Shrinkage Alignment in High-Dimensional Portfolios

Shikun (Barry) Ke Mohammad (Mo) Pourmohammadi*

This draft: November 3, 2025

Abstract

We study how shrinkage affects portfolio efficiency when the number of assets approaches or exceeds the sample size. Standard methods, such as ridge, impose uniform shrinkage, treating all assets as ex-ante identical and creating inefficiency when profitability is heterogeneous. Empirically, this "one-size-fits-all" design produces a hump-shaped relationship between model complexity and out-of-sample Sharpe ratios: adding assets can paradoxically reduce performance. We introduce shrinkage alignment, showing that efficiency requires matching shrinkage strength to each asset's true profitability. Building on this insight, we propose Sharpe Ratio Shrinkage (SRS)—a data-driven approach that aligns shrinkage intensity with empirical Sharpe ratios. SRS outperforms conventional methods under profitability heterogeneity and restores the virtue of complexity in high-dimensional portfolio construction.

JEL Code: G11, G12

Keywords: Shrinkage alignment, portfolio construction, parameter uncertainty, virtue of complexity

*Yale School of Management. Email: barry.ke@yale.edu and mo.pourmohammadi@yale.edu.

We thank Bryan Kelly, Alp Simsek, Eduardo Davila, Will Goetzmann, Toby Moskowitz, Nick Barberis, Theis Jensen, Kaushik Vasudevan, Kangying Zhou, Tim Christensen, Stefano Giglio, and Yale Finance Lunch for their comments.

1 Introduction

The asset pricing research community has generated a staggering variety of return predictors. Advances in data science, machine learning, and alternative data have made it possible to discover thousands of potentially useful trading signals and massively augment the investable asset space. From an investor's standpoint, the rapid expansion of available trading signals presents the challenge of constructing an effective portfolio, particularly when limited training data constrain the estimation of portfolio weights. This naturally leads to a fundamental question: How can an investor build a portfolio from a large cross-section of assets to achieve the highest possible out-of-sample performance?

When the number of assets approaches or exceeds the number of available training observations, portfolio estimation necessitates the use of shrinkage techniques. In this paper, we emphasize the concept of *shrinkage alignment*—the alignment between the strength of shrinkage (i.e., the prior belief about the magnitude of optimal portfolio weights) and the true profitability of the underlying assets. This alignment governs how the portfolio balances the trade-off between the marginal empirical gains from including additional assets and the marginal costs of estimation uncertainty. Empirically, we show that many standard shrinkage methods—such as ridge and LASSO—often fail to strike this balance, leading to deteriorating portfolio performance as the asset universe expands. These methods employ a "one-size-fits-all" shrinkage design, implicitly imposing an equal-prior assumption that treats all assets as ex-ante identical. When the underlying assets exhibit strong heterogeneity, such misalignment between the imposed prior and the assets' true profitability introduces inefficiency—particularly in high-dimensional settings where the choice of prior retains a non-vanishing influence on estimation outcomes.

As a concrete illustration, consider a mean–variance investor constructing an optimal portfolio of N assets based on T observations of asset returns F_t. The assets have an unobservable population mean vector μ and a covariance matrix Σ. A well-known result states that the ordinary least squares (OLS) coefficient from regressing a constant vector on asset returns

$$\widehat{w}^{OLS} = \left(\sum_{t=1}^{T} F_t F_t' \right)^{-1} \sum_{t=1}^{T} F_t, \tag{1}$$

yields a consistent estimator of the true optimal portfolio when $N \ll T$.[1] However, this consistency deteriorates as N grows, and the resulting portfolio weights become unstable. To address this issue, many modern portfolio estimation methods incorporate shrinkage

[1] See Theorem 1 in Britten-Jones (1999).

directly into equation (1). For example, applying ridge shrinkage yields the estimator

$$\widehat{w}^{ridge} = \left(zI + \sum_{t=1}^{T} F_t F_t' \right)^{-1} \sum_{t=1}^{T} F_t, \tag{2}$$

where z controls the strength of shrinkage towards zero. In this estimator, the sample second moment matrix is shrunk toward the identity matrix, with the hyperparameter z controlling the degree of shrinkage.

However, this specification implies that all assets are shrunk uniformly, and such a "one-size-fits-all" approach may be suboptimal when the underlying assets exhibit strong heterogeneity. From a Bayesian perspective, the shrinkage term zI corresponds to a prior belief that the optimal portfolio weights $w^* = \Sigma^{-1}\mu$ are identical across assets, which may misalign with the true configurations of Σ and μ. For instance, consider a universe where some assets exhibit substantially higher Sharpe ratios than others. A uniform ridge penalty would shrink all weights toward zero by the same proportion, penalizing high-Sharpe assets as heavily as low-Sharpe ones. Ideally, the shrinkage intensity should adapt to each asset's ex-ante signal-to-noise ratio; it should shrink less for more informative (profitable) assets and more for noisier ones. While the influence of the prior is negligible when $N \ll T$, in high-dimensional settings, such misalignment between the shrinkage structure and the true data-generating process can induce substantial inefficiencies in out-of-sample portfolio performance.

In the first part of the paper, we empirically demonstrate that misalignment arising from the equal-shrinkage design used in many standard portfolio methods can lead to substantial inefficiency when the underlying assets exhibit strong heterogeneity. Our empirical analysis employs the data-mined strategies constructed in Chen et al. (2022) and Chen and Dim (2023), which are generated by randomly combining firm-level accounting variables or returns through ratios or first differences. This construction provides a purely data-driven approach to discovering trading strategies without economic priors, thereby producing strategies with pronounced heterogeneity in expected returns.

We estimate efficient portfolios using an increasing number of assets (data-mined strategies) N, while keeping the training window length T fixed. To clearly illustrate the potential inefficiency, assets are added in descending order of their in-sample Sharpe ratios. We evaluate several shrinkage-based portfolio estimators with optimally tuned hyperparameters, including ridge, LASSO (Ao et al., 2019), PCA with elastic net regularization (Kozak et al., 2020), and the Universal Portfolio Shrinkage Approximator (UPSA) (Kelly et al., 2024b), among others.

Across all methods, we observe a paradoxical hump-shaped pattern in out-of-sample

portfolio performance: adding more assets initially improves performance, but beyond a certain point, it begins to deteriorate—even when the additional assets are individually highly profitable (i.e., exhibit high Sharpe ratios). This eventual decline in out-of-sample performance suggests an inefficiency in these portfolio methods.

In the second part of the paper, we demonstrate that such inefficiency is caused by shrinkage misalignment. We theoretically analyze the performance of a ridge portfolio in a stylized setting where assets share an identity covariance matrix but differ in their expected returns. In this environment, the ridge portfolio effectively imposes a prior that all assets have identical expected returns and applies uniform shrinkage across portfolio weights, thereby misaligning with the true data-generating process. We explicitly derive the portfolio's expected out-of-sample Sharpe ratio and show that, as more assets with lower (but still positive) expected returns are added, the out-of-sample performance of the ridge portfolio can decline—consistent with the empirical pattern.[2]

Aligning the strength of shrinkage with each asset's true risk-adjusted return is crucial for achieving strong out-of-sample portfolio performance in high-dimensional settings. When the investor does not observe the true risk-adjusted return of each asset, the in-sample Sharpe ratio provides a natural proxy. Building on this idea, we introduce *Sharpe Ratio Shrinkage* (SRS) as a simple and effective remedy for the shrinkage misalignment problem. SRS shrinks portfolio weights in proportion to the inverse of each strategy's empirical squared Sharpe ratio, estimated from the training sample. Intuitively, assets with higher in-sample Sharpe ratios are more likely to be genuinely profitable and therefore receive less shrinkage, whereas assets with lower Sharpe ratios are shrunk more aggressively. From a Bayesian perspective, SRS leverages information from realized data to construct a data-driven prior on the optimal portfolio weights w^*, representing an Empirical Bayes approach that introduces asset-specific shrinkage. In other words, SRS shrinkage encodes a priori skepticism that assets with small (in-sample) Sharpe ratios will have large portfolio weights in the out-of-sample efficient portfolio.

Empirically, we show that SRS substantially improves out-of-sample portfolio performance relative to standard shrinkage methods, particularly in environments with strong cross-sectional heterogeneity. Furthermore, when applied to the principal components of asset returns, SRS consistently outperforms conventional ridge shrinkage.

Although our focus is on portfolio construction, the concept of shrinkage alignment also extends naturally to predictive regressions in high-dimensional settings. In particu-

[2]While our theoretical derivation assumes an identity covariance matrix, in Section 4.1 we conduct simulations that relax this assumption by allowing correlated returns, and we find patterns qualitatively similar to the empirical results.

lar, our analysis clarifies a key aspect of the "virtue of complexity" (VOC) phenomenon documented in the recent financial machine learning literature, which finds that complex, high-dimensional models often outperform simpler, low-dimensional ones. We emphasize that the empirical success of such complex models crucially depends on how well the employed shrinkage method aligns with the structure of the empirical environment.

