NBER WORKING PAPER SERIES

UNIVERSAL PORTFOLIO SHRINKAGE

Bryan T. Kelly
Semyon Malamud
Mohammad Pourmohammadi
Fabio Trojani

Working Paper 32004
http://www.nber.org/papers/w32004

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2023, revised April 2025

Universal Portfolio Shrinkage
Bryan T. Kelly, Semyon Malamud, Mohammad Pourmohammadi, and Fabio Trojani
NBER Working Paper No. 32004
December 2023, revised April 2025
JEL No. C1, C14, C53, C55, C58, G10, G11, G14, G17

ABSTRACT

We introduce a nonlinear covariance shrinkage method for building optimal portfolios. Our universal portfolio shrinkage approximator (UPSA) is given in closed form, is cheap to implement, and improves upon existing shrinkage methods. Rather than annihilating low-variance principal components of returns, UPSA instead reweights components to explicitly optimize expected out-of-sample portfolio performance. We demonstrate robust empirical improvements over alternative shrinkage methods in the literature.

Bryan T. Kelly
Yale School of Management
165 Whitney Ave.
New Haven, CT 06511
and NBER
bryan.kelly@yale.edu

Semyon Malamud
Swiss Finance Institute @ EPFL
Quartier UNIL-Dorigny, Extranef 213
CH - 1015 Lausanne
Switzerland
semyon.malamud@epfl.ch

Mohammad Pourmohammadi
and Swiss Finance Institute
mohammad.pourmohammadi@unige.ch

Fabio Trojani
University of Geneva - Swiss Finance Institute
fabio.trojani@alphacruncher.com

Universal Portfolio Shrinkage

April 18, 2025

Abstract

We introduce a nonlinear covariance shrinkage method for building optimal portfolios. Our universal portfolio shrinkage approximator (UPSA) is given in closed form, is cheap to implement, and improves upon existing shrinkage methods. Rather than annihilating low-variance principal components of returns, UPSA instead reweights components to explicitly optimize expected out-of-sample portfolio performance. We demonstrate robust empirical improvements over alternative shrinkage methods in the literature.

1 Introduction

Efficient portfolios that optimally balance risk and return are central to asset pricing. However, in practically relevant scenarios involving thousands of stocks or hundreds of factors, classical estimators of the Markowitz (1952) portfolio are severely contaminated by noise. Despite their stellar in-sample performance, they typically fail out-of-sample and are often dominated by naïvely diversified portfolios (DeMiguel et al., 2009). When the number of estimated portfolio parameters exceeds the number of observations—as happens in many practical applications—the law of large numbers breaks down, which drives a wedge between in-sample and out-of-sample performance (Didisheim et al., 2024).[1]

[1]For example, when the number of training observations is smaller than the number of assets, the unregularized in-sample Sharpe ratio of the Markowitz portfolio is infinite.

Shrinkage reduces this wedge by tuning the bias-variance tradeoff. Bias sacrifices in-sample estimator performance in order to reduce estimator variance and improve expected out-of-sample performance. Many existing portfolio shrinkage methods tightly constrain admissible forms of shrinkage (e.g., ridge regularizing the covariance matrix or dropping selected principal components as in Kozak et al., 2020) or are implemented through statistical objectives that deviate from the portfolio optimization objective (such as minimizing covariance matrix estimation error as in Ledoit and Wolf, 2017). These design choices can limit the effectiveness of portfolio shrinkage.

We propose the "universal portfolio shrinkage approximator," or UPSA, to overcome the limitations of existing portfolio shrinkage methods. UPSA is a universal method that provides a semi-parametric approximation to a large class of nonlinear shrinkage functions, freeing it from the rigid structure of many shrinkage methods. Second, UPSA chooses the form of shrinkage to directly optimize expected out-of-sample portfolio performance rather than some tangentially relevant statistical objective. Tuning the bias-variance tradeoff based on portfolio performance improves shrinkage efficacy. Additionally, UPSA is available in closed form, computationally cheap to implement, and scalable for truly high-dimensional portfolios in which the number of assets may greatly exceed the number of observations.

1.1 Universal Shrinkage

The nature of optimal portfolio shrinkage developed in this paper begins from the sample Markowitz portfolio. It is convenient to represent the Markowitz solution as a portfolio of principal components (PCs) since these are investable assets with zero correlation in the estimation sample. Among a set of N assets, the Markowitz portfolio assigns weights $\bar{\pi}_i^{\text{Mark}}$ to the i^{th} PC based on its individual risk-return tradeoff. That is, the Markowitz weight on each PC is proportional to its ratio of average return ($\bar{R}_i^{PC}$) to sample variance ($\bar{\lambda}_i$):

$$R_t^{\text{Mark}} = \sum_{i=1}^{N} \bar{\pi}_i^{\text{Mark}} R_{i,t}^{PC} = \sum_{i=1}^{min(N,T)} \frac{\bar{R}_i^{PC}}{\bar{\lambda}_i} R_{i,t}^{PC},$$

where T is the number of data points. When N is large relative to T, estimates of the mean-variance tradeoff are corrupted by noise, giving rise to small sample biases such as the

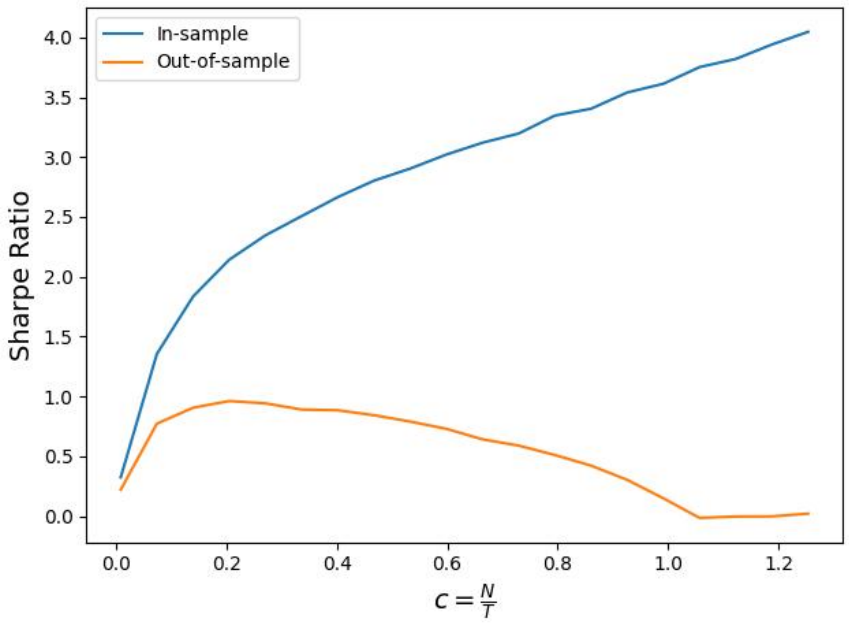

Figure 1: In-sample vs. Out-of-sample Sharpe Ratios without Shrinkage.

Comparison of in-sample and out-of-sample Sharpe ratios for Markowitz portfolios averaged over 100 simulations. The x-axis shows $c = \frac{N}{T}$, where N is the number of factors in the portfolio, T is the number of observations used to build the portfolio, and factors are randomly selected from the 153 factors reported by Jensen et al. (2023). Out-of-sample performance is evaluated via 5-fold cross-validation, using data from November 1971 to December 2022.

tendency to overweight low-variance PCs. Absent shrinkage, these biases can be devastating for the performance of an investor's portfolio.

Figure 1 illustrates how the unregularized Markowitz "plug-in" portfolio suffers in the face of estimation "complexity" defined as $c = N/T$. When $c \approx 0$, the investor is in a data-rich environment where the training observations far outnumber the parameters to be estimated. In this case, the law of large numbers kicks in and the plug-in portfolio recovers the true optimal portfolio. But when the environment is complex ($c \gg 0$), the in-sample portfolio performance is severely upward biased while its out-of-sample performance collapses.

The literature has considered a range of shrinkage-based solutions to improve out-of-sample portfolio performance. One such approach uses ridge shrinkage, which adjusts the Markowitz portfolio weights via "soft" variance thresholding of PCs:

$$\bar{\pi}_i^{\text{Ridge}} = \frac{\bar{R}_i^{PC}}{\bar{\lambda}_i + z} \tag{1}$$

where z is the ridge shrinkage parameter (see e.g. Kozak et al., 2020). Another approach is inspired by the Arbitrage Pricing Theory (APT) of Ross (1976). It applies a "hard" threshold to impose exactly zero weight on low-variance PCs:

$$\bar{\pi}_i^{\text{APT}} = \frac{\bar{R}_i^{PC}}{\bar{\lambda}_i} 1_{\bar{\lambda}_i \geq \kappa}$$

where κ is the hard threshold (e.g. Severini, 2022; Kelly et al., 2023). This enforces an economic restriction that only the factors that are responsible for a large fraction of market risk command a risk premium. In a third approach, Ledoit and Wolf (2004b) shrink portfolio weights linearly to a simple benchmark weight (such as $1/N$):

$$\bar{\pi}_i^{\text{LW}} = (1-\gamma)\frac{\bar{R}_i^{PC}}{\bar{\lambda}_i} + \gamma \frac{1}{N}$$

where γ is the shrinkage parameter.

The goal of UPSA is to generalize these examples and others and achieve an optimal nonlinear portfolio shrinkage function: $\bar{\pi}_i^{\text{UPSA}} = f(\bar{\lambda}_i)\bar{R}_i^{PC}$. The central contribution of our paper is the discovery of a universal approximator for the optimal shrinkage function. Our formulation is interpretable as a shallow neural network approximation to f, hence its universality. Yet it has the computational advantage of being estimable in closed-form as a linear combination of ridge portfolios as in equation (1), each using a different ridge parameter. In addition, we show that UPSA can incorporate certain economic constraints on the shrinkage function, such as strict positivity and monotonicity, with no loss of tractability.

1.2 Portfolio Shrinkage Criterion

In addition to being a universal shrinkage method, UPSA expressly maximizes the core economic objective of interest—expected out-of-sample portfolio performance. This is in contrast to many previously proposed methods that select shrinkage to minimize covariance estimation error, which is only indirectly relevant to the ultimate goal of portfolio optimization. We show that while UPSA relies on spectral shrinkage of the assets' second moment matrix, it also incorporates information from average returns whenever the economic

objective calls for it. We also show that UPSA is easily adaptable to optimizing any quadratic portfolio objective without sacrificing analytical tractability (including, but not restricted to, a Markowitz-type objective).[2]

1.3 Empirical Performance

We investigate the performance of UPSA in forming optimal portfolios from the large set of anomaly factors compiled by Jensen et al. (2023). We compare UPSA to three natural spectral shrinkage benchmarks: (i) a simple ridge-regularized Markowitz portfolio, (ii) a Markowitz portfolio whose plug-in second moment is estimated using the nonlinear spectral shrinkage approach of Ledoit and Wolf (2017), and (iii) a Markowitz portfolio that uses few PCs. In all our experiments, UPSA achieves higher out-of-sample Sharpe ratios than these benchmarks, and its outperformance is particularly pronounced post-2000, when many anomaly portfolios have underperformed. Moreover, UPSA's performance gains are statistically significant across benchmark portfolios with a full sample α t-stat of 3.72 with respect to its closest competitor.

Viewing these portfolios as stochastic discount factors (SDFs), we find that the UPSA SDF yields the smallest out-of-sample pricing errors for anomaly portfolio test assets. In some cases, UPSA delivers a cross-sectional R^2 for test asset average returns that is twice as large as its closest competitor (on a purely out-of-sample basis).

Our empirical results demonstrate how UPSA behaves as a weighted average of ridge-regularized portfolios. By blending different degrees of ridge shrinkage, UPSA is able to apply different shrinkage to different regions of the eigenvalue distribution, improving its empirical performance versus more constrained spectral estimators. When we plot its effective shrinkage function, we find that UPSA applies relatively less shrinkage to smaller principal components than ridge or PCA, indicating that small principal components of returns are more valuable for the investor's portfolio than commonly believed.

We also demonstrate the robustness of UPSA in a variety of dimensions, including for assets with varying degrees of liquidity (among sub-samples of mega, large, and small stocks)

[2]In the context of linear shrinkage for portfolio optimization, Pedersen et al. (2021) emphasize the significance of selecting the correct mean-variance objective.

and in various training sample sizes. We also investigate adding a lasso penalty alongside ridge shrinkage to induce sparsity (i.e., to achieve hard-thresholding of components). With this additional regularization, UPSA continues to deliver statistically significant outperformance compared to competing methods.

1.4 Literature Review

Our work is related to several strands of literature. The first literature studies shrinkage-based estimation of return covariance matrices. A pioneering contribution in this area is the linear shrinkage estimator of Ledoit and Wolf (2004b) with applications to the estimation of minimum variance portfolios (Ledoit and Wolf, 2003) and tracking portfolios (Ledoit and Wolf, 2004a). Another series of influential papers (Ledoit and Wolf, 2012, 2015, 2020) develop various nonlinear spectral shrinkage estimators exploiting the random matrix theory techniques introduced in Ledoit and Péché (2011). We differ from these approaches by pursuing a universal approximator for a broad class of shrinkage functions. Our method has the flexibility of a neural network but is analytically tractable and cheap to compute.

A second related literature studies the role of shrinkage in portfolio optimization and the closely related topic of regularized estimation for SDFs and factor pricing models. Rooted in the principles of the APT, much of this literature achieves regularization through dimension reduction (e.g., Kozak et al., 2018; Kelly et al., 2020; Lettau and Pelger, 2020; Gu et al., 2021; Bryzgalova et al., 2023b; Giglio and Xiu, 2021). A particularly relevant example is Kozak et al. (2020), which build portfolios by shrinking the variance of all PCs by the same tuning parameter. Using a similar but more extensive dataset, we find that UPSA further improves portfolio performance by more extensively leveraging the return and diversification benefits of low-variance PCs. In fact, with UPSA, out-of-sample portfolio performance monotonically increases in the number of PCs.

