

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 May 2026

AQR Capital Management is a global investment management firm that may or may not apply similar investment techniques or methods of analysis as described herein. The views expressed here are those of the authors and not necessarily those of AQR. Semyon Malamud gratefully acknowledges the financial support of the Swiss National Science Foundation (Grant 100018-228042) and the Swiss Finance Institute. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed additional relationships of potential relevance for this research. Further information is available online at <http://www.nber.org/papers/w32004>

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2023 by Bryan T. Kelly, Semyon Malamud, Mohammad Pourmohammadi, and Fabio Trojani. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Universal Portfolio Shrinkage

Bryan T. Kelly, Semyon Malamud, Mohammad Pourmohammadi, and Fabio Trojani

NBER Working Paper No. 32004

December 2023, Revised May 2026

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 uniformly penalizing all principal components of returns or discarding low-variance ones, UPSA instead reweights components to explicitly optimize expected out-of-sample portfolio performance. In empirical applications using a large cross-section of anomaly factors, it delivers robust improvements over alternative shrinkage methods in the literature.

Bryan T. Kelly
Yale University
and NBER
bryan.kelly@yale.edu

Mohammad Pourmohammadi
Yale University
and Swiss Finance Institute
mo.pourmohammadi@yale.edu

Semyon Malamud
Swiss Finance Institute
semyon.malamud@epfl.ch

Fabio Trojani
University of Geneva
fabio.trojani@alphacruncher.com

Universal Portfolio Shrinkage

Bryan Kelly, Semyon Malamud, Mo Pourmohammadi, and Fabio Trojani*

May 21, 2026

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 uniformly penalizing all principal components of returns or discarding low-variance ones, UPSA instead reweights components to explicitly optimize expected out-of-sample portfolio performance. In empirical applications using a large cross-section of anomaly factors, it delivers robust 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

*Bryan Kelly is at AQR Capital Management, Yale School of Management, and NBER. Semyon Malamud is at the Swiss Finance Institute, EPFL, and CEPR, and is a consultant to AQR. Mo Pourmohammadi is at Yale School of Management. Fabio Trojani is at the University of Geneva, a Senior Chair of the Swiss Finance Institute and the AXA Chair in Socioeconomic Risks of Financial Markets at the University of Turin. AQR Capital Management is a global investment management firm that may or may not apply similar investment techniques or methods of analysis as described herein. The views expressed here are those of the authors and not necessarily those of AQR. Semyon Malamud gratefully acknowledges the financial support of the Swiss National Science Foundation (Grant 100018-228042) and the Swiss Finance Institute.

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

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. The method provides a semi-parametric approximation to a large class of nonlinear shrinkage functions, freeing it from the rigid structure of many existing approaches. Rather than targeting tangential statistical objectives, it directly optimizes expected out-of-sample portfolio performance. Additionally, it is available in closed form, computationally efficient, and scalable to truly high-dimensional settings in which the number of assets may greatly exceed the number of observations.

1.1 Universal Shrinkage

The nature of optimal shrinkage developed in this paper begins from the in-sample Markowitz portfolio. It is convenient to represent this 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 portfolio assigns weights $\bar{\pi}_i^{\text{Mark}}$ to the i th PC based on its individual risk-return tradeoff. That is, the weight on each PC is proportional to its ratio of

¹For 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.

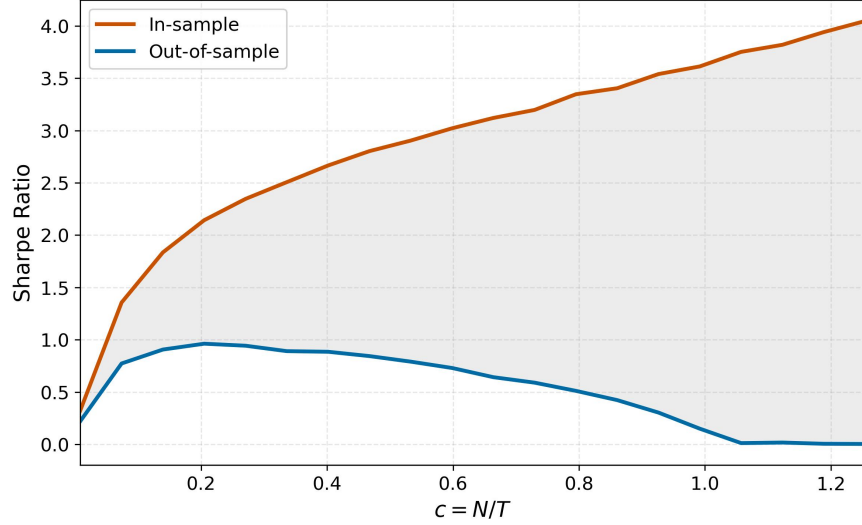


Figure 1: In-sample vs. Out-of-sample Sharpe Ratios without Shrinkage.

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.

average return ($\bar{R}_i^{PC}$) to sample variance ($\bar{\lambda}_i$):

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

where T is the number of observations. When N is large relative to T , estimates of this mean-variance tradeoff are corrupted by noise, giving rise to small-sample biases such as the tendency to overweight low-variance PCs. These distortions may remain hidden in-sample, but they can severely undermine out-of-sample portfolio performance.

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 number of training observations far exceeds the number of 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 ($c \gg 0$), in-sample performance becomes severely upward biased while 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:

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

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 prior examples to capture an optimal nonlinear portfolio shrinkage function, $\bar{\pi}_i^{\text{UPSA}} = f(\bar{\lambda}_i) \bar{R}_i^{PC}$. Because our formulation is interpretable as a shallow neural network approximation to f , it enjoys universal approximation properties ([Hornik et al., 1990](#)). Despite this flexibility, it bypasses typical computational hurdles: the estimator retains a closed-form solution as a linear combination of ridge portfolios (as in equation (1)), each utilizing a different penalty parameter.

Furthermore, UPSA easily accommodates economic shape constraints on the shrinkage function, such as strict positivity and monotonicity, without losing tractability. Importantly, while the estimator fundamentally relies on spectral shrinkage of the assets’ second moment matrix, it simultaneously integrates average return information whenever the economic objective requires it. This allows the entire framework to extend naturally to any quadratic portfolio objective while preserving its analytical convenience.

1.2 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 our method 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, those gains are statistically significant across benchmark portfolios with an α t -statistic of 3.72 against 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).

Inspecting the implied SDF across principal components helps explain these results. For the largest eigenvalue components, UPSA applies similar levels of shrinkage to ridge. The difference emerges in the middle of the spectrum, where it regularizes intermediate components more aggressively. As lower-variance principal components are added, performance deteriorates only gradually, behaving as though the investor had truncated the noisiest components without imposing an explicit cutoff.

When decomposing portfolio weights using the thematic clusters of [Jensen et al. \(2023\)](#), we show that UPSA’s spectral differences induce systematic tilts toward persistent firm fundamentals (e.g., quality, value, and low risk) and away from transient patterns (e.g., seasonality, skewness, and momentum). Using a Shapley decomposition, we show that these fundamental thematic tilts account for UPSA’s superior Sharpe ratio.

UPSA admits a natural Bayesian interpretation. A ridge penalty can be thought of as a belief about expected return uncertainty; by combining a discrete set of shrinkage levels selected from the data, UPSA effectively diversifies across a range of such beliefs rather than committing to a single one. Diversification across beliefs produces a more stable SDF,

leading to lower portfolio turnover, and delivers more robust performance across business cycle regimes, with particularly strong gains during NBER recessions. These patterns are corroborated by simulation evidence and remain robust across variations in asset liquidity and training sample size.

1.3 Literature Review

Our work is related to several strands of literature. The first 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 minimum-variance portfolios ([Ledoit and Wolf, 2003](#)) and tracking portfolios ([Ledoit and Wolf, 2004a](#)). Subsequent papers develop nonlinear spectral shrinkage estimators that exploit random-matrix-theory techniques introduced in [Ledoit and P  ch   \(2011\)](#) (e.g., [Ledoit and Wolf, 2012, 2015, 2020](#)). We differ from these approaches by proposing a universal approximator for a broad class of shrinkage functions. Our method combines the flexibility of a neural-network-type nonlinear mapping with analytical tractability and low computational cost.

A second strand examines shrinkage in portfolio optimization and regularized estimation of SDFs (e.g., [Kozak et al., 2018](#); [Kelly et al., 2019](#); [Lettau and Pelger, 2020](#); [Gu et al., 2021](#); [Kozak et al., 2020](#); [Pedersen et al., 2021](#); [Bryzgalova et al., 2023b](#); [Giglio and Xiu, 2021](#); [Korsaye et al., 2025](#); [Bryzgalova et al., 2025](#)). UPSA generalizes these approaches by directly optimizing a nonlinear spectral shrinkage function for the portfolio’s mean–variance objective; its spectral reweighting implicitly shrinks the mean by tilting toward principal components with higher mean and lower variance.

A growing literature investigates the asset-pricing role of weak factors—factors whose risk premia are difficult to estimate precisely because their sampling variation is small relative to estimation noise (see, e.g., [Bryzgalova et al., 2023a](#); [Preite et al., 2022](#); [Giglio et al., 2025](#)). To address this problem, [Lettau and Pelger \(2020\)](#) propose a PC-based dimension-reduction approach that incorporates information about PC sample means. In contrast, UPSA achieves shrinkage without dimension reduction, reweighting all PCs through eigenvalue shrinkage directly in light of the portfolio-performance objective, so that the resulting shrinkage implicitly reflects information in PC means.

Finally, our paper contributes to a literature on the statistical limits of asset-pricing estimation in high-dimensional settings (Da et al., 2022; Didisheim et al., 2024; Martin and Nagel, 2022; Kelly et al., 2024). UPSA provides an economically motivated shrinkage methodology for such environments, and its ensemble structure connects naturally to Bayesian portfolio choice under prior uncertainty.

1.4 Roadmap

The remainder of the paper is organized as follows. Section 2 introduces the second-moment shrinkage framework and formulates the optimal nonlinear spectral shrinkage problem. Section 3 develops UPSA as an ensemble of ridge portfolios, establishes its universal approximation properties, and derives the optimal ensemble weights using leave-one-out cross-validation. Section 4 presents the empirical analysis, comparing UPSA to ridge, Ledoit–Wolf, and PCA benchmarks using the anomaly factors of Jensen et al. (2023). Section 5 develops the economic interpretation of UPSA: a Bayesian foundation links the ensemble structure to integration over prior beliefs about expected-return uncertainty, and the empirical consequences—including lower turnover, greater recession robustness, and a detailed anatomy of how UPSA allocates risk across the principal component spectrum—are documented. A thematic decomposition translates these spectral differences into portfolio tilts across recognizable asset pricing themes and quantifies their contribution to UPSA’s Sharpe ratio advantage. Section 6 presents simulation evidence that isolates the roles of economic tuning, nonlinear shrinkage, and adaptation to time-varying conditions. Section 7 concludes.

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

solution²

$$\pi_\star = \Sigma^+ \mu, \quad (2)$$

where Σ^+ is the Moore–Penrose generalized inverse of matrix Σ . In contrast to the full information setting, an investor whose information includes only T observations of F_t can instead compute finite-sample moments

$$\bar{\mu} = \frac{1}{T} \sum_{t=1}^T F_t, \quad \bar{\Sigma} = \frac{1}{T} \sum_{t=1}^T F_t F_t'$$

and construct an empirical counterpart of portfolio (2) given by

$$\bar{\pi} = \bar{\Sigma}^+ \bar{\mu} \quad (3)$$

with return $R_t^{\bar{\pi}} = \bar{\pi}' F_t$.

In most practical scenarios, investors find themselves with a large number of assets (e.g., thousands of stocks or hundreds of anomaly factors) and a limited time series of 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: $\bar{\mu} \not\rightarrow \mu$, $\bar{\Sigma} \not\rightarrow \Sigma$ (Didisheim et al., 2024). In such a setting, the estimated Markowitz portfolio (3) is a random quantity, even asymptotically, and can perform badly out-of-sample (see, for example, DeMiguel et al., 2009).

Let $\bar{\Sigma} = \bar{U} \text{diag}(\bar{\lambda}) \bar{U}'$ 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, $\bar{\lambda}$,

²We 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.

without altering its empirical eigenvectors $\bar{U}$:³

$$\bar{\pi}(f) = f(\bar{\Sigma})\bar{\mu}, \text{ s.t. } f(\bar{\Sigma}) = \bar{U} \text{diag}(f(\bar{\lambda}))\bar{U}' . \quad (4)$$

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 shrunk Markowitz portfolio as the return of principal component factors:

$$R_t^{\bar{\pi}(f)} = \sum_{i=1}^K f(\bar{\lambda}_i) \bar{R}_i^{PC} R_{i,t}^{PC} \quad (5)$$

where $K = \min(N, T)$. The weight on an individual PC captures a tradeoff between that PC's return ($\bar{R}_i^{PC}$) and its variance (λ_i).⁴ For example, a common approach to this problem relies on the ridge portfolio estimator,

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

which is defined for some ridge parameter $z > 0$ (see, e.g., [Kozak et al. \(2020\)](#), [Ledoit and Wolf \(2003\)](#), among many others).⁵ 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*

³This class of estimators originates from the seminal work of [Stein \(1986\)](#).

⁴Without 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.

⁵The 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.

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

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

The key new feature of our optimal spectral shrinkage operator f_* 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:⁶

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

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 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.

⁶This approach closely parallels the methodology of [Stein \(1986\)](#). See Lemma 3 and the corresponding proof in the Appendix.

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

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

$\nu(\mathbb{R}_+) = K$. 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:

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

where $\nu_1, \dots, \nu_L > 0$ and $\sum_{i=1}^L \nu_i = K$.

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 $\mathcal{Z} = \{z_1, \dots, z_L\}$ 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:*

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

We call $\mathcal{F}(\mathcal{Z}) = \{f_{\mathcal{Z}, W} : W \in \mathbb{R}_+^L\}$ the **ridge ensemble**.⁷

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.⁸

⁷Without constraints, these shrinkages can cover any continuous function with compact support. We go into detail on this space of admissible shrinkages in Lemma 5 of the Appendix.

⁸The 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 $K = \lim_{\bar{\lambda} \rightarrow \infty} \bar{\lambda} f(\bar{\lambda})$ 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 $z > 0$.

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., 2024) that uses minimal shrinkage:⁹

$$\lim_{z \rightarrow 0} (zI_N + \bar{\Sigma})^{-1} \bar{\Sigma} \bar{\Sigma}^+ \bar{\mu} = \bar{\Sigma}^+ \bar{\mu} \quad (10)$$

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

$$\lim_{z \rightarrow \infty} z(zI_N + \bar{\Sigma})^{-1} \bar{\mu} = \bar{\mu} \quad (11)$$

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_Z in Lemma 1 coincides with the ridge ensemble in equation (9) when ensemble weights are appropriately chosen. To operationalize the ridge ensemble, 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 (5), obtaining

$$R_t^{\bar{\pi}(f_Z, W)} = \sum_{i=1}^L w_i R_t^{\bar{\pi}(f_{z_i})} = \underbrace{W'}_{\text{Ridge Weights}} \underbrace{R_t^{\bar{\pi}(f_Z)}}_{\text{Ridge Returns}},$$

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

$$E[R_t^{\bar{\pi}(f_Z, W)}] - \frac{1}{2} E[(R_t^{\bar{\pi}(f_Z, W)})^2] = W' \underbrace{E[R_t^{\bar{\pi}(f_Z)}]}_{\text{Ridge Means}} - \frac{1}{2} W' \underbrace{E[R_t^{\bar{\pi}(f_Z)} R_t^{\bar{\pi}(f_Z)'}]}_{\text{Ridge Second Moment}} W, \quad (12)$$

⁹Implicitly, we are using here the “empirical” no arbitrage assumption that $\bar{\mu}$ belongs to the range of $\bar{\Sigma}$.

Equation (12) 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.¹⁰ The LOO method drops one training observation t , trains on the remaining observations $1, \dots, t-1, t+1, \dots, T$, and repeats this for each $t = 1, \dots, T$. 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.¹¹ From this procedure, we also obtain LOO estimates of the first and second moments of factor returns, denoted for each $t = 1, \dots, T$, as

$$\bar{\mu}_{T,t} = \frac{1}{T-1} \sum_{\tau \neq t, 1 \leq \tau \leq T} F_{\tau} \quad \text{and} \quad \bar{\Sigma}_{T,t} = \frac{1}{T-1} \sum_{\tau \neq t, 1 \leq \tau \leq T} F_{\tau} F_{\tau}'. \quad (13)$$

From the LOO asset moment estimates, we obtain spectral shrinkage portfolio estimates for each t that are LOO versions of the in-sample estimator (4),

$$\bar{\pi}_{T,t}(f_z) = f_z(\bar{\Sigma}_{T,t})\bar{\mu}_{T,t}. \quad (14)$$

Finally, from (14), we recover the realized LOO portfolio returns $\{R_{T,t}^{\bar{\pi}_{T,t}(f_z)} = \bar{\pi}_{T,t}(f_z)' F_t : t = 1, \dots, T\}$. 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*

¹⁰LOO ridge returns can be computed efficiently from a single matrix inversion via the Sherman-Morrison formula, avoiding the need to re-estimate the model for each left-out observation (see, e.g. Golub et al., 1979).