When the predictive environment features ex-ante homogeneous predictors—such as the random Fourier features studied in Kelly et al. (2022), Kelly et al. (2024a), and Didisheim et al. (2024)—standard ridge shrinkage is sufficient to realize the VOC. In contrast, when predictors exhibit substantial heterogeneity, shrinkage alignment becomes essential to recover the performance gains of complex models. In this sense, our analysis also provides insight into the behavior of high-dimensional models when different predictors contain varying levels of noise[3]. We show that in such heterogeneous environments, ridge shrinkage is no longer optimal, while a shrinkage alignment approach substantially outperforms standard ridge regression and restores the virtue of complexity.

Literature. Our paper relates to three main strands of the literature. The first concerns efficient portfolio estimation (Ledoit and Wolf, 2004; Kan and Zhou, 2007; DeMiguel et al., 2009; Tu and Zhou, 2011; Ao et al., 2019; Kozak et al., 2020; Kan et al., 2024; Yuan and Zhou, 2024; Kelly et al., 2024b). We contribute to this literature by examining the role of shrinkage methods in high-dimensional portfolio estimation. Our results highlight a key challenge faced by many portfolio construction approaches when applied in high-dimensional environments with heterogeneous signal quality: standard shrinkage designs often misalign with the empirical structure of the data. We show that aligning shrinkage with the empirical environment is crucial for improving out-of-sample performance. To this end, we propose the Sharpe Ratio Shrinkage (SRS) method, which provides a simple and effective remedy for shrinkage misalignment and consistently outperforms existing approaches in settings with pronounced cross-sectional heterogeneity.

Second, our paper contributes to the growing literature on machine learning methods in asset pricing, particularly the "virtue of complexity" (VOC) phenomenon (e.g., Kelly et al., 2022, 2024a; Didisheim et al., 2024; Bryzgalova et al., 2024; Cong et al., 2025). We emphasize that whether a complex model delivers superior empirical performance depends critically on how the chosen shrinkage method aligns with the complexity of the estimation environment.

[3] Cartea et al. (2025) present a framework in which adding more predictors uniformly increases the noise level across all predictors, and use this to argue against the virtue of complexity. Kelly and Malamud (2025) point out that such a design offers limited insight into how complex models behave in realistic settings. We consider a more plausible environment where, as investors discover additional signals, newer ones tend to contain less predictability (i.e., more noise) than earlier ones. This framework captures the "diminishing returns" feature of data-driven signal discovery, in which investors first exploit strong predictors and subsequently expand to weaker ones; see Ke (2023).

In many machine learning frameworks, model complexity arises from the construction of a large number of ex-ante homogeneous predictive units—such as neurons in neural networks, basis functions in kernel methods, or splits in tree-based models. In such settings, a "one-size-fits-all" shrinkage design is often appropriate and can yield the desired VOC pattern. By contrast, our empirical analysis identifies cases in which the VOC fails under standard ridge shrinkage but is restored once the shrinkage intensity is aligned with the empirical characteristics of the data. This highlights the broader importance of shrinkage alignment in explaining when and why complex models outperform simpler ones.[4]

Finally, our paper contributes to the literature on high-dimensional learning in financial markets (Martin and Nagel, 2022; Da et al., 2024; Shen and Xiu, 2025; Kelly et al., 2025), by highlighting the potential inefficiencies that arise from misalignment between the true data-generating process and the applied shrinkage method. This issue is closely related to the broader challenge of Bayesian learning in high dimensions. As shown by Diaconis and Freedman (1986) and Ritov et al. (2014), in high-dimensional settings, Bayesian learning can become inconsistent because the prior may "swamp" the information contained in the data. The inefficiency in portfolio performance induced by shrinkage misalignment can be viewed as a manifestation of this "prior swamping" effect. We demonstrate that incorporating information from observed data to align the prior offers a simple and effective remedy for such inefficiency.

2 Empirical Illustration of Shrinkage Misalignment

We begin by demonstrating the inefficiency caused by shrinkage misalignment in an empirical experiment. In our empirical exercise, we gradually increase the number of assets used to build a portfolio while keeping the number of training data fixed. To clearly illustrate the inefficiency, we start with assets with highest profitability and expand into assets with lower profitability. We show that the out-of-sample Sharpe ratio of the ridge portfolio is not monotonically increasing in the number of assets used. While using more assets initially raises portfolio performance, the marginal estimation cost eventually outweighs the marginal benefit from new assets, and the Sharpe ratio declines—even though all assets are individually profitable or have significant alphas relative to the existing portfolio.

Data and Portfolio Estimation Problem. Our asset space is the data-mined strategy universe built in Chen et al. (2022) and Chen and Dim (2023) combined with the anomaly strategies built in Chen and Zimmermann (2021). These strategies can be viewed as the

[4] Kelly and Malamud (2025) discuss the alignment between the eigenvalues of the covariance matrix and the true predictive coefficients. In contrast, our analysis focuses on the alignment between the true predictive coefficients and the shrinkage strength.

output of unconstrained data mining without economic guidance and therefore contains significant heterogeneity in the underlying true profitability.[5] To place the strategies on a comparable footing, we construct volatility-managed versions of all strategies following Moreira and Muir (2017), scaling returns by their trailing 12-month standard deviation.

In each month t, we consider a static portfolio construction problem for N strategies, with the objective of maximizing the quadratic utility

$$\max_{w_t} \quad E[w_t' F_{t+1}] - \frac{1}{2} Var[w_t' F_{t+1}] \tag{3}$$

based on only T past realizations $\{F_\tau\}_{\tau=t-T}^{t}$. This captures a key feature of portfolio estimation in reality: while investors can use many assets to form portfolio, the historical sample available for estimation remains limited. In our main analysis, we estimate portfolios using 360-month rolling windows. Our baseline estimator is the "maximum Sharpe ratio regression" with a ridge penalty (MSRR-ridge following Kelly and Xiu (2023), Didisheim et al. (2024)), given by

$$\widehat{w}_t^{ridge} = \left(zI + \frac{1}{T} \sum_{\tau=t-T}^{t} F_\tau F_\tau' \right)^{-1} \frac{1}{T} \sum_{\tau=t-T}^{t} F_\tau. \tag{4}$$

Intuitively, this regression finds the weights w such that $w' F_t$ is as close as possible (in the L_2 sense) to a positive constant return, equivalent to maximizing the Sharpe ratio.

We study the behavior of the ridge portfolio as the number of assets N increases. To clearly illustrate the potential inefficiency, we expand the strategy set in descending order of historical performance. If the ridge portfolio estimator in equation (4) can efficiently incorporate additional strategies, we should observe a "virtue of complexity" pattern—where increasing the number of assets N consistently improves out-of-sample portfolio performance.

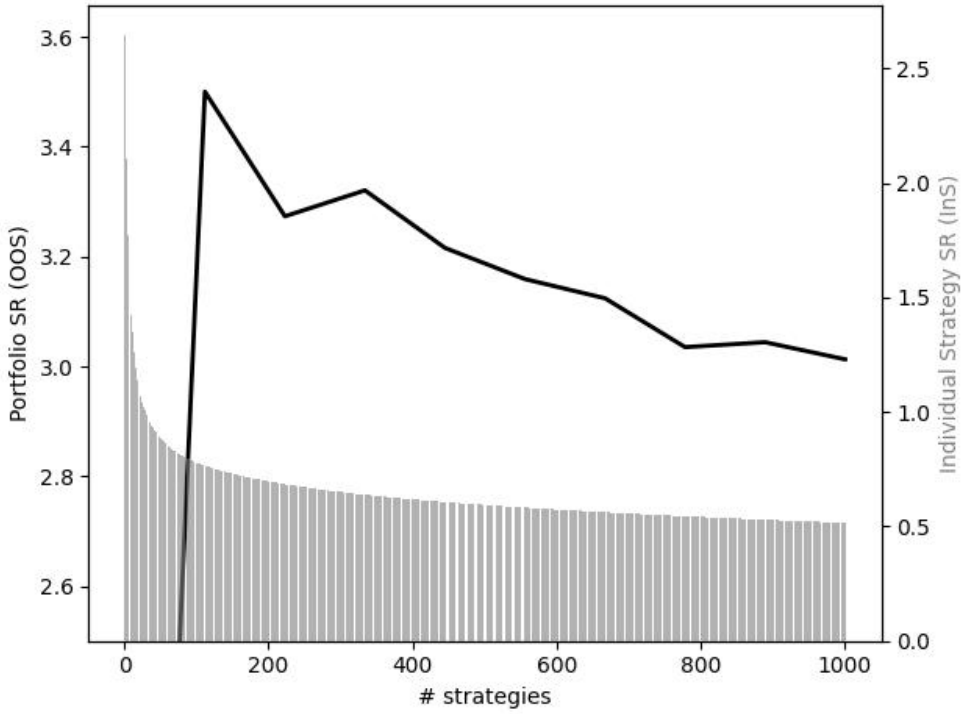
2.1 Main Empirical Results

Figure 1 presents the results. In both panels, the gray bars display the average in-sample Sharpe ratios of individual strategies.[6] By construction, these Sharpe ratios decline with rank, since strategies are added in descending order of in-sample performance. Nevertheless, even the lowest-ranked strategies in the figure exhibit annualized Sharpe ratios above 0.5—comparable to the benchmark value of 0.56 for published anomalies in the Chen–Zimmermann dataset.[7] Hence, all strategies in this experiment are highly profitable.

