Canonical Portfolios: Optimal Asset and Signal Combination

Nick Firoozye
Department of Computer Science
University College London
n.firoozye@cs.ucl.ac.uk

Vincent Tan*
Oxford-Man Institute of Quantitative Finance
University of Oxford
vincent.tan@eng.ox.ac.uk

Stefan Zohren
Oxford-Man Institute of Quantitative Finance
University of Oxford
stefan.zohren@eng.ox.ac.uk

February 2022

Abstract

We present a novel framework for analyzing the optimal asset and signal combination problem. We reformulate the original problem of portfolio selection from a set of correlated assets and signals into one of selecting from a set of uncorrelated trading strategies through the lens of *Canonical Correlation Analysis* of Hotelling (1936). The new environment of uncorrelated trading strategies offers a pragmatic simplification to the inherent complicated structure of our underlying problem that involves interactions between and within variables. We also operationalize our framework to bridge the gap between theory and practice and showcase improved risk-adjusted returns performance of our proposed optimizer over the classic mean-variance optimizer of Markowitz (1952).

JEL Classification: G11, D81, C1.

Keywords: Canonical Correlation Analysis, Covariance and Cross-Covariance Matrix Estimation, Portfolio Selection, Risk Parity, Statistical Arbitrage

*This work was supported by the Oxford-Man Institute of Quantitative Finance. The authors would like to thank Brian Healy for the helpful comments.

1 Introduction

The investment decisions of portfolio managers are typically guided by signals that encompass their view on future returns. 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 requires two inputs: (i) the vector of expected returns and (ii) the covariance matrix of returns.

In many cases, the signals that are used to explain the cross-section of stock returns would receive an equal weight contribution. While in some more elaborate settings, the dynamic relationship between multiple signals and their underlying returns are modeled in a multivariate fashion, and the weights on each signal are assigned accordingly. A simple forecasting approach that captures this relationship between both variables would be to perform a multivariate ordinary least squares (OLS) regression. We shall refer to such modeling as *signal combinations*. These forecasts are then used as *plug-in* estimates in a tactical asset allocation framework.

In this paper, we solve an asset allocation problem, with and without usual portfolio constraints, where signal combinations are utilized and play an inherent part in the overall goal of optimizing the risk-return tradeoff. The utility of such an integrated framework allows us to perform a more holistic analysis that takes into account the multivariate contributions that are embedded in both processes that may impact the allocation of capital. However, this introduces additional challenges to the table. The simultaneous allocation of the asset returns and signals in optimizing a mean-variance objective implies that the covariances of the signals *and* the cross-covariances between both variables now constitute important additional inputs that we have to consider.

Our first contribution is to provide a simplification of this problem by reformulating the set of asset returns and signals into a set of uncorrelated strategy returns. 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) to two sets of random variables. This is pertinent to our case where we have multiple correlated assets and correlated signals that are mutually linked by a joint correlation criterion. CCA allows us to express a generic strategy return in terms of its exposures to uncorrelated sources of strategy returns by reweighting the original set of variables; we coin these weights as *canonical portfolios*.[1]

[1]The 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 Avellaneda and Lee (2010) and Avellaneda et al. (2020). Our work aims to broaden that horizon to *two* sets of variables by also considering the signal as an important input in the overall analysis.

The second contribution is to mobilize our theoretical framework so that it is applicable in practice, even in large dimensions. Unlike the standard mean-variance problem, our framework necessitates two additional inputs; the covariances of signals and the cross-covariances of the asset returns and signals. All three matrices have $O(N^2)$ parameters that require estimation. Given the complexity of our problem, regularization of these large-dimensional objects is paramount to encourage stability in the out-of-sample results. Our prescription for the covariances is to estimate them using structure-free based approaches that fall under the framework of rotational invariance due to Stein (1975, 1986). We then estimate the cross-covariance matrix by addressing the instabilities in the cross-correlation space, as guided by CCA, while preserving the cross-sectional variations from the estimated covariances of the returns and signals. This is achieved by imposing a constant correlation model, which assumes all the self-predictability and cross-predictability of scale invariant variables to be identical. We show that our optimizer responds positively to these innovations that we introduce.

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 and avoid modeling the conditional return distribution. Of relevance to our study is Brandt and Santa-Clara (2006), where the authors analyzed portfolio policies conditioning on state-variables through augmentation of the state-space. However, we provide a counterpoint to the authors on two fronts. First, we choose to emphasize the distinct nature of the asset returns and their associated signals in our optimization. This allows us to analyze the structure of the strategy returns through CCA and elucidate the association between the returns and their signals. Second, since our approach does not expand the dimension of the asset space, our optimizer is more computationally tractable in large dimensions and is less effected by the curse of dimensionality. The price that we have to pay for this is the need to estimate three large-dimensional objects, which we will discuss in this paper.

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 algorithmic trading. While most algorithmic traders will seek to find a good forecast for future returns via methods such as OLS, or in the nonlinear context, via a large suite of machine learning-based methods, Firoozye and Koshiyama (2019, 2020) showed that if the goal is to maximize the Sharpe Ratio of strategy returns, then this is 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.

The remainder of the paper is organized as follows. Section 2 gives a brief description of the dynamic strategy optimization. Section 3 provides the financial interpretation of our optimizer with CCA. Section 4 details our proposed estimation approach. Section 5 describes the empirical methodology and presents the results of the out-of-sample backtest exercise with financial stock returns data. Section 6 concludes. Appendix A–C contain all the figures, tables, and additional mathematical derivations.

2 Setting the Stage

2.1 Notation

In this section, we introduce the necessary notation for our analysis. Let the subscript i index the variables such that $i \in \{1, \ldots, N\}$, where N denotes the dimension of the investment universe and of their respective signals. The subscript t indexes the dates such that $t \in \{1, \ldots, T\}$, where T denotes the sample size. The notation $\mathsf{Cov}(\cdot, \cdot)$ represents the covariance matrix between two random vectors, the notation $\mathsf{Tr}(\cdot)$ represents the trace of a matrix, and the notation $\mathsf{Diag}(\cdot)$ represents the function that sets the off-diagonal elements of a matrix to zero.

Let $r_{t,i}$ be the return for a risky asset i at date t, stacked into a vector $\mathbf{r}_t := (r_{t,1}, \ldots, r_{t,N})'$. Also, let $x_{t,i}$ be the return-predictive signal for asset i at date t, stacked into a vector $\mathbf{x}_t := (x_{t,1}, \ldots, x_{t,N})'$. Their multivariate distributions are assumed to have zero expectations and covariances $\mathsf{Var}(\mathbf{r}_t) := \mathsf{Cov}(\mathbf{r}_t, \mathbf{r}_t) = \boldsymbol{\Sigma}_r$ and $\mathsf{Var}(\mathbf{x}_t) := \mathsf{Cov}(\mathbf{x}_t, \mathbf{x}_t) = \boldsymbol{\Sigma}_x$, respectively, and a cross-covariance $\mathsf{Cov}(\mathbf{r}_t, \mathbf{x}_t) = \boldsymbol{\Sigma}_{rx}$. The composite vector $(\mathbf{r}_t, \mathbf{x}_t)$ has a joint covariance matrix expressed in the following block matrix form

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_r & \boldsymbol{\Sigma}_{rx} \\ \boldsymbol{\Sigma}'_{rx} & \boldsymbol{\Sigma}_x \end{pmatrix}. \tag{1}$$

Due to the symmetric positive definite property, of the covariance matrices of $\boldsymbol{\Sigma}_r$ and $\boldsymbol{\Sigma}_x$, 'square root' factors of the matrices can be found via spectral decompositions; they are defined through $\boldsymbol{\Sigma}_r = \boldsymbol{\Sigma}_r^{1/2}\boldsymbol{\Sigma}_r^{1/2}$ and $\boldsymbol{\Sigma}_x = \boldsymbol{\Sigma}_x^{1/2}\boldsymbol{\Sigma}_x^{1/2}$.

In our notation, we denote population second moments by Greek letters and their estimated counterparts with a hat accent $(\hat{\ })$ on them: for example, $\boldsymbol{\Sigma}_r$ and $\hat{\boldsymbol{\Sigma}}_r$. Their sample counterpart is denoted by a corresponding Latin letter: $\mathbf{S}_r := T^{-1}\sum_{t=1}^{T}\mathbf{r}_t\mathbf{r}'_t$. Typical entries of the second moments $\boldsymbol{\Sigma}_r$ and $\mathbf{S}_r$ are denoted by $[\boldsymbol{\Sigma}_r]_{ij}$ and $[\mathbf{S}_r]_{ij}$, respectively.

In our current set-up, we limit ourselves to work with endogenous returns-based signals, that is, signals that are derived from the asset returns themselves. Every risky asset is assumed to be accompanied by one signal, and hence we necessarily have N signals. The signals use only *lagged* information from the asset returns and are assumed to be weakly stationary.

Remark 1 (Stationarity of signals)**.** Note that while asset returns are often stationary, their constructed signals themselves may not share that property. In this paper, we shall assume that the signals do exhibit weak stationarity, so that we can model their relationships through the covariances or association with the returns through the cross-covariances.

2.2 Dynamic Strategies

We begin our analysis by considering the portfolio weights obtained at time t to be linear in signals, that is, $\mathbf{w}_t := \mathbf{A}'\mathbf{x}_t$, where $\mathbf{A}$ is a $N \times N$ matrix. The weights are dynamic because they are conditional on the strength of the signals but the matrix $\mathbf{A}$ is a static object where each row maps the signal vector into a portfolio weight in each asset. Based on this parameterization, the problem that an investor faces can be formulated as the following minimum variance problem

$$\begin{aligned} &\min_{\mathbf{A}} \frac{1}{2}\mathsf{Var}[\mathbf{x}_t'\mathbf{A}\mathbf{r}_t] \\ &\text{subject to } \mathbb{E}[\mathbf{x}_t'\mathbf{A}\mathbf{r}_t] \geq \mathcal{G}, \end{aligned} \tag{2}$$

where $\mathcal{G}$ denotes the investor's minimum target return. This objective function underscores that the investor chooses to simultaneously allocate between the asset returns and signals to yield a trading strategy that optimizes the strategy returns at time t. This formulation departs from the standard mean-variance optimization (MVO) scheme of Markowitz (1952), which assume that the portfolio weights are static.

Using the theorem of Wick (1950) the resulting optimization can be written as (see Appendix C.1)

$$\begin{aligned} &\min_{\mathbf{A}} \frac{1}{2}\mathsf{Tr}(\boldsymbol{\Sigma}_x\mathbf{A}\boldsymbol{\Sigma}_r\mathbf{A}') + \frac{1}{2}\mathsf{Tr}(\boldsymbol{\Sigma}_{rx}\mathbf{A}\boldsymbol{\Sigma}_{rx}\mathbf{A}) \\ &\text{subject to } \mathsf{Tr}(\mathbf{A}\boldsymbol{\Sigma}_{rx}) \geq \mathcal{G}. \end{aligned} \tag{3}$$

It is possible to solve this optimization problem in closed-form and the details are available in Appendix C.2. However, it is convenient to assume that the squared expectation term in (3), that is $\mathbb{E}[\mathbf{x}_t'\mathbf{A}\mathbf{r}_t]^2 = \mathsf{Tr}(\boldsymbol{\Sigma}_{rx}\mathbf{A}\boldsymbol{\Sigma}_{rx}\mathbf{A})$, is approximately zero.[2] Thus, the following auxiliary

[2]This approximation allows us to express the optimal portfolio solely in terms of the population moments without any reference to a particular decomposition technique.

problem is sufficient for us to study,

$$
\begin{aligned}
&\min_{\mathbf{A}} \frac{1}{2} \operatorname{\mathsf{Tr}}(\boldsymbol{\Sigma}_x \mathbf{A} \boldsymbol{\Sigma}_r \mathbf{A}') \\
\text{subject to} \quad &\operatorname{\mathsf{Tr}}(\mathbf{A}\boldsymbol{\Sigma}_{rx}) \geq \mathcal{G}.
\end{aligned} \tag{4}
$$

Appendix C.3 shows that the solution to this problem is given by

$$
\mathbf{A} = \lambda \cdot \boldsymbol{\Sigma}_x^{-1} \boldsymbol{\Sigma}_{rx}' \boldsymbol{\Sigma}_r^{-1}, \quad \text{where } \lambda := \frac{\mathcal{G}}{\operatorname{\mathsf{Tr}}(\boldsymbol{\Sigma}_{rx} \boldsymbol{\Sigma}_x^{-1} \boldsymbol{\Sigma}_{rx}' \boldsymbol{\Sigma}_r^{-1})}, \tag{5}
$$

where λ is a scaling parameter that is proportional to the target return $\mathcal{G}$. The weight $\mathbf{w}_t$ allocated to each asset conditional on the signal at time t is then

$$
\mathbf{w}_t = \lambda \cdot \boldsymbol{\Sigma}_r^{-1} \boldsymbol{\Sigma}_{rx} \boldsymbol{\Sigma}_x^{-1} \mathbf{x}_t. \tag{6}
$$

In absence of any further constraints on $\mathbf{w}_t$, this is the frictionless portfolio policy, and the size or scale of the portfolio is only determined by the target return constraint. The portfolio takes into account the covariances of both the asset returns and signals as well as the investment opportunities that arise from the cross-covariance matrix. The dynamics of the portfolio is guided by trajectory of the signal $\mathbf{x}_t$ at time t (which may be out-of-sample). However, the solution is not necessarily scale invariant with respect to $\mathcal{G}$, $\boldsymbol{\Sigma}_{rx}$ and $\mathbf{x}_t$. Thus, to remove the portfolio's dependency on the scale of these objects, we set the target return as $\mathcal{G} := \operatorname{\mathsf{Tr}}(\boldsymbol{\Sigma}_{rx})/\sqrt{\mathbf{x}_t'\mathbf{x}_t}$.