¹¹See Lemma 4 in the Appendix for a formal proof.

of out-of-sample ridge portfolio returns $R^{\bar{\pi}(f_{z_1})}, \dots, R^{\bar{\pi}(f_{z_L})}$:

$$\begin{aligned}\bar{\mu}(\mathcal{Z}) &= \left(\frac{1}{T} \sum_{t=1}^T R_{T,t}^{\bar{\pi}(f_{z_i})} \right)_{i=1}^L \in \mathbb{R}^L \\ \bar{\Sigma}(\mathcal{Z}) &= \left(\frac{1}{T} \sum_{t=1}^T R_{T,t}^{\bar{\pi}(f_{z_i})} R_{T,t}^{\bar{\pi}(f_{z_j})} \right)_{i,j=1}^L \in \mathbb{R}^{L \times L}.\end{aligned}\tag{15}$$

Then, from Definition 2, UPSA shrinkage is given by

$$f_{UPSA} = f_{\mathcal{Z}, W_{UPSA}^*},\tag{16}$$

where W_{UPSA}^* solves

$$W_{UPSA}^* = \arg \max_{W \in \mathbb{R}_+^L} \{W' \bar{\mu}(\mathcal{Z}) - 0.5 W' \bar{\Sigma}(\mathcal{Z}) W\}.\tag{17}$$

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 assets. This formulation of the estimator renders UPSA tractable and cheap to compute even in very high-dimensional contexts.¹²

¹²Our UPSA code is available [here](#).

4 Empirical Evidence

4.1 Data and Benchmarks

Our empirical analysis uses monthly returns on 153 characteristic-managed portfolios (“factors”) from [Jensen et al. \(2023\)](#).¹³ The factors are constructed as capped value-weighted long–short portfolios of U.S. stocks and cover the period from November 1971 to December 2022.¹⁴

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 penalties¹⁵ available to UPSA as

$$\mathcal{Z} \equiv \{10^i : i \in \{-10, -9, \dots, -1\}\}.$$

We compare the performance of UPSA with two spectral shrinkages, one sparse shrinkage, and two factor pricing models:¹⁶

- **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, ordered by variance, selected to maximize expected quadratic utility based on LOO cross-validation.
- **FF5**: The Markowitz portfolio constructed from the five factors of [Fama and French \(2015\)](#).

¹³The complete list of these 153 factors and their monthly returns is available at <https://jkpfactors.com>.

¹⁴[Ke and Pourmohammadi \(2025\)](#) apply UPSA to the datasets of [Chen and Dim \(2023\)](#) and [Chen and Zimmermann \(2022\)](#) and show that it remains superior even when compared to a substantially more comprehensive set of benchmarks.

¹⁵Our results are robust to the choice of the grid $\mathcal{Z}$, both in terms of Sharpe ratio and turnover. See Appendix C for details.

¹⁶Appendix E.2 reports Sharpe ratio comparisons with two additional benchmarks: the elastic-net estimator on the space of principal components similar to [Kozak et al. \(2020\)](#) and a PCA method where principal components are sorted by absolute means.

- **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.

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 1.13 for the Fama-French factors and 0.54 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 examine 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

¹⁷To ensure comparability across shrinkage methods, we normalize portfolio weights so that the implied in-sample variance after shrinkage equals the in-sample variance of the unshrunk empirical eigenvalues. Because different shrinkage functions rescale eigenvalues by varying amounts, performance differences without normalization may inadvertently reflect scale rather than the quality of the shrinkage direction. This adjustment directly addresses the scale-equivariant nature of covariance estimation (see [Ledoit and Wolf, 2020](#)).

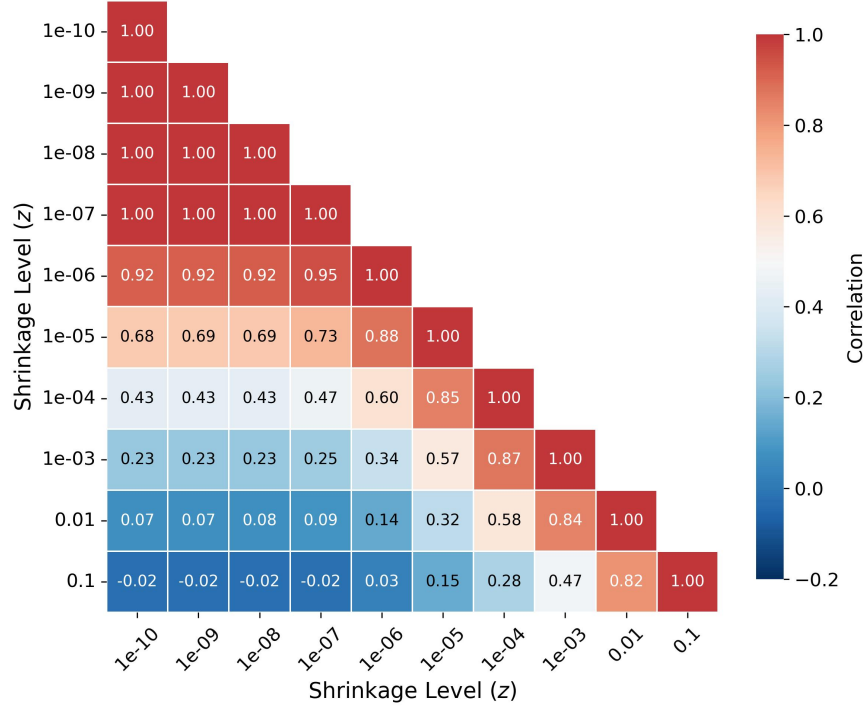


Figure 2: Out-of-sample Return Correlations Across Shrinkage Levels.

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

and statistically significant. Furthermore, the comparative benefits of UPSA are uniform 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:

$$E[F_{i,t+1} M_{t+1}^*] = 0, \quad i = 1, \dots, N, \quad \text{where} \quad M_{t+1}^* = 1 - \pi_{\star}' F_{t+1}.$$

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

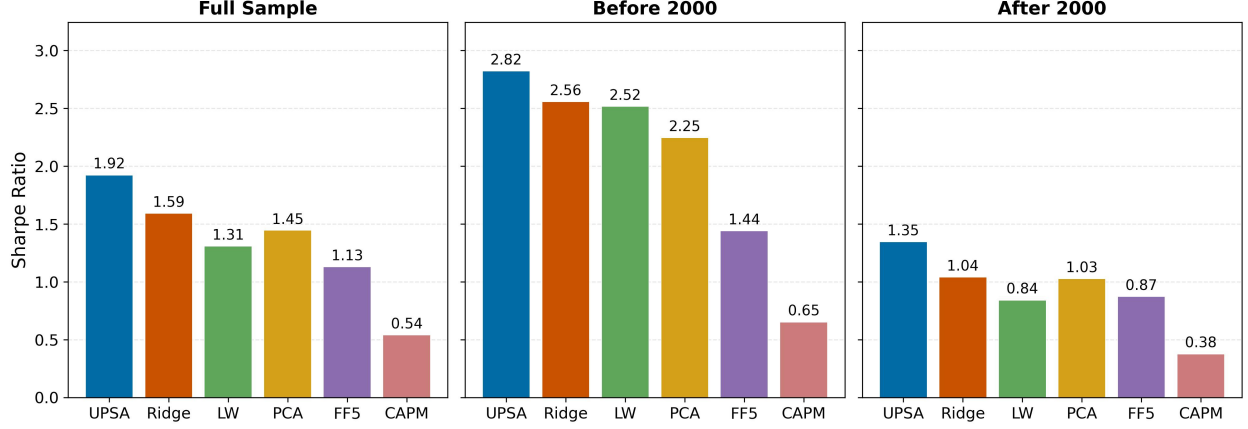


Figure 3: Annualized Out-of-sample Sharpe Ratios for UPSA and Benchmarks.

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

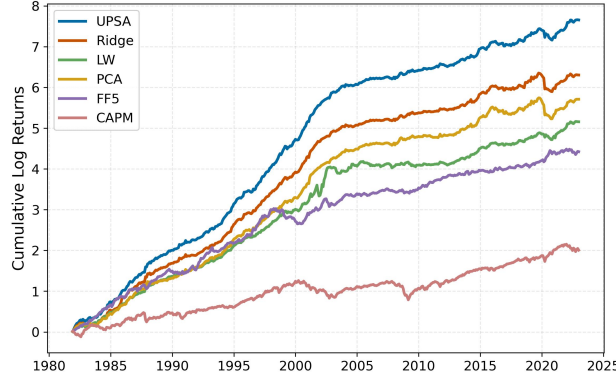


Figure 4: Cumulative Log Returns for SDFs with Different Shrinkage Methods.

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 10% annualized volatility.

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

$$E_{OOS}[F_{i,t+1}] \quad \text{vs.} \quad E_{OOS}[(1 - M_{t+1})F_{i,t+1}] \quad (18)$$

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

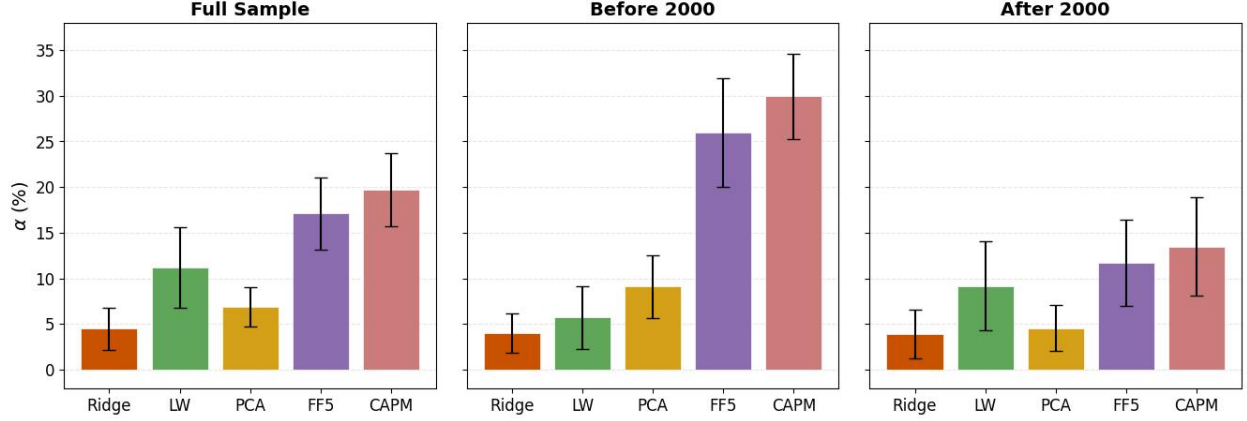


Figure 5: α of UPSA vs. Benchmarks.

Annualized percentage α from regressions of UPSA on each benchmark using HAC standard errors (5 lags). All portfolios are scaled to 10% annualized volatility. The out-of-sample period is November 1981 to December 2022.

(18) 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

$$M_{t+1}^j = 1 - s_j \pi_t^{j'} F_{t+1}$$

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 normalizes each candidate SDF so that it prices its own basis portfolio out-of-sample. Without this adjustment, pricing errors on the test assets would be confounded by scale differences rather than differences in pricing ability.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

¹⁸In particular, $s_j = E_{\text{OOS}}[\pi_t^{j'} F_{t+1}] / E_{\text{OOS}}[(\pi_t^{j'} F_{t+1})^2]$ 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).

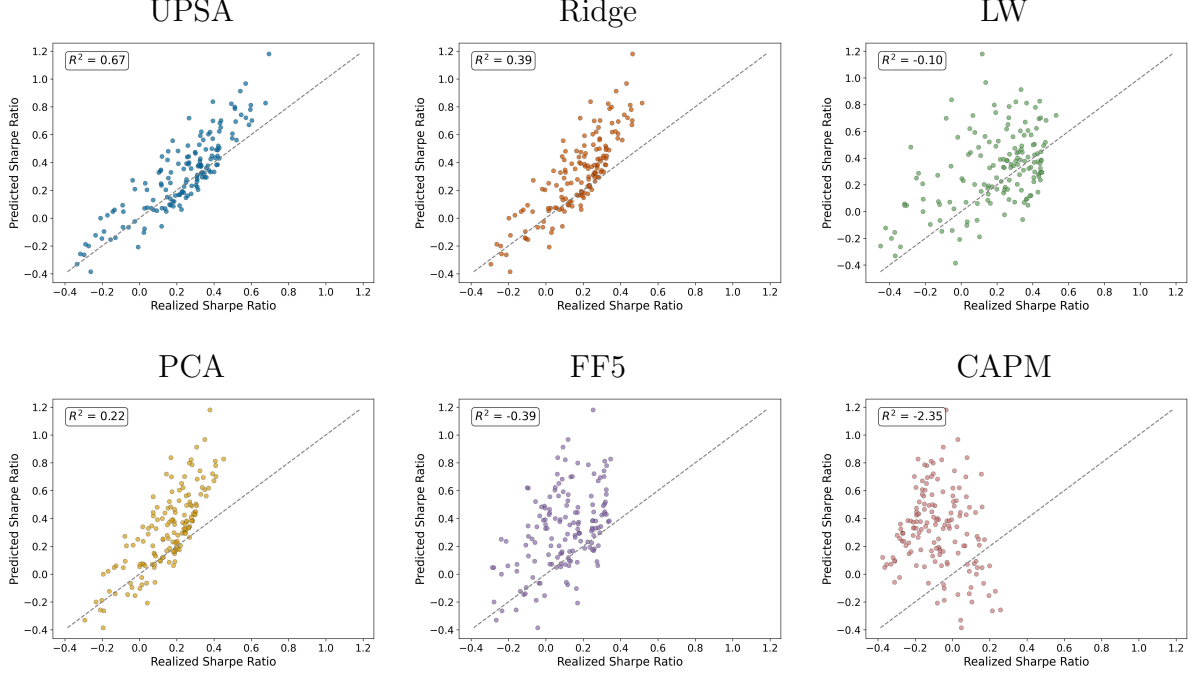


Figure 6: Out-of-sample Pricing Accuracy of Optimal Portfolios.

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, following equation (18). Test assets are standardized to unit volatility so that pricing errors are expressed in Sharpe ratio units. 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.

footing, we normalize each by its sample standard deviation, thus restating pricing errors in terms of Sharpe ratios rather than average returns.¹⁹

Figure 6 reports the results of this analysis. UPSA achieves a high degree of accuracy in pricing test assets, explaining factor realized Sharpe ratios with an out-of-sample R^2 of 67%.²⁰ The next best performers are ridge and PCA, with R^2 values of 39% and 22%, 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 of SDF-based Sharpe ratio

¹⁹An alternative approach evaluates pricing accuracy through a factor model beta representation. In Appendix E.5, we present results using an in-sample beta specification and show that UPSA’s pricing superiority is robust to this alternative evaluation methodology.

²⁰This 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	0.69	0.6	-0.09	0.56	-0.01	-0.13
Profitability	0.53	-0.32	-0.01	-0.64	-1.69	-8.5
Low Risk	0.77	0.85	-1.37	0.76	0.27	-12.14
Value	0.73	0.27	0.27	0.05	-0.3	-14.85
Investment	0.67	0.09	-0.18	-0.29	-0.82	-13.12
Seasonality	0.58	0.44	-0.14	0.36	-1.13	-0.81
Debt Issuance	0.49	0.02	-0.83	-0.26	-1.29	-2.1
Size	0.7	0.62	-1.3	0.53	0.01	-0.1
Accruals	0.5	-0.1	-2.56	-0.64	-0.86	-2.45
Low Leverage	0.3	0.37	-2.16	0.28	-1.26	-1.87
Profit Growth	0.6	0.3	0.72	0.01	-0.54	-0.69
Momentum	0.17	-0.67	0.13	-1.4	-10.04	-19.18
Quality	-0.4	-1.45	-2.54	-1.89	-3.27	-5.0
All	0.67	0.39	-0.1	0.22	-0.39	-2.35

predictions versus realized 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 10 of the 13 themes. The exceptions are the “low risk” theme in which ridge achieves an R^2 of 85%, though UPSA is a close second with a 77% R^2 ; the “low leverage” theme in which ridge achieves 37% versus UPSA’s 30%; and the “profit growth” theme in which LW achieves 72% versus UPSA’s 60%. The most challenging theme for UPSA 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 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 24 and 25 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 different groups, but UPSA is flexible enough to adapt to the group-specific data-generating process.

4.5.2 Rolling Window

We evaluate the performance of UPSA across different rolling windows: $T \in \{60, 240, 360\}$. Figures 26 and 27 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.²¹

These results suggest that even if nonlinear 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.

²¹The 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.

5 Economics of UPSA

UPSA dominates competing models in out-of-sample performance, with ridge as the closest benchmark. This section provides an economic interpretation of UPSA’s outperformance, focusing on how it resolves uncertainty about expected returns, how it allocates risk across principal components, and how these differences manifest in economically interpretable portfolio tilts.

5.1 Bayesian Foundations of UPSA

Section 3 introduced UPSA as the solution to an out-of-sample portfolio utility maximization problem. This section shows that UPSA also admits a natural Bayesian interpretation: it can be viewed as the posterior optimal portfolio of an investor who faces uncertainty about expected returns and whose prior beliefs may vary across principal components of the return covariance matrix.

Consider an investor who observes IID Gaussian returns $F_t \sim N(\mu, \Sigma)$ and knows Σ but faces uncertainty about μ ²². A standard approach (e.g., [Kozak et al., 2020](#); [Pedersen et al., 2021](#)) models this uncertainty through a Gaussian prior

$$\mu \mid \Sigma \sim N(0, g(\Sigma)), \quad (19)$$

where g is a strictly positive matrix function. The function $g(\Sigma)$ determines how prior uncertainty about expected returns varies across the eigen-directions of the covariance matrix. Intuitively, it encodes the investor’s beliefs about which principal components are more likely to contain economically meaningful risk premia.