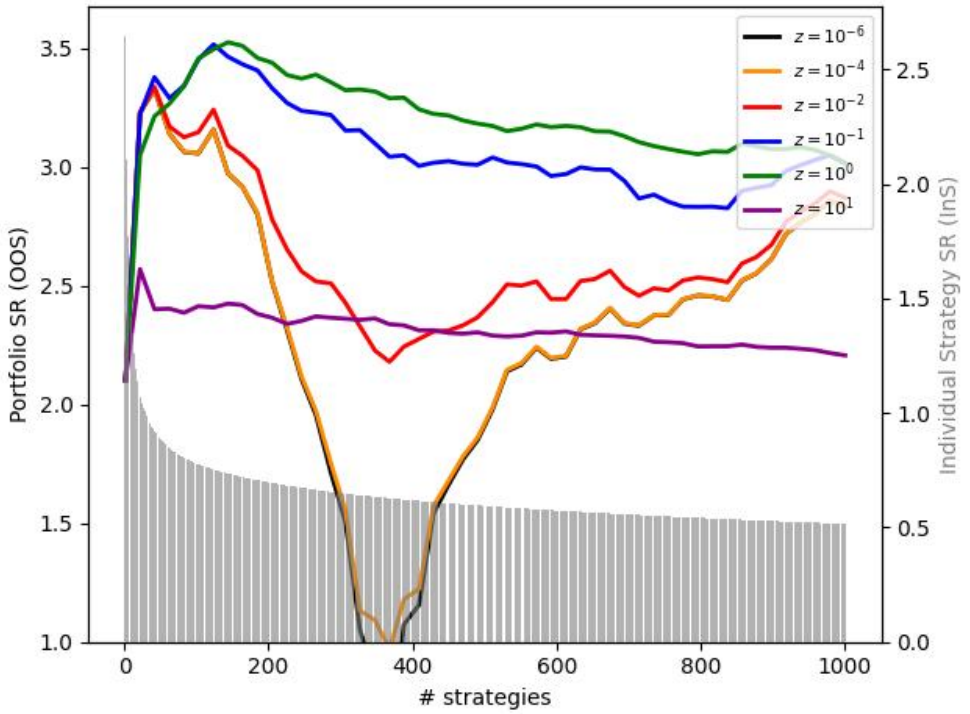
[5] See Chen and Zimmermann (2021), Chen et al. (2022), and Chen and Dim (2023) for full details on the construction and availability of these strategies.

[6] In-sample Sharpe ratios are computed within each rolling window and then averaged across all windows.

[7] See Chen and Zimmermann (2021).



(a) Cross Validated z



(b) Fixed z

Figure 1: Out-of-Sample Sharpe Ratio of MSRR-ridge Portfolios.

Note. This figure shows the out-of-sample Sharpe ratio of MSRR–ridge portfolios as the number of strategies increases in order of their in-sample Sharpe ratios. Panel (a) reports results when the ridge penalty is selected via leave-one-out cross-validation from a grid of shrinkage values $z \in \{10^{-7}, 10^{-6}, \ldots, 10^{3}\}$. Panel (b) reports results for fixed levels of ridge shrinkage z. The gray bars display the average in-sample Sharpe ratios of individual strategies (right axis). Results are computed using a 360-month rolling window over the out-of-sample period January 1994 to December 2022.

Panel (a) of Figure 1 plots the out-of-sample Sharpe ratio of the MSRR–ridge portfolio, where the ridge parameter is selected via cross-validation. A striking non-monotonic pattern emerges: portfolio performance initially improves as additional strategies are included, peaks at an out-of-sample Sharpe ratio of 3.5 when roughly the top 150 strategies are used, and then declines steadily. By the time 1,000 strategies are incorporated, the Sharpe ratio falls to about 3.0—a decline of nearly 20%. Importantly, this deterioration occurs even though each newly added strategy continues to exhibit a high in-sample Sharpe ratio. Panel (b) of Figure 1 further shows that this non-monotonicity in portfolio performance persists across all fixed levels of ridge shrinkage z.

This hump-shaped relationship between the number of strategies and portfolio performance is our central empirical finding. It highlights the limits of the "virtue of complexity" in our setup: adding more individually profitable strategies does not always translate into higher out-of-sample efficiency when the ridge portfolio (4) is used.

Are additional strategies redundant? The counterintuitive decline in the Sharpe ratio as more strategies are added might suggest that the newly included strategies are redundant—that is, their return predictability is already spanned by the earlier ones. To investigate this possibility, we perform an alternative exercise in which strategies are added sequentially based on the t-statistics of their alphas relative to the previously constructed efficient portfolio, measured *out of sample*. This procedure is designed to prioritize strategies that provide large, unspanned sources of return. However, it is important to note that such portfolios are not implementable in real time: determining whether a strategy adds value requires knowledge of its *out-of-sample* alpha with respect to the existing portfolio, which introduces look-ahead bias. Nevertheless, this exercise sheds light on an important conceptual question: even with perfect hindsight—when an investor knows that an asset offers return predictability orthogonal to her existing portfolio—does incorporating it necessarily improve portfolio performance?

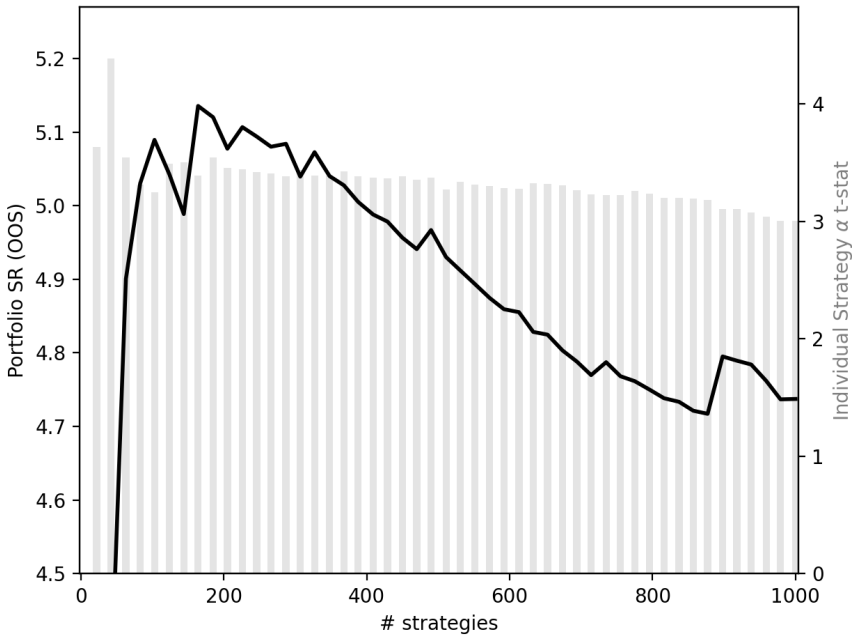


Figure 2: Adding Strategies Based on Alpha t-statistics

Note. This figure shows the out-of-sample Sharpe ratio of the MSRR–ridge portfolio when strategies are added in order of their out-of-sample alphas. The gray bars display the alpha t-statistics of individual strategies with respect to the existing portfolio's out-of-sample returns (right axis). Specifically, we construct the MSRR–ridge portfolio using a given set of strategies and identify additional strategies not used in its construction that exhibit high out-of-sample alphas relative to the portfolio's returns. The black line plots the corresponding out-of-sample Sharpe ratio of the portfolio. The ridge penalty is selected via leave-one-out cross-validation from a grid of shrinkage values $z \in \{10^{-7}, 10^{-6}, \ldots, 10^{3}\}$. Results are computed using a 360-month rolling window over the out-of-sample period January 1994 to December 2022.

Figure 2 shows that the answer is no. The gray bars display the t-statistics of the alphas for each newly added strategy, measured relative to the out-of-sample returns of the existing portfolio.[8] All newly added strategies exhibit alpha t-statistics above 2.0, indicating that each provides statistically significant, incremental profitability not spanned by the current portfolio.

However, when we sequentially incorporate these strategies into portfolio construction, performance follows a distinctly non-monotonic trajectory. Initially, adding new strategies enhances the portfolio's Sharpe ratio, but beyond a certain point, the gains plateau—and eventually, further inclusion of strategies, despite their significant alphas, leads to a decline in overall performance. This pattern suggests that the eventual deterioration in performance cannot be attributed to redundancy among the newly added assets.[9]

[8] At each iteration, we add the strategy with the highest alpha t-statistic relative to the current portfolio's out-of-sample return. Consequently, the reported t-statistics correspond to different portfolios across iterations.

[9] Figure 14 in the Appendix presents a complementary analysis in which strategies are sorted by their

Random Expansion of Assets. What is the reason behind the inefficiency, i.e., the hump-shaped pattern in Figure 1? We show that it arises from a misalignment between the shrinkage strength and underlying asset profitability: the decreasing ordering of asset profitability in Figure 1 misaligns with the shrinkage strength in the ridge portfolio (4), which imposes equal shrinkage strength across assets.

To demonstrate this, we expand the asset space by randomly including strategies in the consideration set. This random ordering effectively ensures that all assets used to construct the portfolio have ex-ante the same profitability. Figure 3 shows that when strategies are added in random order, portfolio Sharpe ratios increase monotonically. In this setting, more assets always help the ridge portfolio, but the resulting curve remains below the decreasing-order case, and the highest Sharpe ratio is achieved with a subset of strategies rather than the full set. From this exercise, we conclude that ridge portfolio estimation fails to generate an efficient portfolio out-of-sample when the underlying asset space contains strong heterogeneity in profitability.