A number of recent papers investigate the asset pricing role of weak factors whose risk premia are too small to be effectively estimated (see, e.g., Bryzgalova et al., 2023a; Preite et al., 2022; Giglio et al., 2021). To address the weak factor problem, Lettau and Pelger (2020) advocate a PC-based dimension reduction that is cleverly augmented with information about PC sample means. In contrast, UPSA shrinks without dimension reduction, instead

re-weighting all PCs through eigenvalue shrinkage and in light of the portfolio performance objective (so that UPSA's shrinkage is also influenced by PC means).

Lastly, our paper relates to a nascent literature on the statistical limits of asset pricing model estimation. This literature discusses how classical statistical theory needs to be adjusted when dealing with such situations of estimation complexity. Da et al. (2022) discusses statistical limits to arbitrage while Didisheim et al. (2024) analyzes the limits to statistical learning inherent in complex empirical settings. Martin and Nagel (2021) emphasize the importance of shrinkage techniques and out-of-sample testing in Bayesian high-dimensional models. Kelly et al. (2022) and Didisheim et al. (2024) highlight the theoretical and empirical advantages of heavily parameterized asset pricing models for improving out-of-sample model performance. Our paper contributes to the literature by introducing a robust shrinkage methodology to mitigate the estimation noise inherent in complex settings.

2 Second Moment Shrinkage

In this section, we outline our main setup. We consider an economy of N assets. We collect their excess returns in a vector $F_t \in \mathbb{R}^N$, with second moment matrix $E[F_t F_t'] = \Sigma \in \mathbb{R}^{N\times N}$, and mean $E[F_t] = \mu \in \mathbb{R}^N$. We consider the standard portfolio choice problem of optimizing expected quadratic utility. With full information, the efficient portfolio is the Markowitz solution[3]

$$\pi_\star \;=\; \Sigma^{-1}\mu \;. \tag{2}$$

In contrast to the full information setting, an investor whose information includes only T observations of F_t can instead compute finite-sample moments $\bar{\mu} = \frac{1}{T}\sum_{t=1}^T F_t$, $\bar{\Sigma} = \frac{1}{T}\sum_{t=1}^T F_t F_t'$ and construct an empirical counterpart of portfolio (2) given by $\bar{\pi} = \bar{\Sigma}^+\bar{\mu}$, with return $R_t^{\bar{\pi}} = \bar{\pi}' F_t$.[4]

In most practical scenarios, investors find themselves with a large number of assets (e.g.,

[3] We use second moments rather than covariance matrices to simplify the algebra, though the formulation is equivalent up to a mean adjustment and easily extends to the covariance-based Markowitz setting.

[4] $\bar{\Sigma}^+$ is the Moore-Penrose generalized inverse of matrix $\bar{\Sigma}$. We work with generalized inverses to cover situations where matrix $\bar{\Sigma}$ may be singular.

thousands of stocks or hundreds of anomaly factors) and a limited time series data (dozens or hundreds of monthly observations). When N is large relative to T, the law of large numbers breaks down and empirical moments do not consistently estimate theoretical moments: $\bar{\mu} \not\to \mu$, $\bar{\Sigma} \not\to \Sigma$ (Didisheim et al., 2024). In such a setting, the estimated Markowitz portfolio (2) is a random quantity, even asymptotically, and can perform badly out-of-sample (see for example DeMiguel et al., 2009).

Let $\bar{\Sigma} = \bar{U}\,\mathrm{diag}(\bar{\lambda})\bar{U}'$ be the spectral decomposition of the empirical second-moment matrix. A standard approach to improve the performance of the Markowitz plug-in solution is to inject bias through a shrinkage function, f, in order to rein in the estimator's variance. In this paper we focus on a common class of so-called "spectral" shrinkage functions that regularize the sample second moment matrix by shrinking its empirical eigenvalues, $\bar{\lambda}$, without altering its empirical eigenvectors $\bar{U}$:[5]

$$\bar{\pi}(f) \;=\; f(\bar{\Sigma})\bar{\mu}\;,\;\; s.t.\; f(\bar{\Sigma}) \;=\; \bar{U}\,\mathrm{diag}(f(\bar{\lambda}))\bar{U}'\;. \tag{3}$$

What does spectral shrinkage imply for the Markowitz portfolio? Define $R_t^{PC} = \bar{U}'F_t$ to be the vector of returns of the associated principal component factors, and let $\bar{R}^{PC} = \bar{E}[R_t^{PC}]$ be their in-sample means. We can rewrite the return of the shrunken Markowitz portfolio as the return of principal component factors:

$$R_t^{\bar{\pi}(f)} \;=\; \sum_{i=1}^{K} f(\bar{\lambda}_i)\bar{R}_i^{PC}R_{i,t}^{PC} \tag{4}$$

where $K = \min(N, T)$. The weights on individual PC portfolio captures a tradeoff between that PC's return ($\bar{R}_i^{PC}$) and its variance (λ_i).[6] For example, a common approach to this problem relies on the ridge portfolio estimator,

$$f_z(\bar{\lambda}_i) = \frac{1}{\bar{\lambda}_i + z},$$

[5] This class of estimators was introduced in the influential paper by Stein (1986).

[6] Without shrinkage and when N is large relative to T, this tradeoff may be severely misestimated, especially for low-variance PC factors, partly because their estimated eigenvalues tend to be systematically biased towards zero.

which is defined for some ridge parameter $z > 0$ (see, e.g., Kozak et al. (2020), Ledoit and Wolf (2003), among many others).[7] As this example shows, portfolio shrinkage is achieved by manipulating how components' risk and return map into portfolio weights.

"One size fits all" shrinkage methods like ridge can underperform in environments with complex eigenvalue structures and heterogeneous risk-return tradeoffs across PCs. One would ideally build a portfolio that more flexibly manipulates the risk-return tradeoff for different components. To this end, we develop a nonlinear spectral shrinkage function that learns from in-sample data how to weight PCs in order to maximize out-of-sample quadratic utility.

Definition 1 (Optimal Nonlinear Portfolio Shrinkage) *Given in-sample return data for N assets, $F_{IS} = \{F_1, \cdots, F_T\}$, let $\bar{\Sigma} = \bar{U} diag(\bar{\lambda}) \bar{U}'$ be the associated sample second moment matrix. An* **optimal spectral shrinkage portfolio estimator** *$\bar{\pi}(f_\star)$ for a class $\mathcal{C}$ of admissible shrinkage functions is defined by a strictly positive function $f_\star : \mathbb{R}_+ \to \mathbb{R}_{++}$ such that $f_\star(\bar{\Sigma})$ solves the out-of-sample (i.e., for $t > T$) expected quadratic utility optimization problem:*

$$f_\star = \arg\max_{f \in \mathcal{C}} E\left[R_t^{\bar{\pi}(f)} | F_{IS}\right] - \frac{1}{2} E\left[(R_t^{\bar{\pi}(f)})^2 | F_{IS}\right] \quad . \tag{5}$$

The key new feature of our optimal spectral shrinkage operator $f_\star$ is its potentially complex nonlinear dependence on the entire in-sample second-moment matrix $\bar{\Sigma}$ as well as on the means through the portfolio utility objective.

With perfect knowledge of population moments Σ and μ, the exact solution to the optimal (but infeasible) spectral shrinkage problem reads:[8]

$$f_\star(\bar{\lambda}_i) = \frac{\bar{U}_i' \pi_\star}{\bar{U}_i' \bar{\mu}} = \frac{\bar{U}_i' \pi_\star}{\bar{R}_i^{PC}} \tag{6}$$

This solution depends on both the population moments and the noisy empirical ones. This infeasible example captures the key insight of UPSA: Effective nonlinear shrinkage should tilt

[7] The ridge portfolio estimator adjusts the Markowitz portfolio weights via a "soft" eigenvalue thresholding, in which the shrinkage factor $\frac{1}{1+z/\lambda_i}$ more aggressively down-weights the portfolio weights of low-variance PCs.

[8] This is similar to the approach Stein (1986). See Lemma 3 and its proof in the Appendix.

portfolio weights toward the true Markowitz portfolio, while accounting for noise distortions in sample means, sample eigenvectors and sample eigenvalues.

3 UPSA

In this section, we formally define the universal portfolio shrinkage approximator (UPSA). Its formulation is surprisingly simple and amounts to an ensemble of ridge portfolios with heterogeneous penalties. Yet despite this simplicity, we prove that UPSA can approximate any nonlinear spectral shrinkage function satisfying basic regularity conditions.

We begin by demonstrating how a basic ridge shrinkage function, $f_z(\lambda) = \frac{1}{\lambda+z}$, can serve as the foundation for more flexible shrinkage functions.

Lemma 1 *Let f be a strictly positive, matrix monotone-decreasing function such that $K = \lim_{\lambda\to\infty} \lambda f(\lambda)$. There then exists a positive finite measure ν on $\mathbb{R}_+$ such that:*

$$f(\lambda) = \int_0^\infty \frac{1}{z+\lambda} d\nu(z) \ , \tag{7}$$

$\nu(\mathbb{R}_+) = K$. Moreover, whenever the grid $\mathcal{Z}$ is sufficiently wide and dense, there exists a function $f_{\mathcal{Z}} \in \mathcal{F}(\mathcal{Z})$ which approximates f uniformly over compact intervals:

$$f_{\mathcal{Z}}(\lambda) = \sum_{i=1}^N \frac{1}{z_i+\lambda} \nu_i \ , \tag{8}$$

where $\nu_1, \ldots, \nu_N > 0$ and $\sum_{i=1}^N \nu_i = K$.

The key insight of Lemma 1 is equation (8). It formally establishes that combinations of basic ridge portfolios are universal approximators for a general nonlinear spectral shrinkage function. Based on this insight, we propose the following definition of the ridge ensemble as a non-negative linear combination of basic ridge portfolios indexed by distinct penalty parameters.

Definition 2 *Given a grid $\mathcal{Z} = \{z_1, \ldots, z_L\}$ of ridge penalties and a vector $W = (w_i)_{i=1}^L$ of*

weights, define for any $\lambda \geq 0$ *the following weighted shrinkage function:*

$$f_{\mathcal{Z},W}(\lambda) := \sum_{i=1}^{L} w_i f_{z_i}(\lambda) = \sum_{i=1}^{L} w_i (z_i + \lambda)^{-1} \ . \tag{9}$$

We call $\mathcal{F}_C(\mathcal{Z}) \ = \ \{f_{\mathcal{Z},W} : \ W \in \mathbb{R}_+^L\}$ *the* **ridge ensemble**.[9]

The constraint that the ensemble weights must be positive ensures that the second moment matrix remains positive definite. Furthermore, the shrinkage function is matrix monotone, meaning that the risk ordering of components is preserved after shrinkage.[10]

Even within the family of Ridge portfolios, varying the ridge parameter z can lead to markedly different portfolio behaviors. At one extreme, we have the "ridgeless" portfolio (Kelly et al., 2022) that uses minimal shrinkage:[11]

$$\lim_{z \to 0} (z I_N + \bar{\Sigma})^{-1} \bar{\Sigma} \bar{\Sigma}^{+} \bar{\mu} \ = \ \bar{\Sigma}^{+} \bar{\mu} \tag{10}$$

which recovers the Markowitz portfolio when $N < T$ and more generally uses the sample second moment to its fullest expression. At the other extreme, we have

$$\lim_{z \to \infty} z (z I_N + \bar{\Sigma})^{-1} \bar{\mu} \ = \ \bar{\mu} \tag{11}$$

which eliminates all second moment information from the portfolio, thus behaving as a "momentum" strategy based only on past average returns. Lemma 1 shows that combining the heterogeneous behaviors of different ridge portfolios achieves more flexible shrinkage than those achieved by any single ridge portfolio.

The approximating function $f_{\mathcal{Z}}$ in Lemma 1 coincides with the ridge ensemble in equation (9) when ensemble weights are appropriately chosen. To operationalize the ridge ensemble,

[9] Without constraints, these shrinkages can cover any continuous function with compact support. We go into detail on this space of admissible shrinkages in Lemma 8 of the Appendix.

[10] The fact that $\bar{\lambda} f(\bar{\lambda})$ in Lemma 1 is bounded ensures that the resulting shrinkage of the Markowitz portfolio weights of PC factors is bounded as a function of the PC factors' variances. In Lemma 1, limit condition $K = \lim_{\bar{\lambda} \to \infty} \bar{\lambda} f(\bar{\lambda})$ fixes the decay of admissible shrinkage functions at $+\infty$, which needs to be hyperbolic at least. This includes, e.g., the decay of pure ridge shrinkages of the form $f(\bar{\lambda}) = (z + \bar{\lambda})^{-1}$ for some $z > 0$.

[11] Implicitly, we are using here the "empirical" no arbitrage assumption that $\bar{\mu}$ belongs to the range of $\bar{\Sigma}$.

we require an estimator of the ensemble weights, which gives rise to UPSA. In particular, we derive UPSA as the optimizer of an out-of-sample utility maximization objective. We first substitute (9) into the PC-based Markowitz shrinkage portfolio of equation (4), obtaining

$$R_t^{\bar{\pi}(f_{\mathcal{Z},W})} = \sum_{i=1}^{L} w_i R_t^{\bar{\pi}(f_{z_i})} = \underbrace{W'}_{\text{Ridge Weights}} \quad \underbrace{R_t^{\bar{\pi}(f_{\mathcal{Z}})}}_{\text{Ridge Returns}}. \tag{12}$$

where, $R_t^{\bar{\pi}(f_{\mathcal{Z}})} = (R_t^{\bar{\pi}(f_{z_i})})_{i=1}^{L}$. We propose tuning the ensemble weights to optimize the out-of-sample quadratic utility objective,

$$E[R_t^{\bar{\pi}(f_{\mathcal{Z},W})}] - \frac{1}{2}E[(R_t^{\bar{\pi}(f_{\mathcal{Z},W})})^2] = W' \underbrace{E[R_t^{\bar{\pi}(f_{\mathcal{Z}})}]}_{\text{Ridge Means}} - \frac{1}{2}W' \underbrace{E[R_t^{\bar{\pi}(f_{\mathcal{Z}})} R_t^{\bar{\pi}(f_{\mathcal{Z}})'}]}_{\text{Ridge Second Moment}} W. \tag{13}$$

Equation (13) shows how the ensemble weight estimation problem amounts to a portfolio choice problem on the space of ridge shrinkage portfolios.

