N$, boils down to maximizing the following univariate problems

$$\max_{L_1, \dots, L_N} \sum_{i=1}^N L_i s_i - \frac{\gamma}{2}(L_i^2 + s_i^2 L_i^2). \quad (57)$$

Our problem can now be easily solved to give us

$$L_i = \frac{1}{\gamma} \frac{s_i}{1 + s_i^2}. \quad (58)$$

Hence, the diagonal elements are a nonlinear function of the canonical correlations. The optimal matrix A in (46) is effectively the optimal scaling of the canonical portfolios of the asset returns and signals. \square

Proof of Proposition 2. Assuming that the second term in Equation (4) is approximately zero, we can rewrite the objective function (5) in unconstrained form as

$$\max_{\tilde{A}} \text{Tr}(\tilde{A}\Sigma'_{\tilde{r}\tilde{x}}) - \frac{\gamma}{2}\text{Tr}(\tilde{A}\tilde{A}'), \quad (59)$$

where we have used a change-of-variables $\tilde{A} := \Sigma_x^{1/2} A \Sigma_r^{1/2}$. The problem (59) can be solved to yield

$$\tilde{A} = \frac{1}{\gamma} \Sigma_x^{-1/2} \Sigma'_{rx} \Sigma_r^{-1/2}. \quad (60)$$

Since the solution is expressed in a different basis, we rescale it back to the original assets through

the following operation

$$A = \Sigma_x^{-1/2} \tilde{A} \Sigma_r^{-1/2} = \frac{1}{\gamma} \Sigma_x^{-1} \Sigma_{rx}' \Sigma_r^{-1}, \quad (61)$$

As a result of this approximation, we now have

$$L_i \approx \frac{1}{\gamma} s_i. \quad (62)$$

Figure 2 contrasts the optimal singular value adjustment in (47) to the linear approximation (62) for $\gamma = 1$. We can see that for small values of s_i , the optimal adjustment is approximately linear while for large values of s_i , the optimal adjustment deflates the values of the canonical correlations. \square

Proof of Proposition 3. We start by writing down the expression for the Lagrangian

$$\text{Tr}(A \Sigma_{rx}) - \frac{\gamma}{2} \text{Tr}(\Sigma_x A \Sigma_r A') + \lambda(1 - \mathbf{1}' A' x_t), \quad (63)$$

$\lambda > 0$ is the Lagrange multiplier. Using the change-of-variables $\tilde{A} := \Sigma_x^{1/2} A \Sigma_r^{1/2}$, we have

$$\text{Tr}(\tilde{A} \Sigma_{\tilde{r}\tilde{x}}) - \frac{\gamma}{2} \text{Tr}(\tilde{A} \tilde{A}') + \lambda(1 - \mathbf{1}' \Sigma_r^{-1/2} \tilde{A}' \Sigma_x^{-1/2} x_t). \quad (64)$$

Performing the first-order conditions

$$[\text{FOC } \tilde{A}] : \Sigma_{\tilde{r}\tilde{x}}' - \gamma \tilde{A} - \lambda \Sigma_x^{-1/2} x_t \mathbf{1}' \Sigma_r^{-1/2} = 0, \quad (65)$$

$$[\text{FOC } \lambda] : 1 - \mathbf{1}' w_t = 0. \quad (66)$$

By reverting the change-of-variables we have made and using the fact that the portfolio policies are linear the signals, the portfolio weights is then

$$w_t^{\text{FI}} = \frac{1}{\gamma} \Sigma_r^{-1} \Sigma_{rx} \Sigma_x^{-1} x_t + \lambda (x_t' \Sigma_x^{-1} x_t) \mathbf{1}' \Sigma_r^{-1}. \quad (67)$$

Solving for λ gives

$$\lambda = \frac{1 - \kappa}{(x_t' \Sigma_x^{-1} x_t) (\mathbf{1}' \Sigma_r^{-1} \mathbf{1})}, \quad (68)$$

where $\kappa := \gamma^{-1} \mathbf{1}' \Sigma_r^{-1} \Sigma_{rx} \Sigma_x^{-1} x_t$. Inserting the λ into (67) and rearranging the expression gives us the optimal portfolio (18). \square

Proof of Proposition 4. If r_{t+1} and x_t are N -dimensional zero-mean Gaussian random variables, then their projections on to the i th canonical portfolios are $u_i' \tilde{r}_{t+1}$ and $v_i' \tilde{x}_t$ are also Gaussian random variables with mean zero and unit variances, respectively and a joint correlation s_i . For the expected

return of the i th canonical portfolio, we have

$$\mathbb{E}[\pi_{ti}] = s_i v_i' \Sigma_{\tilde{r}\tilde{x}}' u_i = s_i^2. \quad (69)$$

Furthermore, the variance of the i th canonical portfolio is

$$\text{Var}[\pi_{ti}] = \mathbb{E}[s_i^2 (v_i' \tilde{x}_t)^2 (u_i' \tilde{r}_{t+1})^2] - s_i^4 \quad (70)$$

$$= s_i^2 (1 + 2s_i^2) - s_i^4 \quad (71)$$

$$= s_i^2 (1 + s_i^2). \quad (72)$$

The second equality follows from the second-moment-based result of the product of two correlated Gaussian variables from [Haldane \(1942, Section 6\)](#). Substituting the expression of the optimal policy A into the definition of the expected value and variance of the portfolio returns, we have:

$$\mathbb{E}[x_t' A r_{t+1}] = \frac{1}{\gamma} \text{Tr}(\Sigma_x^{-1} \Sigma_{rx}' \Sigma_r^{-1} \Sigma_{rx}) = \frac{1}{\gamma} \sum_{i=1}^N s_i^2, \quad (73)$$

$$\text{Var}[x_t' A r_{t+1}] = \frac{1}{\gamma} \text{Tr}(\Sigma_x^{-1} \Sigma_{rx}' \Sigma_r^{-1} \Sigma_{rx}) = \frac{1}{\gamma} \sum_{i=1}^N s_i^2, \quad (74)$$

where we assumed that the squared managed portfolio returns are approximately zero in the expression for the variance term. \square

Proof of Corollary 1. This follows immediately Proposition 4 and the definition of the Sharpe ratio. \square

Proof of Proposition 5. Substituting the second-moment matrix of managed portfolio returns (38) into the expression of the expected portfolio returns from Proposition 1 and using the cyclic property of the trace operator, we have

$$\text{Tr}(\Sigma_x^{-1} \Sigma_{rx}' \Sigma_r^{-1} \Sigma_{rx}) = \frac{1}{\gamma} (\mu_r' \Sigma_r^{-1} \mu_r) (\mu_x' \Sigma_x^{-1} \mu_x) + \frac{1}{\gamma} \sum_{i=1}^N s_i^2, \quad (75)$$

and the claim follows. \square