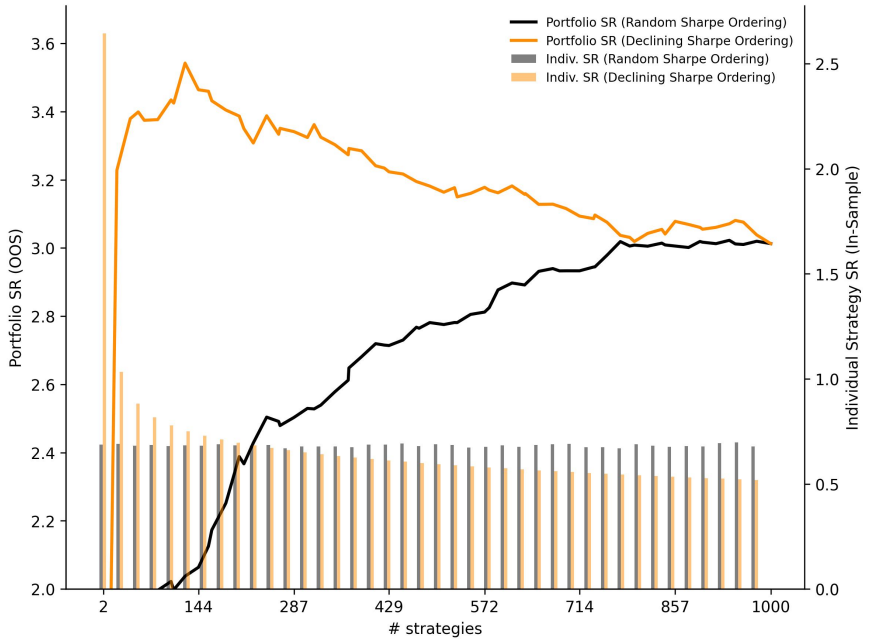


Figure 3: Portfolio Performance with Decreasing vs. Random Sorting of Strategies. This figure plots the out-of-sample Sharpe ratio of MSRR-ridge portfolios as the number of strategies increases. The orange line shows when strategies are added in decreasing order of in-sample Sharpe ratio (as in Figure 1), and the orange bars show the in-sample Sharpe ratio of n-th strategy. The black line shows the average performance when strategies are included in random order within the top 1,000 strategies, and the gray bars show the in-sample Sharpe ratio of n-th strategy averaged across random draws. The ridge penalty is selected via leave-one-out cross-validation from a grid of shrinkage values $z \in \{10^{-7}, 10^{-6}, \ldots, 10^{3}\}$. Results are computed using a 360-month rolling window over the out-of-sample period January 1994 to December 2022.

full-sample Sharpe ratios. We again observe a decline in out-of-sample performance when too many strategies are included, even though each individual strategy exhibits a high full-sample Sharpe ratio.

The monotonic increase in Sharpe ratios under random sorting offers insight into the "virtue of complexity" (VOC) phenomenon. Didisheim et al. (2024) show that portfolios constructed from Random Fourier Feature (RFF) factors—nonlinear transformations of stock characteristics—exhibit steadily improving out-of-sample Sharpe ratios as model complexity increases. Our results suggest that the VOC documented in Didisheim et al. (2024) arises from an implicit alignment between RFF factor generation and ridge shrinkage. At each level of model complexity, RFFs generate factors that are, *ex ante*, homogeneous in their predictive content. Consequently, the "one-size-fits-all" shrinkage design of ridge regression aligns naturally with this property of the empirical factors, leading to superior out-of-sample performance. In contrast, when the underlying assets or predictors exhibit substantial heterogeneity, ridge shrinkage can induce inefficiencies in out-of-sample portfolio performance.[10]

2.2 Shrinkage Misalignment in Other Portfolio Methods

In addition to the MSRR-ridge portfolio, we also consider a wide range of alternative portfolio construction methods suitable in a high-dimensional setting, including the universal portfolio shrinkage method (Kelly et al. (2024b)), the nonlinear shrinkage estimator (Ledoit and Wolf (2020)), portfolio construction based on Principal Component Analysis (PCA) (Kozak et al. (2018)), PCA factors with elastic net (Kozak et al. (2020)), portfolio construction based on LASSO regression (Ao et al. (2019)), and the naive 1/N portfolios (DeMiguel et al. (2009)). We briefly describe each portfolio construction method in Appendix A.1.

Figure 4 delivers a clear, robust finding similar to Figure 1: in every portfolio construction method we consider, the out-of-sample Sharpe ratio falls as more strategies are included—even when each strategy boasts a high individual Sharpe. This consistency across diverse methodologies highlights the inherent inefficiency of aggregating strategies with heterogeneous profitability amid estimation complexity, a core theme of our paper. This inefficiency is related to the equal-prior assumption embedded in these methods. For example, the LASSO portfolio proposed in Ao et al. (2019) corresponds to having a Laplace prior of equal scale hyperparameter across all asset weights, the elastic net portfolio in Kozak et al. (2020) corresponds to having a mixture Normal-Laplace prior on asset weights with the same hyperparameters across assets, and the UPSA in Kelly et al. (2024b) corresponds to using a same Gaussian mixture prior on all asset weights. All of these priors (and shrinkage design) misalign with the strong heterogeneity in the underlying assets in our specific empirical setting, which leads to inefficiency in out-of-sample performance.

[10] In Appendix A.3, we replicate the MSRR–ridge exercise using RFF factors, expanding them in decreasing order of their in-sample Sharpe ratios. We find that individual RFF factors do display heterogeneity in their Sharpe ratios, which leads to a mild decline in portfolio performance; however, the effect is small on average. The fact that RFFs produce less correlated strategies likely contributes to the persistence of the VOC, as discussed further in Section 4.1.

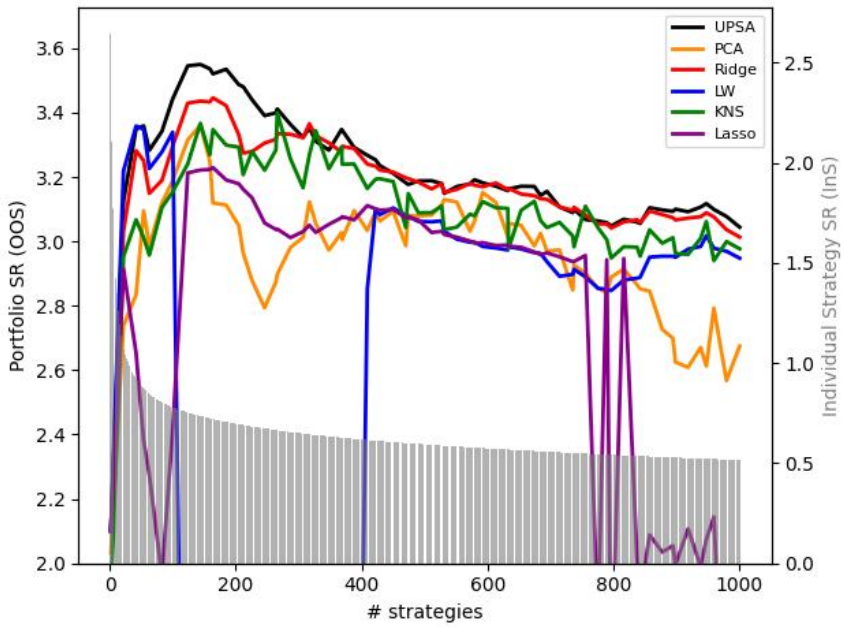


Figure 4: Different Portfolio Construction Methodologies

Note. This figure shows the out-of-sample portfolio performance obtained using alternative portfolio construction methods. The gray bars display the average in-sample Sharpe ratios of individual strategies (right axis), computed for each rank across rolling windows. Details on hyperparameter selection are provided in Appendix A.1. Results are based on a 360-month rolling window over the out-of-sample period January 1994 to December 2022.

2.3 Robustness of the Main Result

We perform a batch of robustness checks for our main results in Appendix A.

Rolling window length. First, we note that finite observations in the training data play an important role in driving the out-of-sample portfolio performance. Figure 12 shows the decline in MSRR-ridge portfolio performance remains robust when we use other rolling window lengths of $T = 120, 180$ and 240 months.

Factor structure in assets. One might suspect that the decline simply reflects a strong factor structure in the strategy returns. Once they are captured by the first few strategies, the additional strategies might add nothing but purely idiosyncratic noise. This is somewhat related to the alpha results of Figure 2, basically assets can add anything in addition to strong factor etc. We show, however, that this is not the case. First, Figure 11 shows that the factor structure in our data is not very strong: it takes more than 200 principal components to explain up to 90% variations in strategy returns, and most strategies are not highly correlated. Second, we also show that the result is robust after removing common factors from the strategies. Figure 13 shows the result for building an efficient portfolio

using the residuals after removing the top 5, 10, 20, and 50 principal components from the strategy set. We notice that even after removing principal components, the residuals carry positive and significant Sharpe ratios. More importantly, the non-monotonic pattern in the out-of-sample portfolio Sharpe ratio is similar—if not stronger—when principal components are removed. This result motivates us to write a model without a strong factor structure in Section 3 to explain our empirical findings.