To implement this framework in practice, we require reliable estimates of ridge portfolio out-of-sample means and second moments. We suggest estimating these out-of-sample moments using classical leave-one-out (LOO) cross-validation.[12] The LOO method drops one training observation t, trains on the remaining observations $1, .., t-1, t+1, ...T$, and repeats this for each $t = 1, ..., T$. LOO exploits the fact that, under the assumption of i.i.d. asset returns F_t, we obtain an unbiased estimate of this out-of-sample portfolio utility as the average utility of each trained portfolio on its corresponding left-out observation.[13] From this procedure, we also obtain LOO estimates of the first and second moments of factor returns, denoted for each $t = 1, ..., T$, as

$$\bar{\mu}_{T,t} = \frac{1}{T-1} \sum_{\tau \neq t, 1 \leq \tau \leq T} F_\tau \quad \text{and} \quad \bar{\Sigma}_{T,t} = \frac{1}{T-1} \sum_{\tau \neq t, 1 \leq \tau \leq T} F_\tau F_\tau'. \tag{14}$$

From the LOO asset moment estimates, we obtain spectral shrinkage portfolio estimates for

[12] In this context, Lemma 7 in the Appendix provides the foundation for an efficient LOO computation by deriving simple expressions for LOO returns in terms of the full-sample returns and the full-sample empirical second-moment matrix $\bar{\Sigma}$.

[13] See Lemma 4 in the Appendix for a formal proof.

each t that are LOO versions of the in-sample estimator (3),

$$\bar{\pi}_{T,t}(f_z) \;=\; f_z(\bar{\Sigma}_{T,t})\bar{\mu}_{T,t}. \tag{15}$$

Finally, from (15), we recover the realized LOO portfolio returns $\{R_{T,t}^{\bar{\pi}_{T,t}(f_z)} = \bar{\pi}_{T,t}(f_z)'F_t : t = 1,\ldots,T\}$. These proxies for out-of-sample returns are thus used to estimate the out-of-sample mean and second moment of the ridge portfolios, as described by the next lemma.

Lemma 2 *Consider the following LOO-based estimators for the means and second moment of out-of-sample ridge portfolio returns $R^{\bar{\pi}(f_{z_1})},\ldots,R^{\bar{\pi}(f_{z_L})}$:*

$$\begin{aligned}
\bar{\mu}(\mathcal{Z}) \;&=\; \left(\frac{1}{T}\sum_{t=1}^{T} R_{T,t}^{\bar{\pi}_{T,t}(f_{z_i})}\right)_{i=1}^{L} \in \mathbb{R}^L \\
\bar{\Sigma}(\mathcal{Z}) \;&=\; \left(\frac{1}{T}\sum_{t=1}^{T} R_{T,t}^{\bar{\pi}_{T,t}(f_{z_i})} R_{T,t}^{\bar{\pi}_{T,t}(f_{z_j})}\right)_{i,j=1}^{L} \in \mathbb{R}^{L\times L}.
\end{aligned} \tag{16}$$

Then, from Definition 2, UPSA shrinkage is given by

$$f_{UPSA} \;=\; f_{Z,W^\star_{UPSA}}, \tag{17}$$

where, $W^\star_{UPSA}$ solves

$$W^\star_{UPSA} \;=\; \arg\max_{W\in\mathbb{R}_+} \left\{W'\bar{\mu}(\mathcal{Z}) \;-\; 0.5W'\,\bar{\Sigma}(\mathcal{Z})\,W\right\}. \tag{18}$$

To summarize, our derivation, culminating in Lemma 2, shows that UPSA is a ridge ensemble that approximates an unknown nonlinear spectral shrinkage function f. The estimator of this ensemble is the solution to an out-of-sample utility maximization problem (in essence, an extension of the usual out-of-sample performance optimization that takes place in cross-validation). UPSA's ridge ensemble weights require estimates of the out-of-sample moments of basic ridge portfolios, and Lemma 2 shows how to estimate these using the LOO methodology. Training UPSA reduces to finding the optimal weight vector W across basic ridge portfolios that maximizes quadratic utility, which is equivalent to solving

a Markowitz portfolio optimization using LOO returns of basic ridge portfolios instead of the original N original assets. This formulation of the estimator renders UPSA tractable and cheap to compute even in very high-dimensional contexts.[14]

4 Empirical Evidence

4.1 Data and Benchmarks

Our empirical analysis uses monthly returns on 153 characteristic-managed portfolios ("factors") from Jensen et al. (2023).[15] The factors are constructed as capped value-weight long-short portfolios of US stocks and cover the period from November 1971 to December 2022.

We estimate portfolio weights using a rolling window of $T = 120$ months, with weights retrained and rebalanced on a monthly basis. We then use the estimated optimal portfolios as SDFs and evaluate their ability to price returns on the underlying factors. We fix the grid of ridge penalties available to UPSA as $\mathcal{Z} \equiv z \in [10^i : i \in \{-10, -9, \cdots, -1\}]$.[16]

We compare the performance of UPSA with two spectral shrinkages, one sparse shrinkage, and two factor pricing models:

- **Ridge**: The single ridge portfolio selected from the penalty grid $\mathcal{Z}$ that maximizes expected quadratic utility based on LOO cross-validation.
- **LW**: The nonlinear shrinkage estimator of Ledoit and Wolf (2020).
- **PCA**: Subset of principal components selected to maximize expected quadratic utility based on LOO cross-validation.
- **FF5**: The Markowitz portfolio constructed from the five factors of Fama and French (2015).

[14] Our UPSA code is available here.

[15] The complete list of these 153 factors and their monthly returns are available at https://jkpfactors.com.

[16] Our results are robust to the choice of the grid $\mathcal{Z}$. Results for alternative grid choices are available upon request. However, optimal outcomes are achieved when the $\mathcal{Z}$ grid spans from the smallest to the largest eigenvalues of the second moment.

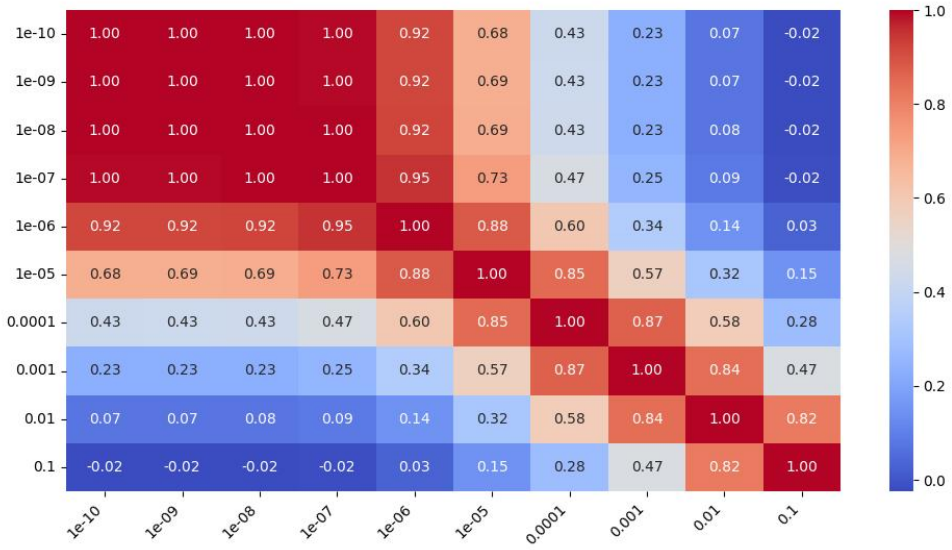

Figure 2: Out-of-sample Return Correlations Across Shrinkage Levels.

Average out-of-sample return correlations between portfolios with different levels of Ridge shrinkage. The out-of-sample period is from November 1981 to December 2022.

- **CAPM**: The market portfolio.

We train all benchmarks with the same rolling 120-month training window used for UPSA.[17]

4.2 Ridge Heterogeneity

Drawing on the insights of Lemma 2, UPSA generates performance benefits by efficiently combining a variety of basic ridge portfolios. Thus, for UPSA to be effective, it must leverage the heterogeneity across basic ridge portfolios to the extent it exists. Figure 2 reports the pairwise correlation of basic ridge portfolio returns using different penalties in the grid $\mathcal{Z}$. There is indeed a high degree of diversity at UPSA's disposal, with correlations as low as -2% for some pairs.

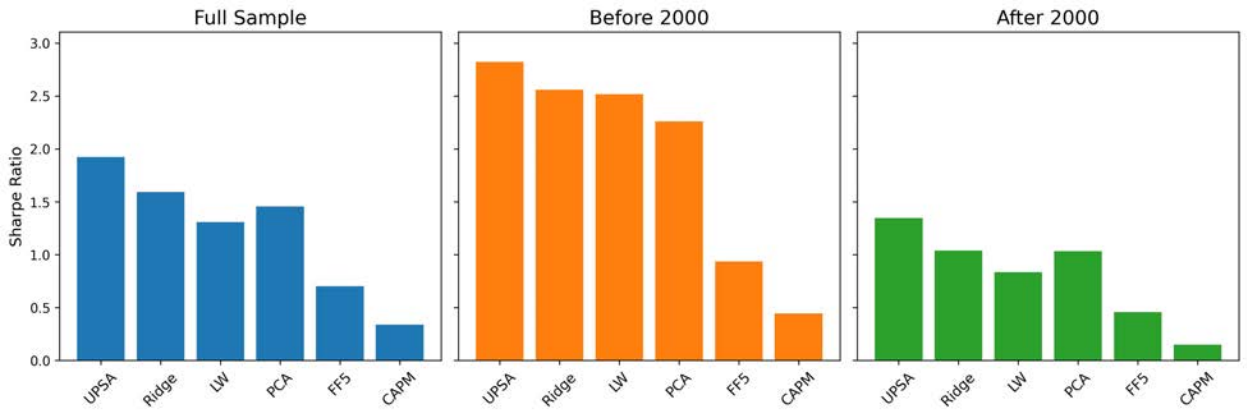

Figure 3: Annual Out-of-sample Sharpe Ratios for UPSA and Benchmarks.

Annual out-of-sample Sharpe ratios for UPSA and benchmark strategies. The full-sample results are computed from November 1981 to December 2022.

4.3 Comparative Performance of UPSA

The primary assessment of UPSA rests on its out-of-sample portfolio performance vis-à-vis competing methods. To this end, the left panel of Figure 3 reports the out-of-sample Sharpe ratio of UPSA and its benchmarks over the full out-of-sample period (beginning in November 1981). UPSA achieves a Sharpe ratio of 1.92, compared to 1.59 for the single best ridge model, 1.31 for LW, and 1.45 for PCA. The Sharpe ratios of 0.7 for the Fama-French factors and 0.34 for the CAPM provide a frame of reference from the factor pricing literature. The remaining two panels of Figure 3 report performance in pre-2000 and post-2000 subsamples. The relative outperformance of UPSA is similar across subsamples. To see performance in further detail, Figure 4 plots the cumulative returns for UPSA and the benchmarks.

Next, we investigate the statistical significance of UPSA's improvement over the benchmarks. Figure 5 reports the alpha of UPSA against each benchmark and associated standard error bars. UPSA's annual alpha versus ridge, its second closest competitor, over the full sample is 4.46% with a t-statistic of 3.72. For other benchmarks, the alpha is even larger and statistically significant. Furthermore, the comparative benefits of UPSA are uniform

[17]To ensure the volatility of returns stays similar and comparable throughout the sample, we normalize portfolio weights so that the implied in-sample variance after shrinkage equals the in-sample variance of the empirical eigenvalues, following the recommendation of Ledoit and Wolf (2020).

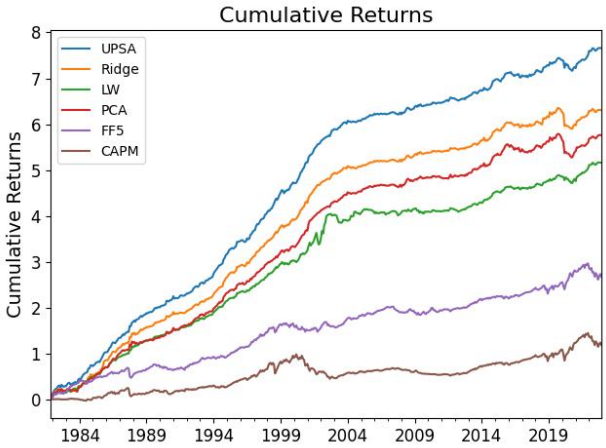

Figure 4: Cumulative Log Returns for SDFs with Different Shrinkage Methods.

Comparison of out-of-sample SDF performance across different shrinkage methods from November 1981 to December 2022. The figure shows cumulative log returns, scaled to an annual volatility of 10%.

across the sample, as evident from the pre-2000 and post-2000 subsample analyses in the center and right columns.