After observing data, the posterior Markowitz portfolio $b = \Sigma^{-1}\mu$ takes a spectral shrinkage form (Lemma 6 in [Appendix B](#)):

$$E[b \mid F_1, \dots, F_T, \Sigma] = \underbrace{\left(\frac{N}{T} \cdot \frac{\text{tr}(\Sigma)}{N} [h(\Sigma)]^{-1} + \Sigma \right)^{-1}}_{=: f(\Sigma)} \bar{E}[F], \quad (20)$$

²²Appendix B extends this result to joint uncertainty over (μ, Σ) using mixtures of Normal–Inverse Wishart priors, where the posterior again takes the form of a ridge ensemble.

where h is a strictly positive matrix function derived from g .²³ Equation (20) is a spectral shrinkage estimator of exactly the form used by UPSA. The complexity ratio N/T naturally enters the posterior: when data are abundant ($N/T \rightarrow 0$), the prior vanishes and the investor recovers the sample Markowitz portfolio; in the complex regime relevant for our empirical setting ($N/T \rightarrow c > 0$), the prior exerts non-negligible regularization, endogenously tying the amount of shrinkage to the difficulty of the estimation problem.

The form of shrinkage depends on the prior function g , or equivalently h , which determines how prior uncertainty varies across principal components. A restrictive parametric specification therefore implies a restrictive shrinkage rule. For example, the power function prior

$$g(\Sigma) = \frac{1}{z \operatorname{tr}(\Sigma)} \Sigma^\eta$$

with $\eta = 2$ yields pure ridge shrinkage:

$$f(\Sigma) = \left(\frac{\operatorname{tr}(\Sigma)}{T} z I_N + \Sigma \right)^{-1}.$$

The case $\eta = 2$ implies that the investor assigns the same prior uncertainty to expected returns along all principal components, producing a single uniform penalty on all eigenvalues.

UPSA generalizes this framework by not committing to a specific parametric form for g . Instead, it learns the shrinkage function f directly from the data by optimizing out-of-sample portfolio performance. The universal approximation property (Lemma 1) guarantees that whenever the optimal Bayesian shrinkage rule is matrix monotonically decreasing, UPSA's ridge ensemble can approximate it:

$$E[b \mid F_1, \dots, F_T, \Sigma] = \left[\int_0^\infty \left(\frac{\operatorname{tr}(\Sigma)}{T} z I_N + \Sigma \right)^{-1} dw(z) \right] \bar{E}[F]. \quad (21)$$

Each value of z corresponds to a different degree of prior uncertainty about expected returns. Small values of z correspond to weaker regularization and greater reliance on the sample mean, while larger values of z reflect stronger prior skepticism about the precision

²³Specifically, $g(\Sigma) = \frac{1}{\operatorname{tr}(\Sigma)} \Sigma h(\Sigma) \Sigma$. See Appendix B for the derivation.

of estimated expected returns. UPSA can thus be interpreted as Bayesian portfolio choice under a mixture of priors, leading to a nonparametric Empirical Bayes learning rule that averages over different degrees of regularization. This rule decides on the appropriate degree of regularization in each eigen-direction of the covariance matrix, rather than imposing a fixed prior belief as ridge does.

5.2 Prior Instability and SDF Dynamics

The Bayesian framework in Section 5.1 interprets ridge as an investor with a single prior belief about expected-return uncertainty, and UPSA as an investor that integrates over a continuum of such priors. Figure 7 illustrates this distinction directly. The heatmap reports the time series of UPSA ensemble weights w_j across the grid of ridge penalties, while the orange line overlays the single penalty selected by cross-validation. The selected ridge penalty is highly unstable, frequently shifting by several orders of magnitude across adjacent windows. UPSA instead maintains stable exposure across a range of shrinkage levels, avoiding these discrete reallocations.

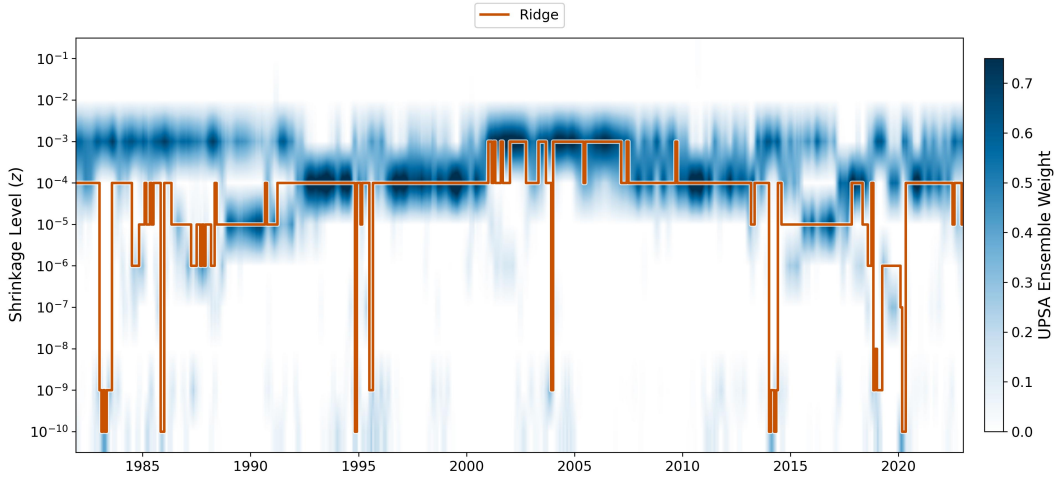


Figure 7: UPSA Ensemble Weights vs. Ridge Penalty Selection.

Time series of the ridge penalty selected by leave-one-out cross-validation (orange line) overlaid on UPSA ensemble weights across the shrinkage grid. UPSA weights are determined using Lemma 2. The out-of-sample period is November 1981 to December 2022.

This time-series instability translates directly into higher turnover.²⁴ Figure 8 reports the

²⁴We define turnover as $\text{Turnover}_t = \frac{1}{2} \frac{\sum_{i=1}^N |\bar{\pi}_{i,t} - \bar{\pi}_{i,t-1}|}{\sum_{i=1}^N |\bar{\pi}_{i,t-1}|}$, normalizing by the previous period's gross exposure to ensure comparability across methods with different portfolio scales.

12-month rolling average. Although ridge exhibits lower turnover in roughly three-quarters of months, its average is about 30% higher and its standard deviation nearly five times larger.²⁵

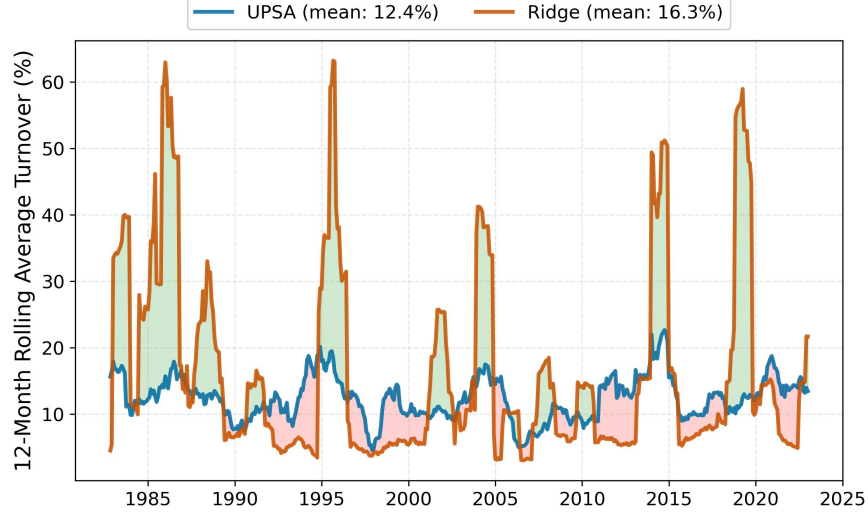


Figure 8: 12-Month Rolling Average Turnover.

12-month rolling average turnover for UPSA and ridge. Green shading indicates periods in which ridge turnover exceeds UPSA; red shading marks the reverse. Ridge penalties are selected using leave-one-out cross-validation, and UPSA weights are determined using Lemma 2. The out-of-sample period is November 1981 to December 2022.

The cost of this prior instability is not just higher turnover. When the covariance structure shifts, as it often does in recessions, the single prior selected in the previous estimation window can quickly become stale. The performance consequences are substantial: the Sharpe ratio gap between UPSA and ridge more than doubles in recessions ($\Delta = 0.57$) relative to expansions ($\Delta = 0.25$). Ridge’s Sharpe ratio drops by 53% from expansions to recessions, compared with only 30% for UPSA. The difference is driven primarily by the mean-return channel. Because UPSA integrates across shrinkage levels, it is less exposed to regime-dependent misalignment in the mean estimate, whereas ridge’s single point estimate becomes more fragile in recessions (Appendix E.6).

Importantly, this performance gap is driven entirely by how each method resolves uncertainty over time, rather than a fundamental disagreement about the unconditional level of shrinkage. As Figure 9 demonstrates, the two approaches agree on average: the time-

²⁵Turnover spikes are amplified when cross-validation selects very small penalties, leaving the portfolio only weakly regularized and highly sensitive to estimation noise.

averaged UPSA weights closely match the empirical distribution of ridge penalties selected by cross-validation. This confirms that the core advantage of UPSA lies in its Bayesian foundation: by integrating over penalty uncertainty rather than taking a fragile stand on a single prior, it delivers a more stable and economically robust portfolio.

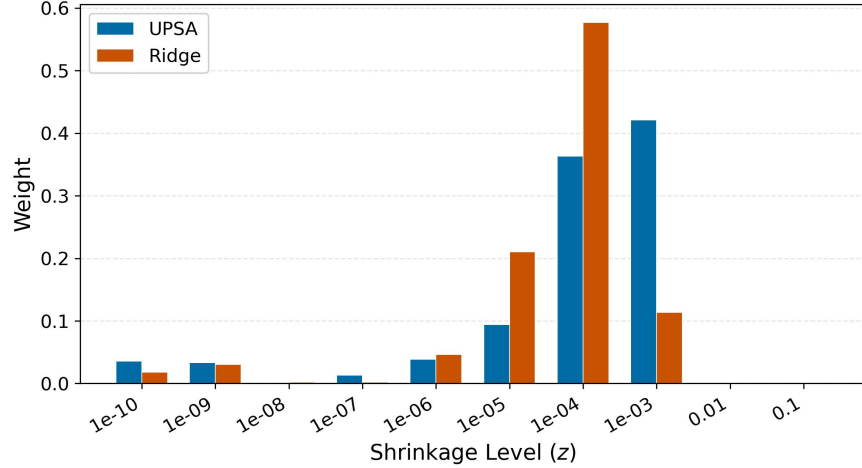


Figure 9: Distribution of Penalty Weights.

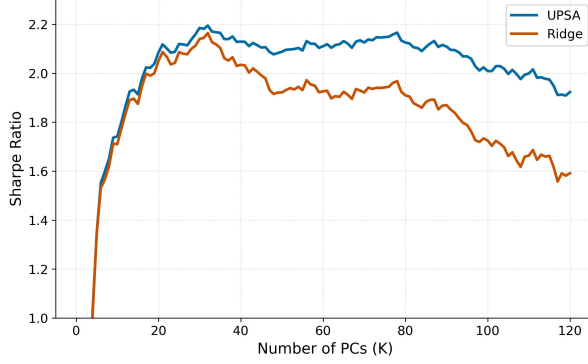
Histogram of ridge penalties selected by cross-validation compared with the average UPSA ensemble weights across the same grid. UPSA weights are determined using Lemma 2. The out-of-sample period is November 1981 to December 2022.

5.3 The Anatomy of Nonlinear Shrinkage

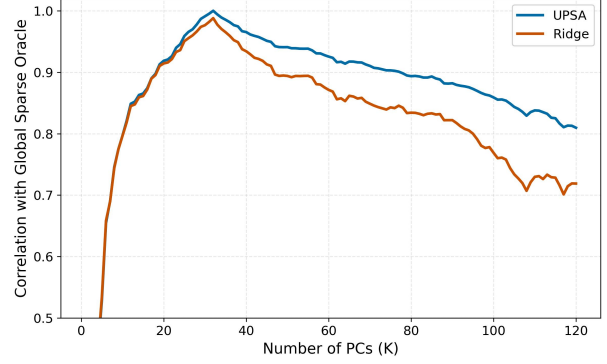
UPSA and ridge investors also differ in how they treat the eigenvalue spectrum. Ridge applies a uniform penalty across all eigenvalues, while UPSA varies shrinkage across the spectrum, which matters most for low-variance components where estimation noise is largest. This difference becomes most visible when we examine how performance evolves as additional principal components are incorporated. Figure 10 shows how SDF performance changes as PCs are incrementally included, ordered from largest to smallest variance. Both SDFs peak at $K = 32$, achieving Sharpe ratios of 2.19 (UPSA) and 2.16 (Ridge).²⁶

The key difference emerges beyond the peak. As lower-variance PCs are added, ridge deteriorates steadily, falling from 2.16 to 1.59 when all 120 PCs are included, a decline of

²⁶The oracle’s $K^* = 32$ is notably larger than the 3–5 dominant principal components that standard factor models typically rely on (Lettau and Pelger, 2020; Kelly et al., 2023; Kozak et al., 2020), suggesting that SDF construction benefits from incorporating components that individually explain only a small fraction of return variance.



(a) Out-of-sample Sharpe Ratio



(b) Correlation with (Infeasible) Sparse Oracle

Figure 10: Out-of-sample Performance by Number of Principal Components.

PCs are ordered by decreasing variance explained (eigenvalue rank). Panel (a): Annualized out-of-sample Sharpe ratio as PCs are incrementally added from the largest to the smallest eigenvalue. Panel (b): Correlation between the K -PC portfolio and the (infeasible) sparse oracle—the subset of K^* PCs that maximizes the out-of-sample Sharpe ratio. The out-of-sample period is November 1981 to December 2022.

roughly 26%. UPSA degrades far more gradually, from 2.19 to 1.92, a drop of only 12%. UPSA behaves as though the investor had truncated low-signal components *ex ante*, without imposing an explicit cutoff.

Panel (b) offers a complementary view, comparing each method to the (infeasible) sparse oracle: the portfolio constructed from the $K^* = 32$ PCs that maximize the out-of-sample Sharpe ratio.²⁷ Ridge’s correlation with the oracle drops to 0.72 at $K = 120$, while UPSA maintains a correlation of 0.81, indicating that it attenuates the influence of low-signal components even when they are retained. This pattern points to a fundamental difference in how the two methods allocate shrinkage across the eigenvalue spectrum.

Figure 11 makes this mechanism explicit. It illustrates how shrinkage varies across the eigenvalue spectrum via the function $f^{-1}(\bar{\lambda})$, which maps sample eigenvalues to their shrunk counterparts. For the largest eigenvalues, the two methods behave similarly. Differences emerge further along the spectrum, where UPSA applies stronger regularization to intermediate eigenvalues while continuing to shrink the smallest ones. This flexibility allows the estimator to dampen the influence of noisier components without discarding them entirely, helping explain the greater robustness observed in Figure 10.

²⁷PCs are ordered by decreasing variance (eigenvalue rank). Appendix E.1 repeats this analysis with PCs sorted by absolute sample mean return, yielding qualitatively similar results with even higher oracle correlations (0.91 for UPSA vs. 0.84 for ridge at $K = 120$).

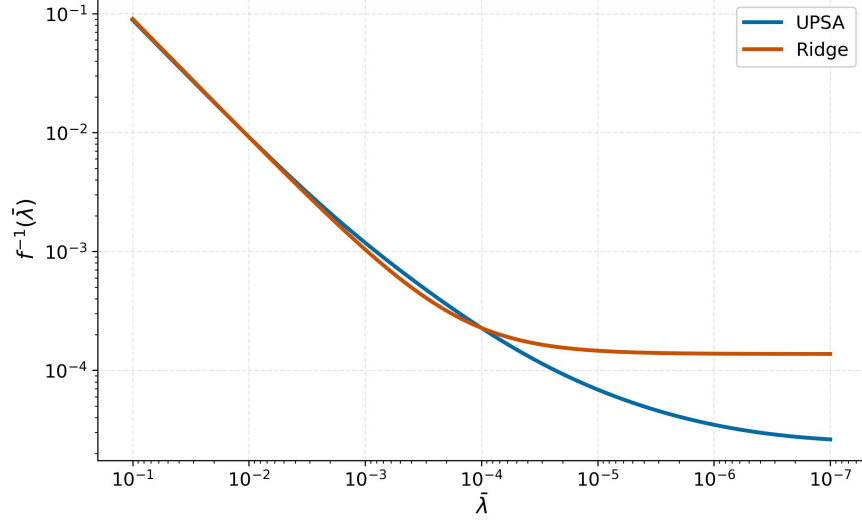


Figure 11: Shrinkage Functions for Ridge and UPSA.

Time-averaged inverse shrinkage function $f^{-1}(\lambda)$ mapping sample eigenvalues to their shrunk counterparts (Definition 2). Ridge: $f_z^{-1}(\lambda) = \lambda + z$, with z selected by leave-one-out cross-validation. UPSA: $f_{\mathcal{Z},W}^{-1}(\lambda) = [\sum_{i=1}^L w_i (\lambda + z_i)^{-1}]^{-1}$, with ensemble weights W determined by Lemma 2. The out-of-sample period is November 1981 to December 2022.

A natural question raised by Figure 10 is whether explicitly selecting a subset of PCs could further improve performance. Empirically, augmenting ridge and UPSA with a lasso penalty produces only modest changes in Sharpe ratios while generating highly unstable factor selection across rolling windows: the number of selected PCs fluctuates substantially over time, especially during periods of market stress.²⁸

5.4 Thematic Decomposition of SDF Loadings

The spectral differences between UPSA and ridge translate into differences in the economic composition of the SDF. We map SDF weights to the thematic clusters of Jensen et al. (2023) to characterize how each SDF allocates across asset pricing themes.