Remark 2 (Implied Regression)**.** Observe that the portfolio weights $\mathbf{w}_t$ can be seen as a solution which has been obtained from a two-stage process of forecasting and asset allocation. In particular, one would first perform a multivariate regression of the asset returns on the signals in order to obtain a $N \times 1$ vector of cross-sectional predictive signals, and then plug it into a standard Markowitz scheme to arrive at the vector of allocations. However, this implied regression idea is silent about the fact that both the allocation of capital over both the underlying returns and the signals is very much a *joint* selection process.

2.3 Portfolio Constraints

In quantitative equity investing, it is commonplace to impose constraints on the portfolio weights. Typical constraints would have either the weights sum to one ("fully invested") or to zero ("zero-investment"). A fully-invested portfolio enforces a budget constraint while a dollar-neutral portfolio requires the dollar amount of all long positions to equal to that of all short positions. These equality constraints can be included into our optimization problem and given that they are linear in the policy matrix $\mathbf{A}$, the portfolio weights can be fortunately solved in closed-form.

In the fully-invested case, we can formulate the problem as follows

$$\begin{aligned} &\min_{\mathbf{A}} \frac{1}{2}\mathsf{Tr}(\boldsymbol{\Sigma}_x \mathbf{A} \boldsymbol{\Sigma}_r \mathbf{A}') \\ &\text{subject to } \ \mathsf{Tr}(\mathbf{A}\boldsymbol{\Sigma}_{rx}) \geq \mathcal{G}, \ \text{ and, } \ \mathbf{1}'\mathbf{A}'\mathbf{x}_t = 1, \end{aligned} \tag{7}$$

where $\mathbf{1}$ denotes the vector of ones of dimension N. Appendix C.4 shows that the problem has the following analytical solution

$$\begin{aligned} \mathbf{w}_t^{\mathrm{FI}} &= (1-\lambda^{\mathrm{FI}})\frac{\boldsymbol{\Sigma}_r^{-1}\mathbf{1}}{\mathbf{1}'\boldsymbol{\Sigma}_r^{-1}\mathbf{1}} + \lambda^{\mathrm{FI}}\frac{\boldsymbol{\Sigma}_r^{-1}\boldsymbol{\Sigma}_{rx}\boldsymbol{\Sigma}_x^{-1}\mathbf{x}_t}{\mathbf{1}'\boldsymbol{\Sigma}_r^{-1}\boldsymbol{\Sigma}_{rx}\boldsymbol{\Sigma}_x^{-1}\mathbf{x}_t}, \\ \text{where } \lambda^{\mathrm{FI}} &:= \frac{\mathcal{G}ab - b^2}{ac - b^2}, \\ \text{with } a &:= (\mathbf{1}'\boldsymbol{\Sigma}_r^{-1}\mathbf{1})(\mathbf{x}_t'\boldsymbol{\Sigma}_x^{-1}\mathbf{x}_t), \\ b &:= \mathbf{1}'\boldsymbol{\Sigma}_r^{-1}\boldsymbol{\Sigma}_{rx}\boldsymbol{\Sigma}_x^{-1}\mathbf{x}_t, \ \text{ and} \\ c &:= \mathsf{Tr}(\boldsymbol{\Sigma}_x^{-1}\boldsymbol{\Sigma}_{rx}'\boldsymbol{\Sigma}_r^{-1}\boldsymbol{\Sigma}_{rx}). \end{aligned} \tag{8}$$

We can see that the portfolio weights are expressed as a convex linear combination of the global minimum variance portfolio and the tangency portfolio.[3] This solution has a similar form to the one obtained from the standard MVO found in Huang and Litzenberger (1988, Sections 3.8 and 3.9).

In addition, the problem for the zero-investment case can be formulated as

$$\begin{aligned} &\min_{\mathbf{A}} \frac{1}{2}\mathsf{Tr}(\boldsymbol{\Sigma}_x \mathbf{A} \boldsymbol{\Sigma}_r \mathbf{A}') \\ &\text{subject to } \ \mathsf{Tr}(\mathbf{A}\boldsymbol{\Sigma}_{rx}) \geq \mathcal{G}, \ \text{ and, } \ \mathbf{1}'\mathbf{A}'\mathbf{x}_t = 0. \end{aligned} \tag{9}$$

The general solution to (9) has the following form:

$$\begin{aligned} \mathbf{w}_t^{\mathrm{ZI}} &= \lambda^{\mathrm{ZI}}\left[\frac{\boldsymbol{\Sigma}_r^{-1}\boldsymbol{\Sigma}_{rx}\boldsymbol{\Sigma}_x^{-1}\mathbf{x}_t}{\mathbf{1}'\boldsymbol{\Sigma}_r^{-1}\boldsymbol{\Sigma}_{rx}\boldsymbol{\Sigma}_x^{-1}\mathbf{x}_t} - \frac{\boldsymbol{\Sigma}_r^{-1}\mathbf{1}}{\mathbf{1}'\boldsymbol{\Sigma}_r^{-1}\mathbf{1}}\right], \\ \text{where } \lambda^{\mathrm{ZI}} &:= \mathcal{G}\left[\frac{\mathsf{Tr}(\boldsymbol{\Sigma}_{rx}\boldsymbol{\Sigma}_x^{-1}\boldsymbol{\Sigma}_{rx}'\boldsymbol{\Sigma}_r^{-1})}{\mathbf{1}'\boldsymbol{\Sigma}_r^{-1}\boldsymbol{\Sigma}_{rx}\boldsymbol{\Sigma}_x^{-1}\mathbf{x}_t} - \frac{\mathbf{1}'\boldsymbol{\Sigma}_r^{-1}\boldsymbol{\Sigma}_{rx}\boldsymbol{\Sigma}_x^{-1}\mathbf{x}_t}{(\mathbf{1}'\boldsymbol{\Sigma}_r^{-1}\mathbf{1})(\mathbf{x}_t'\boldsymbol{\Sigma}_x^{-1}\mathbf{x}_t)}\right]^{-1}. \end{aligned} \tag{10}$$

The zero-investment weights prescribes going long one dollar in the tangency portfolio and short one dollar in the global minimum variance portfolio, up to some scaling factor λ^{ZI}.

[3] The tangency portfolio is a portfolio of risky assets that has the highest Sharpe Ratio. It is the solution (4) but with weights summing to one and it does not depend on $\mathcal{G}$.

2.4 Relation to Existing Literature

At this juncture, it is important to highlight the differences of our work from those of the existing literature. We first note that our optimizer can be seen as a natural generalization to the classic MVO of Markowitz (1952) up to a scaling factor when the cross-covariance matrix and covariance matrix of signals are precisely the identity matrix. A slight difference is that, even under these restrictions, our framework still yields an optimal portfolio that is dynamic; this is in contrasts to one obtained from the static formulation of the classical MVO.

We also note that the objective function under our consideration in (4) is similar to the quadratic utility function of Brandt and Santa-Clara (2006), and both formulations lead to the same solution. To see the latter, note that we can also express the optimal policy matrix $\mathbf{A}$ from (5) in vectorized form as

$$\begin{aligned}\mathsf{vec}(\mathbf{A}) &= \lambda \cdot \mathsf{vec}(\boldsymbol{\Sigma}_x^{-1}\boldsymbol{\Sigma}_{rx}'\boldsymbol{\Sigma}_r^{-1}) \\ &= \lambda(\boldsymbol{\Sigma}_x^{-1} \otimes \boldsymbol{\Sigma}_r^{-1})\mathsf{vec}(\boldsymbol{\Sigma}_{rx}') \\ &= \lambda(\boldsymbol{\Sigma}_x \otimes \boldsymbol{\Sigma}_r)^{-1}\mathsf{vec}(\boldsymbol{\Sigma}_{rx}') \\ &= \lambda \cdot \mathbb{E}[(\mathbf{x}_t\mathbf{x}_t') \otimes (\mathbf{r}_t\mathbf{r}_t')]^{-1}\mathbb{E}[\mathbf{x}_t \otimes \mathbf{r}_t],\end{aligned} \tag{11}$$

where $\mathsf{vec}(\cdot)$ is an operator that stacks the columns of matrix $\mathbf{A}$ into a vector of dimension N^2, and $\otimes$ is the Kronecker product of two matrices. This is precisely the solution that one obtains by rewriting the strategy returns in the objective function (2) as $\mathbf{x}_t'\mathbf{A}\mathbf{r}_t = \mathsf{vec}(\mathbf{A})(\mathbf{x}_t \otimes \mathbf{r}_t)$, and optimizing over $\mathsf{vec}(\mathbf{A})$ with respect to the conditional product returns $\mathbf{x}_t \otimes \mathbf{r}_t$.[4]

Brandt and Santa-Clara (2006) 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 combining them in one step, thus reducing the noise in the allocation process. However, there are two potential issues with this approach. First, it results in a severe expansion of the problem dimension. The implications are that the problem becomes more exposed to the curse of dimensionality, and it 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 the optimal solution since it treats both the returns and the signals as homogeneous objects. This is in contrast to the solution that we derived from Equation (5), which highlights the distinct role of the signals and returns, as well as their interactions in the final investment decision.

Finally, our work is also related to that of Kelly et al. (2020) on cross-predictability.

[4] A subtle difference worth pointing out is that Brandt and Santa-Clara (2006) models the portfolio policy to be *affine* in the state-variables, and hence the allocations are also driven by the traditional asset classes themselves. Here, we work in a slightly restricted setting where the portfolio policy is only driven by the trading strategies themselves.

In their work, the authors optimize an objective function that is subject to a robust risk constraint that controls for leverage of $\mathbf{A}$. As a result, the cross-covariance is the only object that they have to confront. However, this departs from our setting on two fronts. First, we choose to actually minimize the variance of the portfolio subject to a target return constraint in a similar spirit to the Markowitz (1952) problem. Second, our application of CCA suggests that the cross-correlation matrix is the more appropriate object of analysis, not the cross-covariance as the latter potentially mixes up the non-trivial correlations that are embedded in both variables that may obscure the inference.

3 Strategy Diversification

3.1 Canonical Correlation Analysis

The goal of CCA is to perform dimension reduction on two different data sets that comprise of 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 pair variables, the canonical variates, 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 solve the classical 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; Hardoon et al., 2004), or through singular value decomposition (SVD) (Healy, 1957). We adopt the latter approach in this brief exposition of the classical subject but emphasize a sequence of two change of basis operations to arrive at a reduced problem whose financial interpretation we provide in the next subsection.