4.4 Asset Pricing Implications: The UPSA SDF

Classical asset pricing theory establishes an equivalence between the mean-variance efficient portfolio and the tradable SDF (Hansen and Jagannathan, 1991). By direct calculation, the infeasible portfolio $\pi_\star = \Sigma^{-1}\mu$ from equation (2) defines the unique tradable SDF that prices assets with zero error:

$$E[F_{i,t+1}\, M^*_{t+1}] = 0, i = 1, ..., N \quad \text{where} \quad M^*_{t+1} \;=\; 1 \;-\; \pi'_\star F_{t+1}.$$

A similar calculation shows that the unregularized in-sample Markowitz portfolio achieves zero in-sample pricing error by construction. But, just as in the case of its out-of-sample Sharpe ratio in Figure 1, the Markowitz portfolio typically prices assets exceedingly poorly out-of-sample. Given that portfolio shrinkage aims to improve out-of-sample performance, we now investigate whether UPSA's superior portfolio returns also lead to lower out-of-sample pricing errors. Specifically, we compare

$$E_{OOS}[F_{i,t+1}] \quad \text{vs.} \quad E_{OOS}[(1 - M_{t+1})F_{i,t+1}] \tag{19}$$

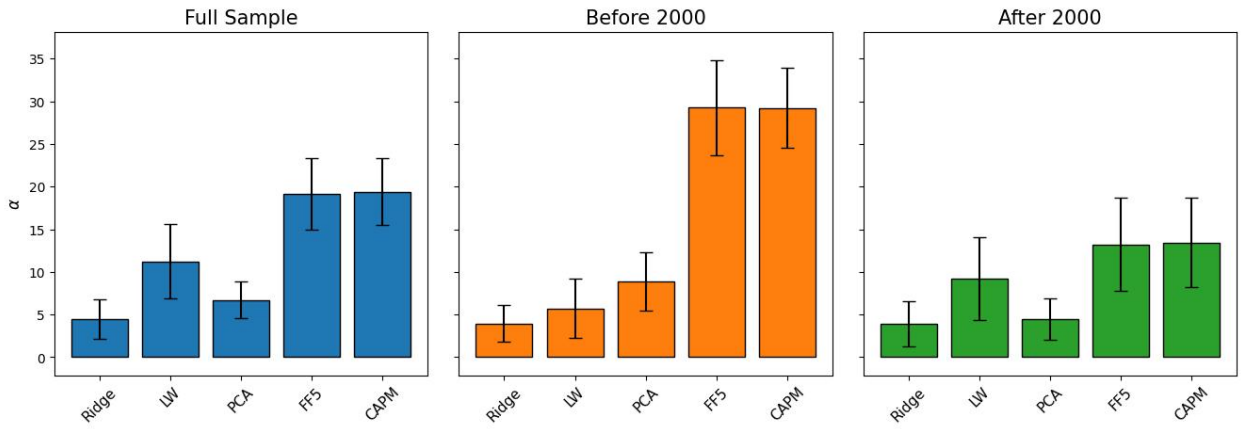

Figure 5: α of UPSA Vs. Benchmarks.

Annual percentage α from the regression of UPSA on benchmarks using heteroscedasticity-adjusted standard errors (5 lags). All portfolios have been scaled to an annual volatility of 10%. The full sample results are from November 1981 to December 2022.

where M_{t+1} denotes a candidate SDF, $F_{i,t+1}$ is a test asset, and E_{OOS} denotes the sample average over the out-of-sample realizations of the SDF portfolio. The first object in equation (19) is simply the out-of-sample average return of the test asset. The second object is the SDF-implied out-of-sample expected return for the test asset. An SDF with smaller out-of-sample pricing errors will exhibit a closer correspondence between these two objects.

The SDF corresponding to a candidate optimal portfolio j is computed as

$$M_{t+1}^{j} = 1 - s_f \pi_t^{j'} F_{t+1} \tag{20}$$

where π_t^j is the optimal portfolio prescribed by model $j \in \{\text{UPSA, ridge, LW, FF5, CAPM}\}$ and estimated using data through date t. The scaling factor s_j ensures that SDF j is correctly scaled to minimize pricing errors.[18] For the test assets F_{t+1}, we use the same set of 153 JKP factors that we use to construct the optimal UPSA, ridge, and LW portfolios. To put test assets on equal footing, we normalize each by its sample standard deviation, thus restating pricing errors in terms of Sharpe ratios rather than average returns.

Figure 6 reports the results of this analysis. UPSA achieves a high degree of accuracy

[18] In particular, $s_j = E_{\text{OOS}}[R_t^f]/E_{\text{OOS}}[(R_t^f)^2]$ is the ratio of the out-of-sample mean return of the portfolio to its out-of-sample second moment. This represents the optimal mean-second-moment weight when the portfolio is treated as a single factor for building an SDF (see Chapter 6 of Cochrane (2009) for a discussion of single-factor pricing models).

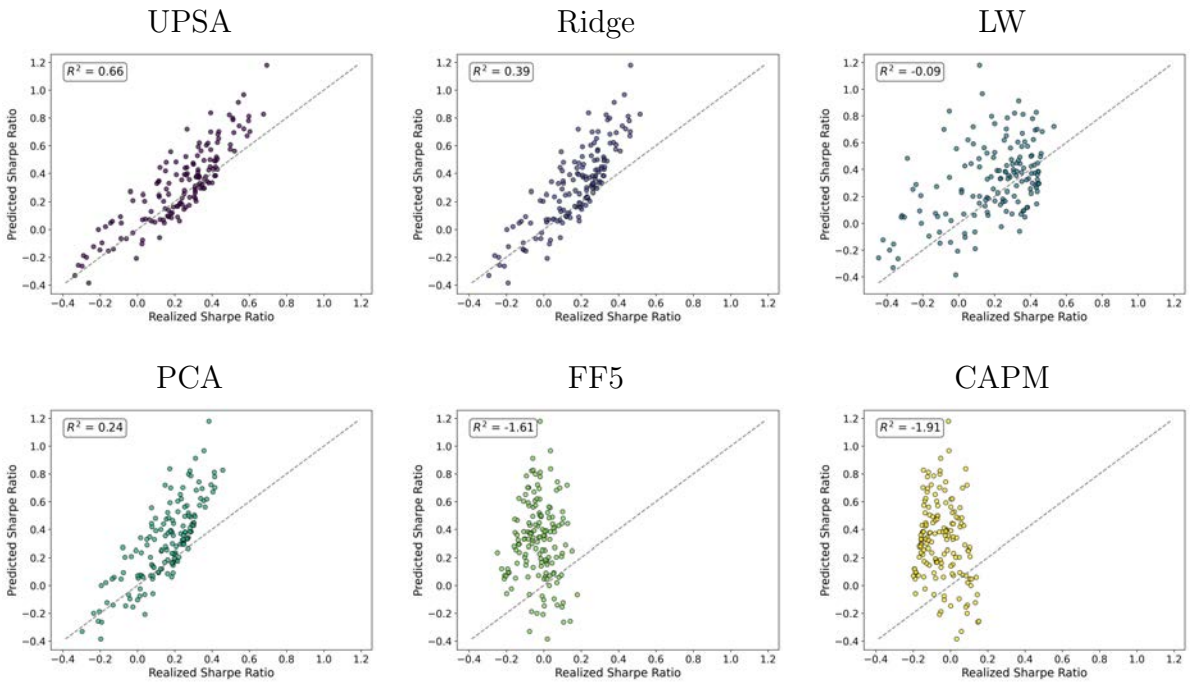

Figure 6: Out-of-sample Pricing Accuracy of Optimal Portfolios.

Comparison of SDF-based out-of-sample Sharpe ratio predictions (vertical axis) versus realized out-of-sample Sharpe ratios (horizontal axis) for test assets defined as the 153 JKP factors. R^2 represents fit versus the 45-degree line. Candidate SDFs are derived from UPSA, ridge, LW, PCA, FF5, and CAPM optimal portfolios. The out-of-sample period is November 1981 to December 2022.

in pricing test assets, explaining factor realized Sharpe ratios with an out-of-sample R^2 of 66%.[19] The next best performers are ridge, and PCA with R^2 values of 39% and 24% respectively. The remaining SDFs fare poorly at pricing the test assets and produce negative out-of-sample R^2.

Jensen et al. (2023) categorizes the 153 JKP factors into 13 themes. We further extend the analysis of Figure 6 by calculating the out-of-sample R^2 SDF-based out-of-sample Sharpe ratio predictions versus realized out-of-sample Sharpe ratios within each theme. For reference, the last row reports the R^2 for all factors combined, corresponding to the numbers reported in Figure 6. UPSA provides a more accurate out-of-sample fit to factor Sharpe ratios within 12 of the 13 themes. The exception is the "low risk" theme in which ridge achieves an R^2 of 85%, though UPSA is a close second with a 76% R^2. The most challenging theme for UPSA

[19]This R^2 is relative to the 45-degree line, thus it imposes an intercept of zero and a slope coefficient of one when comparing predicted and realized Sharpe ratios.

Table 1: Out-of-sample Pricing Accuracy By Factor Theme

This table calculates the R^2 of SDF-based out-of-sample Sharpe ratio predictions versus realized out-of-sample Sharpe ratios within each of the 13 JKP factor themes. R^2 represents fit versus the 45-degree line, as described in the analysis of Figure 6.

Factor Theme	UPSA	Ridge	LW	PCA	FF5	CAPM
Skewness	0.68	0.60	-0.02	0.57	-0.01	-0.24
Profitability	0.53	-0.32	0.01	-0.62	-6.65	-6.83
Low Risk	0.76	0.85	-1.29	0.76	-6.32	-6.69
Value	0.73	0.27	0.30	0.08	-9.23	-12.92
Investment	0.67	0.09	-0.21	-0.25	-8.49	-11.47
Seasonality	0.58	0.44	-0.14	0.36	-0.72	-0.64
Debt Issuance	0.49	0.02	-0.84	-0.23	-1.98	-2.16
Size	0.70	0.62	-1.31	0.54	-0.12	0.02
Accruals	0.50	-0.10	-2.53	-0.59	-2.04	-2.40
Low Leverage	0.29	0.37	-2.17	0.29	-0.66	-1.02
Profit Growth	0.60	0.30	0.70	0.03	-0.35	-0.53
Momentum	0.17	-0.67	0.13	-1.31	-11.29	-12.92
Quality	-0.40	-1.45	-2.48	-1.86	-5.00	-4.35
All	0.66	0.39	-0.09	0.24	-1.61	-1.91

to price was the "quality" theme, where the R^2 was -40%; however, alternative benchmarks performed even worse in this case.

4.5 Robustness

In this section, we provide additional robustness evidence.

4.5.1 Adding Lasso Regularization

Motivated by classic Arbitrage pricing theory (Ross, 1976; Chamberlain and Rothschild, 1982), it is plausible to argue that any candidate SDF should be sparse in the space of PCs. Hence, we follow the intuition in Kozak et al. (2020) and implement a sequential elastic-net

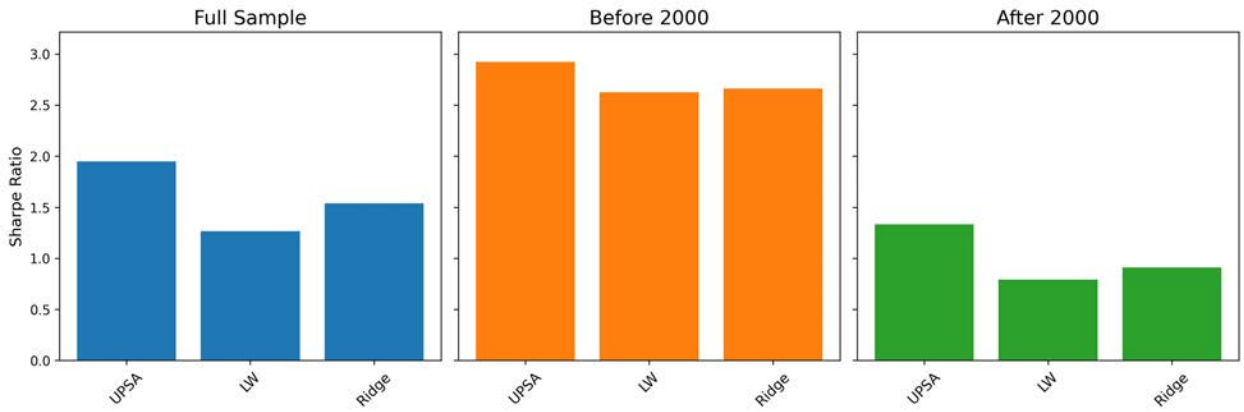

Figure 7: Annual Out-of-sample Sharpe Ratios with Lasso-Augmented Shrinkage.

Annual out-of-sample Sharpe ratios for shrinkage benchmarks—UPSA, LW, and ridge—with an additional Lasso penalty (Lemma 6 of the Appendix). The Lasso penalty is selected from the same grid as for the ridge using leave-one-out cross-validation. The out-of-sample period is November 1981 to December 2022.

shrinkage procedure.[20] Specifically, we first apply our spectral shrinkage methods—UPSA, Ridge, and LW—and then apply an additional Lasso[21] penalty in a second stage.[22] The results are shown in Figure 7. The results suggest that the Sharpe ratios are not materially affected by the introduction of sparsity in PCs. Specifically, UPSA increases slightly to 1.95, while Ridge and LW decrease to 1.53 and 1.26, respectively.

4.5.2 Stratification on Size Groups

Following (Jensen et al., 2023), we form capped value-weighted factors within size groups. At each period, stocks are classified using NYSE breakpoints into three groups—Mega (top 20%), Large (80th–50th percentile), and Small (50th–20th percentile), excluding micro and nano stocks for liquidity reasons. Figures 14 and 15 of the Appendix show the Sharpe ratio and α within size groups. UPSA has statistically significant alphas and larger Sharpe ratios across all groups, with the second-best benchmark differing by size: ridge for Mega, PCA for Large, and LW for Small. This suggests that different forms of shrinkage may be beneficial in

[20] Quaini and Trojani (2022) demonstrates that, under certain technical conditions, sequentially applying ridge followed by Lasso is equivalent to Elastic Net shrinkage. This approach is both lossless and computationally convenient.

[21] Least Absolute Shrinkage and Selection Operator (Lasso; (Tibshirani, 1996)).

[22] For a formal derivation of this approach in our UPSA setting, see Lemmas 5 and 6 in the Appendix.

different groups, but UPSA is flexible enough to adapt to the group-specific data-generating process.

4.5.3 Rolling Window

We evaluate the performance of UPSA across different rolling windows: $t = \{60, 240, 360\}$. Figures 16 and 17 in the Appendix report the corresponding Sharpe ratios and alphas. UPSA outperforms across all horizons, with particularly strong results for the 360-month window, achieving a statistically significant alpha with a t-statistic of 2.4 and a Sharpe ratio gain of 12%. For the 60-month window, UPSA performs best, though its performance is similar to that of LW. In the 240-month case, UPSA again leads, with performance comparable to ridge.[23]

These results suggest that even if non-linear shrinkage is not strictly optimal, it can perform on par with ridge. Likewise, if the primary noise source in the portfolio problem arises from the second moment matrix rather than the means (a situation in which LW thrives), UPSA performs as well as LW.