To isolate differences in composition rather than scale, we normalize both SDFs to 10% annualized volatility. For each period, the share of absolute SDF weight allocated to theme k is $s_t^k = \sum_{i \in k} |\bar{\pi}_{i,t}| / \sum_j |\bar{\pi}_{j,t}|$. Figure 12 reports the time-averaged allocation. Panel (a)

²⁸See Appendix D, Figures 19 and 20. The turnover cost is substantial: adding lasso increases mean monthly turnover by 57% for UPSA and 52% for ridge (Figure 21).

shows that both SDFs agree on the broad ranking: investment, value, low risk, and quality receive the largest shares.

Panel (b) shows how the allocations diverge. UPSA allocates relatively more to quality, low risk, and profitability, themes linked to persistent firm fundamentals. Ridge tilts toward seasonality, skewness, and momentum, themes driven by transient return dynamics. These differences are consistent with the aforementioned fact that UPSA favors more stable and lower turnover allocations.

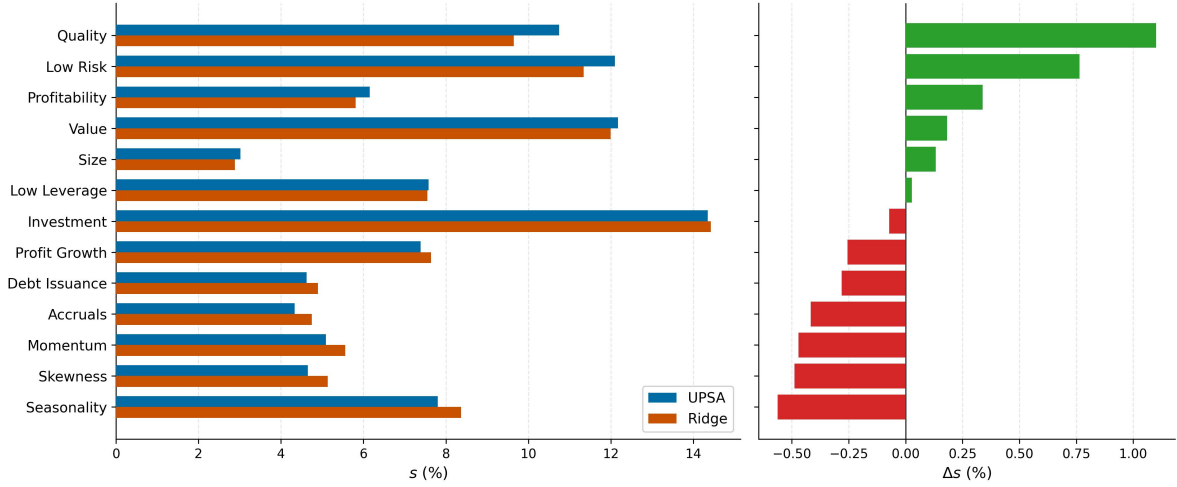


Figure 12: Thematic Portfolio Allocation: UPSA vs. Ridge.

Panel (a): Time-averaged share of absolute portfolio weight allocated to each JKP theme cluster (Jensen et al., 2023) for UPSA and ridge. Panel (b): Difference in allocation share, $s(\text{UPSA}) - s(\text{Ridge})$, in percentage points. Themes are sorted by the difference. Both portfolios are normalized to 10% annualized volatility. The out-of-sample period is November 1981 to December 2022.

To assess how these allocation differences contribute to the Sharpe ratio gap, we decompose it using Shapley values (Shapley, 1953). For each theme k , the Shapley value averages its marginal contribution to the Sharpe ratio difference $\Delta SR = SR(\text{UPSA}) - SR(\text{Ridge})$ over all possible orderings in which themes can be added:

$$\phi_k = \sum_{S \subseteq \mathcal{K} \setminus \{k\}} \frac{|S|! (K - |S| - 1)!}{K!} [\Delta SR(S \cup \{k\}) - \Delta SR(S)], \quad (22)$$

where $\mathcal{K}$ is the set of all themes, $K = |\mathcal{K}|$, and $\Delta SR(S)$ denotes the Sharpe ratio gap computed using only the return contributions of factors belonging to themes in S . By construction, the Shapley values sum exactly to the total gap: $\sum_k \phi_k = \Delta SR(\mathcal{K})$.

Figure 13 reports the results. The themes where UPSA allocates relatively more—value, quality, and low risk (Figure 12)—are precisely those with the largest positive Shapley contributions. Negative contributions are limited and small in magnitude. The thematic decomposition confirms that UPSA’s nonlinear shrinkage produces an SDF that systematically reallocates toward persistent, fundamental-related themes, and that these tilts account for a substantial share of UPSA’s out-of-sample outperformance.

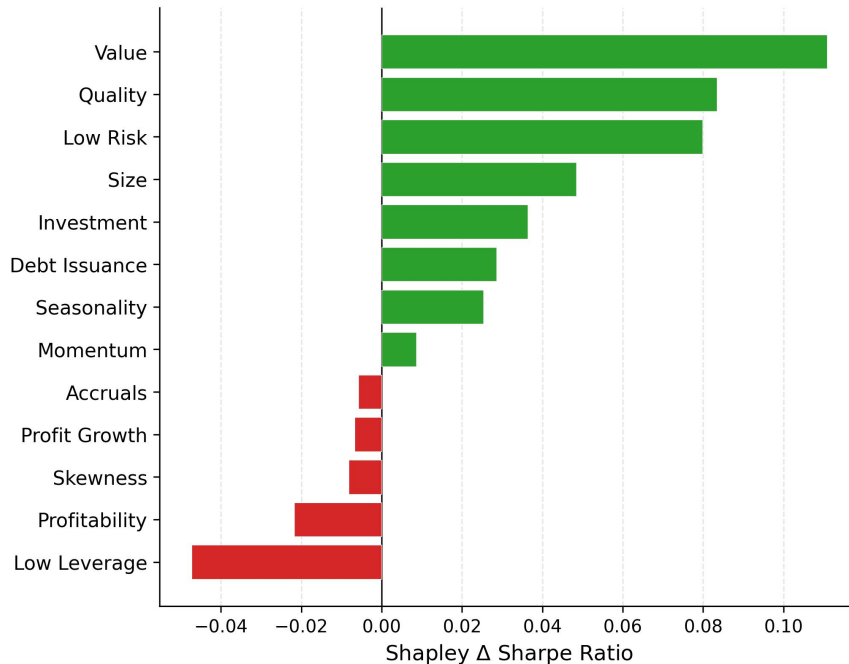


Figure 13: Shapley Decomposition of Δ Sharpe Ratio by Theme.

Shapley decomposition of the Sharpe ratio gap between UPSA and ridge by theme (Equation (22)). Bars show each theme’s contribution to ΔSR , which sums to the total gap. Positive values indicate contributions to UPSA’s outperformance; negative values indicate offsets. Both portfolios are scaled to 10% annualized volatility. The out-of-sample period is November 1981 to December 2022.

6 UPSA Through the Lens of Simulation

We use simulations to isolate the mechanisms behind UPSA’s outperformance. The design separates two key ingredients: tuning shrinkage to a portfolio objective and allowing shrinkage to vary flexibly across the eigenvalue spectrum.

We simulate data from

$$F_t = \mu + (\Sigma - \mu\mu')^{1/2}X_t, \quad X_t \sim \mathcal{N}(0, I), \quad (23)$$

with diagonal $\Sigma = \text{diag}(\lambda)$, so the infeasible optimal Markowitz weights are $\pi_{*,i} = \mu_i/\lambda_i$. The eigenvalue distribution has 10% large ($\lambda = 10$), 80% medium ($\lambda = 1$), and 10% small ($\lambda = 0.1$) components. Expected returns are

$$\mu_i = \begin{cases} \frac{1}{\sqrt{N}} & \text{if } \lambda_i = 10, \\ 0 & \text{if } \lambda_i = 1, \\ \frac{0.5}{\sqrt{N}} & \text{if } \lambda_i = 0.1. \end{cases} \quad (24)$$

The asset market thus consists of three types of factors: concentrated risks ($\lambda_i = 10$) commanding a significant return premium, a large set of moderate-variance factors ($\lambda_i = 1$) with no premium, and a minority of low-risk factors ($\lambda_i = 0.1$) with non-negligible expected returns resembling near-arbitrage opportunities. These near-arbitrage factors are difficult to estimate because their low variance makes them easily obscured by higher-variance components.

We produce 10,000 simulated samples with $N = 150$ assets and $T = 600$ observations, roughly matching our empirical dimensions in Section 4. The first half of each sample is used for training and the second half for out-of-sample testing. We compare model performance based on the out-of-sample Sharpe ratio $SR(R_t) = \bar{E}[R_t]/\sqrt{\bar{E}[R_t^2]}$.²⁹

6.1 Economic vs. Statistical Tuning

We first isolate the role of the tuning objective by comparing two ridge shrinkage approaches that differ only in their calibration objective. Ridge Statistical minimizes second moment

²⁹By 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.

estimation error:³⁰

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

Ridge Economic instead selects the shrinkage parameter to maximize out-of-sample quadratic utility:³¹

$$\max_f E[R_t^{\bar{\pi}(f)} | F_{IS}] - \frac{1}{2} E[(R_t^{\bar{\pi}(f)})^2 | F_{IS}] \quad \text{s.t.} \quad f(\bar{\lambda}_i) = f_z(\bar{\lambda}_i) = \frac{1}{z + \bar{\lambda}_i}. \quad (25)$$

At the population level, the low-variance PCs are the most profitable, delivering the highest optimal weights $\pi_{*,i}$. To maximize out-of-sample performance, the shrinkage function must tilt portfolio weights toward these components despite substantial finite-sample noise. Statistical shrinkage, however, corrects sample eigenvalues toward their population values, remaining agnostic to expected returns.

Figure 14 confirms this distinction. Statistical ridge shrinks eigenvalues toward the bulk of the population distribution, while economic ridge does the opposite and tilts more towards the smaller, high-return PCs. Economic tuning achieves a higher average Sharpe ratio (0.28 vs. 0.22) and outperforms in 87% of simulations, confirming that the choice of objective is a primary driver of improvement, even before allowing for nonlinear shrinkage.

6.2 Linear vs. Nonlinear Shrinkage

We next hold the economic objective (25) fixed and compare Ridge (linear shrinkage) with UPSA (nonlinear shrinkage maximizing out-of-sample quadratic utility as in Definition 1).³² An effective shrinkage rule should apply minimal shrinkage to the low-variance, high-return PCs to preserve their profitability, aggressively shrink the mid-variance PCs since they do not contribute to the efficient portfolio, and apply moderate shrinkage to the high-variance PCs, which are less noisy and moderately informative. Ridge cannot address all three cases

³⁰The shrinkage parameter is selected in closed-form using the methodology of Ledoit and Wolf (2003), which is designed to minimize Frobenius norm loss.

³¹The shrinkage parameter is selected via leave-one-out cross-validation over a logarithmic grid spanning 10^{-4} to 10^4 .

³²Both methods use the same leave-one-out cross-validation procedure and logarithmic tuning grid.

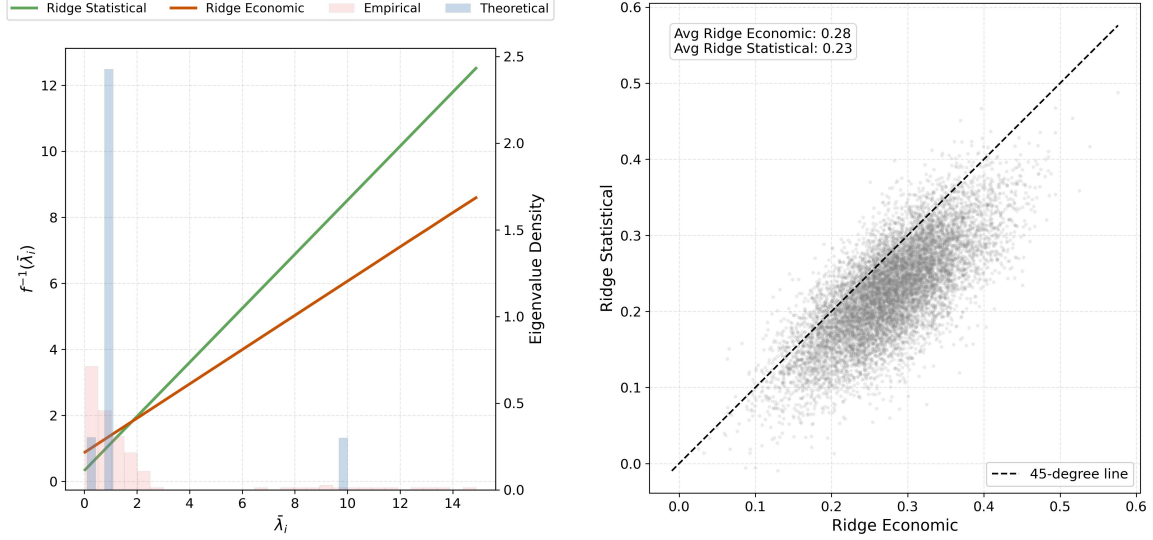


Figure 14: Tuning to Economic vs. Statistical Objectives.

Data are simulated using the data-generating process (23)–(24) 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.

simultaneously: if z is large enough to suppress the uninformative mid-variance PCs, it over-shrinks the profitable low-variance PCs; if z is small enough to preserve them, it under-regularizes the noisy middle of the spectrum.

UPSA resolves this tradeoff by dynamically reallocating capital across the spectrum. As shown in Figure 15, UPSA assigns larger portfolio weights to principal components that pair small eigenvalues with high expected returns, strictly limits exposure to mid-range components, and maintains baseline weights on the largest eigenvalues owing to their high signal-to-noise ratio. Ultimately, this structural flexibility yields a higher average Sharpe ratio (0.33 vs. 0.28) and drives outperformance in 96% of the simulations.

The simulation also shows that only a small number of shrinkage regimes are needed to approximate the optimal rule. Because the data-generating process contains three distinct eigenvalue regions, a Bayesian investor who knew this structure would require a prior mixture with only two components (Section 5.1). Consistent with this intuition, most gains are achieved with a small grid, and additional grid points provide negligible improvements (Figure 30 in the Appendix).

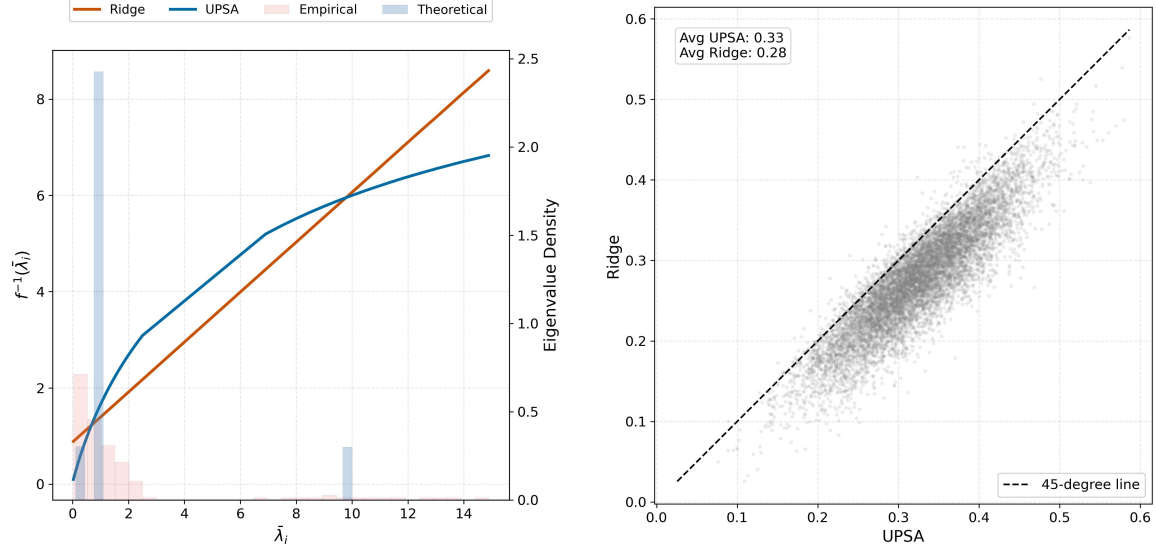


Figure 15: Linear vs. Nonlinear Shrinkage with Economic Objectives.

Data are simulated using the data-generating process (23) and (24) for $N = 150$ assets, $T = 600$ observations, and 10,000 simulations. The left panel shows the empirical eigenvalue distribution versus the shrunk eigenvalues for the linear (ridge) and nonlinear (UPSA) 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.