Let $\mathbf{r}$ and $\mathbf{x}$ be random vectors with population covariance and cross-covariances. Given that scale invariance is an important property of correlation, the first step is to orthogonalize the random vectors $\mathbf{r}$ and $\mathbf{x}$ such that their covariance matrices are the identity matrices. This can be achieved through the following linear transformations $\tilde{\mathbf{r}} := \mathbf{\Sigma}_r^{-1/2}\mathbf{r}$ and $\tilde{\mathbf{x}} := \mathbf{\Sigma}_x^{-1/2}\mathbf{x}$. The transformed objects are now invariant to coordinate changes and their joint covariance matrix is:

$$\begin{pmatrix} \mathbb{I}_N & \mathbf{\Sigma}_r^{-1/2}\mathbf{\Sigma}_{rx}\mathbf{\Sigma}_x^{-1/2} \\ \mathbf{\Sigma}_x^{-1/2}\mathbf{\Sigma}_{rx}'\mathbf{\Sigma}_r^{-1/2} & \mathbb{I}_N \end{pmatrix}, \tag{12}$$

where $\mathbb{I}_N$ is an identity matrix of dimensions $N \times N$. It is helpful to introduce the following object

$$\boldsymbol{\Sigma}_{\tilde{r}\tilde{x}} := \boldsymbol{\Sigma}_x^{-1/2}\boldsymbol{\Sigma}_{rx}'\boldsymbol{\Sigma}_r^{-1/2}, \tag{13}$$

as the cross-covariance matrix between the transformed variables $\tilde{\mathbf{r}}$ and $\tilde{\mathbf{x}}$ or the cross-correlation matrix between the original variables $\mathbf{r}$ and $\mathbf{x}$.

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

Finally, since the solution is expressed in a different coordinate system than our original problem, we have to translate back to our original coordinate system. By inverting the change of variables we made, the ith canonical variate can be written as $(\mathbf{u}_i'\boldsymbol{\Sigma}_r^{-1/2}\mathbf{r}, \mathbf{v}_i'\boldsymbol{\Sigma}_x^{-1/2}\mathbf{x})$. This pair can be seen as a linear combination of the original variables with coefficients given by $(\boldsymbol{\Sigma}_r^{-1/2}\mathbf{u}_i, \boldsymbol{\Sigma}_x^{-1/2}\mathbf{v}_i)$, 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 moments $\boldsymbol{\Sigma}_r, \boldsymbol{\Sigma}_x$, and $\boldsymbol{\Sigma}_{rx}$, with their sample counterparts $\mathbf{S}_r, \mathbf{S}_x$, and $\mathbf{S}_{rx}$ computed from random samples $\mathbf{r}_1, \ldots, \mathbf{r}_T$ and $\mathbf{x}_1, \ldots, \mathbf{x}_T$. Let $((\hat{s}_1, \ldots, \hat{s}_N); (\hat{\mathbf{u}}_1, \ldots, \hat{\mathbf{u}}_N); (\hat{\mathbf{v}}_1, \ldots, \hat{\mathbf{v}}_N))$ be a system of singular values (sorted in increasing order) and singular vectors of the sample cross-correlation matrix $\mathbf{S}_{\tilde{r}\tilde{x}}$. Then the sample canonical variates of $\mathbf{r}$ and $\mathbf{x}$ are given by $\hat{\mathbf{u}}_1'\tilde{\mathbf{r}}, \ldots, \hat{\mathbf{u}}_N'\tilde{\mathbf{r}}$ and $\hat{\mathbf{v}}_1'\tilde{\mathbf{x}}, \ldots, \hat{\mathbf{v}}_N'\tilde{\mathbf{x}}$, respectively.

However, one should be skeptical of sample canonical correlations as they are not reflective of the 'true' canonical correlations. Indeed, the 'true' canonical correlation of the ith canonical variate pair $(\hat{\mathbf{u}}_i'\tilde{\mathbf{r}}, \hat{\mathbf{v}}_i'\tilde{\mathbf{x}})$ is $\hat{\mathbf{v}}_i'\boldsymbol{\Sigma}_{\tilde{r}\tilde{x}}'\hat{\mathbf{u}}_i$, as opposed to $\hat{s}_i = \hat{\mathbf{v}}_i'\mathbf{S}_{\tilde{r}\tilde{x}}'\hat{\mathbf{u}}_i$. 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 true correlation matrix of $\tilde{\mathbf{r}}$ and $\tilde{\mathbf{x}}$ is the zero matrix. Figure 1 shows the distribution of sample canonical correlations for various values of the ratio $N/T \in \{0.5, 0.05, 0.005\}$. We obtained this figure by using the asymptotic formula derived by Wachter (1980) restricted to the setting where the dimensions of both variables are the same. From the Figure 1, we observe that the largest sample canonical correlations are biased upwards away from zero and the severity of the bias worsens as N/T increases. This distributional shape is expected to hold when N and T are approximately large and it only depends on the ratio N/T; this is true regardless of any particular realization of the sample cross-correlation matrix.

3.2 Reformulation as Canonical Portfolios

The strategy returns $\mathbf{x}_t'\mathbf{A}\mathbf{r}_t$ at time t obtained with (5) 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 to.

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

$$\mathbf{x}_t'\mathbf{A}\mathbf{r}_t = \lambda \cdot \mathbf{x}_t'\boldsymbol{\Sigma}_x^{-1}\boldsymbol{\Sigma}_{rx}'\boldsymbol{\Sigma}_r^{-1}\mathbf{r}_t = \lambda \cdot \tilde{\mathbf{x}}_t'\tilde{\boldsymbol{\Sigma}}_{\tilde{r}\tilde{x}}\tilde{\mathbf{r}}_t = \tilde{\mathbf{x}}_t'\mathbf{B}\tilde{\mathbf{r}}_t, \tag{14}$$

where $\mathbf{B} := \boldsymbol{\Sigma}_x^{1/2}\mathbf{A}\boldsymbol{\Sigma}_r^{1/2}$ is a policy matrix in the new basis. Since the decorrelated objects $\tilde{\mathbf{r}}_t$ and $\tilde{\mathbf{x}}_t$ span the same universe as the original variables, 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 (14), we can see that the strategy return depends intricately on the synthetic asset returns and synthetic signals, which are coupled through their cross-correlations. The strategy returns do not depend on whether we swap the role of the asset returns of signals; what matters is how they are correlated and their bi-directional relationships. Since a cross-correlation 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.

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

$$\mathbf{x}_t'\mathbf{A}\mathbf{r}_t = \lambda \sum_{i=1}^{N} s_i(\mathbf{v}_i'\tilde{\mathbf{x}}_t)(\mathbf{u}_i'\tilde{\mathbf{r}}_t). \tag{15}$$

Thus, we see that a generic strategy return can be viewed either as a combination of the original securities and signals that has been optimally blended with the policy matrix $\mathbf{A}$, or as a combination of the uncorrelated long-short portfolios weighted by their singular values.

Economically, we interpret the change of basis as an operation that reorganizes the original set of N synthetic assets and N synthetic signals into a set of N uncorrelated strategies. Each of these strategies is expressed as certain weighted combination of the original assets and signals. The particular combinations are determined by the canonical directions, which we shall refer to as *canonical portfolios*.

Furthermore, we can work out the expected returns and variance corresponding to these canonical long-short portfolios. Denote $\pi_i := (\mathbf{v}_i'\tilde{\mathbf{x}}_t)(\mathbf{u}_i'\tilde{\mathbf{r}}_t)$, to be the return that the ith canonical portfolio generates. Since $\mathbf{r}_t$ and $\mathbf{x}_t$ are N-dimensional Gaussian variables, then their ith projections on to their respective canonical portfolios given $\mathbf{u}_i'\tilde{\mathbf{r}}_t$ and $\mathbf{v}_i'\tilde{\mathbf{x}}_t$ are also Gaussian variables with correlation s_i. Then using a result from Firoozye and Koshiyama (2020), the expected return and Sharpe Ratio of the ith canonical portfolio are:

$$\mathbb{E}[\pi_i] = \mathbf{v}_i'\mathbf{\Sigma}_{\tilde{r}\tilde{x}}'\mathbf{u}_i = s_i, \quad \text{and} \quad \mathsf{SR}_i := \frac{\mathbb{E}[\pi_i]}{\sqrt{\mathsf{Var}[\pi_i]}} = \frac{s_i}{\sqrt{1+s_i^2}}. \tag{16}$$

The correlation and Sharpe Ratio of these strategies can be ranked according to singular values s_i, which correspond to the correlation of these canonical portfolios. The higher the ith singular value, the higher the return of the ith canonical portfolio and hence, its Sharpe Ratio. Put 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 portfolios such as those constructed by PCA.

Circling back to Equation (15), the source of strategy return can be seen to be distributed across the canonical portfolios of $\mathbf{\Sigma}_{\tilde{r}\tilde{x}}$. This reinforces the economic notion of diversification: having exposure to different uncorrelated canonical portfolios is akin to putting each of your eggs into a different baskets. We see that the capital assigned to each canonical portfolio is proportional to its original returns $\mathbf{r}_t$ and signal $\mathbf{x}_t$, and proportional to the correlation s_i. Put it differently, we want to assign capital to strategies that have high predictive power with high correlations but are also orthogonal to each other.

3.3 In-Sample, True, and Out-of-Sample Sharpe Ratio

With our CCA decomposition, we can also gain some insight into how the canonical portfolios impact the risk-adjusted returns of a portfolio. To facilitate the analysis in this section, suppose that the random vectors $\tilde{\mathbf{r}}$, and $\tilde{\mathbf{x}}$, as well as the random samples $\tilde{\mathbf{r}}_t$, and $\tilde{\mathbf{x}}_t$ for $t = 1, \ldots, T$ have identity covariances. This means that the cross-covariance matrix coincide with the cross-correlation matrix.

We define the true, in-sample, and out-of-sample Sharpe Ratio of the strategy returns as

$$\mathsf{SR} := \frac{\mathsf{Tr}(\mathbf{B}\boldsymbol{\Sigma}_{\tilde{r}\tilde{x}})}{\sqrt{\mathsf{Tr}(\mathbf{B}\mathbf{B}')}}, \quad \widehat{\mathsf{SR}} := \frac{\mathbb{E}[\mathsf{Tr}(\hat{\mathbf{B}}\mathbf{S}_{\tilde{r}\tilde{x}})]}{\sqrt{\mathbb{E}[\mathsf{Tr}(\hat{\mathbf{B}}\hat{\mathbf{B}}')]}}, \quad \text{and} \quad \mathsf{SR}^{\circ} := \frac{\mathbb{E}[\mathsf{Tr}(\hat{\mathbf{B}}\boldsymbol{\Sigma}_{\tilde{r}\tilde{x}})]}{\sqrt{\mathbb{E}[\mathsf{Tr}(\hat{\mathbf{B}}\hat{\mathbf{B}}')]}}, \tag{17}$$

where $\hat{\mathbf{B}}$ is the policy matrix that replaces the population moments in the transformed policy matrix $\mathbf{B}$ with sample-based estimates. The Sharpe Ratio SR°, is an out-of-sample performance measure that is conditional on the in-sample policy matrix.

The in-sample, true, and out-of-sample Sharpe Ratios are related to the canonical correlations as follows (see Appendix C.5):

$$\mathsf{SR} = \sqrt{\sum_{i=1}^{N} s_i^2}, \quad \widehat{\mathsf{SR}} = \sqrt{\sum_{i=1}^{N} \mathbb{E}[\hat{s}_i^2]}, \quad \text{and} \quad \mathsf{SR}^{\circ} = \frac{\sum_{i=1}^{N} \mathbb{E}[\hat{s}_i s_i^{\circ}]}{\sqrt{\sum_{i=1}^{N} \mathbb{E}[\hat{s}_i^2]}}, \tag{18}$$

where $s_i^{\circ} := \hat{\mathbf{u}}_i' \boldsymbol{\Sigma}_{\tilde{r}\tilde{x}} \hat{\mathbf{v}}_i$ is the out-of-sample correlation associated with the vectors $\hat{\mathbf{u}}_i$ and $\hat{\mathbf{v}}_i$. We see that the Sharpe Ratio of the strategy returns is the sum of the contributions from the canonical correlations; the larger the expected return of the ith canonical portfolio, the larger its contribution towards the overall Sharpe Ratio. While both the true and in-sample Sharpe Ratios are always non-negative, the out-of-sample one may take on negative values.