Ex-post optimal shrinkage hyperparameter. In the main analysis, we use cross-validation to select the ridge shrinkage hyperparameter z. One might suspect that the decline in portfolio performance arises from the increasing difficulty of identifying the optimal hyperparameter as the number of strategies grows. Figure 14 shows that the decline in the Sharpe ratio persists even when using the *ex post* optimal value of z. This finding indicates that hyperparameter selection is not driving the result: even an investor with perfect hindsight about the best shrinkage parameter would still observe declining portfolio performance as additional strategies are included.

Alphas. Finally, Figure 15 shows that the portfolio using 100 strategies has a significantly positive alpha with respect to the portfolio using 1000 strategies, with the highest individual Sharpe ratios in-sample. This result again highlights our main finding: using more strategies to construct a portfolio can hurt portfolio performance.

3 A Simple Model of Shrinkage Misalignment

In this section, we study a simple model that formalizes the idea of shrinkage misalignment and sheds light on the empirical patterns we document in Section 2. We theoretically characterize the behavior of the MSRR–ridge portfolio as the number of assets increases in order of their Sharpe ratios, while the size of the training sample remains fixed. Our analysis introduces heterogeneity in the true profitability of the underlying assets and derives an explicit condition under which the virtue-of-complexity phenomenon arises under ridge shrinkage. We show that when asset profitability is highly heterogeneous, expanding the portfolio by adding assets in descending order of their Sharpe ratios can generate a hump-shaped pattern in out-of-sample performance—closely mirroring the empirical results in Figure 1. This heterogeneity will later interact with uniform shrinkage to generate misalignment and the hump-shaped pattern we observe empirically.

3.1 Model Setup

We consider the problem of building a portfolio from a set of N assets. Each asset is indexed by n, and its excess returns at time t are denoted by $F_{n,t}$. Denote $F_t^N = [F_{1,t}, F_{2,t}, \cdots F_{N,t}]' \in \mathbb{R}^N$, we make the following assumption regarding its distribution:

Assumption 1 (Data Generating Process). We assume that $\{F_t^N\}$ are i.i.d. with $\mathbb{E}[F_t^N] = [\lambda_1, \lambda_2 \cdots \lambda_N] := \lambda^N, \text{Var}(F_t^N) = I_N$. Furthermore, we assume they have finite higher moments, in particular for $\forall n, t$, we assume $\mathbb{E}[F_{n,t}^3] = 0, \mathbb{E}[F_{n,t}^4] < \infty$. Finally, we denote $\gamma := \sum_{n=1}^N \lambda_n^2$ be the (squared) L_2 norm of vector λ^N, which we assume to be bounded, i.e. $\gamma = O(1)$.

The data generating process in Assumption 1 is purposefully kept restrictive for a simpler exposition of our theoretical results.[11] Our assumption of identity covariance rules out any low-dimensional factor structure, such that there is no small subset of latent factors that can fully capture the cross-section of expected excess returns across assets. Each of the N assets captures an equal and distinct portion of return variation, mirroring the "anomaly zoo" that investors face.[12] One can view this setup as one with no strong factors, where the investor builds a portfolio of "weak factors" (Lettau and Pelger (2020); Giglio et al. (2025)). When the assets do have a factor structure, our setup can be interpreted as focusing on the idiosyncratic component of the assets after removing the common factors. This perspective aligns with that of an investor who cares only about the expected return beyond compensation for common risk—that is, the alphas—and seeks to combine these alphas to construct an arbitrage portfolio (see Ingersoll Jr (1984); Kelly et al. (2019); Kim et al. (2021)).

The key ingredient in Assumption 1 is λ_n, which is assumed to be different across n and introduces heterogeneity in the true profitability of each asset. Given the identity covariance matrix, the n-th asset has a (population) Sharpe ratio of λ_n. As we will see later, γ captures the total amount of the Sharpe ratio that the investor has access to. The condition $\gamma = O(1)$ means the population efficient portfolio has a bounded population Sharpe ratio when $N \to \infty$, which guarantees the no-arbitrage condition. Furthermore, the bounded L_2 norm implies that the majority of individual strategies must have a small λ_n, similar to a "weak signal" setting (Shen and Xiu (2025)).

[11] See Didisheim et al. (2024) for a micro-founded setup where the assets are managed portfolios based on asset-level signals, and correlated assets.

[12] By contrast, Liao et al. (2023) study a setting with a strong latent factor structure and show that adding noise aids implicit shrinkage in forecasting problems.

Portfolio Estimation in High-dimension. We consider the MSRR-ridge portfolio estimator as introduced in Section 2.1,

$$\hat{w}(z,c) = \left(\bar{E}_T^c[F_t^N F_t^{N\prime}] + z\, I_N\right)^{-1} \bar{E}_T^c[F_t^N]. \tag{5}$$

where $\bar{E}_T^c\left[F_t^N\right]$ and $\bar{E}_T^c\left[F_t^N F_t^{N\prime}\right]$ denote the sample first and second moments using only T return observations,

$$\bar{E}_T^c\left[F_t^N\right] \;=\; \frac{1}{T}\sum_{s=1}^{T} F_s^N, \qquad \bar{E}_T^c\left[F_t^N F_t^{N\prime}\right] \;=\; \frac{1}{T}\sum_{s=1}^{T} F_s^N F_s^{N\prime}.$$

We use $c = N/T$ to denote the estimation "complexity." If the investor has access to infinite in-sample data, $c = 0$ and the empirical estimates converge to the true population moments. However, in the real world, the investor only has a finite amount of in-sample data to estimate the empirical moments, and it is not feasible to generate more in-sample data in the time series. Therefore, we think of T as being finite and fixed, giving rise to $c > 0$ and increasing in N. With a slight abuse of notation, we denote $\gamma(c) = \sum_{n=1}^{N}(\lambda_n)^2$ as the squared L_2 norm of λ^N when $c = N/T$. The ridge penalty $z > 0$ stabilizes the inversion by shrinking the second moment estimate towards an identity matrix.

3.2 Theoretical Portfolio Performance

We next characterize the out-of-sample Sharpe ratio of the MSRR-ridge portfolio

$$R_{t+1}(z;c) = \widehat{w}(z;c)' F_{t+1}^N$$

as a function of estimation complexity c and λ^N. To aid the discussion, we first consider the *infeasible* ridge portfolio, where the true moments from 1 are plugged directly into the portfolio solution (5):

$$\widehat{w}^{infeas}(z,c) = \frac{\lambda^N}{1+z+\|\lambda^N\|^2} \tag{6}$$

This portfolio is infeasible because it relies on the population moments of the strategies. Nevertheless, studying the infeasible portfolio provides a useful benchmark that highlights key implications for the feasible portfolio. The following proposition characterizes the expected return, second moment, and Sharpe ratio of the infeasible portfolio:

Proposition 1 (Expected return and variance of the infeasible ridge portfolio)
The infeasible ridge portfolio for strategies F_t^N, i.e. $R_{t+1}^{infeas}(z;c) = \widehat{w}^{infeas}(z,c)' F_{t+1}^N$, has

expected return

$$E[R_{t+1}^{infeas}(z;c)] \equiv \mathcal{E}(z;\gamma(c)) = \frac{\gamma(c)}{1+z+\gamma(c)} \tag{7}$$

and second moment

$$\text{Var}[R_{t+1}^{infeas}(z;c)] \equiv \mathcal{V}(z;\gamma(c)) = \frac{\gamma(c)}{(1+z+\gamma(c))^2}. \tag{8}$$

Hence the Squared Sharpe ratio is

$$\frac{\mathbb{E}[R_{t+1}^{infeas}(z;c)]^2}{\text{Var}[R_{t+1}^{infeas}(z;c)]} \equiv \text{SR}(R_{t+1}^{infeas}(z;c))^2 = \gamma(c) \tag{9}$$

Proposition 1 shows that when there's no estimation uncertainty and the investor can recover the true moments, the expected out-of-sample MSRR-ridge portfolio performance is determined by the total amount of asset Sharpe ratio γ that the investor has access to. As $\lambda^N > 0$ for all assets, the Sharpe ratio of the infeasible portfolio is strictly increasing when more assets are used in building the portfolio.

We then turn to characterize the expected out-of-sample behavior of the feasible ridge portfolio in (6). When T is finite and fixed while N is large, the investor faces estimation complexity, such that the empirical moments deviate from the true moments. In Proposition 3 in Appendix B.1, we characterize the asymptotic behavior of the feasible portfolio when both T and N are large and correspond to a complexity level of c. The result closely resembles Theorem 3 in Didisheim et al. (2024), where the feasible portfolio moments are related to the $\mathcal{E}(z;\gamma(c))$ and $\mathcal{V}(z;\gamma(c))$ function but with some complexity adjustment. The asymptotic analysis allows us to explicitly characterize what happens when complexity increases due to an increase in N, and in Section 4.1 we show in simulation that our asymptotic theory captures portfolio performance in finite samples as well.

How does using more assets affect the expected out-of-sample Sharpe ratio of the feasible portfolio? Surprisingly, we show that there exist scenarios that the out-of-sample performance might actually decrease when more assets are used. To show this, we first note that under Assumption 1, the out-of-sample Sharpe ratio is increasing in ridge shrinkage z.