6.3 Adapting to Time-Varying Uncertainty

We extend the data-generating process to isolate the effect of time-varying expected returns. The eigenvalue structure is held fixed while expected returns on the high-variance PCs switch between two regimes:

$$\mu_i^{(s)} = \begin{cases} \frac{1}{\sqrt{N}} & \text{if } \lambda_i = 10 \text{ and } s = 1, \\ \frac{5}{\sqrt{N}} & \text{if } \lambda_i = 10 \text{ and } s = 2, \\ 0 & \text{if } \lambda_i = 1, \\ \frac{0.5}{\sqrt{N}} & \text{if } \lambda_i = 0.1, \end{cases} \quad (26)$$

where $s \in \{1, 2\}$ denotes the latent state. State 1 represents a recession-like environment in which concentrated-risk factors earn a moderate premium; State 2 represents normal times in which these same factors earn a substantially larger premium. The regime evolves as a

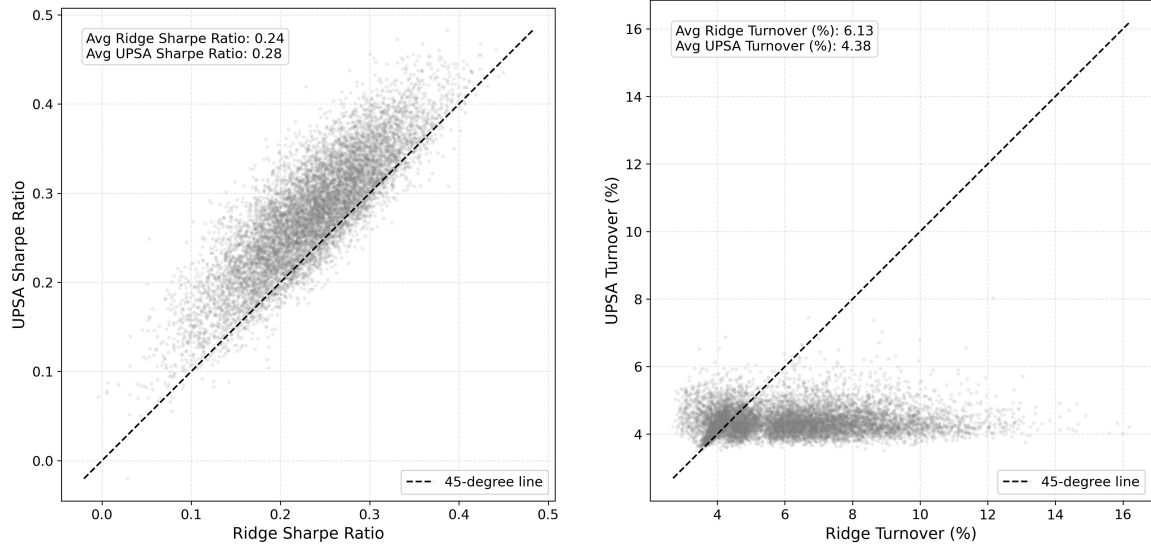


Figure 16: UPSA vs. Ridge Under Markov-Switching Regimes.

Data are simulated from the Markov-switching data-generating process (26) with $N = 150$ assets, $T = 600$ observations, a rolling window of 300, and 10,000 independent paths. The left panel plots UPSA vs. ridge out-of-sample Sharpe ratios; the right panel plots average portfolio turnover. Each point represents one simulated path. Points above (below) the 45-degree line indicate UPSA outperformance (underperformance).

two-state Markov chain: at each period, the state switches with probability $p = 0.05$ and persists with probability $1 - p$, implying an expected regime duration of roughly 20 periods.

Figure 16 summarizes the results. UPSA delivers a higher Sharpe ratio in 88% of simulations (mean 0.28 vs. 0.24) and lower turnover in 74% (mean 4.4% vs. 6.1%). Because the covariance structure is identical across regimes, these differences isolate how the two estimators adapt to changes in expected returns.

Appendix Figures 31–33 examine a representative simulation path. The patterns closely mirror the empirical evidence in Section 5.2: UPSA gradually reallocates weight across the penalty grid as regimes change, whereas ridge alternates between extreme penalty values. The simulation shows that time variation in expected returns alone is sufficient to generate the joint improvement in Sharpe ratios and turnover documented in the data.

7 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. The estimator adapts to changing market conditions and allocates risk across principal components more effectively than standard spectral shrinkage methods. In particular, UPSA regularizes intermediate components more aggressively while allowing the most informative components to remain active, effectively tailoring shrinkage to each component’s signal-to-noise ratio without relying on arbitrary truncation rules. This flexibility admits a natural Bayesian interpretation in which the investor integrates over beliefs about expected-return uncertainty rather than committing to a single shrinkage level. Empirically, this leads to more stable portfolio allocations, lower turnover, and greater resilience during recessions. Our findings underscore the importance of designing shrinkage strategies that are flexible and aligned with the portfolio’s economic 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.
- , **Markus Pelger, and Jason Zhu**, “Forest through the trees: Building cross-sections of stock returns,” *The Journal of Finance*, 2025, 80 (5), 2447–2506.
- , **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.

- Chen, Andrew Y. and Chukwuma Dim**, “High-Throughput Asset Pricing,” *arXiv preprint*, 2023, *arXiv:2311.10685*. Available at <https://arxiv.org/abs/2311.10685>.
- Chen, Andrew Y and Tom Zimmermann**, “Open source cross-sectional asset pricing,” *Critical Finance Review*, 2022, 11 (2), 207–264.
- 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,” *The Journal of Finance*, 2025, 80 (1), 259–319.
- Golub, Gene H, Michael Heath, and Grace Wahba**, “Generalized cross-validation as a method for choosing a good ridge parameter,” *Technometrics*, 1979, 21 (2), 215–223.
- 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.
- Hornik, Kurt, Maxwell Stinchcombe, and Halbert White**, “Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks,” *Neural networks*, 1990, 3 (5), 551–560.
- 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.

- Ke, Shikun (Barry) and Mohammad Pourmohammadi**, “Shrinkage Alignment in High-Dimensional Portfolios,” 2025. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5723922.
- Kelly, Bryan, Semyon Malamud, and Lasse Heje Pedersen**, “Principal portfolios,” *The Journal of Finance*, 2023, 78 (1), 347–387.
- , **Seth Pruitt, and Yinan Su**, “Characteristics are covariances: A unified model of risk and return,” *Journal of Financial Economics*, 2019, 134 (3), 501–524.
- Kelly, Bryan T, Semyon Malamud, and Kangying Zhou**, “The virtue of complexity in return prediction,” *The Journal of Finance*, 2024, 79 (1), 459–503.
- Korsaye, Sofonias Alemu, Alberto Quaini, and Fabio Trojani**, “Smart stochastic discount factors,” *Management Science*, 2025.
- 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 P     , “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.
- L     , Karl, “     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*, 2022, 145 (1), 154–177.
- Murphy, Kevin P, “Conjugate Bayesian analysis of the Gaussian distribution,” *def*, 2007, 1 (2  2), 16.
- 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.
- Shapley, Lloyd S., “A value for n-person games,” in Harold W. Kuhn and Albert W. Tucker, eds., *Contributions to the Theory of Games II*, Princeton University Press, 1953, pp. 307–317.
- 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

This appendix collects the proofs of all results stated in the main text. We prove the universal approximation properties of UPSA (Lemmas 1 and 5), the characterization of the optimal full-information shrinkage function (Lemma 3), and the unbiasedness of LOO cross-validation (Lemma 4).

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 the 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 $\bar{\lambda} \geq 0$:

$$g(\bar{\lambda}) = a + \bar{\lambda}b + \int_0^\infty \frac{\bar{\lambda}}{z + \bar{\lambda}} d\mu(z) = a + \bar{\lambda}(b + \mu(\mathbb{R}_+)) - \int_0^\infty \frac{z}{z + \bar{\lambda}} d\mu(z). \quad (27)$$

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

$$0 = \lim_{\bar{\lambda} \rightarrow \infty} g(\bar{\lambda}) = \lim_{\bar{\lambda} \rightarrow \infty} (a + \bar{\lambda}(b + \mu(\mathbb{R}_+))) , \quad (28)$$

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

$$f(\bar{\lambda}) = \int_0^\infty \frac{z}{z + \bar{\lambda}} d\mu(z) =: \int_0^\infty \frac{1}{z + \bar{\lambda}} d\nu(z) , \quad (29)$$

for a positive measure ν on $\mathbb{R}_+$ having Radon-Nikodym 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:

$$K \geq \sup_{\bar{\lambda} \geq 0} \bar{\lambda}f(\bar{\lambda}) \geq \lim_{\bar{\lambda} \rightarrow \infty} \int_0^\infty \frac{\bar{\lambda}}{z + \bar{\lambda}} d\nu(z) = \int_0^\infty d\nu(z) , \quad (30)$$

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

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:

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

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 $\mathbb{R}_+ \setminus [z_{\max}, \infty) = \bigcup_{i=1}^L [z_{i-1}, z_i)$, where $z_0 := 0$ and $z_L := z_{\max}$, such that $|z_i - z_{i-1}| < \epsilon$ for any $i = 1 \dots L$. Consider now a piecewise 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 $z \in [z_{i-1}, z_i)$ and $g(\bar{\lambda}, z) = \frac{1}{z_{\max} + \bar{\lambda}}$ if $z \geq z_{\max}$. Using this function, we obtain:

$$\int_0^\infty g(\bar{\lambda}, z) d\nu(z) = \sum_{i=1}^L \frac{1}{z_i + \bar{\lambda}} \nu([z_{i-1}, z_i)) + \frac{1}{z_{\max} + \bar{\lambda}} \nu([z_{\max}, \infty)) , \quad (31)$$

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

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

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. $\square$

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 $i = 1, \dots, K := \min(N, T)$:

$$f_\star(\bar{\lambda}_i) = \frac{\bar{U}'_i \pi_\star}{\bar{U}'_i \bar{\mu}}. \quad (32)$$

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:

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

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

$$\mathbf{f}_\star(\bar{\lambda}) \circ \bar{U}' \bar{\mu} = (\bar{U}' \Sigma \bar{U})^{-1} \bar{U}' \mu = \bar{U}' \Sigma^{-1} \mu = \bar{U}' \pi_\star, \quad (33)$$

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

$$f_\star(\bar{\lambda}_i) = \frac{\bar{U}'_i \pi_\star}{\bar{U}'_i \bar{\mu}}. \quad (34)$$

This concludes the proof. □

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

$$R_{T+1,t}(f) := \bar{\pi}_{T+1,t}(f)' F_t ; \quad t = 1, \dots, T+1, \quad (35)$$

denote the sequence of LOO portfolio returns. Then,

$$U_{LOO}^{OOS}(f) := \frac{1}{T+1} \sum_{t=1}^{T+1} U(R_{T+1,t}(f)) \quad (36)$$

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

$$E[U(R_\tau(f))] = E[U_{LOO}^{OOS}(f)] ,$$

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 $\tau \in \{1, \dots, T\}$. 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 $\tau' \in \{1, \dots, T\}$:

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

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. $\square$

Lemma 5 *Any continuous function f can be uniformly approximated over compact intervals by a function $f_Z \in \mathcal{F}(\mathcal{Z})$, whenever the grid $\mathcal{Z}$ is sufficiently wide and dense.*

Proof of Lemma 5. 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

$$\Theta_{z_1}(x) - \Theta_{z_2}(x) = (z_2 - z_1)\Theta_{z_1}(x)\Theta_{z_2}(x), \quad (37)$$

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. $\square$

B Bayesian Foundations: Technical Details

This appendix provides the formal details underlying the Bayesian interpretation of UPSA developed in Section 5.1. Section B.1 derives the posterior Markowitz portfolio under expected return uncertainty and discusses its economic implications. Section B.2 extends the analysis to joint mean and covariance uncertainty. Proofs are collected in Section B.3.

B.1 Expected Return Uncertainty

Let $F_1, \dots, F_T$ be IID Gaussian in $\mathbb{R}^N$ with expectation μ and positive definite covariance matrix Σ :

$$F_t \mid \mu, \Sigma \sim N(\mu, \Sigma). \quad (38)$$

The unknown Markowitz portfolio is $b := \Sigma^{-1}\mu$. We model expected return uncertainty with a Gaussian prior whose dispersion depends on Σ :

$$\mu \mid \Sigma \sim N(0, g(\Sigma)), \quad (39)$$

where g is a strictly positive matrix function.

This prior has direct implications for the Markowitz portfolio through two economically interpretable quantities. First, the prior expected squared Sharpe ratio:

$$E[\mu' \Sigma^{-1} \mu \mid \Sigma] = \text{tr}(\Sigma^{-1} g(\Sigma)) = \sum_{i=1}^N \frac{g(d_i)}{d_i}, \quad (40)$$

and second, the prior expected squared norm of the portfolio weights:

$$E[b' b \mid \Sigma] = \text{tr}(\Sigma^{-2} g(\Sigma)) = \sum_{i=1}^N \frac{g(d_i)}{d_i^2}, \quad (41)$$

where $d_1, \dots, d_N$ are the eigenvalues of Σ . These show that g implicitly governs how different principal components contribute to the portfolio's performance and leverage.

Lemma 6 *Given the Gaussian likelihood (38) and prior (39), the posterior Markowitz portfolio is:*

$$E[b \mid F_1, \dots, F_T, \Sigma] = \left(\frac{1}{T} \Sigma[g(\Sigma)]^{-1} \Sigma + \Sigma \right)^{-1} \bar{E}[F]. \quad (42)$$

To connect this to UPSA, we factorize g as $g(\Sigma) = \frac{1}{\text{tr}(\Sigma)} \Sigma h(\Sigma) \Sigma$ for some strictly positive matrix function h . This specific factorization provides a clean economic mapping: since the optimal portfolio weights are $b = \Sigma^{-1} \mu$, specifying the prior variance of μ as proportional to $\Sigma h(\Sigma) \Sigma$ implies that the prior variance of the weights b is proportional to $\Sigma^{-1} (\Sigma h(\Sigma) \Sigma) \Sigma^{-1} = h(\Sigma)$. This ensures that the prior uncertainty about expected returns directly translates into the desired spectral shrinkage on the portfolio weights, yielding the exact form stated in equation (20) of the main text.

Under this factorization, the prior expected portfolio norm simplifies to

$$E[b'b \mid \Sigma] = \frac{\sum_{i=1}^N h(d_i)}{\sum_{i=1}^N d_i}, \quad (43)$$

so the monotonicity of h directly governs whether high-variance or low-variance PCs contribute more to the Markowitz portfolio norm *a priori*.

Established parametric prior specifications in the literature include matrix power functions $g(\Sigma) = \frac{1}{z \text{tr}(\Sigma)} \Sigma^\eta$ for $z > 0$ and $\eta \geq 0$, giving $h(d) = \frac{1}{z} d^{\eta-2}$. This function is decreasing when $\eta \leq 2$ and increasing when $\eta \geq 2$. The boundary case $\eta = 2$ makes all PCs contribute equally to the portfolio norm, producing ridge shrinkage:

$$f(\Sigma) = \left(\frac{\text{tr}(\Sigma)}{T} z I_N + \Sigma \right)^{-1}.$$

More generally, whenever the posterior shrinkage f is matrix monotonically decreasing, the universal approximation property (Lemma 1 of the main text) yields the UPSA representation in equation (21).

B.2 Joint Mean and Covariance Uncertainty

The analysis above conditions on a known covariance matrix. We now show that UPSA emerges even when the investor faces joint uncertainty about both μ and Σ , modeled through a mixture of L conjugate Normal–Inverse Wishart priors (see, e.g., [Murphy, 2007](#)):

$$\mu \mid \Sigma \sim N(0, \Sigma/\tau), \quad (44)$$

$$\Sigma \sim W^{-1}(m_j, \Lambda_j), \quad (45)$$

with probability $\omega_j > 0$ for each $j = 1, \dots, L$, where $\tau > 0$, $m_j > 0$, and Λ_j are $N \times N$ positive definite matrices. Each mixture component represents a distinct hypothesis about the scale of covariance uncertainty.

Lemma 7 *Given the Gaussian likelihood (38) and the mixture prior (44)–(45), the posterior Markowitz portfolio is:*

$$E[b \mid F_1, \dots, F_T] = \sum_{j=1}^L \nu_j \left(\frac{1}{T} \Lambda_j + \bar{E}[FF'] \right)^{-1} \bar{E}[F], \quad (46)$$

with strictly positive weights

$$\nu_j = \frac{q_j (m_j + T)}{\tau + T \left(1 - \bar{E}[F]' \left(\frac{1}{T} \Lambda_j + \bar{E}[FF'] \right)^{-1} \bar{E}[F] \right)}, \quad (47)$$

where $q_1, \dots, q_L$ are strictly positive posterior mixing probabilities.

Setting the hyper-parameters $\Lambda_j = Nz_j I_N$ for a grid of constants $0 < z_1 < \dots < z_L$, equation (46) becomes

$$E[b \mid F_1, \dots, F_T] = \sum_{j=1}^L \nu_j \left(\frac{N}{T} z_j I_N + \bar{E}[FF'] \right)^{-1} \bar{E}[F], \quad (48)$$

which is precisely a UPSA ridge ensemble.

The posterior ensemble weights ν_j are endogenous: they are determined jointly by the prior (through ω_j and m_j) and the data (through the posterior mixing probabilities q_j).

When the mixture reduces to a single component ($\omega_k = 1$, $\omega_j = 0$ for $j \neq k$), the posterior collapses to a single ridge portfolio, confirming that ridge is a special case corresponding to a non-mixing prior.

Even when the prior on expected returns is uninformative ($\tau \rightarrow 0$), a UPSA ensemble emerges with weights

$$\nu_j = \frac{q_j \left(1 + \frac{N}{T} \cdot \frac{m_j}{N}\right)}{1 - \bar{E}[F]' \left(\frac{N}{T} z_j I_N + \bar{E}[FF']\right)^{-1} \bar{E}[F]}. \quad (49)$$

In the standard data regime ($N/T \rightarrow 0$), covariance uncertainty vanishes and the weights are proportional to the posterior mixing probabilities q_j . In the complex regime ($N/T \rightarrow c > 0$), the weights tilt toward components with larger m_j (greater prior covariance uncertainty), proportionally to complexity N/T .

It is worth contrasting this foundation with that of Section B.1. There, UPSA arises from a nonlinear prior on expected returns—function $g(\Sigma)$ introduces differential uncertainty across PCs, and the posterior shrinkage inherits this nonlinearity. Here, UPSA arises from uncertainty about the *scale* of the covariance matrix in a mixture model, even when the prior on expected returns is flat across PCs. Indeed, prior (44)–(45) implies a flat contribution of different principal components to the squared Sharpe ratio:

$$E[\mu' \Sigma^{-1} \mu] = \frac{N}{\tau}, \quad (50)$$

but a decreasing contribution to the Markowitz portfolio norm:

$$E[b'b] = \frac{1}{\tau} \sum_{i=1}^N \frac{1}{d_i}. \quad (51)$$

These are economically distinct mechanisms, yet both produce the same UPSA functional form.