We can also make further statements about the relationship between the different Sharpe Ratios. Indeed, by using convexity arguments, one can show that, $\widehat{\mathsf{SR}} \geq \mathsf{SR}^{\circ}$; see Appendix C.6.[5] Hence, on average, the in-sample Sharpe Ratio is 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 Sharpe Ratio are more in sync, we have to shrink the in-sample singular values and align singular vectors closer towards the truth.

4 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 large-dimensional objects are generally unknown to us, and so in order to render our framework to practice, we have to estimate them with real data. We will tackle these problems in the following subsections.

[5]In fact, a more precise relationship between the in-sample and out-of-sample Sharpe Ratios and their dependency on the problem dimension and number of observations have been established by Benaych-Georges et al. (2019, Proposition 2.9).

4.1 Covariances

The challenge of estimating the covariance matrix of financial returns $\mathbf{\Sigma}_r$ 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 obervation led Michaud (1989) refer to mean-variance optimizers as "error maximization".

There have been many proposals put forth to address the issues in the sample covariance matrix, many of which can be classified as either structured or structure-free. Structure-based estimators aim to reduce the effective number of risk parameters by incorporating prior knowledge into the estimation process such as sparsity (Bickel and Levina, 2008), graph models (Rajaratnam et al., 2008), or factor structure (Fan et al., 2012).

In constrast, structure-free estimators does not require us to take a stance on the structure of the covariance. They typically belong to the framework of rotational invariance of Stein (1975, 1986), which postulates that the sample eigenvectors should be preserved but sample eigenvalues are allowed to be modified. An example that falls under this class of estimators is the linear shrinkage of Ledoit and Wolf (2004), which shrinks the sample eigenvalues towards their grand mean. A more elaborate extension upon this framework that allows for greater flexibility in the shrinkage would be the analytical-based approach from Ledoit and Wolf (2012, 2015, 2020), and Bun et al. (2017) or numerical-based approach of Abadir et al. (2014) and Lam (2016). In particular, the former approach leverages upon the machinery of large-dimensional asymptotics to consistently estimate the optimal shrinkage function, while the latter appeals to brute-force spectral decompositions for this purpose. In this paper, we shall remain agnostic to the underlying structure of the financial returns and confine ourselves to structure-free approaches to obtain an estimate $\hat{\mathbf{\Sigma}}_r$ for the covariance of returns.

On the other hand, there is less research emphasis around the covariances of signals $\mathbf{\Sigma}_x$ even though, generically speaking, they suffer from the same problems. This can be attributed to the fact that the construction of the vector of signals is largely idiosyncratic and based on individual expectations of how the assets will perform. As such, every different signal design would correspond to a covariance that would have to be tailored specifically to the nature of that signal. Given the lack of universal structure for signal construction, one can resort to structure-free approaches[6] to obtain an estimate $\hat{\mathbf{\Sigma}}_x$ for the covariance of signals.

[6]This also includes the identity matrix as the simplest choice.

4.2 Cross-Covariances

We can also introduce an estimator for the cross-covariance matrix $\mathbf{\Sigma}_{rx}$. However, instead of regularizing cross-covariance matrix directly, we propose to regularize the cross-*correlation* matrix. There are two reasons for this. First, our CCA framework implies that the cross-correlation matrix is the central object for analysis 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 policy matrix for our problem (4) (under an appropriate change of basis) is indeed the cross-correlation matrix, it makes sense for us to focus our analysis on this object. Second, by regularizing the cross-covariances directly, one may inadvertently destroy important cross-sectional variations that are embedded in both variables.

To this end, we start by considering a special case where the population covariances of the asset returns and signals are precisely the identity matrix, so that the cross-covariance matrix coincides with the cross-correlation matrix. We propose to set our cross-correlation estimator to be precisely the constant correlation model, that is

$$\hat{\mathbf{\Sigma}}_{\tilde{r}\tilde{x}} := \hat{\varphi}\mathbb{I}_N + \hat{\rho}\mathbb{J}_N, \tag{19}$$

where $\mathbb{J}_N := \mathbf{11'} - \mathbb{I}_N$ is a matrix with ones in the off-diagonals and zeros on the diagonals, and $\hat{\varphi}$ and $\hat{\rho}$ are two scalars values whose values can be estimated consistently from the sample cross-correlation matrix $\mathbf{S}_{\tilde{r}\tilde{x}}$; see Appendix C.7. Given that the proposed target matrix is symmetric, we are enforcing a symmetry in the values of the off-diagonal elements of the cross-correlation matrix; that is, the magnitude and sign predictability of signal i on asset j is equal to that of signal j on asset i. We believe that this strict form of regularization is a reasonable yet parsimonous restriction to impose into our problem in order to reduce the number of parameters from N^2 to 2. This also amounts to constraining the left and right singular vectors to be equal, and letting the singular values take on two values. We check if this assumption is plausible with the empirical experiments in Section 5.3.

In a more general setting where the population covariances are likely to be different from the identity matrix, it is more appropriate to consider a regularized version of the sample cross-correlation matrix of the following form

$$\hat{\mathbf{\Sigma}}_r^{-1/2}\mathbf{S}_{rx}\hat{\mathbf{\Sigma}}_x^{-1/2}. \tag{20}$$

The pre- and post-multiplication of the sample cross-covariance matrix by the estimated covariances help to prevent the singularity issues that might arise in the cross-correlation matrix when $N > T$. At the same time, it allows the covariances of the synthetic assets $\tilde{\mathbf{r}}_t$

and signals $\tilde{\mathbf{x}}_t$ to be closer to the identity matrix. The parameters of $\hat{\boldsymbol{\Sigma}}_{\tilde{r}\tilde{x}}$ will now be based upon the regularized sample cross-correlation matrix (20).

Once the cross-correlation matrix has been estimated, we rebuild the estimated cross-covariance matrix according to

$$\hat{\boldsymbol{\Sigma}}_{rx} := \hat{\boldsymbol{\Sigma}}_r^{1/2} \hat{\boldsymbol{\Sigma}}_{\tilde{r}\tilde{x}} \hat{\boldsymbol{\Sigma}}_x^{1/2}. \tag{21}$$

From this, we can see that the cross-covariance matrix is allowed to have cross-covariations that are inherited from the covariances of the asset returns and the signals but when they are appropriately normalized, the cross relationships become homogenized. Hence, cross-asset signal-return predictability is still being leveraged upon in our framework.

Remark 3 (Regularized CCA)**.** The regularized sample cross-correlation matrix (20) has close similarity to that of Vinod (1976), where the author proposed ridge regression as a means of regularizing the sample cross-correlation matrix. This approach requires tuning two hyperparameters, which are estimated through cross-validation on a two-dimensional surface. Generally, this is not computationally appealing in our setting where we have to perform numerous optimizations to update the portfolio weights over time. For this purpose, we resort to analytical based methods to address the instabilities in the covariances.

Remark 4 (Positive semi-definiteness)**.** Note that by modifying the cross-covariances, we may run into the situation where the composite covariance $\hat{\boldsymbol{\Sigma}}$ consisting of the estimated objects will no longer be positive semi-definite. The following proposition provides a condition for which this requirement is satisfied.

Proposition 1. The resulting estimated covariance matrix $\hat{\boldsymbol{\Sigma}}$ of the composite vector $(\mathbf{r}_t, \mathbf{x}_t)$ under our proposed regularization schemes is a positive-definite symmetric matrix.

Proof. From Albert (1969), the covariance matrix $\hat{\boldsymbol{\Sigma}}$ is positive-semidefinite if and only if $\hat{\boldsymbol{\Sigma}}_r$ and the Schur complement $\hat{\boldsymbol{\Sigma}}_x - \hat{\boldsymbol{\Sigma}}'_{rx} \hat{\boldsymbol{\Sigma}}_r^{-1} \hat{\boldsymbol{\Sigma}}_{rx}$ is positive-semidefinite. Based on the definition of $\hat{\boldsymbol{\Sigma}}_{rx}$ given in (21), this reduces to asking if the difference $\mathbb{I}_N - \hat{\boldsymbol{\Sigma}}_{\tilde{r}\tilde{x}}$ is positive semidefinite. Since our proposed cross-correlation estimator for $\hat{\boldsymbol{\Sigma}}_{\tilde{r}\tilde{x}}$ in (19) is a symmetric matrix, a spectral decomposition can be performed and given that the identity matrix is invariant under rotations, this condition can be verified if the eigenvalues of $\hat{\boldsymbol{\Sigma}}_{\tilde{r}\tilde{x}}$ are smaller than one. $\square$

4.3 Regularized Canonical Portfolios

We assemble all of the three estimated covariances, $\hat{\boldsymbol{\Sigma}}_r$, $\hat{\boldsymbol{\Sigma}}_x$, and $\hat{\boldsymbol{\Sigma}}_{rx}$ from Section 4.1 and 4.2 and feed them into the linear portfolio policies with respect to the unconstrained portfolio (6), the fully-invested portfolio (8), and dollar-neutral portfolio (10). This provides us with feasible weights $\hat{\mathbf{w}}_t$ that can be computed from real data. We christen all of our linear portfolio policies that specifically deploys the cross-covariance matrix (21) with a constant correlation model (20) as *regularized canonical portfolio* (RCP). Our proposed RCPs distinguish themselves from other linear portfolio policies within our framework that attempt to modify the cross-covariance matrix, and the classic MVO of Markowitz (1952) (which implicitly assumes that the cross-covariance and the covariance of signals are identity covariances).

5 Empirical Analysis

5.1 Data and Portfolio Construction Rules

For our empirical analysis, we obtain daily stock returns data from the Center for Research in Security Prices (CRSP), starting in 01/01/1995 and ending in 31/12/2020. Our analysis is based on the stocks from the NYSE, AMEX and NASDAQ stock exchanges. The size of the investment universe we consider is $N \in \{30, 50, 100, 250, 500\}$. The portfolios that we construct are rebalanced on a daily basis using a rolling window scheme where only past information is used to avoid a look-ahead bias. In particular, the covariance matrices are estimated using a in-sample period of size, $T = 252$. The strategy returns are evaluated using an out-of-sample period of length $T_{\text{out}} = 1$. The out-of-sample period starts from 28/01/1999 to 31/12/2019. This provides us with a total of $h = 5265$ days of consecutive, nonoverlapping observations for which the portfolios are rebalanced on. For convenience, let $\tau_j = T + j \times T_{\text{out}} + 1$ index the first day in the out-of-sample period for a given day $j \in \{0, \ldots, h-1\}$.

We obtain a well-defined investment universe which we can estimate the covariances and cross-covariance on using the following rules, similar to that of Engle et al. (2019). For each rebalancing date $j = 0, 1, \ldots, h-1$ (using a zero-based indexing), we first select the stocks that have complete data over in-sample period and out-of-sample period. Then, we search for pairs of highly correlated stocks (that is, those with a sample correlation that exceeds 0.95) and remove the stock with the lower volume in each pair. From this remaining set of stocks, we select the largest N stocks, as measured by their market capitalization on the rebalancing day h, to include in our investment universe. This construction helps mitigate the survivorship bias and allows our portfolio universe to evolve gradually over time.

5.2 Signal Construction

We consider reversal-based strategies due to their stationary behavior along the time series dimension. From the academic perspective, there have been several studies that postulate the economic rationale for such strategies as a measure of liquidity-provision of returns; for example, see Campbell et al. (1993), Pástor and Stambaugh (2003), and Nagel (2012).