Lemma 1 (Optimal shrinkage for identity covariance matrix)
Under Assumption 1, for fixed c and $\gamma(c)$, we have

$$\lim_{z\to\infty} \text{SR}\big(R_{t+1}^{\text{feas}}(z,c)\big) \;=\; \max_{z\geq 0} \text{SR}\big(R_{t+1}^{\text{feas}}(z,c)\big). \tag{10}$$

Intuitively, when the investor uses a larger ridge shrinkage, she shrinks her covariance estimates away from the sample covariance to an identity matrix, but under Assumption 1, the identity matrix is the true population covariance matrix. This implies that an investor using $z \to \infty$ coincidentally recovers the true covariance matrix, making it the optimal shrinkage to use in the feasible MSRR-ridge portfolio.[13]

Lemma 1 is a convenient result that drastically simplifies the characterization of the out-of-sample Sharpe ratio of the optimal feasible portfolio under Assumption 1. It also aids interpretation: by focusing on the $z \to \infty$ case, we remove the noisy covariance estimation from the portfolio problem, meaning that the remaining estimation uncertainty comes from the empirical mean estimates. The following proposition characterizes the optimal feasible MSRR-ridge portfolio Sharpe ratio, and the condition under which using more assets benefits the portfolio:

Proposition 2 (Optimal feasible portfolio Sharpe ratio)
Let $F_t^N \in \mathbb{R}^N$ and λ^N satisfy Assumption 1, and define $c = N/T$. Set the feasible ridge portfolio return as

$$R_{t+1}^{\text{feas}}(z,c) \;=\; \hat{w}(z,c)'\, F_{t+1}^N,$$

where $\hat{w}_T(z,c)$ is from (5). In the double-asymptotic limit $N,T \to \infty$ with $N/T \to c$, the out-of-sample Sharpe ratio at extreme shrinkage ($z \to \infty$) is

$$\lim_{z\to\infty} \text{SR}\big(R_{t+1}^{\text{feas}}(z,c)\big)^2 \;=\; \frac{\gamma(c)}{1+\dfrac{c}{\gamma(c)}}. \tag{11}$$

Consequently, the Sharpe ratio increases with c only if

$$\frac{\partial \gamma}{\partial c} \;>\; \frac{\gamma(c)}{\gamma(c)+2c} \;=\; MC\big(c,\gamma(c)\big). \tag{12}$$

If this condition fails, using more assets (increasing c) reduces the out-of-sample Sharpe ratio.

Proposition 2 states that using more assets does not guarantee higher out-of-sample portfolio performance. Whether including an additional asset in the portfolio improves overall performance depends on the trade-off between the additional Sharpe ratio it captures versus the heavier parameterization that leads to higher estimation complexity. In (12), $\frac{\partial \gamma}{\partial c}$ captures the marginal benefit of the additional Sharpe ratio from new assets, while $MC(c,\gamma(c))$

[13] The $z \to \infty$ shrinkage is also optimal if we consider Frobenius-norm loss (Ledoit and Wolf (2004)).

captures the marginal cost of the estimation complexity. The key observation is that the marginal cost is never zero with positive c, which means that if the additional strategy has a positive but lower Sharpe ratio, the overall Sharpe ratio of the portfolio will decline.

3.3 Effect of Shrinkage Misalignment on Portfolio Performance

Proposition 2 shows that the behavior of the optimal MSRR-ridge portfolio depends on the function $\gamma(c)$, which is given by λ^N that describes how profitability (in the form of Sharpe ratios for individual assets) is distributed across assets. This gives us a convenient way to capture the (mis)alignment between individual asset profitability and shrinkage strength.

The ridge shrinkage zI in (5) implies that all assets are shrunk by the same degree. The distribution of λ^N that perfectly aligns with this equal-strength shrinkage would be the case where all assets have ex-ante the same profitability, i.e. $\lambda_n = \frac{\bar{\gamma}}{\sqrt{T}}$ for all n. This implies that $\gamma(c) = c\bar{\gamma}^2$ is a linear function in complexity level c (and in N). On the other hand, if there is strong heterogeneity in λ_n's, the shape of λ^N and the strength of shrinkage will be misaligned. This will lead to out-of-sample inefficiency, such that using more assets actually worsens Sharpe ratio of the MSRR-ridge portfolio. A specific setup to observe this pattern is when the set of assets expands with decreasing profitability, i.e. $\lambda_n > \lambda_{n+1}$ for all n, which mimics the descending sorting experiment we conduct in Section 2.1. The next Lemma shows the effect of shrinkage alignment on the behavior of high-dimensional portfolios by characterizing the portfolio's out-of-sample Sharpe ratios when more assets are used.

Lemma 2 (Effect of shrinkage alignment on MSRR-ridge portfolio)
Let λ_n be the n'th asset's Sharpe ratio that satisfies Assumption 1.

- *Suppose all assets have equal Sharpe ratios ($\lambda_n = \frac{\bar{\gamma}}{\sqrt{T}}$) so that $\gamma(c) = c\bar{\gamma}^2$. Then $\max_{z\geq 0} \mathrm{SR}\big(R^{\mathrm{feas}}_{t+1}(z,c)\big)$ is always increasing in c.*
- *Suppose assets have Sharpe ratios $\{\lambda_i\}_{n=1}^N$ that are strictly decreasing ($\lambda_1 > \lambda_2 > \cdots > \lambda_N$) with sufficient decay such that $\gamma(c)$ satisfies (52). Then there exists an optimal complexity level c^* such that $\max_{z\geq 0} \mathrm{SR}\big(R^{\mathrm{feas}}_{t+1}(z,c)\big)$ is increasing for $c < c^*$ and decreasing in $c > c^*$.*

The first part of Lemma 2 shows that when all assets have the same ex-ante Sharpe ratio, the distribution of λ aligns with the shrinkage strength in the cross-section of assets. As a result, the MSRR-ridge portfolio achieves the virtue of complexity property, such that using more assets is always beneficial out-of-sample. Intuitively, the alignment between one-size-fits-all shrinkage and the assets' true profitability allows the model to strike the correct bias-variance tradeoff for any number of assets.

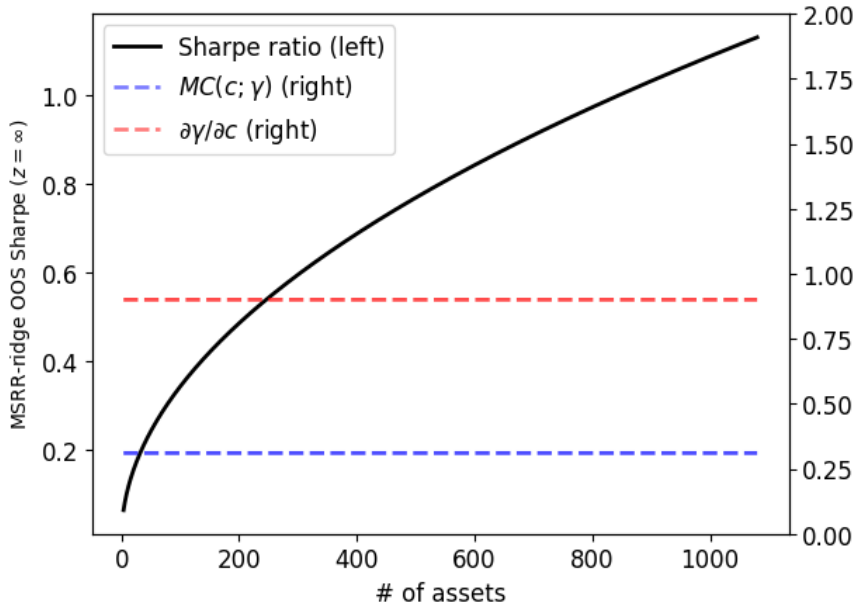
The second part of Lemma 2 shows what happens when there is a misalignment between shrinkage strength and the assets' true profitability. In this case, MSRR-ridge portfolio shrinks all assets weights in the same fashion–regardless their true profitability, which means high Sharpe ratio assets are shrunk by the same degree as low Sharpe ratio assets. This creates difficulty in striking the correct bias-variance tradeoff. In particular, when the asset Sharpe ratio is declining fast as c grows, each additional asset brings lower benefits while still increases estimation complexity by the same amount. If the decline in strategy profitability is strong, there exists an optimal complexity level such that the virtue of complexity disappears.

This result can be visualized in Figure 5, where we draw the theoretical out-of-sample Sharpe ratio of MSRR-ridge portfolio, as well as the marginal benefit and cost of complexity as derived in Proposition 3. In panel (a), we consider when λ_n is equal across n. In this case, the marginal benefit $(\partial\gamma(c)/\partial c)$ is always higher than the marginal cost $MC(c;\gamma)$, which leads to a continuously improved portfolio Sharpe ratio when more assets are used. In panel (b), we consider a case where λ_n declined exponentially. In this case, using more assets is useful initially for the MSRR-ridge portfolio, but eventually the estimation cost outweighs the additional profitability as less profitable assets are considered, which leads to an eventual decline in portfolio performance.