5 Understanding UPSA's Outperformance

Thus far, we have seen that UPSA dominates competing models in terms of out-of-sample portfolio performance, with simple ridge being the next closest competitor. This section provides a more detailed view of the source of UPSA's improvement over the ridge.

5.1 A Visualization

First, Figure 8 reports the time series of ridge penalties selected by the basic ridge method using LOO cross-validation. Evidently, basic ridge suffers from instability in its penalty selection, often changing by several orders of magnitude from one training sample to the next.

[23] The LW analytical approach (Ledoit and Wolf, 2020) is particularly sensitive to small eigenvalues and tends to underperform when the number of assets is close to the number of observations. In the 240-month setting, its out-of-sample Sharpe ratio is close to zero. We recommend omitting the smallest eigenvalues when applying this methodology in practice.

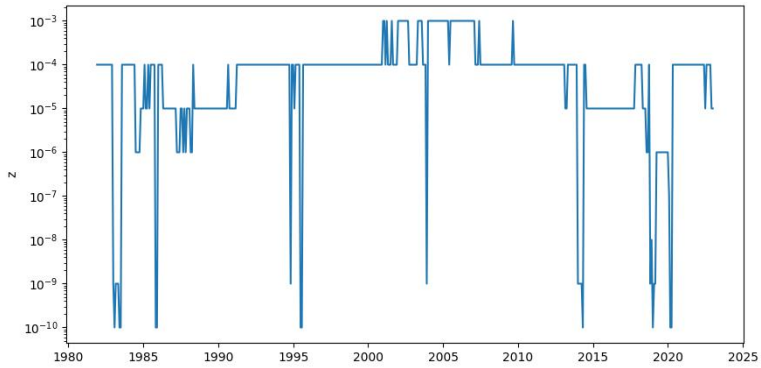

Figure 8: Time Series of Ridge Shrinkage.

Time series of chosen ridge shrinkage. Ridge shrinkage is chosen using leave-one-out cross-validation. The out-of-sample period is November 1981 to December 2022.

UPSA achieves nonlinear shrinkage through a linear combination of simple ridge shrinkages, thus it too is potentially exposed to occasional large changes in the best single ridge shrinkage. This is demonstrated in Figure 9. The three-dimensional portion of this plot shows the time series of UPSA weights across the grid of shrinkage parameter values. The orange line at the top overlays the line from Figure 8 to illustrate the single shrinkage value selected by basic ridge each period. Basic ridge is constrained to choose a single shrinkage value, forcing it to oscillate over time. Meanwhile, UPSA selects a convex combination of these same ridge values, thus allowing it to benefit from contributions of both high and low shrinkage values over time.

Figure 10 shows that, on average, UPSA and basic ridge have a similar allocation across the grid of ridge penalties. The difference is that basic ridge selects only one penalty value at each point in time. By blending the different ridge shrinkages at each point in time, UPSA enjoys the benefit of "shrinkage diversification." As Section 3 shows, diversifying over shrinkages in fact accomplishes nonlinear spectral shrinkage. Thus, Figure 9 offers visual insight into how nonlinear shrinkage improves portfolio optimization.

Finally, Figure 11 shows the eigenvalue shrinkage achieved by UPSA versus ridge (averaged over the rolling training samples). The horizontal axis shows the sample eigenvalues, and the vertical axis shows the shrunken eigenvalues. This figure visualizes the key nonlinear form achieved by UPSA's spectral shrinkage. In particular, it uses similar shrinkage to that

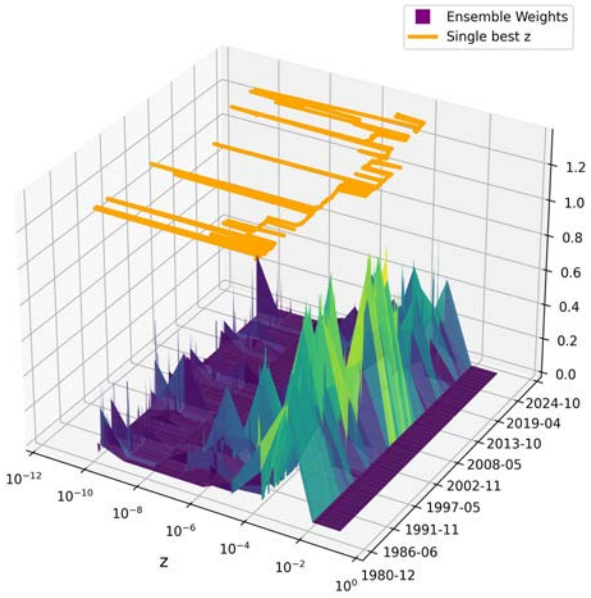

Figure 9: Single Best z Vs. UPSA Ensemble Weights.

Time series of chosen ridge shrinkage overlayed on UPSA ensemble weights. Optimal ridge shrinkage is chosen using leave-one-out cross-validation, and ensemble weights are determined using Lemma 2. The out-of-sample period is November 1981 to December 2022.

of ridge for large eigenvalues. However, it shrinks small eigenvalues far less than ridge, essentially extracting more Sharpe ratio from small principal components. The linearity of the ridge constrains how it trades off the shrinkage of large and small eigenvalues. Thanks to its nonlinearity, UPSA is not bound by this constraint and can enjoy the best of both worlds (heavy shrinkage of large eigenvalues and light shrinkage of small ones).

5.2 UPSA Through the Lens of Simulation

In this section, we present a simulation analysis to help interpret our empirical findings. These simulations illustrate how UPSA's two key attributes, i) tuning the shrinkage function to directly optimize a portfolio objective and ii) allowing for flexible nonlinear shrinkage functions, contribute to its outperformance versus competing methods.

We simulate data from the following simple model:

$$F_t = \mu \; + \; (\Sigma - \mu\mu')^{1/2} X_t, \tag{21}$$

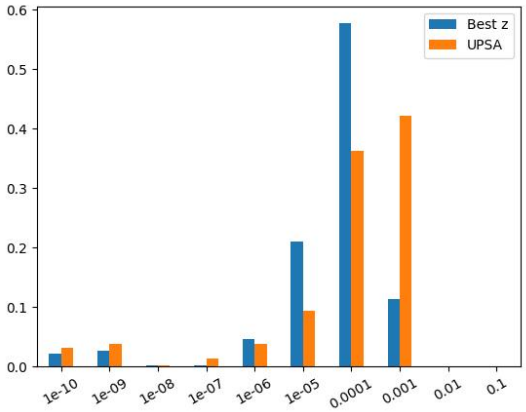

Figure 10: Histogram of Selected Ridge Penalties and UPSA Weights.

Histogram of selected ridge penalties overlapped with the average UPSA ensemble weights. The out-of-sample period is November 1981 to December 2022.

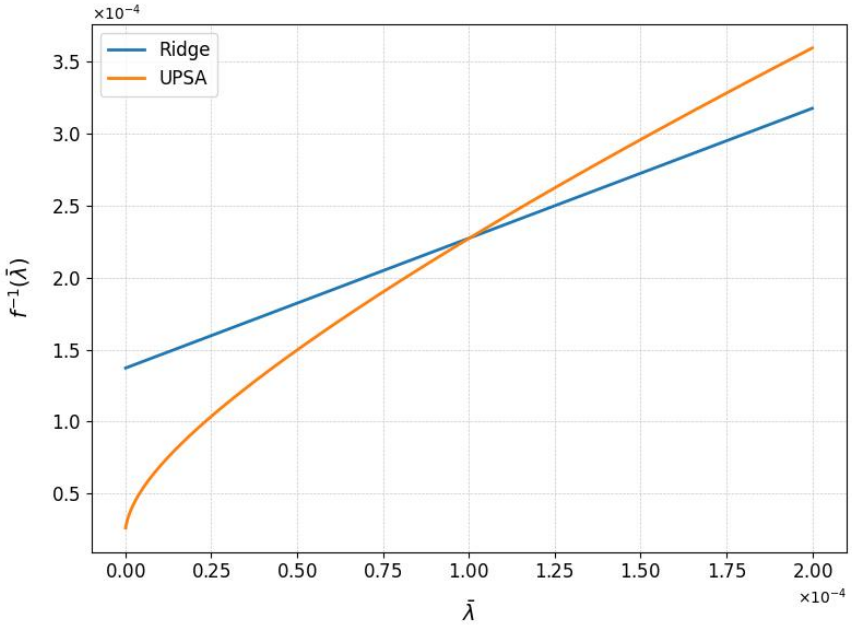

Figure 11: Historical Average Shrinkage Functions for Ridge and UPSA.

Historical averages of ridge and UPSA shrinkage functions. Ridge penalties are selected using leave-one-out cross-validation, and UPSA weights are determined using Lemma 2. Both are computed over the period from December 1981 to December 2022.

where $X_t \sim \mathbb{N}(0, I_{N\times N})$, $\Sigma \in \mathbb{R}^{N\times N}$ is the second moment matrix of returns, and $\mu \in \mathbb{R}^N$ is the mean vector. For simplicity, we assume the second moment matrix is diagonal, $\Sigma = \text{diag}(\lambda)$, hence the population eigenvectors constitute an identity matrix ($U = I_{N\times N}$).

The infeasible optimal Markowitz portfolio weights ($\pi_{*,i} = \mu_i/\lambda_i$) are the PC mean second moment tradeoffs.

We assume that the eigenvalue distribution for the asset return second moment matrix has 10% of the eigenvalues equal to 10, 80% equal to 1, and the remaining 10% equal to 0.1. Expected returns are specified as

$$\mu_i = \begin{cases} \frac{1}{\sqrt{N}} & \text{if } \lambda_i = 10, \\ 0 & \text{if } \lambda_i = 1, \\ \frac{0.5}{\sqrt{N}} & \text{if } \lambda_i = 0.1 \end{cases} \tag{22}$$

The economic interpretation of this generating process is that the asset market consists of three types of factors. Factors for which $\lambda_i = 10$ represent concentrated risks in the market and thus command a significant return premium. Then there is a large number of factors with moderate variance ($\lambda_i = 1$) and no premium. Finally, there is a minority of factors with low risk ($\lambda_i = 0.1$) and non-negligible expected returns, thus resembling near-arbitrage opportunities. These near-arbitrage factors are challenging to estimate because their variance is overshadowed by other factors with much higher risk.

Based on this specification, we produce 10,000 simulated samples each with $N = 150$ assets and $T = 600$ observations, roughly matching our empirical sample dimensions in Section 4. In each simulation, we use the first half of the time series to train models and the second half to test their out-of-sample performance. We compare model performance based on the (modified) out-of-sample Sharpe ratio $SR(R_t) = \bar{E}[R_t]/\sqrt{\bar{E}[R_t^2]}$, where $\bar{E}[R^i]$ is the i^{th} empirical moment of asset returns R_t.[24]

5.2.1 Tuning to an Economic Objective Versus a Statistical Objective

Our first set of simulations highlights the importance of tuning the degree of shrinkage to the economic portfolio objective, rather than to an indirectly related statistical objective. To draw out this comparison, we compare two portfolio shrinkage approaches. Both use standard

[24]By Jensen's inequality, this uncentered Sharpe ratio is always below one. We adopt it to keep values in a bounded, interpretable range and ensure consistency with our shrinkage objective.

linear ridge shrinkage, but one is tuned to optimize expected out-of-sample quadratic utility (as in UPSA), while the other is tuned to minimize expected out-of-sample second-moment error and is thus only indirectly related to portfolio performance (as in LW).[25] Specifically, we have use of the following approaches:

- **Ridge Statistical:** Ridge shrinkage to minimize second moment matrix estimation error, using the methodology in Ledoit and Wolf (2003):[26]

$$\min_f \|\bar{U}\, f(\bar{\lambda})\, \bar{U}' - \Sigma\| \quad \text{s.t.} \quad f(\bar{\lambda}_i) = f_z(\bar{\lambda}_i) = \frac{1}{z + \bar{\lambda}_i}.$$

- **Ridge Economic:** Ridge shrinkage to maximizes out-of-sample quadratic utility chosen with LOO:

$$\max_f E\big[R_t^{\bar{\pi}(f)} \mid F_{IS}\big] \; - \; \tfrac{1}{2}\, E\big[(R_t^{\bar{\pi}(f)})^2 \mid F_{IS}\big] \quad \text{s.t.} \quad f(\bar{\lambda}_i) = f_z(\bar{\lambda}_i) = \frac{1}{z + \bar{\lambda}_i}. \tag{23}$$

At the population level, the smaller PCs are the most profitable, delivering the highest Sharpe ratios. To maximize out-of-sample performance, the shrinkage function must therefore tilt portfolio weights toward these smaller PCs, despite the presence of substantial finite-sample noise. Statistical shrinkage methods, however, will focus solely on adjusting noisy empirical eigenvalues toward their population counterparts, agnostic of the means of those PCs.

This distinction is evident in Figure 12, where statistical ridge shrinks empirical eigenvalues toward the bulk of the population eigenvalues, typically around 10 and 0.1. In contrast, the ridge with an economic objective places greater weight on the smaller, high-return PCs and less on the large eigenvalue components. As a result, the average out-of-sample Sharpe ratio improves to 0.27 under the economic shrinkage, compared to 0.21 with the statistical approach.

[25] The LW methodology was developed to specifically tune to this particular statistical objective and is not immediately adaptable for tuning to an economic objective.

[26] Matrix estimation error is measured with the Frobenius norm. We make use of the SKlearn package.