To construct a signal $\mathbf{x}_t$, we use a variant of the cross-sectional short-term reversal strategy of Lehmann (1990), which we refer to as a reversal. In particular, the signal of asset i is computed as the negative of a simple moving average of the returns over past 21 days. The panel of time series signals are then cross-sectionally normalized in every period. That is, we normalize each signal by first centering it cross-sectionally and then dividing by sum of squared deviations from the mean of all stocks. A winsorization of the signals is then applied at each period so that the portfolio is insensitive to outliers; following Chincarini and Kim (2006, page 180), we convert the normalized signals with absolute values greater than 3 into 3 or -3. We collect the individual reversals of the N stocks from the portfolio universe to yield a predictive signal of the returns $\mathbf{r}_t$ at time t.

5.3 Cross-Correlation Instabilities

We check if our hypothesis of negligible cross-correlations between the asset returns and the signals is plausible by estimating the regularized sample cross-correlation matrix in (20) for $N \in \{100, 500\}$ using an sample of size $T = 252$, based on the rolling window scheme outlined in Section 5.1. In this section, the covariances are obtained using the linear shrinkage of Ledoit and Wolf (2004). At each period, we compute the median and 25th and 75th percentiles of the diagonal and off-diagonal elements from sample cross-correlations. Figure 2 plots the time series evolution of these statistics that summarize the diagonal and off-diagonal elements.

As one can observe, the median of the diagonal elements (orange line) of the cross-correlation matrix are a factor of at least 100 times higher than the off-diagonal ones (blue line). This is not unexpected since by construction, we have constructed the signals such that they predict their corresponding return series. We also see that the median of the off-diagonal elements is consistently close to zero. By contrast, the median values of the diagonal elements are highly time-varying but also positive and small in magnitude. This reflects some self-association between the synthetic returns and their corresponding synthetic signals, which we can exploit.

Also, note that in both the time series statistics, we see that there is quite a bit of cross-sectional dispersion around their median values. These measurement errors may result in instabilities for our portfolio, and unreported results show that using the sample cross-correlations (or its regularized variants that only shrink the singular values while maintain the singular vectors) directly display poor out-of-sample properties. Thus, in order to ensure all of our canonical portfolios exhibit stable out-of-sample results, we collapse these cross-sectional dispersions of the diagonal and off-diagonal elements towards their respective means; this is achieved with the RCP from Section 4.3.

5.4 Candidate Portfolios

We restrict our empirical analysis to the default portfolio selection problem without weight constraints. The following portfolios are considered in our study:

- BSV: the portfolio (6) where $\hat{\mathbf{\Sigma}}_r$, $\hat{\mathbf{\Sigma}}_x$ and $\hat{\mathbf{\Sigma}}_{rx}$ are estimated with the identity matrix $\mathbb{I}_N$. This portfolio is based on the proposal of Brandt et al. (2009).
- MVO-SC: the portfolio (6) where $\hat{\mathbf{\Sigma}}_r$ is estimated with the sample covariance matrix estimator itself, and $\hat{\mathbf{\Sigma}}_{rx}$ and $\hat{\mathbf{\Sigma}}_x$ with the identity matrix $\mathbb{I}_N$. This portfolio is based on the proposal of Markowitz (1952).
- MVO-LS: the portfolio (6) where $\hat{\mathbf{\Sigma}}_r$ is estimated with the linear shrinkage estimator of Ledoit and Wolf (2004), with $\hat{\mathbf{\Sigma}}_{rx}$ and $\hat{\mathbf{\Sigma}}_x$ with the identity matrix $\mathbb{I}_N$. This portfolio is based on the proposal of Markowitz (1952).
- MVO-NL: the portfolio (6) where $\hat{\mathbf{\Sigma}}_r$ is estimated with the nonlinear shrinkage estimator of Ledoit and Wolf (2020), with $\hat{\mathbf{\Sigma}}_{rx}$ and $\hat{\mathbf{\Sigma}}_x$ with the identity matrix $\mathbb{I}_N$. This portfolio is based on the proposal of Markowitz (1952).
- RCP-SC: the portfolio (6) where $\hat{\mathbf{\Sigma}}_r$ and $\hat{\mathbf{\Sigma}}_x$ are estimated with the sample covariance matrices, and $\hat{\mathbf{\Sigma}}_{rx}$ with (21).
- RCP-LS: the portfolio (6) where $\hat{\mathbf{\Sigma}}_r$ and $\hat{\mathbf{\Sigma}}_x$ are estimated with the linear shrinkage estimator of Ledoit and Wolf (2004), and $\hat{\mathbf{\Sigma}}_{rx}$ with (21).
- RCP-NL: the portfolio (6) where $\hat{\mathbf{\Sigma}}_r$ and $\hat{\mathbf{\Sigma}}_x$ are estimated with the nonlinear shrinkage estimator of Ledoit and Wolf (2020), and $\hat{\mathbf{\Sigma}}_{rx}$ with (21).

5.5 Performance Analysis

To evaluate the performance of the different portfolios portfolio, we compute three out-of-sample performance measures annualized over the 252 trading days:

- AV: Annualized average of the $h = 5265$ out-of-sample (daily) returns:

$$\mathrm{AV} = 252 \times \frac{1}{hT_{\text{out}}} \sum_{j=0}^{h-1} \sum_{\tau=\tau_j}^{\tau_j+T_o-1} \hat{\mathbf{w}}'_{\tau_j} \mathbf{r}_{\tau_j}. \tag{22}$$

- SD: Annualized standard deviation of the $h = 5265$ out-of-sample (daily) returns:

$$\text{SD} = \sqrt{252} \times \sqrt{\frac{1}{hT_{\text{out}}} \sum_{j=0}^{h-1} \sum_{\tau=\tau_j}^{\tau_j + T_{\text{out}}} \left(\hat{\mathbf{w}}'_{\tau_j} \mathbf{r}_{\tau_j} \right)^2}. \tag{23}$$

- IR: Information ratio given by the ratio AV/SD.

The primary metric for which we evaluate the all the portfolio performances will be the out-of-sample information ratio.

We summarize the main results from Table 1 as follows:

- For a chosen covariance estimator, our RCP optimizer consistently outperforms the classic MVO over different investment universe sizes N. This informs us that taking into account cross-sectional information that arises from both the signals and their cross-relationships with the underlying returns can enhance the performances. In general, these improvements can range from 10% to 25% when shrinkage is applied.
- The performances of both RCP and MVO schemes are generally better with larger size portfolios N though there does not seem to be significant gains of our portfolio for above 250.
- The application of shrinkage generally improve the performances of RCP and MVO schemes for large values of N with NL offering the best improvement. This is reassuring for us as our RCP optimizer can benefit from the advances of large dimensional covariance modeling.
- BSV is consistently outperformed in terms of IR by all other portfolios with the exception of MVO-SC and RCP-SC for $N \in \{250, 500\}$. Indeed, this is the setting where the sample covariance matrices of the asset returns and signals approaches singularity.

Additionally, we report the following portfolio weight statistics:

- TO: Average (daily) turnover as $\frac{1}{h-1} \sum_{j=0}^{h-2} \|\mathbf{w}_{\tau_j+1} - \mathbf{w}_{\tau_j}\|_1$ where $|| \cdot ||_1$ denotes the ℓ_1-norm.
- GL: Average (daily) gross leverage as $\frac{1}{h} \sum_{j=0}^{h-1} \|\mathbf{w}_{\tau_j}\|_1$
- PL: Average (daily) proportion of leverage as $\frac{1}{hN} \sum_{j=0}^{h-1} \sum_{i=1}^{N} \chi\{w_{\tau_j,i} < 0\}$ where $\chi\{\cdot\}$ denotes the indicator function

Note that these weight statistics are not our primary interest since our problem is not optimized to account for these measures. Nevertheless, we report these results to give a better overview of our methods.

The results from Table 2 are summarized below:

- It can be seen that the turnover for our RCP optimizer tends to be more than twice as large as that from the standard MVO. Thus, this will likely diminish the realized returns after accounting for the costs. The turnover also increases when N becomes larger.
- The gross leverage is seen to be higher for RCP over MVO for all portfolio size, N.
- We can see that the application of shrinkage of the covariances can also reduce the turnover and gross leverage for both MVO and RCP when compared to using the sample covariance matrix.

Remark 5 (Robustness Checks)**.** We have performed several robustness checks to examine whether the outperformance of the RCP scheme over MVO is robust to different adjustments in the empirical set-up; these results are not detailed here for the brevity of this report but we provide the key findings. In particular, we repeated the analysis for (i) different subperiods and (ii) estimation window length. We find that the results are robust for a longer estimation window length of size $T = 504$. However, the subperiod analysis show that while the outperformance of RCP hold for majority of the subperiods over the out-of-sample period, we find that the RCP scheme underperformed the standard MVO in the recent five years. This can be seen in the portfolio return trajectory for $N \in \{100, 500\}$ in Figure 3 and 4, respectively.

6 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 framework generalizes the classic Markowitz (1952) mean-variance optimization (MVO) scheme to account for cross-predictability and uncertainty in the signals. The application of CCA helps to lift the veil of complexity from our problem by breaking down all of the asset returns and signals into independent long–short canonical portfolios. Each of those canonical portfolios is ranked from the one with the smallest correlation to the one with the highest. Our proposed optimizer builds upon Brandt and Santa-Clara (2006) by taking all of these independent trading strategies and scaling them up according to the strength of their canonical correlation; the canonical portfolios with the highest returns get scaled up the most.

Concurrently, the issues that plague CCA in large dimensions naturally spillover into our problem. By recognizing them, we are able to propose solutions to deliver stable out-of-sample results. In particular, we apply shrinkage to the eigenvalues of the sample covariances of the returns and signals while maintaining their eigenvectors, and impose a constant-correlation structure on the cross-correlation matrix to address the instabilities in our problem. By employing these innovations, we are able to demonstrate that our proposed method, regularized canonical portfolio (RCP), can display outperformance over the classic mean-variance optimizer (MVO). However, we stress that a more careful design of the covariance of the signals or cross-covariance matrix of signals and returns may be needed in order to ensure consistent improvement over the classic MVO across different regimes.

Finally, our proposed modeling framework is not set in stone and is flexible enough to accommodate further improvements. Indeed, nonlinear extensions (for example, kernel CCA or nonlinear CCA) such that the portfolio policy depends nonlinearly on the signal is a promising avenue and will be pursued in a forthcoming work, as well the consideration of designing signals and portfolios with maximal mutual information. One can also investigate the application covariances that models the dynamic structure of financial returns through the recent advances from Engle et al. (2019) or Tan and Zohren (2021) on these portfolios. Additionally, one may consider more elaborate settings that allows for a heterogeneous set of signals, or implements constraints on the portfolio weights such as long-only portfolios or transaction costs. Last but not least, by recasting the problem into a CCA framework, it enables researchers to leverage upon the insights and techniques from the rich literature of CCA that has been expanded by developments in machine learning.

A Figures

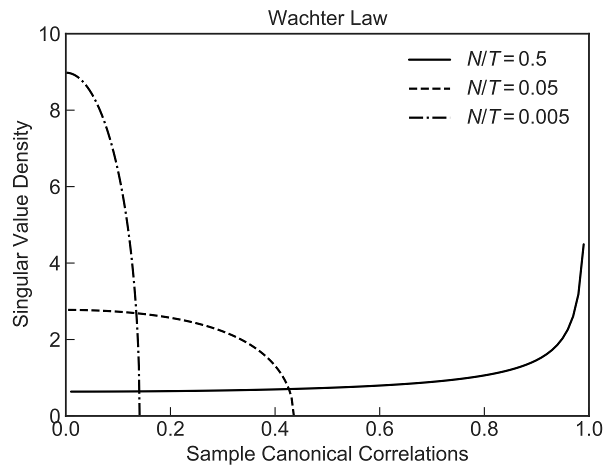

Figure 1: Limiting spectral density of the sample canonical correlations of Wachter (1980) for different ratios, N/T. This assumes identical dimensions for both variables.