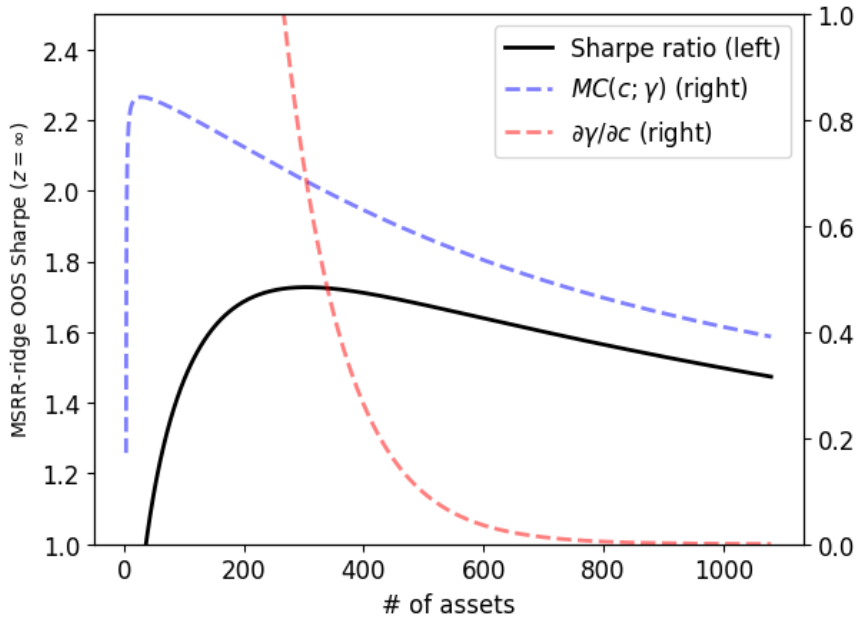
In Figure 6, we plot the theoretical mean, volatility, and Sharpe ratio of the MSRR–ridge portfolio for different values of z under an exponentially declining sequence of asset Sharpe ratios, and compare them with the empirical patterns in Figure 1. The theory closely replicates the empirical behavior: the decline in Sharpe ratio is driven by a fall in expected returns rather than rising volatility.

As Kelly and Malamud (2025) emphasize, this pattern reflects implicit shrinkage—as model complexity grows with a fixed ridge parameter, the effective degree of regularization increases automatically. The result is lower variance but also stronger bias in expected returns. Sustaining performance under increasing complexity thus requires that newly added assets provide sufficiently large marginal returns and are not excessively shrunk.

Figure 6 illustrates that, although variance continues to fall (a hallmark of implicit shrinkage), the mean return declines even faster, so the rise in bias dominates and the portfolio's Sharpe ratio deteriorates.



(a) Alignment



(b) Misalignment

Figure 5: Theoretical Out-of-sample Sharpe Ratio: Alternative Calibrations

Note. This figure shows the expected out-of-sample Sharpe ratio of the MSRR-ridge portfolio as characterized in Proposition 2. For the n'th asset, we set $\lambda_n = 0.05$ in panel (a) and $\lambda_n = a \cdot e^{-bn}$ with $a = 0.2$ and $b = 0.005$ in panel (b). The red dashed line plots the marginal benefit of data mining $\partial\gamma/\partial c$ and the blue dashed line plots the marginal cost of data mining $MC(c;\gamma)$ as c increases. The black line plots the expected out-of-sample MSRR-ridge portfolio Sharpe ratio.

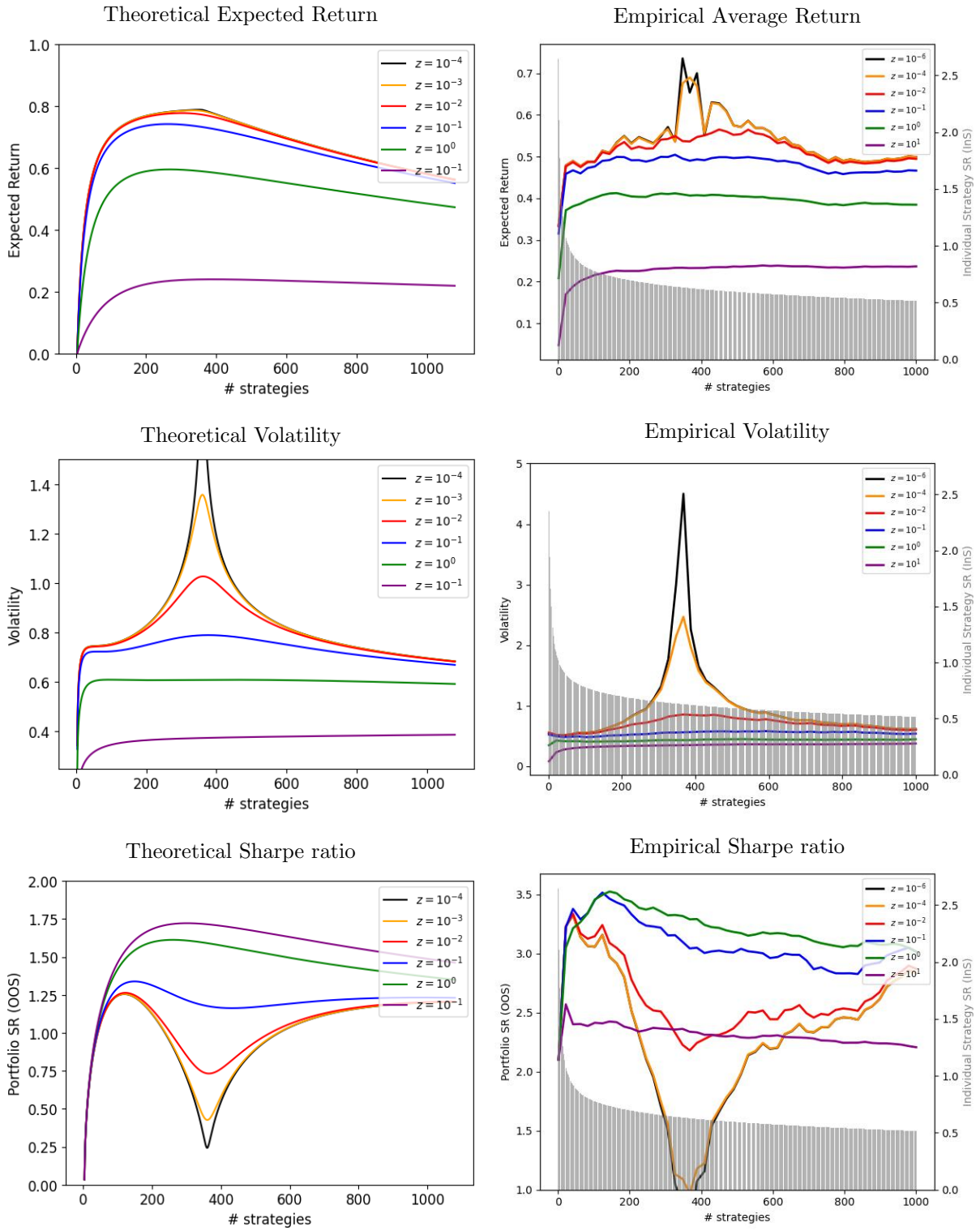


Figure 6: Theoretical and Empirical Expected Return, Volatility and Sharpe ratios with diffferent z

Note. This figure shows the expected return, volatility, and Sharpe ratio as characterized in Proposition 3 and the empirical counterparts in Section 2.1. For theoretical computation, the n'th asset has $\lambda_n = a \cdot e^{-bn}$ with $a = 0.2$ and $b = 0.005$.

Connection with the virtue of complexity. Our theoretical results essentially establish a necessary condition for the virtue of complexity phenomenon. Conceptually, using complex models always achieves better performance if the marginal economic benefit of increasing complexity is higher than the marginal cost of less precise parameter estimation. In many empirical designs using machine learning methods, this condition is likely to be satisfied, such as when researchers build tangency portfolios based on kernel methods (Kozak (2020)), neural networks (Chen et al. (2024), Didisheim et al. (2024)), or tree-based methods (Bryzgalova et al. (2024), Cong et al. (2025)). In these setups, the models take stock-level characteristics as fixed inputs and use complex machine learning models to approximate the unknown and likely nonlinear data-generating process.

These machine learning methods expand model complexity by adding more ex-ante identical predictive units—such as neurons in neural networks, basis functions in kernel methods, or splits in tree-based models. In such settings, the marginal contribution of each additional unit remains roughly constant: the first neuron in a neural network has the same expected approximation ability as the thousandth. This corresponds to the first case characterized in Lemma 2, where a uniform shrinkage scheme like ridge naturally aligns with the underlying predictive environment. Consequently, we should expect the virtue of complexity to hold for these models when standard shrinkage techniques are employed.[14]

On the other hand, if the complex predictive environment contains strongly heterogeneous predictors, a one-size-fits-all shrinkage will be suboptimal. One example for such strong heterogeneity in predictors is undisciplined data mining. Investors have long explored accounting-based variables like valuation ratios or profitability to predict stock returns. These variables are evident predictors because there's a direct economic connection between these fundamentals and a firm's future return. In contrast, many recent alternative datasets, such as web traffic, satellite images, and transcript data, are more challenging to process, have a lower signal-to-noise ratio, and only have an indirect connection to future stock returns. As a result, it is reasonable to expect the true predictability from data-mined signals will exhibit strong heterogeneity. Our result implies that with data-mined predictors, standard shrinkage methods are likely to be suboptimal and may not deliver virtuous out-of-sample performance when more data-mined predictors are used.