B.3 Proofs of Bayesian Lemmas

Proof of Lemma 6. Given hyper-parameters $\mu_0 = 0$ and $\Sigma_0 = g(\Sigma)$ in prior (39), the posterior expected return is given in closed-form by standard conjugate Gaussian results (see, e.g., [Murphy, 2007](#)):

$$\begin{aligned} E[\mu \mid F_1, \dots, F_T, \Sigma] &= (\Sigma_0^{-1} + T\Sigma^{-1})^{-1}(\Sigma_0^{-1}\mu_0 + T\Sigma^{-1}\bar{E}[F]) \\ &= \left(\frac{1}{T}\Sigma\Sigma_0^{-1} + I_N\right)^{-1} \bar{E}[F]. \end{aligned}$$

Therefore, the expected posterior Markowitz portfolio is:

$$E[b \mid F_1, \dots, F_T, \Sigma] = \Sigma^{-1}E[\mu \mid F_1, \dots, F_T, \Sigma] = \left(\frac{1}{T}\Sigma[g(\Sigma)]^{-1}\Sigma + \Sigma\right)^{-1} \bar{E}[F].$$

This concludes the proof. □

Proof of Lemma 7. Under IID assumption (38) for observations $\mathbf{F} := (F_1, \dots, F_T)$, the Gaussian likelihood factorizes as $L(\mathbf{F} \mid \mu, \Sigma) = \prod_{t=1}^T f(F_t \mid \mu, \Sigma)$, where $f(\cdot \mid \mu, \Sigma)$ is the Gaussian density. The conjugate Normal-Inverse Wishart mixture prior for (μ, Σ) is such that:

$$\begin{aligned} \mu \mid \Sigma &\sim N(0, \Sigma/\tau), \\ \Sigma &\sim W^{-1}(m_j, \Lambda_j), \end{aligned}$$

with probability $\omega_j > 0$ for each $j = 1, \dots, L$. Therefore, the posterior density $\pi(\mu, \Sigma \mid \mathbf{F})$ satisfies:

$$\pi(\mu, \Sigma \mid \mathbf{F}) \propto L(\mathbf{F} \mid \mu, \Sigma) g(\mu, \Sigma) = \sum_{j=1}^L \omega_j L(\mathbf{F} \mid \mu, \Sigma) g_j(\mu \mid \Sigma) h_j(\Sigma),$$

where $g(\mu, \Sigma)$ is the joint prior density of (μ, Σ) , $g_j(\mu \mid \Sigma)$ the density of $N(0, \Sigma/\tau)$, and $h_j(\Sigma)$ the density of $W^{-1}(m_j, \Lambda_j)$. Consequently, the posterior is a mixture of L posteriors

$\pi_j(\mu, \Sigma \mid \mathbf{F})$, each arising from a standard Normal–Inverse Wishart conjugate update:

$$\pi_j(\mu, \Sigma \mid \mathbf{F}) \propto L(\mathbf{F} \mid \mu, \Sigma) g_j(\mu \mid \Sigma) h_j(\Sigma).$$

By standard conjugate results (see, e.g., [Murphy, 2007](#)), under posterior π_j we have:

$$\begin{aligned} \mu \mid \mathbf{F}, \Sigma &\sim N(\mu_T, \Sigma/\tau_T), \\ \Sigma \mid \mathbf{F} &\sim W^{-1}(m_{jT}, \Lambda_{jT}), \end{aligned}$$

where

$$\begin{aligned} m_{jT} &= m_j + T, \\ \Lambda_{jT} &= \Lambda_j + T(\bar{E}[FF'] - \bar{E}[F]\bar{E}[F]') + \frac{\tau T}{\tau + T} \bar{E}[F]\bar{E}[F]', \\ \mu_T &= \frac{T}{\tau + T} \bar{E}[F], \\ \tau_T &= \tau + T. \end{aligned}$$

This gives the posterior Markowitz portfolio:

$$E_\pi[\Sigma^{-1} \mu \mid \mathbf{F}] = E_\pi[\Sigma^{-1} E_\pi[\mu \mid \mathbf{F}, \Sigma] \mid \mathbf{F}] = E_\pi[\Sigma^{-1} \mid \mathbf{F}] \mu_T = \frac{T}{\tau + T} E_\pi[\Sigma^{-1} \mid \mathbf{F}] \bar{E}[F]. \quad (52)$$

We next compute $E_\pi[\Sigma^{-1} \mid \mathbf{F}]$. The posterior distribution of Σ is a mixture of inverse Wishart distributions with mixing weights

$$q_j = \frac{\omega_j \int L(\mathbf{F} \mid \mu, \Sigma) g_j(\mu \mid \Sigma) h_j(\Sigma) d\mu d\Sigma}{\sum_{\ell=1}^L \omega_\ell \int L(\mathbf{F} \mid \mu, \Sigma) g_\ell(\mu \mid \Sigma) h_\ell(\Sigma) d\mu d\Sigma}; \quad j = 1, \dots, L.$$

The posterior distribution of Σ^{-1} is therefore a mixture of Wishart distributions with mixing weights q_j and parameters $(m_{jT}, \Lambda_{jT}^{-1})$. Using the fact that $E[\Sigma^{-1}] = m \Lambda^{-1}$ when $\Sigma^{-1} \sim W(m, \Lambda^{-1})$, we obtain:

$$E_\pi[\Sigma^{-1} \mid \mathbf{F}] = \sum_{j=1}^L q_j m_{jT} \left(\Lambda_j + T \bar{E}[FF'] - \frac{T^2}{\tau + T} \bar{E}[F]\bar{E}[F]' \right)^{-1}.$$

Applying the Sherman–Morrison formula to handle the rank-one adjustment by $-\frac{T^2}{\tau+T}\bar{E}[F]\bar{E}[F]'$:

$$\left(\Lambda_j + T\bar{E}[FF'] - \frac{T^2}{\tau+T}\bar{E}[F]\bar{E}[F]'\right)^{-1}\bar{E}[F] = \frac{(\Lambda_j + T\bar{E}[FF'])^{-1}\bar{E}[F]}{1 - \frac{T^2}{\tau+T}\bar{E}[F]'(\Lambda_j + T\bar{E}[FF'])^{-1}\bar{E}[F]}.$$

Therefore, using identity (52):

$$\begin{aligned} E_\pi[\Sigma^{-1}\mu \mid \mathbf{F}] &= \frac{T}{\tau+T}E_\pi[\Sigma^{-1} \mid \mathbf{F}]\bar{E}[F] \\ &= \frac{T}{\tau+T}\sum_{j=1}^L \frac{q_j(m_j+T)}{p_{jT}(\Lambda_j)}(\Lambda_j + T\bar{E}[FF'])^{-1}\bar{E}[F] \\ &= \frac{1}{\tau+T}\sum_{j=1}^L \frac{q_j(m_j+T)}{p_{jT}(\Lambda_j)}\left(\frac{\Lambda_j}{T} + \bar{E}[FF']\right)^{-1}\bar{E}[F], \end{aligned}$$

where:

$$p_{jT}(\Lambda_j) = 1 - \frac{T}{\tau+T}\bar{E}[F]'\left(\frac{1}{T}\Lambda_j + \bar{E}[FF']\right)^{-1}\bar{E}[F].$$

For hyper-parameter choices $\Lambda_j = Nz_jI_N$ with $0 < z_1 < \dots < z_L$, this gives:

$$E_\pi[\Sigma^{-1}\mu \mid \mathbf{F}] = \sum_{j=1}^L \nu_j \left(\frac{N}{T}z_jI_N + \bar{E}[FF']\right)^{-1}\bar{E}[F],$$

with the UPSA portfolio weights:

$$\nu_j := \frac{1}{\tau+T} \cdot \frac{q_j(m_j+T)}{p_{jT}(\Lambda_j)}.$$

This concludes the proof. □

C Sensitivity to the Ridge Grid

The performance of UPSA depends on the grid $\mathcal{Z}$ of ridge penalties used to construct the basis functions. As shown in equations (11) and (10), the grid must span values well above the largest eigenvalue (strong regularization, approaching identity shrinkage) to well below the

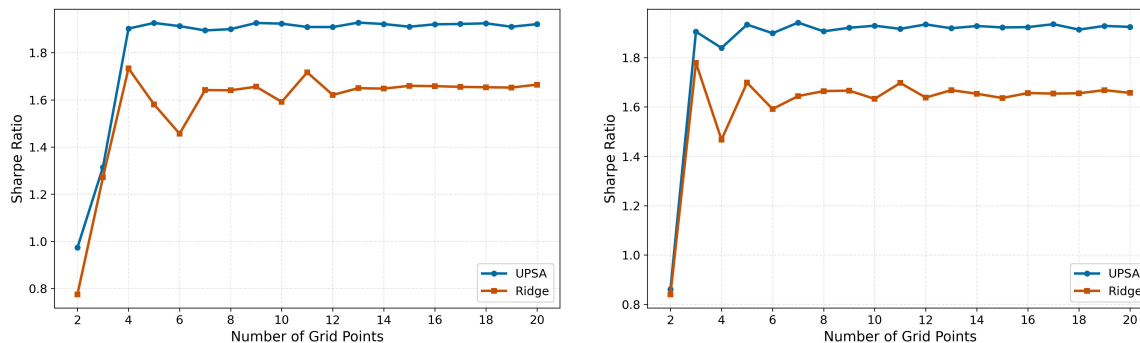


Figure 17: Sensitivity of UPSA to the Number and Placement of Ridge Grid Points.

The left panel varies the number of logarithmically spaced ridge penalties between the fixed endpoints used in the main analysis (10^{-10} to 10^{-1}). The right panel constructs the grid between the smallest and largest empirical eigenvalues of the covariance matrix. The out-of-sample period is November 1981 to December 2022.

smallest (weak regularization, approaching the ridgeless limit), with logarithmically spaced points in between.

Figure 17 assesses sensitivity via two experiments. The left panel fixes the endpoints at 10^{-10} and 10^{-1} and varies the number of intermediate grid points; the right panel instead spans the empirical eigenvalue range. In both cases, UPSA consistently outperforms ridge, and its out-of-sample Sharpe ratio stabilizes once the grid contains roughly four points. These results confirm that UPSA’s performance is not sensitive to the number or placement of ridge penalties.

Figure 18 shows that this robustness extends to turnover. UPSA’s average monthly turnover stabilizes at approximately 12.5% once the grid contains three or more points, and remains flat regardless of further grid refinement. Ridge turnover is both higher (15–17%) and more volatile across grid sizes, reflecting the instability of the cross-validated penalty parameter documented in Section 5.2.

One could also impose an ℓ_1 penalty on the ensemble weights W to let the data select a sparse subset of grid points. However, the results above suggest this is unnecessary, as the estimated weights already concentrate on a small number of penalties.

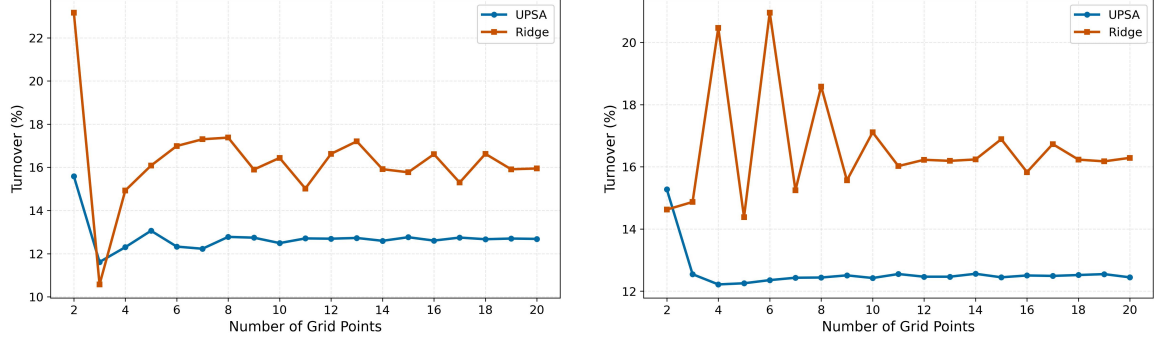


Figure 18: Turnover Sensitivity to the Number and Placement of Ridge Grid Points.

Average monthly turnover (%) as a function of the number of ridge grid points for UPSA and Ridge. The left panel uses a fixed grid (10^{-10} to 10^{-1}); the right panel uses the empirical eigenvalue range. The out-of-sample period is November 1981 to December 2022.

D 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 shrinkage procedure.³³ Specifically, we first apply our spectral shrinkage methods—UPSA and Ridge—and then apply an additional Lasso³⁴ penalty in a second stage. For any spectral shrinkage function f , the matrix $f^{-1}(\bar{\Sigma})$ is diagonal in PC space, so the elastic-net problem separates across coordinates and admits a closed-form solution via soft-thresholding (Tibshirani, 1996, Section 10).

D.1 Effect on Portfolio Performance

Figure 19 reports the annualized out-of-sample Sharpe ratios of UPSA and ridge with and without Lasso augmentation. Introducing an additional ℓ_1 penalty has little effect on performance. UPSA’s Sharpe ratio increases slightly from 1.92 to 1.94, while ridge declines

³³Quaini 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.

³⁴Least Absolute Shrinkage and Selection Operator (Lasso; (Tibshirani, 1996)).

marginally from 1.59 to 1.54. These results indicate that imposing explicit sparsity through a Lasso penalty does not materially change the portfolios' out-of-sample Sharpe ratios.

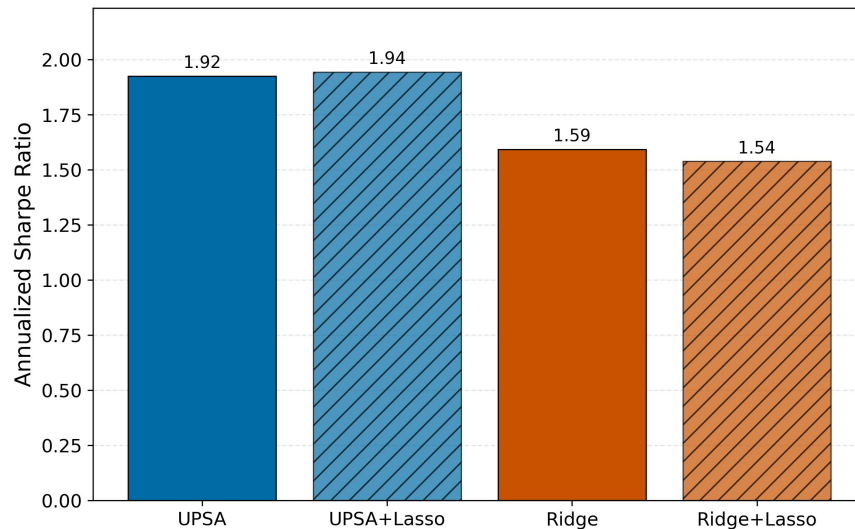


Figure 19: Annualized Sharpe Ratios: Base vs. Lasso-Augmented Portfolios.

Annualized out-of-sample Sharpe ratios for UPSA and ridge, with (hatched) and without (solid) an additional Lasso penalty. The Lasso penalty is selected from the same grid as the ridge penalty using leave-one-out cross-validation. The out-of-sample period is November 1981 to December 2022.

D.2 Instability of Factor Selection

Figure 20 plots the number of non-zero PCs retained by lasso-augmented ridge and UPSA over time. The number of selected components varies substantially across rolling windows. In many periods both methods retain a large share of the available $N = 120$ PCs, while in other periods the number of selected components declines markedly. The median number of retained PCs is roughly 80, although the count occasionally falls to around 17–18.

This variability helps explain why augmenting UPSA with a Lasso penalty has little impact on portfolio performance. As shown in Section 5.3 (Figure 10), UPSA's Sharpe ratio remains relatively stable across a wide range of included PCs. As a result, fluctuations in the number of retained components translate into only modest changes in the resulting portfolios.

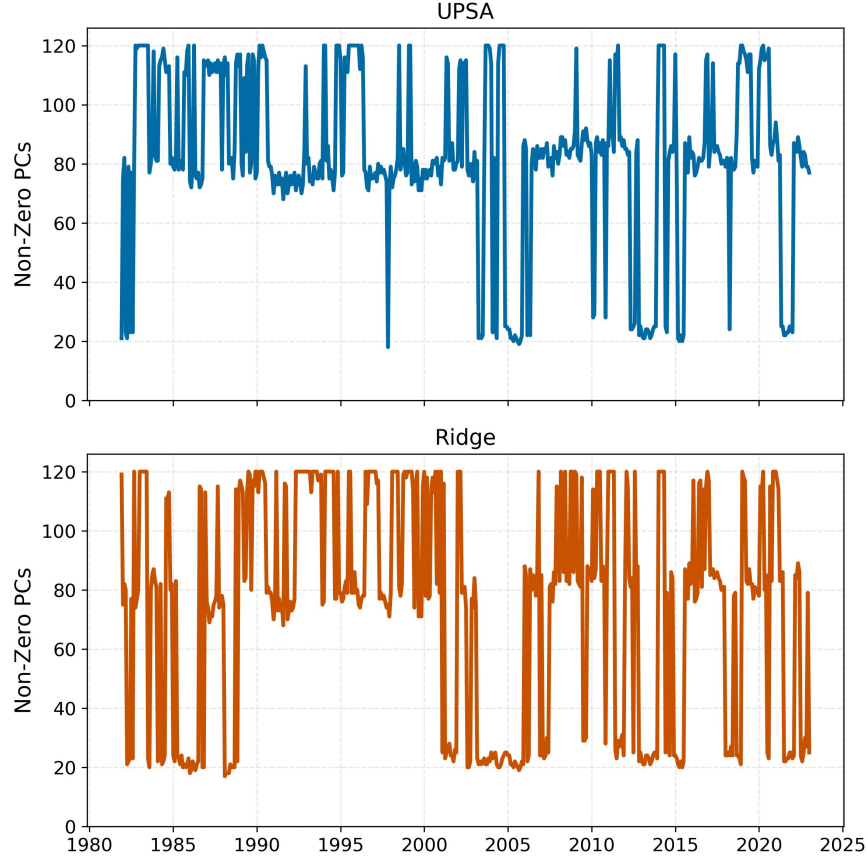


Figure 20: Number of Non-Zero PCs Retained by Lasso Over Time.