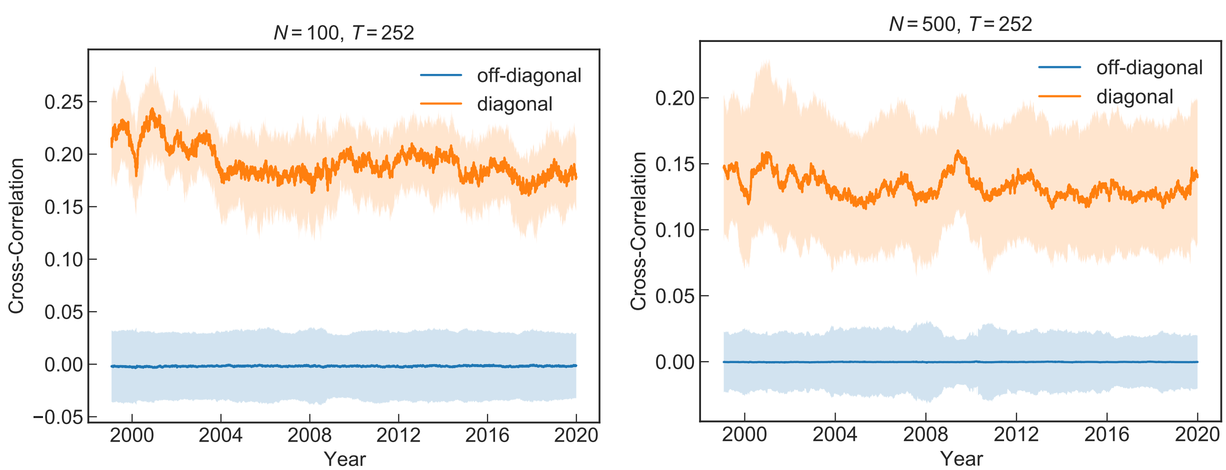

Figure 2: Evolution of the median of the diagonal elements (orange line) and the off-diagonal elements (blue line) of the regularized sample cross-correlations matrix (20) along with the 25th and 75th percentiles. The left and right figures correspond to universe sizes of $N = 100$ and $N = 500$, respectively with $T = 252$.

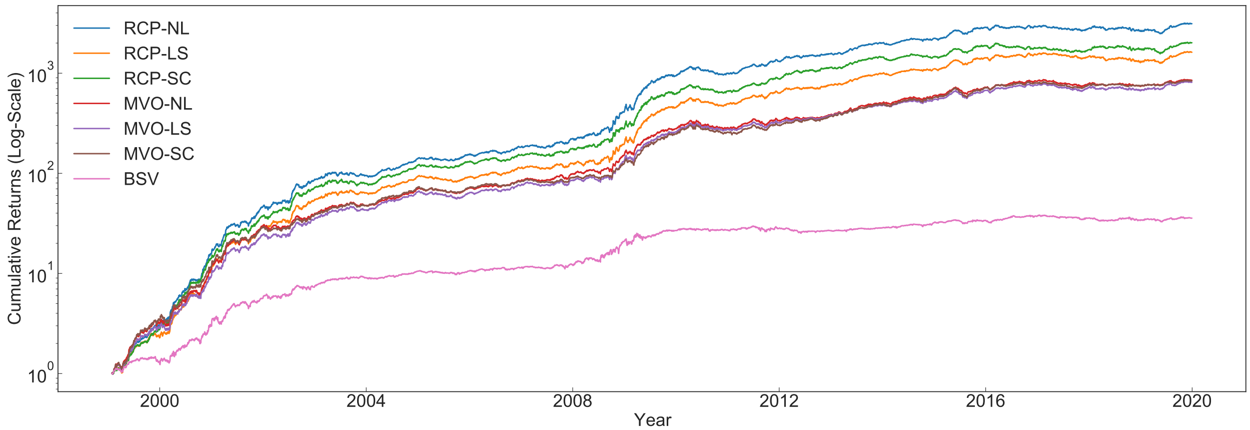

Figure 3: Portfolio return trajectory of the RCP, MVO and BSV schemes for different estimated covariances with universe size $N = 100$.

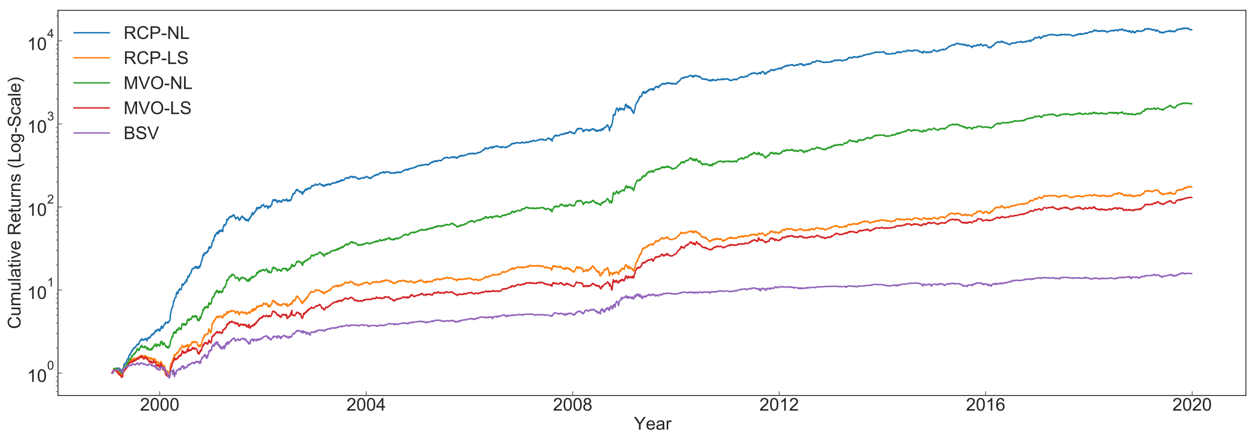

Figure 4: Portfolio return trajectory of the RCP, MVO and BSV schemes for different estimated covariances with universe size $N = 500$.

B Tables

Out-of-Sample Period: 28/01/1999 to 31/12/2019							
	BSV	MVO-SC	MVO-LS	MVO-NL	RCP-SC	RCP-LS	RCP-NL
				$N = 30$			
AV	38.70	20.17	22.14	21.56	52.82	46.22	46.39
SD	46.98	15.38	16.86	17.11	33.49	30.26	29.93
IR	0.82	1.31	1.31	1.26	**1.58**	1.53	1.55
				$N = 50$			
AV	48.39	18.61	21.85	22.53	66.48	54.13	55.22
SD	54.74	13.24	15.54	16.09	34.50	30.10	29.02
IR	0.88	1.41	1.41	1.40	**1.93**	1.80	1.90
				$N = 100$			
AV	84.90	24.62	32.40	37.25	95.51	75.12	73.03
SD	73.37	11.67	15.43	17.62	40.22	32.54	29.13
IR	1.16	2.11	2.10	2.11	2.37	2.31	**2.51**
				$N = 250$			
AV	118.95	-2.64	27.17	38.90	148.45	83.50	77.84
SD	113.82	19.97	12.16	16.42	281.62	35.26	25.99
IR	1.05	-0.13	2.23	2.37	0.53	2.37	**3.00**
				$N = 500$			
AV	141.36	NaN	18.24	52.05	NaN	71.14	60.79
SD	155.22	NaN	11.79	22.33	NaN	43.52	20.60
IR	0.91	NaN	1.55	2.33	NaN	1.63	**2.95**

Table 1: A summary of the annualized out-of-sample average returns (AV), standard deviation of returns (SD) and information ratio (IR) for each combination of investment universe size N, covariance matrix estimators, and optimizers. The covariance matrices are estimated using a lookback window of $T = 252$ days. The daily out-of-sample returns starts from 28/01/1999 and ends in 31/12/2019. The most competitive IR value is highlighted in **bold**.

Out-of-Sample Period: 28/01/1999 to 31/12/2019							
	BSV	MVO-SC	MVO-LS	MVO-NL	RCP-SC	RCP-LS	RCP-NL
$N = 30$							
TO	1.52	1.19	1.22	1.21	3.78	2.87	3.07
GL	4.24	3.05	3.16	3.14	6.93	5.74	5.84
PL	14.97	15.19	15.20	15.17	15.14	15.14	15.13
$N = 50$							
TO	1.95	1.39	1.44	1.43	6.18	3.95	4.36
GL	5.44	3.49	3.72	3.68	9.37	7.11	7.07
PL	24.87	25.29	25.30	25.26	25.18	25.21	25.18
$N = 100$							
TO	2.75	1.65	1.81	1.79	12.32	5.87	6.14
GL	7.65	3.97	4.63	4.57	14.28	9.27	8.25
PL	49.64	50.41	50.28	50.27	50.09	50.18	50.19
$N = 250$							
TO	4.32	9.57	1.97	3.75	218.47	8.41	8.00
GL	12.02	8.88	5.10	6.02	155.55	13.19	8.85
PL	124.18	124.87	125.04	124.56	124.95	125.24	124.65
$N = 500$							
TO	6.01	NaN	2.56	4.50	NaN	13.72	11.44
GL	16.87	NaN	6.37	8.79	NaN	18.88	12.71
PL	247.39	NaN	247.70	247.26	NaN	249.28	247.99

Table 2: A summary of the out-of-sample turnover (TO), gross leverage (GL), and proportional leverage (PL) for each combination of investment universe size N, covariance matrix estimators, and optimizers. The covariance matrices are estimated using a lookback window of $T = 252$ days. The daily out-of-sample returns starts from 28/01/1999 and ends in 31/12/2019.

C Mathematical Derivations

C.1 Derivation of Mean-Variance Objective

In this section, 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 strategy returns is

$$\mathbb{E}[\mathbf{x}'\mathbf{A}\mathbf{r}] = \mathbb{E}[\mathsf{Tr}(\mathbf{A}\mathbf{r}\mathbf{x}')] = \mathsf{Tr}(\mathbf{A}\boldsymbol{\Sigma}_{rx}). \tag{24}$$

In order to analyze the second-order terms, we need to use a theorem from Isserlis (1918) or Wick (1950). We only need the following results from the one-dimensional setting.

Theorem 1 (Wick)**.** Let z_1, z_2, z_3, and z_4 be jointly normal variables with mean zero. Then we have the following results:

$$\begin{aligned}
\mathbb{E}[z_1] &= 0 \\
\mathbb{E}[z_1 z_2] &= \mathsf{Cov}(z_1, z_2) \\
\mathbb{E}[z_1 z_2 z_3] &= 0 \\
\mathbb{E}[z_1 z_2 z_3 z_4] &= \mathbb{E}[z_1 z_2]\mathbb{E}[z_3 z_4] + \mathbb{E}[z_1 z_3]\mathbb{E}[z_2 z_4] + \mathbb{E}[z_1 z_4]\mathbb{E}[z_2 z_3]
\end{aligned} \tag{25}$$

The last term of (25) takes all partitions of size two of the four variables, which gives us three separate terms.

By recasting fourth-order using the covariance, we can express the variance as

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

where $A_{i,j}$ is the (i, j) entry of $\mathbf{A}$. Reverting to matrix notation, we have

$$\mathsf{Var}[\mathbf{x}'\mathbf{A}\mathbf{r}] = \mathsf{Tr}(\boldsymbol{\Sigma}_x\mathbf{A}\boldsymbol{\Sigma}_r\mathbf{A}') + \mathsf{Tr}(\boldsymbol{\Sigma}_{rx}\mathbf{A}\boldsymbol{\Sigma}_{rx}\mathbf{A}). \tag{26}$$

C.2 Derivation of Mean-Variance Policy Matrix

In order to solve the full mean-variance objective function (3), 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 $\mathbf{W}_r$ and $\mathbf{W}_x$ be matrices whose columns contain the so-called ith canonical directions $\mathbf{\Sigma}_r^{-1/2}\mathbf{u}_i$ and $\mathbf{\Sigma}_x^{-1/2}\mathbf{v}_i$, respectively, and $\mathbf{D} := \mathsf{Diag}(s_1, s_2, \dots, s_N)$ be a diagonal matrix containing the canonical correlations. We now state the main theorem:

Theorem 2. Suppose $\mathbf{x}_t$ and $\mathbf{r}_t$ are N-dimensional jointly Gaussian variables. Then the objective function (2) is maximized at

$$\mathbf{A} = \mathbf{W}_x \mathsf{Diag}(L_1, \dots, L_N)\mathbf{W}_r', \tag{27}$$

where diagonal matrix consist of elements that are nonlinear functions of the canonical correlations given by

$$L_i = \lambda \frac{s_i}{1 + s_i^2}, \tag{28}$$

for $i = 1, \dots, N$, and some scaling parameter $\lambda > 0$. The optimal matrix $\mathbf{A}$ based on Equation (27) is effectively the optimal scaling of the canonical variates of the asset returns and signals.