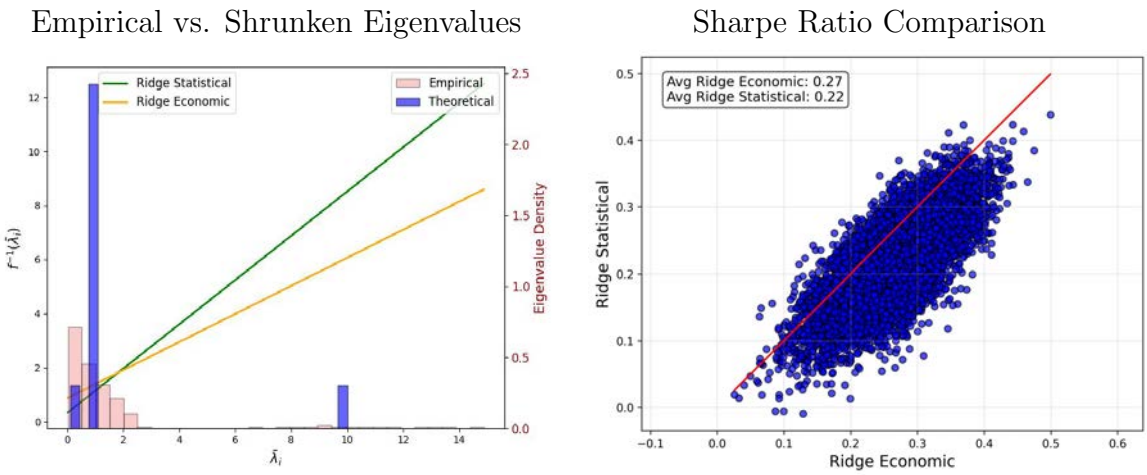

Figure 12: Tuning to Economic Versus Statistical Objectives.

Data is simulated using the data-generating process (21)–(22) for $N = 150$ assets, $T = 600$ observations, and 10,000 simulations. The left panel shows the empirical eigenvalue distribution versus the shrunken eigenvalues for the statistical and economic tuning approaches (averaged across simulations). The right panel shows a scatter plot of the out-of-sample Sharpe ratio of portfolios built using linear economic tuning (horizontal axis) or linear statistical tuning (vertical axis), where each point represents a simulated data set.

5.2.2 Linear vs. Nonlinear Shrinkage with Economic Objectives

Next, we highlight the gains of flexible nonlinear spectral shrinkage (as in UPSA) versus linear shrinkage (as in ridge) while holding the economic tuning approach fixed. Specifically, we compare the following two shrinkage approaches:

- **Ridge:** Linear shrinkage maximizing the out-of-sample quadratic utility (23).
- **UPSA:** Nonlinear shrinkage maximizing the out-of-sample quadratic utility in Definition 1.

The true data-generating process features three types of risk–return trade-offs: (i) PCs with low variance but high risk-return tradeoff, (ii) PCs with moderate variance and zero expected returns, and (iii) PCs with high variance and moderate risk-return tradeoff. An effective shrinkage methodology should apply minimal shrinkage to the low-variance, high-return PCs to preserve their profitability, aggressively shrink the mid-variance PCs since they contribute little to the efficient portfolio, and apply moderate shrinkage to the high-variance PCs, which are less noisy and moderately informative. With its simplistic and

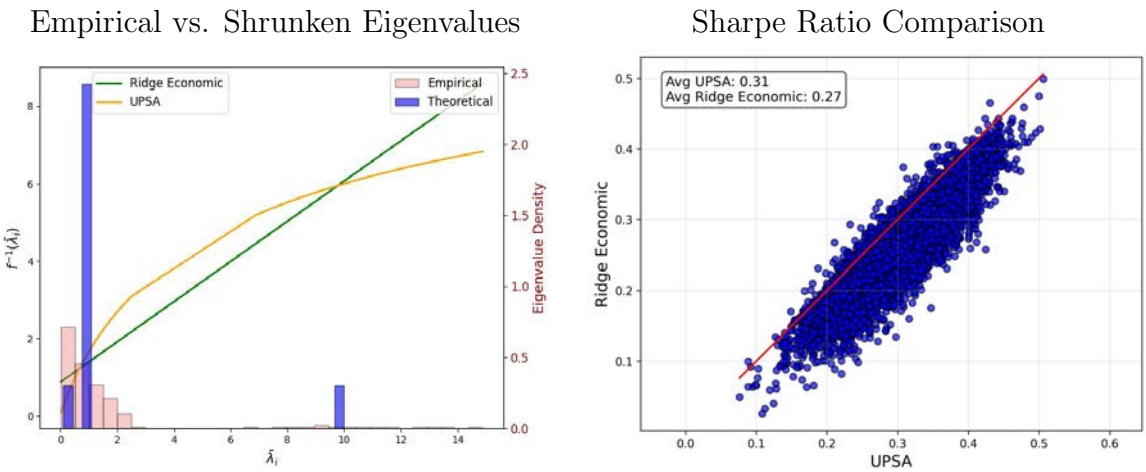

Figure 13: Tuning to Economic Versus Statistical Objectives.

Data is simulated using the data-generating process (21) and (22) for $N = 150$ assets, $T = 600$ observations, and 10,000 simulations. The left panel shows the empirical eigenvalue distribution versus the shrunken eigenvalues for the statistical and economic tuning approaches (averaged across simulations). The right panel shows a scatter plot of the out-of-sample Sharpe ratio of portfolios built using linear economic tuning (horizontal axis) or nonlinear economic tuning (vertical axis), where each point represents a simulated data set.

linear shrinkage structure, ridge lacks the flexibility to address all three cases simultaneously, whereas UPSA can.

Figure 13 illustrates this behavior. Under UPSA, the small but high-return eigenvalues are shrunk less, enabling the model to retain their contribution to portfolio performance. The mid-range eigenvalues, which offer little return benefit, receive stronger shrinkage, while the largest eigenvalues are again shrunk less due to their relatively high signal-to-noise ratio. The nonlinear shrinkage structure of UPSA allows it to consistently outperform ridge, achieving a higher average out-of-sample Sharpe ratio of 0.31 compared to ridge's 0.27.

6 Conclusions

We introduce the Universal Portfolio Shrinkage Approximator (UPSA), an objective-specific and adaptive shrinkage methodology. UPSA directly optimizes out-of-sample portfolio performance and achieves a balance between risk and return through a flexible, closed-form nonlinear shrinkage function.

Empirically, UPSA consistently outperforms existing methods, achieving higher Sharpe

ratios and lower pricing errors. It adapts dynamically to varying market conditions, efficiently allocating risk across principal components without relying on arbitrary sparsity constraints. Our findings underscore the importance of designing shrinkage strategies that are flexible and aligned with the portfolio's specific objective, rather than based on ad hoc statistical criteria.

References

Bryzgalova, Svetlana, Jiantao Huang, and Christian Julliard, "Bayesian solutions for the factor zoo: We just ran two quadrillion models," *The Journal of Finance*, 2023, *78* (1), 487–557.

___, Victor DeMiguel, Sicong Li, and Markus Pelger, "Asset-Pricing Factors with Economic Targets," *Available at SSRN 4344837*, 2023.

Chamberlain, Gary and Michael Rothschild, "Arbitrage, factor structure, and mean-variance analysis on large asset markets," 1982.

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

Da, Rui, Stefan Nagel, and Dacheng Xiu, "The Statistical Limit of Arbitrage," Technical Report, Chicago Booth 2022.

DeMiguel, Victor, Lorenzo Garlappi, Francisco J Nogales, and Raman Uppal, "A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms," *Management Science*, 2009, *55*, 798–812.

Didisheim, Antoine, Shikun Barry Ke, Bryan T Kelly, and Semyon Malamud, "APT or "AIPT"? The Surprising Dominance of Large Factor Models," Technical Report, National Bureau of Economic Research 2024.

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

Giglio, Stefano and Dacheng Xiu, "Asset pricing with omitted factors," *Journal of Political Economy*, 2021, *129* (7), 1947–1990.

___, ___, and Dake Zhang, "Test assets and weak factors," Technical Report, National Bureau of Economic Research 2021.

Gu, Shihao, Bryan Kelly, and Dacheng Xiu, "Autoencoder asset pricing models," *Journal of Econometrics*, 2021, *222* (1), 429–450.

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

Jensen, Theis Ingerslev, Bryan Kelly, and Lasse Heje Pedersen, "Is there a replication crisis in finance?," *The Journal of Finance*, 2023, *78* (5), 2465–2518.

Kelly, Bryan, Semyon Malamud, and Lasse Heje Pedersen, "Principal portfolios," *The Journal of Finance*, 2023, *78* (1), 347–387.

___, Seth Pruitt, and Yinan Su, "Instrumented Principal Component Analysis," *Working paper*, 2020.

Kelly, Bryan T, Semyon Malamud, and Kangying Zhou, "The virtue of complexity in return prediction," Technical Report, National Bureau of Economic Research 2022.

Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh, "Interpreting factor models," *The Journal of Finance*, 2018, *73* (3), 1183–1223.

___, ___, and ___, "Shrinking the cross-section," *Journal of Financial Economics*, 2020, *135* (2), 271–292.

Ledoit, Olivier and Michael Wolf, "Improved estimation of the covariance matrix of stock returns with an application to portfolio selection," *Journal of Empirical Finance*, 2003, *10*, 603–621.

___ and ___, "Honey, I shrunk the sample covariance matrix," *Journal of Portfolio Management*, 2004, *30*, 110–119.

___ and ___, "A well-conditioned estimator for large-dimensional covariance matrices," *Journal of multivariate analysis*, 2004, *88* (2), 365–411.

___ and ___, "Nonlinear shrinkage estimation of large-dimensional covariance matrices," *The Annals of Statistics*, 2012, *40* (2), 1024–1060.

___ and ___, "Spectrum estimation: A unified framework for covariance matrix estimation and PCA in large dimensions," *Journal of Multivariate Analysis*, 2015, *139*, 360–384.

___ and ___, "Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets Goldilocks," *The Review of Financial Studies*, 2017, *30* (12), 4349–4388.

___ **and** ___ , "Analytical nonlinear shrinkage of large-dimensional covariance matrices," *The Annals of Statistics*, 2020, *48* (5), 3043–3065.

___ **and Sandrine Péché**, "Eigenvectors of some large sample covariance matrix ensembles," *Probability Theory and Related Fields*, 2011, *150*, 233–264.

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

Löwner, Karl, "Über monotone matrixfunktionen," *Mathematische Zeitschrift*, 1934, *38* (1), 177–216.

Markowitz, Harry, "Portfolio Selection," *The Journal of Finance*, 1952, *7* (1), 77–91.

Martin, Ian WR and Stefan Nagel, "Market efficiency in the age of big data," *Journal of Financial Economics*, 2021.

Pedersen, Lasse Heje, Abhilash Babu, and Ari Levine, "Enhanced portfolio optimization," *Financial Analysts Journal*, 2021, *77* (2), 124–151.

Preite, Massimo Dello, Raman Uppal, Paolo Zaffaroni, and Irina Zviadadze, "What is Missing in Asset-Pricing Factor Models?," 2022.

Quaini, Alberto and Fabio Trojani, "Proximal Estimation and Inference," *arXiv preprint arXiv:2205.13469*, 2022.

Ross, Stephen A., "The Arbitrage Theory of Capital Asset Pricing," *Journal of Economic Theory*, 1976, *13*, 341–360.

Rudin, Walter, *Principles of Mathematical Analysis*, 3 ed., New York: McGraw-Hill, 1976. See Chapter 7 for the Stone-Weierstrass Theorem.

Severini, Thomas A, "Some properties of portfolios constructed from principal components of asset returns," *Annals of Finance*, 2022, *18* (4), 457–483.

Stein, Charles, "Lectures on the theory of estimation of many parameters," *Journal of Soviet Mathematics*, 1986, *34*, 1373–1403.

Tibshirani, Robert, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 1996, *58* (1), 267–288.

A Proofs

Proof of Lemma 1. To prove the first statement of the lemma, let f be a real-valued, non-negative, matrix monotone decreasing function on $\mathbb{R}_+$ so that $g := -f$ is negative and matrix monotone increasing. Using (Löwner, 1934) Theorem, it follows that there exist constants $a \in \mathbb{R}$, $b > 0$ and a positive finite measure μ on $\mathbb{R}_+$ such that, for $\bar{\lambda} \geq 0$:

$$g(\bar{\lambda}) = a + \bar{\lambda} b + \int_0^\infty \frac{\bar{\lambda}}{z + \bar{\lambda}} d\mu(z) = a + \bar{\lambda}(b + \mu(\mathbb{R}_+)) - \int_0^\infty \frac{z}{z + \bar{\lambda}} d\mu(z) \,. \tag{24}$$

By monotone convergence, $0 = \lim_{\bar{\lambda} \to \infty} \int_0^\infty \frac{z}{z+\bar{\lambda}} d\mu(z)$. Therefore,

$$0 = \lim_{\bar{\lambda} \to \infty} g(\bar{\lambda}) = \lim_{\bar{\lambda} \to \infty} \left(a + \bar{\lambda} \left(b + \mu(\mathbb{R}_+) \right) \right) \,, \tag{25}$$

which implies $a = 0$ and $b + \mu(\mathbb{R}_+) = 0$. This gives the representation:

$$f(\bar{\lambda}) = \int_0^\infty \frac{z}{z + \bar{\lambda}} d\mu(z) =: \int_0^\infty \frac{1}{z + \bar{\lambda}} d\nu(z) \,, \tag{26}$$

for a positive measure ν on $\mathbb{R}_+$ having Radon-Nykodim derivative $\frac{d\nu}{d\mu}(z) = z$ with respect to μ. In particular, we obtain that $\int_0^\infty z^{-1} d\nu(z) = 1$ when $f(0) = 1$. Furthermore, $\bar{\lambda} f(\bar{\lambda})$ is bounded by assumption. Hence, there exists a constant $K > 0$ such that:

$$K \geq \sup_{\bar{\lambda} \geq 0} \bar{\lambda} f(\bar{\lambda}) \geq \lim_{\bar{\lambda} \to \infty} \int_0^\infty \frac{\bar{\lambda}}{z + \bar{\lambda}} d\nu(z) = \int_0^\infty d\nu(z) \,, \tag{27}$$

using in the last identity the monotone convergence theorem. In particular, we obtain that $\nu(\mathbb{R}_+) = K$ whenever $K = \lim_{\bar{\lambda} \to \infty} \bar{\lambda} f(\bar{\lambda})$.