Number of non-zero principal components retained by ℓ_1 -penalized ridge and UPSA. The number of selected PCs varies across rolling windows, with periods in which most components are retained and others in which the selected set becomes substantially smaller. The Lasso penalty is selected by leave-one-out cross-validation. The out-of-sample period is November 1981 to December 2022.

The turnover cost of explicit sparsity is substantial. Figure 21 compares the 12-month rolling average turnover for each method with and without the lasso penalty. Adding lasso increases UPSA's mean monthly turnover from 12.4% to 19.5% and ridge's from 16.3% to 24.7%. The turnover spikes are particularly pronounced for ridge, where lasso-augmented turnover exceeds 100% during periods of market stress. These spikes reflect the instability of discrete factor selection: as the set of retained PCs changes across adjacent estimation windows, the portfolio undergoes large reallocations.

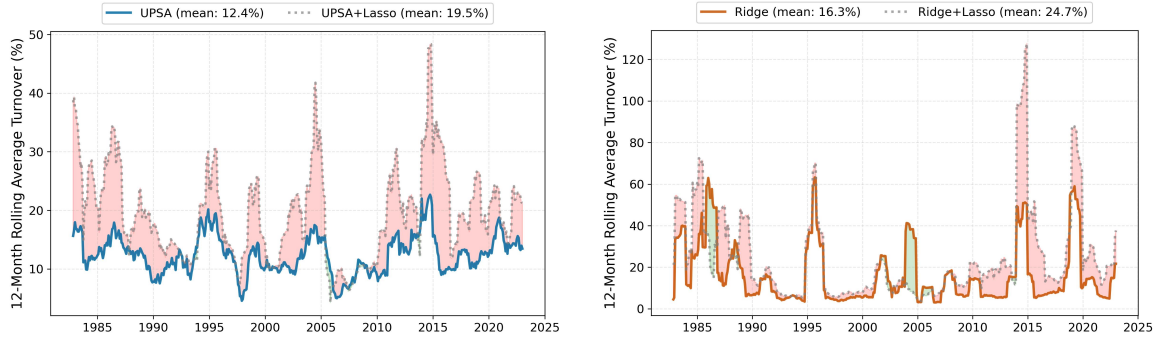


Figure 21: 12-Month Rolling Average Turnover for Base vs. Lasso-Augmented Portfolios.

12-month rolling average turnover for UPSA (left) and ridge (right), with (grey dotted) and without (solid) an additional Lasso penalty. Red shading indicates periods in which the lasso-augmented portfolio exhibits higher turnover. The Lasso penalty is selected by leave-one-out cross-validation. The out-of-sample period is November 1981 to December 2022.

Taken together, UPSA implicitly achieves the regularization benefits that explicit ℓ_1 penalties aim to provide, without the instability of discrete factor selection.

E Additional Empirics and Robustness

This appendix presents supplementary empirical results that complement the analysis in Sections 4 and 5. We begin with a mean-sorted variant of the principal component analysis (Section E.1), followed by a comparison of UPSA against additional benchmarks (Section E.2). Sections E.3 and E.4 document robustness across size-sorted stock universes and alternative rolling window lengths. Section E.5 provides a beta-based pricing evaluation, and Section E.6 examines performance across NBER business cycle regimes.

E.1 Mean-Sorted Principal Components

Figure 10 in the main text orders PCs by decreasing variance (eigenvalue rank). An alternative ordering, by decreasing absolute sample mean return, places the PCs most likely to carry a risk premium first. Figure 22 repeats the analysis under this ordering. The results are qualitatively similar but saturation happens sooner, around $K = 15$ PCs.

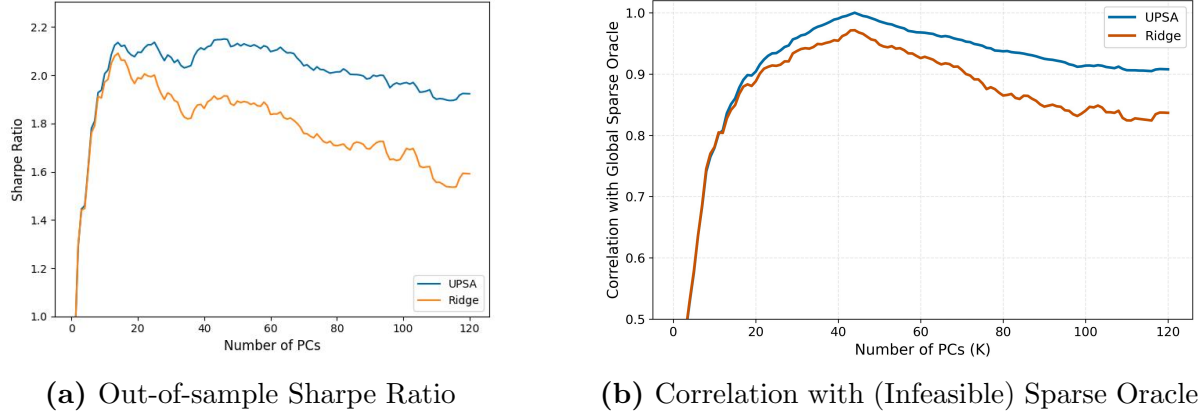


Figure 22: Out-of-sample Performance by Number of Principal Components (Mean-Sorted).

PCs are ordered by decreasing absolute sample mean return. Panel (a): Annualized out-of-sample Sharpe ratio as PCs are incrementally added. Panel (b): Correlation between the K -PC portfolio and the (infeasible) sparse oracle—the subset of K^* PCs that maximizes the out-of-sample Sharpe ratio. The out-of-sample period is November 1981 to December 2022.

E.2 Other Benchmarks

We compare UPSA against two additional benchmarks not considered in the main text:

- **KNS:** Elastic-net shrinkage applied to principal components following [Kozak et al. \(2020\)](#). This method jointly applies ridge and Lasso penalties to the SDF weights in PC space, selecting both penalty parameters by LOO cross-validation.
- **PCA SR:** A subset-selection approach that retains only those principal components with the highest absolute sample mean returns. The number of retained PCs is selected by LOO cross-validation.

Both methods are cross-validated using LOO and trained on the same rolling 120-month window as in the main analysis. Figure 23 reports the results. UPSA delivers the highest out-of-sample Sharpe ratio, achieving 1.92 in the full sample compared to 1.61 for KNS and 1.32 for PCA SR.

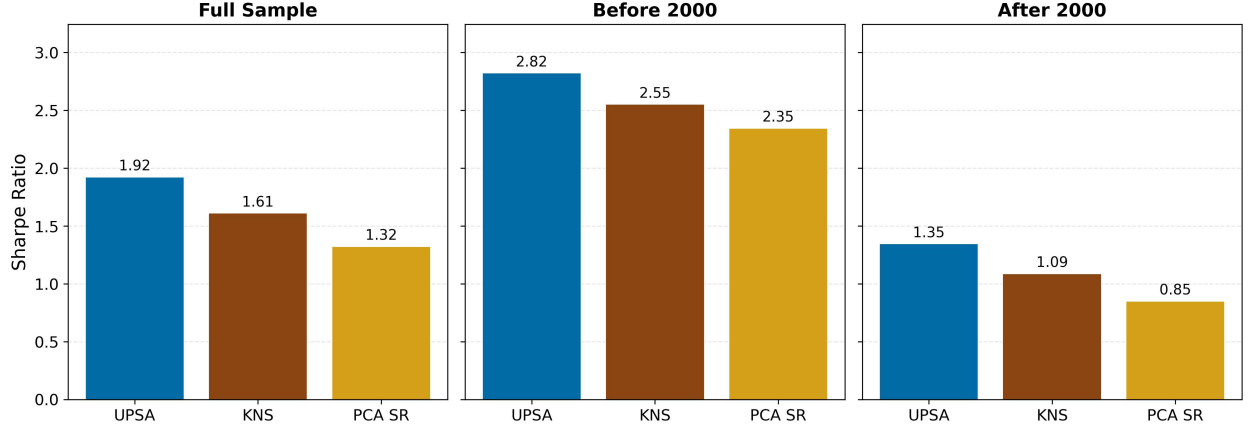


Figure 23: Annualized Out-of-sample Sharpe Ratios: UPSA and Additional Benchmarks.

Annualized out-of-sample Sharpe ratios for UPSA and additional benchmark strategies across three sub-periods. All methods use leave-one-out cross-validation with a 120-month rolling window. The out-of-sample period is November 1981 to December 2022.

E.3 Robustness Across Size Groups

To assess whether UPSA’s performance is driven by a particular segment of the cross-section, we re-estimate all methods on three size-sorted stock universes: mega-cap, large-cap, and small-cap. The main analysis uses the full cross-section of value-weighted anomaly portfolios; here we restrict to size sub-samples while maintaining the same 120-month rolling window and LOO cross-validation.

Figures 24 and 25 report Sharpe ratios and regression alphas, respectively. UPSA delivers the highest Sharpe ratio in all three size groups—1.36 (mega), 1.38 (large), and 2.28 (small)—with the largest margin of relative outperformance among large-cap stocks. The alpha estimates remain positive and economically meaningful across all size groups, confirming that UPSA’s advantage is not concentrated in any single liquidity segment.

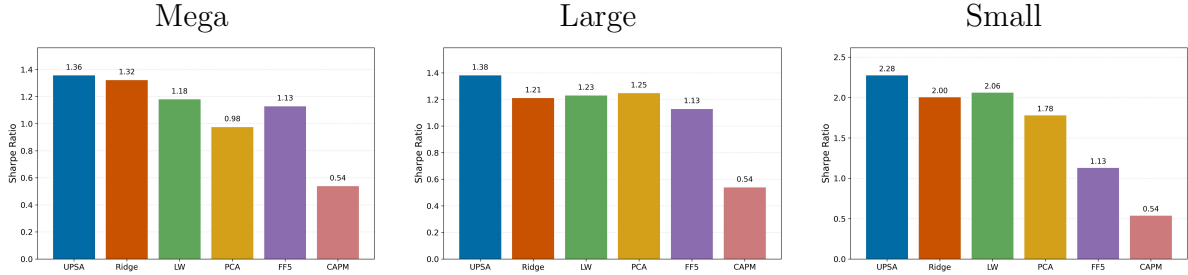


Figure 24: Annualized Sharpe Ratios Across Size Groups.

Annualized out-of-sample Sharpe ratios for UPSA and benchmarks across mega-cap, large-cap, and small-cap stock universes. All methods use a 120-month rolling window with LOO cross-validation. The out-of-sample period is November 1981 to December 2022.

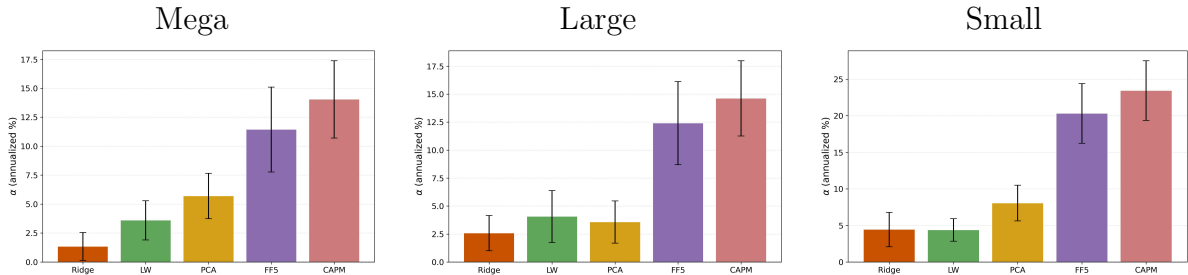


Figure 25: α Across Size Groups.

Annualized percentage α from regressions of UPSA on each benchmark using HAC standard errors (5 lags). All portfolios are scaled to 10% annualized volatility. The out-of-sample period is November 1981 to December 2022.

E.4 Robustness Across Rolling Windows

The main analysis uses a 120-month (10-year) rolling window for estimation. Here we examine sensitivity to this choice by varying the window length across 60, 240, and 360 months. Shorter windows allow the model to adapt more quickly to changing market conditions but increase estimation noise; longer windows provide more stable estimates but may miss structural shifts.

Figures 26 and 27 report the results. UPSA maintains the highest or near-highest Sharpe ratio across all window lengths. Notably, at the 240-month window, LW produces a negative Sharpe ratio.³⁵ The alpha estimates remain consistently positive for UPSA across

³⁵LW shrinks eigenvalues toward their population counterparts without regard to expected returns, leaving the portfolio exposed to noisy small PCs (see Section 4.3).

all specifications, indicating that its outperformance is robust to the investor's choice of estimation horizon.

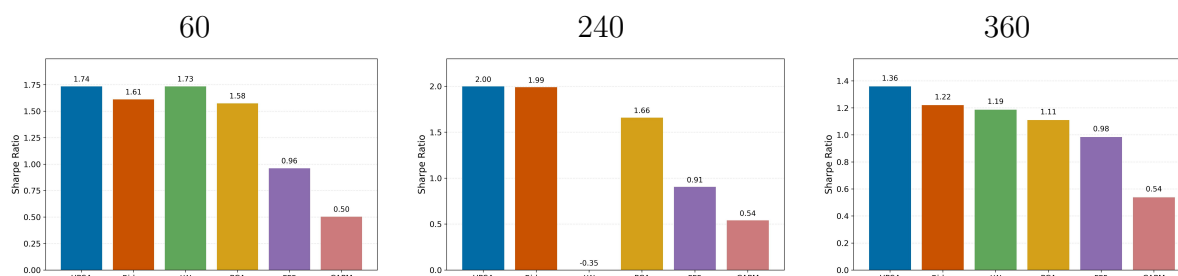


Figure 26: Annualized Sharpe Ratios Across Rolling Windows.

Annualized out-of-sample Sharpe ratios for UPSA and benchmarks across three rolling window lengths: 60 months (November 1976–December 2022), 240 months (November 1991–December 2022), and 360 months (November 2001–December 2022).

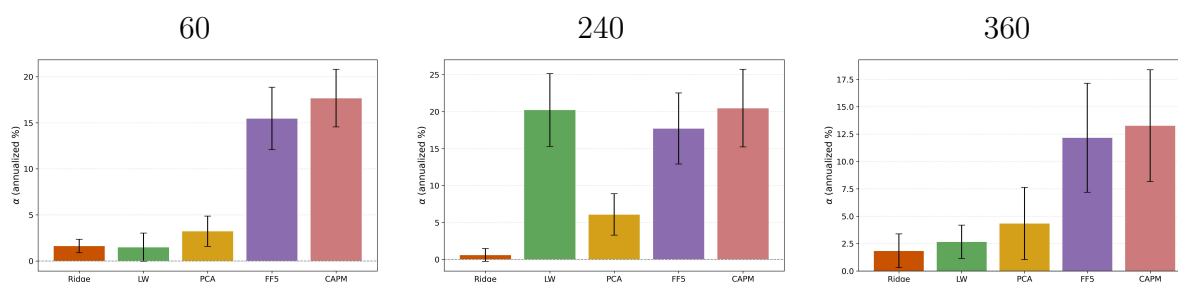


Figure 27: α Across Rolling Windows.

Annualized percentage α from regressions of UPSA on each benchmark using HAC standard errors (5 lags). All portfolios are scaled to 10% annualized volatility. Rolling window lengths: 60 months (November 1976–December 2022), 240 months (November 1991–December 2022), and 360 months (November 2001–December 2022).

E.5 Beta-Based Pricing Errors

The pricing accuracy analysis of Section 4.4 evaluates model-implied expected returns using the direct out-of-sample SDF representation in equation (18). A classical alternative is to express expected returns through a factor-model beta representation.

Consider a candidate SDF portfolio j with out-of-sample returns

$$R_{t+1}^j = (\bar{\pi}_t^j)' F_{t+1},$$

where $\bar{\pi}_t^j$ denotes the portfolio weights on the base assets and j indexes the shrinkage rule used to estimate the SDF. In a beta representation, the expected return of each test asset should satisfy

$$E_t^j[F_{i,t+1}] = \beta_{i,t}^j E_t[R_{t+1}^j],$$

where $\beta_{i,t}^j$ measures the exposure of asset i to the SDF portfolio.

A natural estimator of $\beta_{i,t}^j$ uses the in-sample second moment matrix $\bar{\Sigma}$ (defined in Section 2):

$$\beta_{i,t}^j = \frac{[\bar{\Sigma} \bar{\pi}_t^j]_i}{\bar{\pi}_t^{j'} \bar{\Sigma} \bar{\pi}_t^j}. \quad (53)$$

Model-implied expected returns are then computed as

$$\hat{E}_t^j[F_{i,t+1}] = \beta_{i,t}^j E_t^{OOS}[R_{t+1}^j],$$

and pricing accuracy is evaluated via the R^2 relative to the 45-degree line, following the methodology of Section 4.4.

Figure 28 reports scatter plots of model-implied versus realized out-of-sample expected returns using specification (53). Table 2 reports the corresponding R^2 statistics by factor theme. The results are broadly consistent with the direct SDF-based analysis in Figure 6. UPSA attains positive R^2 in 12 of 13 factor themes, while ridge, LW, and PCA remain

competitive in selected categories. FF5 and CAPM yield uniformly negative R^2 , confirming that UPSA's superior pricing performance is robust to the choice of evaluation methodology.