Proof. Without loss of generality, let us reparameterize the policy matrix in terms of the canonical directions as $\mathbf{A} = \mathbf{W}_x \mathbf{L} \mathbf{W}_r'$ where $\mathbf{L}$ is a $N \times N$ matrix that we now have to optimize on. The expected value of the strategy returns at time t can be written as

$$\begin{aligned} \mathbb{E}[\mathbf{x}_t' \mathbf{A} \mathbf{r}_t] &= \mathsf{Tr}(\mathbf{A}\mathbf{\Sigma}_{rx}) & (29) \\ &= \mathsf{Tr}(\mathbf{L}\mathbf{D}), & (30) \end{aligned}$$

and the variance as

$$\begin{aligned} \mathsf{Var}[\mathbf{x}_t' \mathbf{A} \mathbf{r}_t] &= \mathsf{Tr}(\mathbf{\Sigma}_x \mathbf{A} \mathbf{\Sigma}_r \mathbf{A}') + \mathsf{Tr}(\mathbf{\Sigma}_{rx} \mathbf{A} \mathbf{\Sigma}_{rx} \mathbf{A}) & (31) \\ &= \mathsf{Tr}(\mathbf{L}\mathbf{L}') + \mathsf{Tr}(\mathbf{D}\mathbf{L}\mathbf{D}\mathbf{L}). & (32) \end{aligned}$$

Putting all together, the objective function (3) can be written in unconstrained form as:

$$\min_{\mathbf{L}} \frac{1}{2} \left(\mathsf{Tr}(\mathbf{L}\mathbf{L}') + \mathsf{Tr}(\mathbf{D}\mathbf{L}\mathbf{D}\mathbf{L})\right) + \lambda(\mathcal{G} - \mathsf{Tr}(\mathbf{L}\mathbf{D})), \tag{33}$$

where $\lambda > 0$ is the shadow cost of violating the target return constraint. Using the rules for matrix derivatives from Lütkepohl (1997), we take first order conditions with respect to the matrix $\mathbf{L}$ to get

$$\mathbf{L} + \mathbf{D}\mathbf{L}'\mathbf{D} - \lambda\mathbf{D} = 0. \tag{34}$$

Since the matrix $\mathbf{D}$ is diagonal and thus symmetric, we necessarily have the following relationship

$$\lambda\mathbf{D} = \mathbf{L} + \mathbf{D}\mathbf{L}'\mathbf{D} = \mathbf{L}' + \mathbf{D}\mathbf{L}\mathbf{D}. \tag{35}$$

Let $\mathbf{M} := \mathbf{L} - \mathbf{L}'$, which is an anti-symmetric matrix (that is, $\mathbf{M}' = -\mathbf{M}$) and has diagonal elements zero. Then we have

$$\mathbf{M} = \mathbf{D}\mathbf{M}\mathbf{D}. \tag{36}$$

If we restrict our attention to the off-diagonal elements of (36), we see that $M_{i,j} = s_i s_j M_{i,j}$ for all $i \neq j$, so $(1 - s_i s_j)M_{i,j} = 0$. Therefore, either $s_i s_j = 1$ for all $i \neq j$ or $M_{i,j} = 0$. Consequently, $\mathbf{M}$ is identically zero and hence, $\mathbf{L}$ must be a symmetric matrix.

The benefit of having the variable matrix $\mathbf{L}$ to be symmetric is that the condition (34) can now be written as

$$\mathbf{L} + \mathbf{D}\mathbf{L}\mathbf{D} = \lambda\mathbf{D}. \tag{37}$$

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

$$\min_{L_1,\ldots,L_N} \sum_{i=1}^{N} \frac{1}{2}(L_i^2 + s_i^2 L_i^2) - \lambda L_i s_i. \tag{38}$$

Our problem can now be easily solved to give us

$$L_i = \lambda \frac{s_i^2}{1 + s_i^2} = \lambda \cdot \mathsf{SR}_i^2. \tag{39}$$

Hence, the diagonal elements is a nonlinear function of the canonical correlations that also depends on the Sharpe Ratio of the ith canonical portfolio from Equation (16). $\square$

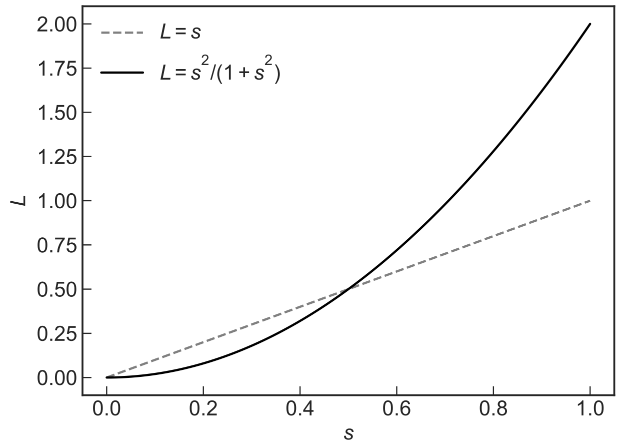

Figure 5: Optimal singular value adjustment versus linear approximation.

C.3 Alternate Derivation of Mean-Variance Policy Matrix

It is convenient to assume that the squared expectation term $\mathbb{E}[\mathbf{x}'\mathbf{A}\mathbf{r}]^2$ is zero. Under this assumption, we can rewrite the objective function (4) in unconstrained form as

$$\min_{\mathbf{B}} \frac{1}{2}\mathsf{Tr}(\mathbf{B}\mathbf{B}') - \lambda \mathsf{Tr}(\mathbf{B}\mathbf{\Sigma}'_{\tilde{r}\tilde{x}}), \tag{40}$$

where we have used a change-of-variables $\mathbf{B} := \mathbf{\Sigma}_x^{1/2}\mathbf{A}\mathbf{\Sigma}_r^{1/2}$. The problem (40) can be solved to yield

$$\mathbf{B} = \lambda \cdot \mathbf{\Sigma}_x^{-1/2}\mathbf{\Sigma}'_{rx}\mathbf{\Sigma}_r^{-1/2}. \tag{41}$$

Since the solution is based on transformed variables, we rescale it back to the original assets through the following operation

$$\mathbf{A} = \mathbf{\Sigma}_x^{-1/2}\mathbf{B}\mathbf{\Sigma}_r^{-1/2} = \lambda \cdot \mathbf{\Sigma}_x^{-1}\mathbf{\Sigma}'_{rx}\mathbf{\Sigma}_r^{-1}, \tag{42}$$

As a result of this approximation, we now have

$$L_i \approx \lambda s_i. \tag{43}$$

Figure 5 contrasts the singular value adjustment in (28) to the linear approximation (43) for $\lambda = 1$. We can see that for small (large) values of s_i, a optimal approximation downweighs (upweighs) the original singular values.

C.4 Derivation of the Optimal Portfolio Policy under Fully-Invested Constraint

In this section, we provide the derivation of the optimal portfolio only for the fully-invested problem (7) since the derivation for the zero-investment (9) follows a similar reasoning. We start by writing down the expression for the Lagrangian

$$\frac{1}{2}\,\mathsf{Tr}(\boldsymbol{\Sigma}_x\mathbf{A}\boldsymbol{\Sigma}_r\mathbf{A}') + \lambda_1(\mathcal{G} - \mathsf{Tr}(\mathbf{A}\boldsymbol{\Sigma}_{rx})) + \lambda_2(1 - \mathbf{1}'\mathbf{A}'\mathbf{x}_t), \tag{44}$$

where λ_1 and λ_2 are the Lagrange multipliers. Using the change-of-variables $\mathbf{B} := \boldsymbol{\Sigma}_x^{1/2}\mathbf{A}\boldsymbol{\Sigma}_r^{1/2}$ that we introduced in Section C.3, we have

$$\frac{1}{2}\,\mathsf{Tr}(\mathbf{B}\mathbf{B}') + \lambda_1(\mathcal{G} - \mathsf{Tr}(\mathbf{B}\boldsymbol{\Sigma}_{\tilde{r}\tilde{x}})) + \lambda_2(1 - \mathbf{1}'\boldsymbol{\Sigma}_r^{-1/2}\mathbf{B}'\boldsymbol{\Sigma}_x^{-1/2}\mathbf{x}_t). \tag{45}$$

Performing the first-order conditions

$$[\text{FOC } \mathbf{B}]:\ \mathbf{B} - \lambda_1\boldsymbol{\Sigma}'_{\tilde{r}\tilde{x}} - \lambda_2\boldsymbol{\Sigma}_x^{-1/2}\mathbf{x}_t\mathbf{1}'\boldsymbol{\Sigma}_r^{-1/2} = 0 \tag{46}$$

$$[\text{FOC } \lambda_1]:\ \mathcal{G} - \mathsf{Tr}(\mathbf{B}\boldsymbol{\Sigma}_{\tilde{r}\tilde{x}}) = 0 \tag{47}$$

$$[\text{FOC } \lambda_2]:\ 1 - \mathbf{1}'\boldsymbol{\Sigma}_r^{-1/2}\mathbf{B}'\boldsymbol{\Sigma}_x^{-1/2}\mathbf{x}_t = 0 \tag{48}$$

Solving for $\mathbf{B}$ in terms of λ_1, λ_2,

$$\mathbf{B} = \lambda_1\boldsymbol{\Sigma}'_{\tilde{r}\tilde{x}} + \lambda_2\boldsymbol{\Sigma}_x^{-1/2}\mathbf{x}_t\mathbf{1}'\boldsymbol{\Sigma}_r^{-1/2}. \tag{49}$$

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

$$\mathbf{w} = \lambda_1\boldsymbol{\Sigma}_r^{-1}\boldsymbol{\Sigma}_{rx}\boldsymbol{\Sigma}_x^{-1}\mathbf{x}_t + \lambda_2(\mathbf{x}'_t\boldsymbol{\Sigma}_x^{-1}\mathbf{x}_t)\mathbf{1}'\boldsymbol{\Sigma}_r^{-1}. \tag{50}$$

We can solve for λ_1 and λ_2 by substituting the solution for $\mathbf{B}$ in Equation (47) and (48):

$$1 = \lambda_1\mathbf{1}'\boldsymbol{\Sigma}_r^{-1}\boldsymbol{\Sigma}_{rx}\boldsymbol{\Sigma}_x^{-1}\mathbf{x}_t + \lambda_2(\mathbf{1}'\boldsymbol{\Sigma}_r^{-1}\mathbf{1})(\mathbf{x}'_t\boldsymbol{\Sigma}_x^{-1}\mathbf{x}_t) \tag{51}$$

$$\mathcal{G} = \lambda_1\mathsf{Tr}(\boldsymbol{\Sigma}_x^{-1}\boldsymbol{\Sigma}'_{rx}\boldsymbol{\Sigma}_r^{-1}\boldsymbol{\Sigma}_{rx}) + \lambda_2\mathbf{1}'\boldsymbol{\Sigma}_r^{-1}\boldsymbol{\Sigma}_{rx}\boldsymbol{\Sigma}_x^{-1}\mathbf{x}_t. \tag{52}$$

Define the following constants

$$a := (\mathbf{1}'\boldsymbol{\Sigma}_r^{-1}\mathbf{1})(\mathbf{x}'_t\boldsymbol{\Sigma}_x^{-1}\mathbf{x}_t),\ \ b := \mathbf{1}'\boldsymbol{\Sigma}_r^{-1}\boldsymbol{\Sigma}_{rx}\boldsymbol{\Sigma}_x^{-1}\mathbf{x}_t,\ \text{ and }\ c := \mathsf{Tr}(\boldsymbol{\Sigma}_x^{-1}\boldsymbol{\Sigma}'_{rx}\boldsymbol{\Sigma}_r^{-1}\boldsymbol{\Sigma}_{rx}). \tag{53}$$

We have a system of two equations with two unknowns, which can be solved to yield

$$\lambda_1 = \frac{\mathcal{G}a - b}{ac - b^2}, \quad \lambda_2 = \frac{c - \mathcal{G}b}{ac - b^2} \tag{54}$$

Note that $\lambda_1 b + \lambda_2 a = 1$. Hence, by defining $\lambda^{\mathrm{FI}} := \lambda_1 b$, we obtain the expression stated in Equation (8).