We next show the uniform approximation property of UPSA. To this end, note first that for any $z_1, z_2 \geq 0$ and $\bar{\lambda} \geq \bar{\lambda}_{\min} > 0$ we have:

$$\left| \frac{1}{z_1 + \bar{\lambda}} - \frac{1}{z_2 + \bar{\lambda}} \right| \leq \frac{|z_2 - z_1|}{(z_1 + \bar{\lambda})(z_2 + \bar{\lambda})} \leq \frac{|z_2 - z_1|}{\bar{\lambda}^2} < \frac{1}{\bar{\lambda}^2_{\min}} |z_2 - z_1| \,.$$

Let further $\epsilon > 0$ be arbitrary and $z_{\max} > 0$ be such that $\frac{1}{z+\bar{\lambda}} < \epsilon$ for any $\bar{\lambda} \geq 0$ and $z \geq z_{\max}$. There then exists a partition $\mathbb{R}_+ \backslash [z_{\max}, \infty) = \bigcup_{i=1}^N [z_{i-1}, z_i)$, where $z_0 := 0$ and $z_N := z_{\max}$, such that $|z_i - z_{i-1}| < \epsilon$ for any $i = 1 \ldots N$. Consider now a piece-wise constant

function $g(\bar{\lambda}, z)$ with respect to variable z, which is defined for any $\bar{\lambda} \geq 0$ by $g(\bar{\lambda}, z) = \frac{1}{z_i + \bar{\lambda}}$ if $z \in [z_i, z_{i-1})$ and $g(\bar{\lambda}, z) = \frac{1}{z_{\max} + \bar{\lambda}}$ if $z \geq z_{\max}$. Using this function, we obtain:

$$\int_0^\infty g(\bar{\lambda}, z) d\nu(z) = \sum_{i=1}^{n} \frac{1}{z_i + \bar{\lambda}} \nu([z_{i-1}, z_i)) + \frac{1}{z_{\max} + \bar{\lambda}} \nu([z_{\max}, \infty)) \ , \tag{28}$$

which is by definition a function in the UPSA family. Furthermore, since $\bar{\lambda}_{\min} < 1$, without loss of generality, following inequalities hold:

$$\begin{aligned}
\left| f(\bar{\lambda}) - \int_0^\infty g(\bar{\lambda}, z) d\nu(z) \right| &\leq \int \left| \frac{1}{z + \bar{\lambda}} - g(\bar{\lambda}, z) \right| d\nu(z) \\
&= \sum_{i=1}^{N} \int_{z_{i-1}}^{z_i} \left| \frac{1}{z + \bar{\lambda}} - \frac{1}{z_i + \bar{\lambda}} \right| d\nu(z) \\
&\quad + \int_{z_{\max}}^{\infty} \left| \frac{1}{z + \bar{\lambda}} - \frac{1}{z_{\max} + \bar{\lambda}} \right| d\nu(z) \\
&\leq \epsilon \frac{\nu([0, \bar{\lambda}_{\max}))}{\bar{\lambda}_{\min}^2} + \epsilon \nu([\bar{\lambda}_{\max}, \infty)) \\
&\leq \epsilon \frac{\nu(\mathbb{R}_+)}{\bar{\lambda}_{\min}^2} \ .
\end{aligned}$$

Note that ϵ was chosen arbitrary and that this bound holds uniformly for any $\bar{\lambda} \geq \bar{\lambda}_0$, given a fixed but otherwise arbitrary $\bar{\lambda}_0 > 0$. Therefore, the stated uniform approximation property of UPSA for decreasing matrix monotone shrinkage functions follows. This concludes the proof. □

Lemma 3 *If return process (F_t) is identically and independently distributed over time with expectation μ and second moment matrix Σ, the solution to the optimization problem in Definition 1 under full information, i.e., under knowledge of both μ and Σ, is such that for any $i = 1, \ldots, K := \min(N, T)$:*

$$f_\star(\bar{\lambda}_i) = \frac{\bar{U}_i' \pi}{\bar{U}_i' \bar{\mu}}. \tag{29}$$

Proof. For a fixed $t > T$, we first write explicitly the criterion that has to be optimized in Definition 1, while recalling that returns are identically and independently distributed over

time:

$$\begin{aligned} U(\bar{\pi}(f)|F_{IS}) &:= E[R_t^{\bar{\pi}(f)}|F_{IS}] - \frac{1}{2}E\left[(R_t^{\bar{\pi}(f)})^2|F_{IS}\right] \\ &= \bar{\mu}' f(\bar{\Sigma})\mu - \frac{1}{2}\bar{\mu}' f(\bar{\Sigma})\Sigma f(\bar{\Sigma})\bar{\mu} \\ &= (\bar{U}'\bar{\mu})'\,\mathrm{diag}(f(\bar{\lambda}))\bar{U}'\mu - \frac{1}{2}(\bar{U}'\bar{\mu})'\,\mathrm{diag}(f(\bar{\lambda}))\bar{U}'\Sigma\bar{U}\,\mathrm{diag}(f(\bar{\lambda}))\bar{U}'\bar{\mu} \\ &= (\boldsymbol{f}(\bar{\lambda}) \circ \bar{U}'\bar{\mu})'\bar{U}'\mu - \frac{1}{2}(\boldsymbol{f}(\bar{\lambda}) \circ \bar{U}'\bar{\mu})'\bar{U}'\Sigma\bar{U}(\boldsymbol{f}(\bar{\lambda}) \circ \bar{U}'\bar{\mu}) \end{aligned}$$

where $\boldsymbol{f}(\bar{\lambda}) := (f(\bar{\lambda}_1), \ldots, f(\bar{\lambda}_K))'$ and $\circ$ denotes the Hadamar product. It follows that the optimal shrinkage $\boldsymbol{f}_\star(\bar{\lambda})$ is such that:

$$\boldsymbol{f}_\star(\bar{\lambda}) \circ \bar{U}'\bar{\mu} = (\bar{U}'\Sigma\bar{U})^{-1}\bar{U}'\mu = \bar{U}'\Sigma^{-1}\mu = \bar{U}'\pi_\star \;, \tag{30}$$

where $\pi_\star$ denotes the (unfeasible) population Markowitz portfolio. Assuming that all PC's have a nonzero average return, it then follows for any $i = 1, \ldots, K$:

$$f_\star(\bar{\lambda}_i) = \frac{\bar{U}_i'\pi_\star}{\bar{U}_i'\bar{\mu}}. \tag{31}$$

This concludes the proof. □

Lemma 4 *Let $\{F_1, \ldots, F_T, F_{T+1}\}$ be an exchangeable random sequence and*

$$R_{T+1,t}(f) \;:=\; \bar{\pi}_{T+1,t}(f)'F_t \;;\; t = 1, \ldots, T+1 \;, \tag{32}$$

denote the sequence of LOO portfolio returns. Then,

$$U_{LOO}^{OOS}(f) \;:=\; \frac{1}{T+1}\sum_{t=1}^{T+1} U(R_{T+1,t}(f)) \tag{33}$$

is an unbiased estimator of the out-of-sample portfolio's expected utility:

$$E[U(R_\tau(f))] \;=\; E\left[U_{LOO}^{OOS}(f)\right] \;, \tag{34}$$

for any $\tau > T$.

Proof of Lemma 4. Exchangeability implies that the distribution of $((F_s)_{s\neq\tau,1\leq s\leq T}, F_\tau)$ is independent of $\tau \in \{1,\ldots,T\}$. Therefore, the distribution of return $R_{T,\tau}(f) = \bar{\pi}_{T,\tau}(f)'F_\tau$ does also not depend on τ. This gives, for any $\tau' \in \{1,\ldots T\}$:

$$E\left[U_{LOO}^{OOS}(f)\right] = E\left[\frac{1}{T}\sum_{\tau=1}^{T} U(R_{T,\tau}(f))\right] = E[U(R_{T,\tau'}(f))] = E[U(R_{T,T}(f))] \ .$$

Since $E[U(R_{T,T}(f)]$ is the out-of-sample portfolio expected utility criterion implied by all data available up to time $T-1$ for factor return F_T, the proof is complete. □

Lemma 5 *Given a generalized Ridge penalty of the form $g_V(\pi) = \frac{1}{2}\pi' V\pi$, consider the associated generalized Elastic net-type penalized portfolio optimization problem:*

$$\bar{\pi}_V(\alpha) = \arg\min_{\pi}\left\{-\bar{\mu}'\pi + \frac{1}{2}\pi'\bar{\Sigma}\pi + g_V(\pi) + \alpha||\pi||_1\right\} \ .$$

Using the optimal generalized ridge portfolio $\bar{\pi}_V(0) = (\bar{\Sigma}+V)^{-1}\bar{\mu}$, this optimal portfolio can be equivalently written as:

$$\bar{\pi}_V(\alpha) = \arg\min_{\pi}\left\{\frac{1}{2}(\bar{\pi}_V(0)-\pi)'(\bar{\Sigma}+V)(\bar{\pi}_V(0)-\pi) + \alpha||\pi||_1\right\} \ .$$

By setting $\bar{\Sigma}+V = f^{-1}(\bar{\Sigma})$, the case of the nonlinear shrinkage optimal portfolio reads as follows:

$$\bar{\pi}(f,\alpha) = \arg\min_{\pi}\left\{\frac{1}{2}(\bar{\pi}(f)-\pi)'f^{-1}(\bar{\Sigma})(\bar{\pi}(f)-\pi) + \alpha||\pi||_1\right\} \ .$$

Proof. Given penalty $g_V(\pi) = \frac{1}{2}\pi' V\pi$, the associated generalized Elastic net-type portfolio reads:

$$\bar{\pi}_V(\alpha) = \left\{-\bar{\mu}'\pi + \frac{1}{2}\pi'(\bar{\Sigma}+V)\pi + \alpha||\pi||_1\right\} \ .$$

The generalized ridge solution follows for $\alpha = 0$ as:

$$\bar{\pi}_V(0) = (\bar{\Sigma}+V)^{-1}\bar{\mu} \tag{35}$$

We can now write the generalized Elastic Net solution as a penalized minimum distance

correction of the generalized ridge solution. To this end, consider following quadratic objective function:

$$\begin{aligned} ||\bar{\pi}_V(0) - \pi||^2_{\bar{\Sigma}+V} &:= (\bar{\pi}_V(0) - \pi)'(\bar{\Sigma} + V)(\bar{\pi}_V(0) - \pi) \\ &= \bar{\pi}_V(0)'(\bar{\Sigma} + V)\bar{\pi}_V(0) + \pi'(\bar{\Sigma} + V)\pi - 2\bar{\mu}'\pi \ . \end{aligned}$$

This yields:

$$\bar{\pi}_V(\alpha) = \arg\min_{\pi} \left\{ \frac{1}{2}||\bar{\pi}_V(0) - \pi||^2_{\bar{\Sigma}+V} + \alpha||\pi||_1 \right\} \ .$$

For the case of the nonlinear shrinkage estimator, set $\bar{\Sigma} + V = f^{-1}(\bar{\Sigma})$. This yields, for the corresponding generalized Elastic Net optimal portfolio:

$$\bar{\pi}(f, \alpha) = \arg\min_{\pi} \left\{ \frac{1}{2}||\bar{\pi}(f) - \pi||^2_{f^{-1}(\bar{\Sigma})} + \alpha||\pi||_1 \right\} \ ,$$

concluding the proof. □

Lemma 6 *The nonlinear shrinkage Elastic Net optimal portfolio in the space of PC factors is the solution of the optimization problem:*

$$\bar{\pi}^C(f, \alpha) = \arg\min_{\pi^C} \left\{ \frac{1}{2}||\bar{\pi}^C(f) - \pi^C||^2_{\text{diag}(f^{-1}(\bar{\lambda}))} + \alpha||\pi^C||_1 \right\} \ ,$$

where $\text{diag}(f^{-1}(\bar{\lambda}))$ *is the diagonal matrix containing the eigenvalues of* $\bar{\Sigma}$ *after shrinkage and* $\bar{\pi}^C(f) = \text{diag}(f(\bar{\lambda}))\bar{U}'\bar{\mu}$ *if the nonlinearly shrunk optimal PC portfolio. Using the standard soft-thresholding operator* $\mathcal{S}(z, \lambda) = sgn(z)\max(0, |z| - \lambda)$*, the Elastic Net optimal portfolio is given in closed-form for any* $i = 1, \ldots, \min(N, T)$ *with the soft-thresholding formula:*

$$\bar{\pi}^C_i(f, \alpha) = \mathcal{S}(\bar{\pi}^C_i(f), \alpha f(\bar{\lambda}_i)) \ . \tag{36}$$

Proof. With Lemma 5, the nonlinear shrinkage Elastic net optimal portfolio of principal

component factors can be written as:

$$\begin{aligned}\bar{\pi}^C(f,\alpha) &= \arg\min_{\pi^C}\left\{\frac{1}{2}||\bar{\pi}^C(f)-\pi^C||^2_{\mathrm{diag}(f^{-1}(\bar{\lambda}))}+\alpha||\pi^C||_1\right\} \\ &= \arg\min_{\pi^C}\left\{\frac{1}{2}\sum_i\frac{(\bar{\pi}^C_i(f)-\pi^C_i)^2}{f(\bar{\lambda}_i)}+\alpha\sum_i|\pi^C_i|\right\}\ ,\end{aligned}$$

where

$$\bar{\pi}^C_i(f)=\mathrm{diag}(f(\bar{\lambda}))\bar{U}'\bar{\mu}\ .$$

Whenever $\bar{\pi}^C_i(f,\alpha)>0$ for some i, the associated optimality condition reads:

$$\begin{aligned}0 &= \bar{\pi}^C_i(f,\alpha)-\bar{\pi}^C_i(f)+\alpha f(\bar{\lambda}_i) \\ &\Longleftrightarrow \\ \bar{\pi}^C_i(f,\alpha) &= \bar{\pi}^C_i(f)-\alpha f(\bar{\lambda}_i)\end{aligned}$$