Quantitatively, UPSA achieves the highest overall beta-based R^2 at 73%, followed by Ridge at 63%, PCA at 55%, and LW at 52%. FF5 and CAPM produce negative R^2 values of -59% and -197% , respectively. At the theme level, all four shrinkage-based methods achieve strong pricing accuracy in profit growth ($R^2 \geq 86\%$) and remain competitive across most themes. The most challenging theme is value, where UPSA produces a near-zero R^2 (-1%), though the shrinkage-based SDFs substantially outperform FF5 and CAPM in this category. Overall, the beta-based evaluation confirms the conclusions of the direct SDF analysis: shrinkage-based methods deliver materially better out-of-sample pricing accuracy than traditional factor models, with UPSA achieving the strongest aggregate performance.

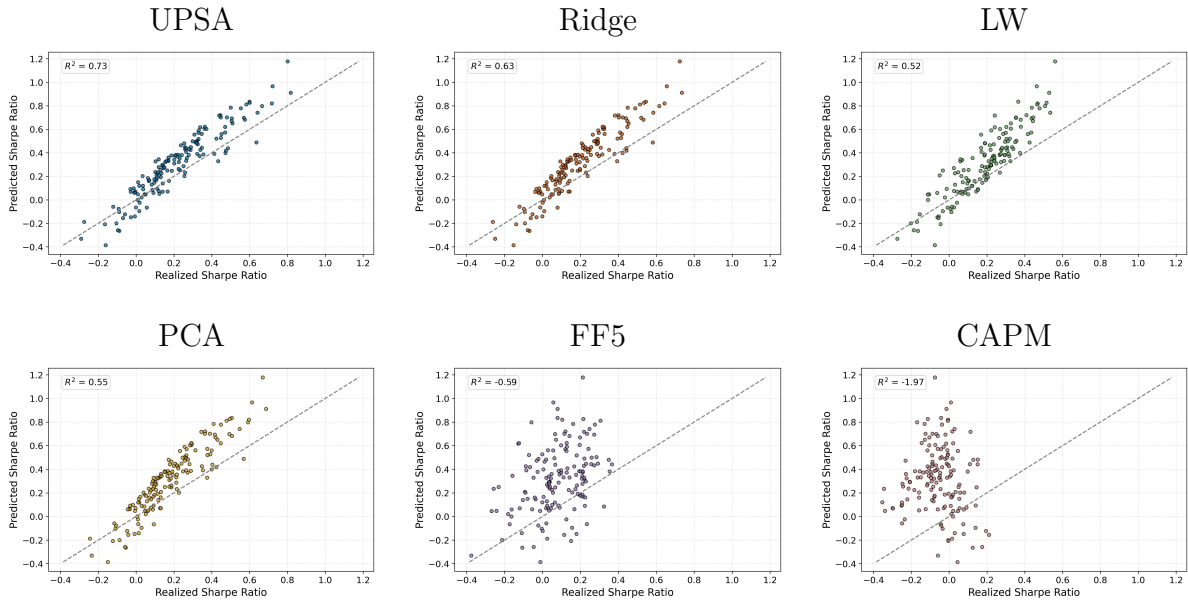


Figure 28: Beta-Based Out-of-sample Pricing Accuracy of Optimal Portfolios.

Scatter plots of beta-based out-of-sample Sharpe ratio predictions (vertical axis) versus realized out-of-sample Sharpe ratios (horizontal axis) using the in-sample beta specification (53). Test assets are the 153 JKP factors, standardized to unit volatility so that pricing errors are expressed in Sharpe ratio units. R^2 measures fit relative to the 45-degree line. Candidate SDFs are derived from UPSA, ridge, LW, PCA, FF5, and CAPM. The out-of-sample period is November 1981 to December 2022.

Table 2: Out-of-sample Pricing Accuracy By Factor Theme (In-Sample Second-Moment Beta)

This table evaluates the accuracy of SDF-based out-of-sample Sharpe ratio predictions for the 153 JKP factors, utilizing the in-sample second-moment beta representation. R^2 represents fit versus the 45-degree line within each JKP theme, as described in the analysis of Figure 28.

Factor Theme	UPSA	Ridge	LW	PCA	FF5	CAPM
Skewness	0.79	0.74	0.64	0.67	-0.22	-0.38
Profitability	0.04	-0.19	-0.04	-0.48	-2.74	-6.52
Low Risk	0.2	-0.03	0.71	-0.19	-0.87	-8.0
Value	-0.01	-0.45	0.36	-0.75	-0.99	-7.89
Investment	0.49	0.19	0.3	-0.02	-1.2	-6.0
Seasonality	0.74	0.71	0.44	0.68	-1.14	-0.75
Debt Issuance	0.87	0.74	0.13	0.64	-2.24	-3.1
Size	0.78	0.75	0.61	0.67	-0.87	-0.2
Accruals	0.9	0.81	0.48	0.74	-1.86	-3.58
Low Leverage	0.77	0.71	0.37	0.67	-0.73	-0.62
Profit Growth	0.86	0.9	0.92	0.9	-0.97	-0.86
Momentum	0.65	0.42	-0.76	0.25	-2.41	-7.9
Quality	0.35	0.09	-0.57	-0.16	-2.25	-4.39
All	0.73	0.63	0.52	0.55	-0.59	-1.97

E.6 Business Cycle Performance

To assess robustness across macroeconomic regimes, we normalize all portfolios to 10% annualized volatility over the full sample and compare performance during NBER recessions and expansions. Figure 29 reports the results.

UPSA is the most recession-resilient estimator across all three metrics. Its Sharpe ratio declines by 30% from expansion to recession (2.01 to 1.40), compared with 53% for ridge (1.77 to 0.83), 55% for LW, and 50% for PCA. The mean return panel reveals why: UPSA’s annualized return is virtually flat across regimes (18.8% in recessions vs. 19.3% in expansions), while all other methods earn substantially less in recessions. Ridge’s mean return falls from 16.2% to 13.2%, and CAPM turns negative (−11.2%). On the volatility side, while most SDFs experience realized risk above their 10% unconditional averages during

recessions, the increase for UPSA is notably more contained than for other shrinkage-based methods. Specifically, UPSA’s realized volatility rises to 13.5%, whereas the next-best alternative—LW—reaches 14.1%. This relative stability suggests that integrating over a continuum of shrinkage levels effectively hedges against the heightened parameter uncertainty and regime shifts characteristic of economic downturns, providing a smoother risk profile than methods tethered to a single point-selected penalty.

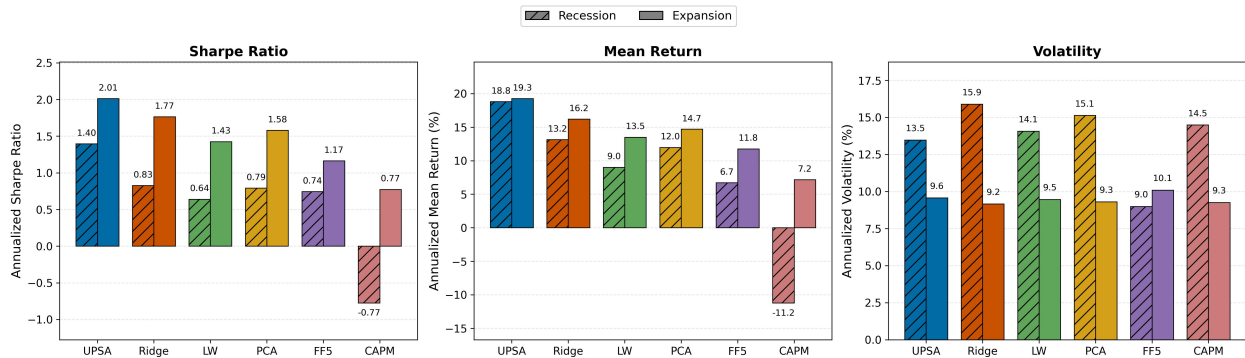


Figure 29: SDF Performance by NBER Regime.

Annualized Sharpe ratios, mean returns, and volatilities of candidate SDF portfolios during NBER recessions (hatched bars) and expansions (solid bars). All SDFs are normalized to 10% annualized volatility over the full out-of-sample period. 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 (48 recession months, 446 expansion months).

F Simulation: Supplementary Figures

F.1 Grid-Size Robustness

Figure 30 reports the mean out-of-sample Sharpe ratio as a function of the number of grid points in the simulation setting of Section 6.

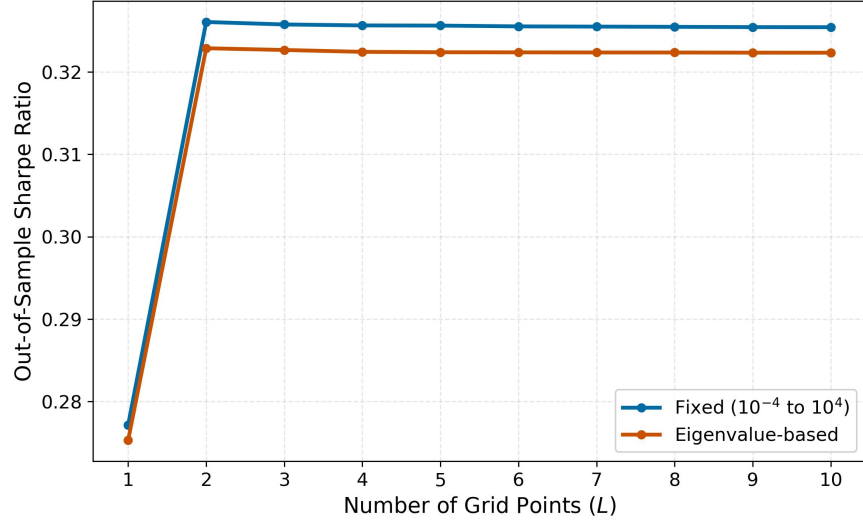


Figure 30: Grid-Size Robustness of UPSA.

Data are simulated using the data-generating process (23)–(24) for $N = 150$ assets and $T = 600$ observations. The figure reports the mean out-of-sample Sharpe ratio across simulations as a function of the number of grid points L . Two grid types are shown: a fixed logarithmic grid spanning 10^{-4} to 10^4 and an eigenvalue-based grid that adapts its range to the empirical eigenvalue distribution.

F.2 Time-Varying Simulation: Additional Figures

Figures 31, 32, and 33 display the ensemble weights, rolling turnover, and penalty distributions for a representative Markov-switching simulation path.

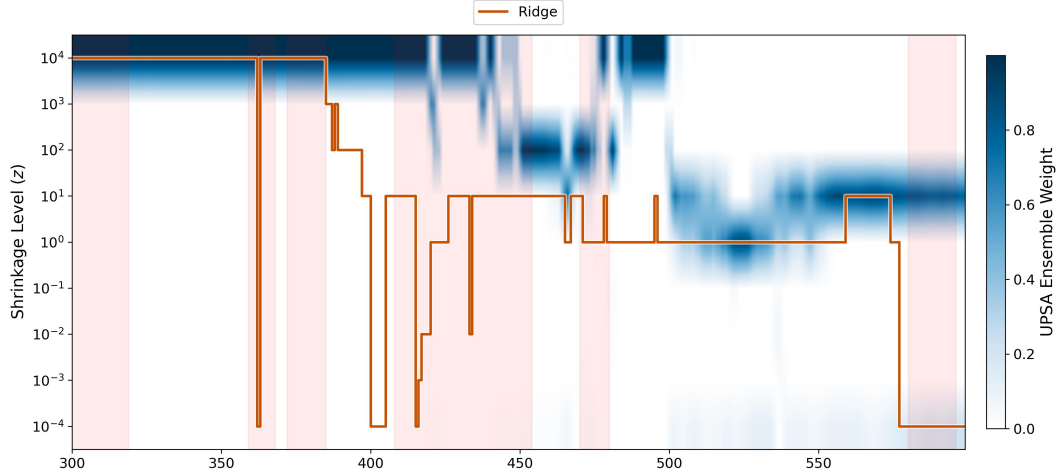


Figure 31: UPSA Ensemble Weights and Ridge Penalty Over Time.

Time series of the ridge penalty selected by leave-one-out cross-validation (orange line) overlaid on UPSA ensemble weights across the shrinkage grid, for a representative Markov-switching simulation. UPSA weights are determined using Lemma 2. Pink shading marks periods in Regime 2, where the high-variance PCs earn a larger premium. Data are simulated from the data-generating process (26) with $N = 150$ assets, $T = 600$ observations, and a rolling window of 300.

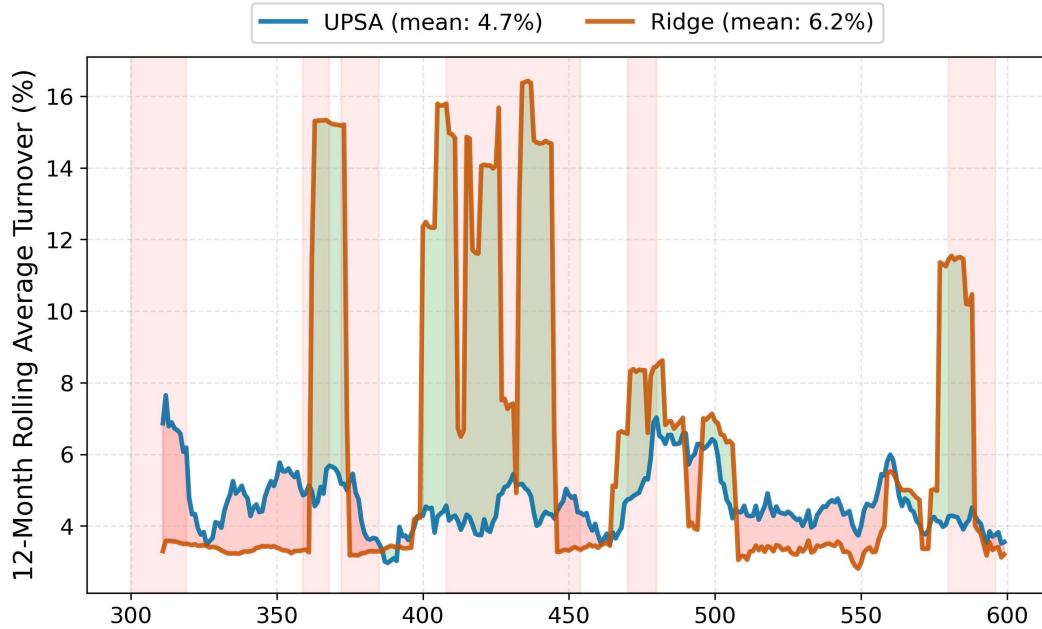


Figure 32: Rolling Turnover for UPSA vs. Ridge Under Regime Switching.

12-month rolling average turnover for UPSA and ridge from a representative Markov-switching simulation. Ridge penalties are selected using leave-one-out cross-validation, and UPSA weights are determined using Lemma 2. Pink shading marks periods in Regime 2, where the high-variance PCs earn a larger premium. Green shading marks periods in which ridge turnover exceeds UPSA. Data are simulated from the data-generating process (26) with $N = 150$ assets, $T = 600$ observations, and a rolling window of 300.

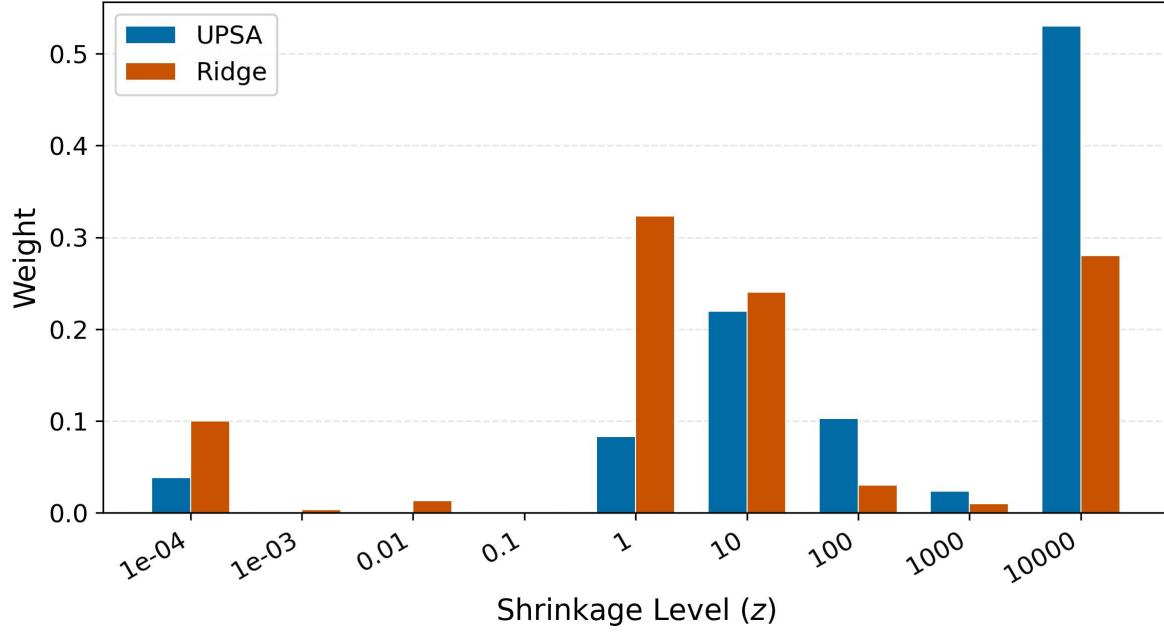


Figure 33: Distribution of Penalty Weights, Markov-Switching Simulation.

Histogram of ridge penalties selected by cross-validation compared with the average UPSA ensemble weights across the same grid, for a representative Markov-switching simulation. UPSA weights are determined using Lemma 2. Data are simulated from the data-generating process (26) with $N = 150$ assets, $T = 600$ observations, and a rolling window of 300.