C.5 Canonical Correlations and Sharpe Ratio

In this section, we continue to assume that the squared expectation of the strategy returns $\mathbb{E}[\mathbf{x}'\mathbf{A}\mathbf{r}]^2$ is zero. Following the definition of the true Sharpe Ratio of the strategy returns from (17) and substituting the expression of $\mathbf{B}$, we have:

$$\mathsf{SR} = \frac{\mathsf{Tr}(\boldsymbol{\Sigma}'_{\tilde{r}\tilde{x}}\boldsymbol{\Sigma}_{\tilde{r}\tilde{x}})}{\sqrt{\mathsf{Tr}(\boldsymbol{\Sigma}'_{\tilde{r}\tilde{x}}\boldsymbol{\Sigma}_{\tilde{r}\tilde{x}})}} = \sqrt{\sum_{i=1}^{N} s_i^2}. \tag{55}$$

Likewise, in-sample Sharpe Ratio can be obtained by replacing the population moments with sample-based estimates:

$$\widehat{\mathsf{SR}} = \frac{\mathbb{E}[\mathsf{Tr}\,(\mathbf{S}'_{\tilde{r}\tilde{x}}\mathbf{S}_{\tilde{r}\tilde{x}})]}{\sqrt{\mathbb{E}[\mathsf{Tr}\,(\mathbf{S}'_{\tilde{r}\tilde{x}}\mathbf{S}_{\tilde{r}\tilde{x}})]}} = \sqrt{\sum_{i=1}^{N} \mathbb{E}[\hat{s}_i^2]}. \tag{56}$$

The out-of-sample Sharpe Ratio is conditional on the in-sample cross-correlation matrix is given by

$$\mathsf{SR}^{\circ} = \frac{\mathbb{E}\left[\mathsf{Tr}\,(\mathbf{S}'_{\tilde{r}\tilde{x}}\boldsymbol{\Sigma}_{\tilde{r}\tilde{x}})\right]}{\sqrt{\mathbb{E}\left[\mathsf{Tr}\,(\mathbf{S}'_{\tilde{r}\tilde{x}}\mathbf{S}_{\tilde{r}\tilde{x}})\right]}} = \frac{\sum_{i=1}^{N} \mathbb{E}[\hat{s}_i s_i^{\circ}]}{\sqrt{\sum_{i=1}^{N} \mathbb{E}[\hat{s}_i^2]}}. \tag{57}$$

C.6 Bias of In-Sample Sharpe Ratio

Note that the scalar function $g(\cdot)$ given by $g(\mathbf{S}_{\tilde{r}\tilde{x}}) := \mathsf{Tr}(\mathbf{S}'_{\tilde{r}\tilde{x}}\mathbf{S}_{\tilde{r}\tilde{x}})$ is a convex function. Thus, using Jensen's inequality, we obtain

$$\mathbb{E}[\mathsf{Tr}\,(\mathbf{S}'_{\tilde{r}\tilde{x}}\mathbf{S}_{\tilde{r}\tilde{x}})] \geq \mathsf{Tr}(\mathbb{E}[\mathbf{S}_{\tilde{r}\tilde{x}}]'\mathbb{E}[\mathbf{S}_{\tilde{r}\tilde{x}}]) = \mathsf{Tr}\,(\boldsymbol{\Sigma}'_{\tilde{r}\tilde{x}}\boldsymbol{\Sigma}_{\tilde{r}\tilde{x}}), \tag{58}$$

where we assume that the sample cross-covariance matrix of the synthetic assets and synthetic signals is an unbiased estimator of $\mathbf{\Sigma}_{\tilde{r}\tilde{x}}$. Hence,

$$\sum_{i=1}^{N} \mathbb{E}[\hat{s}_i^2] \geq \sum_{i=1}^{N} s_i^2. \tag{59}$$

Furthermore, since $\mathsf{Tr}\left(\mathbf{\Sigma}'_{\tilde{r}\tilde{x}}\mathbf{\Sigma}_{\tilde{r}\tilde{x}}\right) = \mathbb{E}[\mathsf{Tr}\left(\mathbf{S}'_{\tilde{r}\tilde{x}}\mathbf{\Sigma}_{\tilde{r}\tilde{x}}\right)]$, we have

$$\sum_{i=1}^{N} s_i^2 = \mathbb{E}\left[\sum_{i=1}^{N} \hat{s}_i s_i^\circ\right]. \tag{60}$$

Hence, from the expressions of the in-sample and out-of-sample Sharpe Ratios in Equation (56) and (57), we have $\widehat{\mathsf{SR}} \geq \mathsf{SR}^\circ$.

C.7 Derivation of Parameters of Constant Correlation Model

Consider the optimization problem:

$$\min_{\varphi,\rho} \|\mathbb{I}_N + \rho\mathbb{J}_N - \mathbf{\Sigma}_{\tilde{r}\tilde{x}}\|_{\mathrm{F}}, \tag{61}$$

where $\|\cdot\|_{\mathrm{F}}$ is the Frobenius norm. We can re-write the objective function as follows:

$$\|\varphi\mathbb{I}_N + \rho\mathbb{J}_N - \mathbf{\Sigma}_{\tilde{r}\tilde{x}}\|_{\mathrm{F}}^2 = \|\varphi\mathbb{I}_N + \rho\mathbb{J}_N\|_{\mathrm{F}}^2 - 2\langle\mathbf{\Sigma}_{\tilde{r}\tilde{x}}, \varphi\mathbb{I}_N + \rho\mathbb{J}_N\rangle \tag{62}$$

$$= N\varphi^2 + N(N-1)\rho^2 + 2\varphi\sum_{i=1}^{N}[\mathbf{\Sigma}_{\tilde{r}\tilde{x}}]_{ii} + 2\rho\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}[\mathbf{\Sigma}_{\tilde{r}\tilde{x}}]_{ij} \tag{63}$$

By performing first order conditions, we have

$$[\text{FOC } \varphi]:\ 2N\varphi - 2\sum_{i=1}^{N}[\mathbf{\Sigma}_{\tilde{r}\tilde{x}}]_{ii} = 0 \tag{64}$$

$$[\text{FOC } \rho]:\ 2N(N-1)\rho - 2\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}[\mathbf{\Sigma}_{\tilde{r}\tilde{x}}]_{ij} = 0 \tag{65}$$

With two equations and two unknowns, we can solve to get the following optimal values:

$$\varphi = \frac{1}{N}\sum_{i=1}^{N}[\mathbf{\Sigma}_{\tilde{r}\tilde{x}}]_{ii} \qquad \text{and} \qquad \rho = \frac{1}{N(N-1)}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}[\mathbf{\Sigma}_{\tilde{r}\tilde{x}}]_{ij}. \tag{66}$$

Since the target parameters φ and ρ are expressed in relation to the population cross-covariance of the synthetic asset and synthetic signals, we estimate these parameters by their consistent estimators through the use of the sample cross-covariance matrix:

$$\hat{\varphi} = \frac{1}{N}\sum_{i=1}^{N}[\mathbf{S}_{\tilde{r}\tilde{x}}]_{ii}, \qquad \text{and} \qquad \hat{\rho} = \frac{1}{N(N-1)}\sum_{i=1}^{N}\sum_{j=1, j\neq i}^{N}[\mathbf{S}_{\tilde{r}\tilde{x}}]_{ij}. \tag{67}$$

We see that the parameters of the target matrix φ and ρ are the cross-sectional mean of the diagonal and off-diagonal elements of the cross-covariance matrix, respectively.

References

Abadir, K. M., Distaso, W., and Žikeš, F. (2014). Design-free estimation of variance matrices. *Journal of Econometrics*, 181(2):165–180.

Albert, A. (1969). Conditions for positive and nonnegative definiteness in terms of pseudoinverses. *SIAM Journal on Applied Mathematics*, 17(2):434–440.

Avellaneda, M., Healy, B., Papanicolaou, A., and Papanicolaou, G. (2020). Principal eigenportfolios for us equities. *Available at SSRN 3738769*.

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*.

Bickel, P. J. and Levina, E. (2008). Covariance regularization by thresholding. *The Annals of Statistics*, 36(6):2577–2604.

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.

Bun, J., Bouchaud, J.-P., and Potters, M. (2017). Cleaning large correlation matrices: Tools from random matrix theory. *Physics Reports*, 666:1–109.

Campbell, J. Y., Grossman, S. J., and Wang, J. (1993). Trading volume and serial correlation in stock returns. *The Quarterly Journal of Economics*, 108(4):905–939.

Chincarini, L. B. and Kim, D. (2006). *Quantitative Equity Portfolio Management: An Active Approach to Portfolio Construction and Management*. New York: McGraw-Hill.

Engle, R. F., Ledoit, O., and Wolf, M. (2019). Large dynamic covariance matrices. *Journal of Business & Economic Statistics*, 37(2):363–375.

Fan, J., Liao, Y., and Mincheva, M. (2012). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society. Series B, Statistical methodology*, 75(4).

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.

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.

Hardoon, D. R., Szedmak, S., and Shawe-Taylor, J. (2004). Canonical correlation analysis: An overview with application to learning methods. *Neural Computation*, 16(12):2639–2664.

Healy, M. (1957). A rotation method for computing canonical correlations. *Mathematics of Computation*, 11(58):83–86.

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.

Jobson, J. D. and Korkie, B. (1980). Estimation for markowitz efficient portfolios. *Journal of the American Statistical Association*, 75(371):544–554.

Kelly, B. T., Malamud, S., and Pedersen, L. H. (2020). Principal portfolios. Technical report, available at SSRN3623983.

Lam, C. (2016). Nonparametric eigenvalue-regularized precision or covariance matrix estimator. *The Annals of Statistics*, 44(3):928–953.

Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2):365–411.

Ledoit, O. and Wolf, M. (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. *The Annals of Statistics*, 40(2):1024–1060.

Ledoit, O. and Wolf, M. (2015). Spectrum estimation: A unified framework for covariance matrix estimation and PCA in large dimensions. *Journal of Multivariate Analysis*, 139:360–384.

Ledoit, O. and Wolf, M. (2020). Analytical nonlinear shrinkage of large-dimensional covariance matrices. *The Annals of Statistics*, 48(5):3043–3065.

Lehmann, B. N. (1990). Fads, martingales, and market efficiency. *The Quarterly Journal of Economics*, 105(1):1–28.

Lütkepohl, H. (1997). Handbook of matrices. *Computational Statistics and Data Analysis*, 2(25):243.

Markowitz, H. (1952). Portfolio selection. *Journal of Finance*, 7:77–91.

Michaud, R. (1989). The Markowitz optimization enigma: Is optimized optimal? *Financial Analysts Journal*, 45:31–42.

Nagel, S. (2012). Evaporating liquidity. *The Review of Financial Studies*, 25(7):2005–2039.

Partovi, M. H. and Caputo, M. (2004). Principal portfolios: Recasting the efficient frontier. *Economics Bulletin*, 7(3):1–10.

Pástor, L. and Stambaugh, R. F. (2003). Liquidity risk and expected stock returns. *Journal of Political Economy*, 111(3):642–685.

Rajaratnam, B., Massam, H., and Carvalho, C. M. (2008). Flexible covariance estimation in graphical Gaussian models. *The Annals of Statistics*, 36(6):2818–2849.

Stein, C. (1975). Estimation of a covariance matrix. Rietz lecture, 39th Annual Meeting IMS. Atlanta, Georgia.

Stein, C. (1986). Lectures on the theory of estimation of many parameters. *Journal of Mathematical Sciences*, 34(1):1373–1403.

Tan, V. and Zohren, S. (2021). Estimation of large noisy exponential weighted moving average covariances. *Available at SSRN 3745692.*

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.

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.