Analogously, whenever $\bar{\pi}^C_i(f,\alpha)<0$:

$$\bar{\pi}^C_i(f,\alpha)=\bar{\pi}^C_i(f)+\alpha f(\bar{\lambda}_i)\ .$$

In all other cases, $\bar{\pi}^C_i(f,\alpha)=0$. Summarizing, we have shown that the optimal portfolio $\bar{\pi}^C(f,\alpha)$ has components $i=1,\ldots,\min(N,T)$ given in closed-form by:

$$\bar{\pi}^C_i(f,\alpha)=\begin{cases}\bar{\pi}^C_i(f)-\alpha f(\bar{\lambda}_i) & \text{if } \bar{\pi}^C_i(f)>\alpha f(\bar{\lambda}_i)\\ \bar{\pi}^C_i(f)+\alpha f(\bar{\lambda}_i) & \text{if } \bar{\pi}^C_i(f)<-\alpha f(\bar{\lambda}_i)\\ 0 & \text{else}\end{cases}$$

Therefore, $\bar{\pi}^C_i(f,\alpha)$ is computable with the soft-thresholding formula:

$$\bar{\pi}^C_i(f,\alpha)=\mathcal{S}(\bar{\pi}^C_i(f),\alpha f(\bar{\lambda}_i)) \tag{37}$$

using the soft-thresholding operator $\mathcal{S}(z,\lambda)=\mathrm{sgn}(z)\max(0,|z|-\lambda)$. This concludes the proof. □

Lemma 7 (LOO Ridge Performance) *$R_{T,t}(f_z)$ admits the following representation:*

$$R_{T,t}(f_z) = \underbrace{\frac{1}{1-\psi_t(z)}}_{complexity\ \ multiplier} \left(R_t(f_z) \ - \ \underbrace{\psi_t(z)}_{overfit}\right) \tag{38}$$

where $R_t(f_z)$ is the in-sample portfolio return at time $t \in \{1,\ldots,T\}$:

$$R_t(f_z) \ = \ \bar{\pi}(f_z)' F_t, \ . \tag{39}$$

Furthermore,

$$\psi_t(z) \ = \ \frac{1}{T} F_t'(zI + \bar{\Sigma})^{-1} F_t \tag{40}$$

Proof of Lemma 7. By the Sherman–Morrison formula,

$$\psi_t(z) \ = \ \frac{1}{T} F_t'(zI + \bar{\Sigma})^{-1} F_t = \frac{T^{-1} F_t'(zI + \bar{\Sigma}_{T,\tau})^{-1} F_t}{1 + T^{-1} F_t'(zI + \bar{\Sigma}_{T,\tau})^{-1} F_t} < 1 \ . \tag{41}$$

For brevity, we use in the sequel the notation $f_z(\Sigma) = (zI_N + \Sigma)^{-1}$ for any symmetric positive semi-definite $N \times N$ matrix Σ. With this notation, the definition of $R_t(f_z)$ and $R_{T,t}(f_z)$ yields:

$$R_t(f_z) = \bar{\pi}'(f_z) F_t = \bar{\mu}' f_z(\bar{\Sigma}) F_t \ , \tag{42}$$

and

$$R_{T,t}(f_z) = \bar{\pi}'_{T,t}(f_z) F_t = \bar{\mu}'_{T,t} f_z(\bar{\Sigma}_{T,t}) F_t = \frac{1}{T} \sum_{\tau \neq t} F_\tau' f_z(\bar{\Sigma}_{T,t}) F_t \ . \tag{43}$$

Since,

$$\bar{\Sigma}_{T,t} + zI_N = \bar{\Sigma} + zI_N - \frac{1}{T} F_t F_t' \ ,$$

we next compute the right-hand side of identity (43) using the Sherman–Morrison formula,

for $A = \bar{\Sigma} + zI$, $u = F_t$, and $v = -\frac{1}{T}F_t$. This gives:

$$f_z(\bar{\Sigma}_{T,t}) = f_z(\bar{\Sigma}) + \frac{1}{T}\frac{f_z(\bar{\Sigma})F_tF_t'f_z(\bar{\Sigma})}{1-\frac{1}{T}F_t'f_z(\bar{\Sigma})F_t} \ .$$

Multiplying both sides of this last equation by F_t', it further follows:

$$\begin{aligned} F_t'f_z(\bar{\Sigma}_{T,t}) = \ & F_t'f_z(\bar{\Sigma}) + \frac{1}{T}\frac{F_t'f_z(\bar{\Sigma})F_tF_t'f_z(\bar{\Sigma})}{1-\frac{1}{T}F_t'f_z(\bar{\Sigma})F_t} \\ = \ & \frac{F_t'f_z(\bar{\Sigma})}{1-\frac{1}{T}F_t'f_z(\bar{\Sigma})F_t} \\ = \ & \frac{1}{1-\psi_t(z)}F_t'f_z(\bar{\Sigma}) \ . \end{aligned}$$

From equation (43), we finally obtain:

$$\begin{aligned} R_{T,t}(f_z) \ &= \ \frac{1}{T}\sum_{\tau\neq t}^{T} F_t'f_z(E_{T,t}[FF'])F_\tau \\ &= \ \frac{1}{T}\sum_{\tau\neq t}^{T} \frac{1}{1-\psi_t(z)}F_t'f_z(\bar{\Sigma})F_\tau \\ &= \ \frac{1}{1-\psi_t(z)}F_t'f_z(\bar{\Sigma})\frac{1}{T}\sum_{\tau\neq t}^{T} F_\tau \\ &= \ \frac{1}{1-\psi_t(z)}\left(F_t'f_z(\bar{\Sigma})\bar{\mu} - \frac{1}{T}F_t'f_z(\bar{\Sigma})F_t\right) \\ &= \ \frac{1}{1-\psi_t(z)}(R_t(f_z) - \psi_t(z)) \ . \end{aligned}$$

This concludes the proof. □

Lemma 8 *Any continuous function f can be uniformly approximated over compact intervals by a function $f_{\mathcal{Z}} \in \mathcal{F}(\mathcal{Z})$, whenever the grid Z is sufficiently wide and dense.*

Proof of Lemma 8. The proof relies on an application of the Stone-Weierstrass Theorem; see, e.g., Rudin (1976). Consider the algebra of functions generated by the ridge ensemble $\{\Theta_z : z > 0\}$, where $\Theta_z(x) := (z+x)^{-1}$ for any $x \geq 0$. These functions are bounded, strictly monotonically decreasing, and continuous on $\mathbb{R}_+$. Using the identity

$$\Theta_{z_1}(x) - \Theta_{z_2}(x) \ = \ (z_2 - z_1)\Theta_{z_1}(x)\Theta_{z_2}(x), \tag{44}$$

it follows that on any compact interval $[a, b]$, the linear span of the ridge ensemble is dense in the algebra generated by the ridge ensemble. Moreover, it is easy to see that the ridge ensemble separates points on any compact interval $[a, b]$, and it vanishes nowhere. As a consequence, for any compact interval $[a, b]$, the algebra generated by the ridge ensemble is dense in $C(a, b)$ – by the Stone-Weierstrass Theorem – and the claim follows. □

B Additional Empirics

B.1 Robustness Results Plots

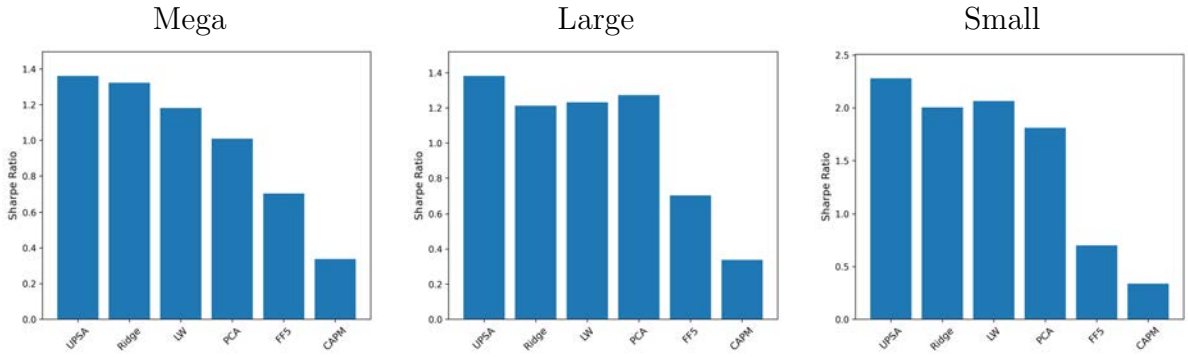

Figure 14: Sharpe Ratio of UPSA and Benchmarks Across Size Groups.

Annual out-of-sample Sharpe ratios for UPSA and benchmarks, across size groups. Out-of-sample results are from November 1981 to December 2022.

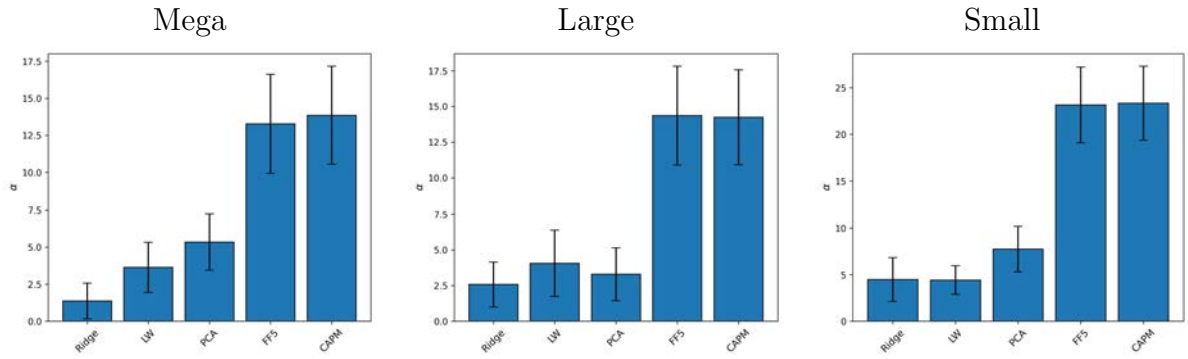

Figure 15: α of UPSA Vs. Benchmarks Across Size Groups.

Annual percentage α from regression of UPSA on benchmarks using heteroscedasticity-adjusted standard errors (5 lags). All portfolios have been scaled to an annual volatility of 10%. Out-of-sample results are from November 1981 to December 2022.

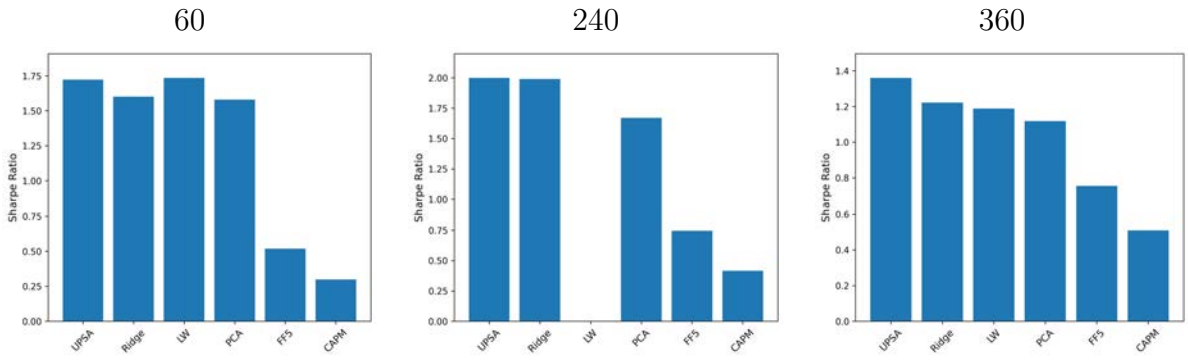

Figure 16: Sharpe Ratio of UPSA and Benchmarks Across Rolling Windows.

Annual out-of-sample Sharpe ratios for UPSA and benchmarks. Out-of-sample estimates are based on three rolling window configurations: 60 months (November 1976–December 2022), 240 months (November 1991–December 2022), and 360 months (November 2001–December 2022).

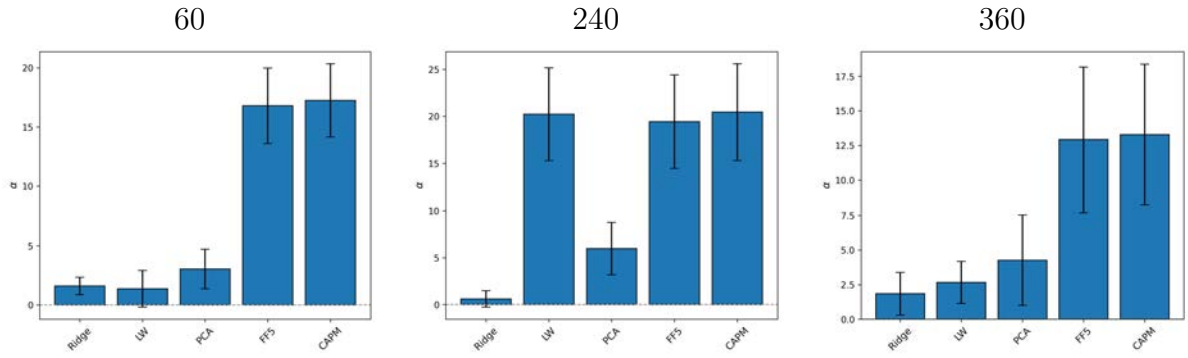

Figure 17: α of UPSA Vs. Benchmarks Across Rolling Windows.

Annual percentage α estimates from regressions of UPSA on benchmark portfolios using heteroscedasticity-adjusted standard errors (5 lags). All portfolios are scaled to an annual volatility of 10%. Out-of-sample estimates are based on three rolling window configurations: 60 months (November 1976–December 2022), 240 months (November 1991–December 2022), and 360 months (November 2001–December 2022).