3.4 Shrinkage Alignment

Observing the potential consequences of shrinkage misalignment in high-dimensional portfolios, a natural question is: what can an investor do to address this issue? In this section,

[14] Alternatively, one can view predictive units (such as neurons) as containing randomly assigned predictability. Under this interpretation, expanding the number of predictive units resembles the random-ordering experiment in Section 2.

we propose a simple extension to the MSRR-ridge portfolio. Instead of shrinking the second moment towards an identity matrix, we propose to shrink towards a diagonal matrix whose values align with the inverse of individual asset Sharpe ratio, which can be estimated using training data.

To see the motivation behind this choice, it is useful to think about the Bayesian interpretation of (16). Suppose we have an MSRR regression as in Britten-Jones (1999),

$$1 = w^{N\prime} F^N_{t+1} + \epsilon_{t+1} \tag{13}$$

with $\epsilon_{t+1} \sim N(0, \sigma^2_\epsilon)$ and the investor's prior on w^N is $w^N \sim N(0, Q_N)$, the posterior mode estimate of w_t given data $\{F_\tau\}^T_{\tau=t-T}$ would be

$$\widehat{w}^N_t = \left(\sigma^2_\epsilon Q_N^{-1} + \bar{E}^c_T\left[F^N_t F^{N\prime}_t\right]\right)^{-1} \bar{E}^c_T\left[F^N_t\right] \tag{14}$$

Here, the prior variance Q_N captures the investor's belief about the true efficient portfolio weights. It also defines the *geometry* of shrinkage. By choosing Q_N appropriately, one can penalize more heavily in directions associated with low Sharpe assets and more lightly where Sharpe ratios are high. In a setting where all assets are independent, as described in Assumption 1, it is easy to see that the "correct belief" that aligns with the true data generating process would be

$$Q_N = \text{diag}([\lambda^2_1, \lambda^2_2, \cdots, \lambda^2_N]) = \text{diag}(\lambda^N) \tag{15}$$

where λ^2_n is the true population squared Sharpe ratio of asset n. Given this prior, the weight of assets with higher true Sharpe ratio (higher λ) gets shrunk less, therefore preserving their high profitability in the portfolio. The following proposition shows that $\widehat{w}^N_t$ in (14) with such Q_N achieves optimal out-of-sample performance:

This construction naturally connects our generalized ridge framework to *Automatic Relevance Determination* (ARD), a Bayesian approach that places independent Gaussian priors on coefficients with component-specific variances (MacKay, 1994; Neal, 1996). In ARD, each coefficient β_j follows

$$w_j \sim \mathcal{N}(0, \lambda_j^{-2}),$$

where the precision parameter λ_j governs the amount of shrinkage applied. Coefficients associated with higher λ_j are pushed closer to zero, while those with smaller λ_j remain active—allowing the model to "automatically" determine relevance.

In practice, the investor does not observe the true values of λ_n, but can estimate them

from in-sample data using the squared in-sample Sharpe ratios. Accordingly, the choice of D in (17) is designed to align the shrinkage strength with each asset's estimated Sharpe ratio. In high-dimensional settings, this idea parallels results in the regression literature that use random matrix theory (e.g., Dobriban and Wager (2018), Wu and Xu (2020)), which show that aligning shrinkage with the distribution of true coefficients can substantially reduce mean-squared error. By analogy, we conjecture that non-uniform shrinkage can restore portfolio efficiency by tailoring the bias–variance tradeoff to the cross-sectional distribution of Sharpe ratios—thereby recovering the diversification benefits that naïve ridge shrinkage suppresses.[15] Therefore, the portfolio estimator we propose can be viewed as the generalized ridge estimator:

$$\widehat{w}_t^{SRS} = \left(zD + \bar{E}_T^c\left[F_t^N F_t^{N\prime}\right]\right)^{-1} \bar{E}_T^c\left[F_t^N\right] \tag{16}$$

where D is a custom shrinkage target. If $D = I$, then we have the MSRR-ridge portfolio as in (4). Instead, we propose to set

$$D = \text{diag}([1/\widehat{SR}_1^2, 1/\widehat{SR}_2^2, \cdots, 1/\widehat{SR}_N^2]) \tag{17}$$

where $\widehat{SR}_n$ is the in-sample estimate of n'th asset's Sharpe ratio. This MSRR with Sharpe ratio shrinkage (MSRR-SRS) estimator incorporates information about individual asset's Sharpe ratio in the shrinkage design. z is again the hyperparameter that controls the overall shrinkage strength, which we tune using cross validation.

4 Simulations

4.1 Shrinkage Alignment in Portfolio Construction

In this section, we present simulation results. The motivation for the simulations is twofold: first, to demonstrate that our large-N, T asymptotics in Proposition 2 approximate the behavior of feasible portfolios in finite samples; and second, to examine the behavior of MSRR-SRS and MSRR-ridge portfolios when strategies are correlated.[16] For the latter, we show that once the correlation structure is accounted for, our simulation produces a pattern that closely matches the main empirical result Section 2.1 .

[15] When true Sharpe ratios are correlated across assets, one can extend this framework by allowing off-diagonal terms in D to capture cross-asset dependencies. The results in Wu and Xu (2020) imply that using the true covariance of λ_n yields optimal out-of-sample performance.

[16] Didisheim et al. (2024) characterizes the performance of feasible portfolios under a general covariance structure; however, it is difficult to analyze its properties in closed form.

We generate asset returns following

$$F_t = \lambda + \Sigma^{1/2} X_t, \tag{18}$$

where $X_t \sim \mathcal{N}(0, I_{N\times N})$, $\Sigma \in \mathbb{R}^{N\times N}$ denotes the covariance matrix of returns, and $\lambda \in \mathbb{R}^N$ is the vector of expected returns. We simulate 100 realizations, each consisting of N strategies and $2T = 720$ time periods. Our simulation design mirrors the empirical setup in Section 2. In each realization, the first half of the sample ($T = 360$) is used for model training, and the second half is reserved for out-of-sample evaluation. We gradually increase the number of assets included in the MSRR-ridge portfolio estimation and track out-of-sample portfolio performance as N grows from 1 to 1000.

4.1.1 Identity Covariance

In the first setup, we align the simulation exactly with our theoretical model by setting $\Sigma = I_{N\times N}$. We set the Sharpe ratio of the n'th asset to be

$$\lambda_n = a \cdot e^{-bi}, \tag{19}$$

Panel (a) in Figure 7 shows the simulation result for $a = 0.2$ and $b = 0.005$. We find that the simulated portfolio performance (black line) closely aligns with the theoretical performance predicted by Proposition 2 (yellow line), suggesting that the large N, T asymptotics we take in the theory section approximate the results in finite samples well.

4.1.2 Matching the Pattern in Section 2 and MSRR-SRS

Although the identity covariance setting is simple enough to characterize portfolio performance in closed forms and to illustrate the driving forces, it has difficulty in matching an important empirical feature: in Figure 1, all assets have high Sharpe ratios, whereas in Figure 7 panel (a) the simulated asset Sharpe ratios decline very fast. In this section, we match this empirical feature by introducing correlations between assets. Intuitively, considering nonzero correlation increases the complexity of portfolio estimation, because now the investor needs to estimate the full $N(N+1)/2$ parameters in the covariance matrix. Therefore, the marginal cost of an additional asset increases, and we should expect to see the portfolio Sharpe ratio decrease when additional assets have high individual Sharpe ratios.

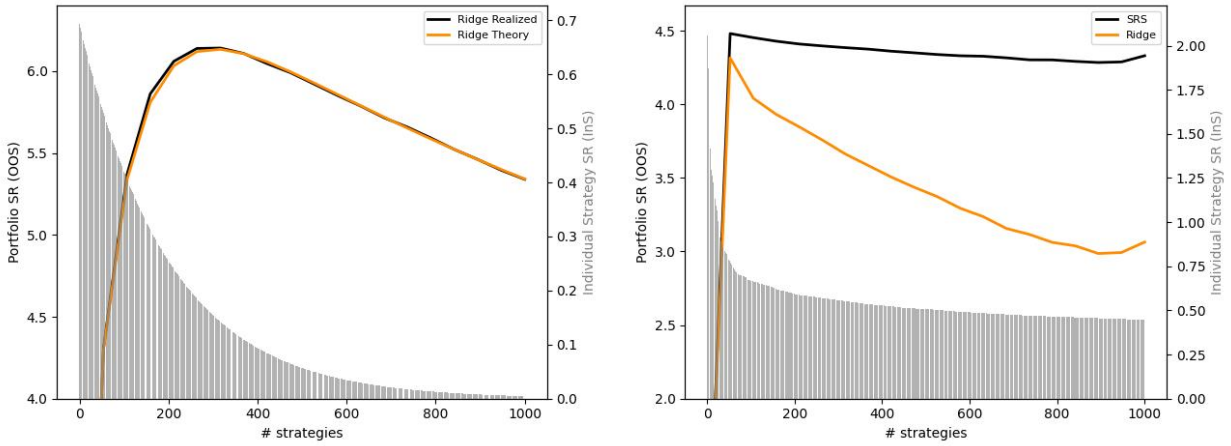
We replicate the cross-sectional pattern of asset returns observed in the data by simulating assets with empirically observed Sharpe ratios and imposing a structured correlation matrix. Specifically, we assign each simulated asset a λ_i using the realized Sharpe ratio. We then

construct the covariance matrix Σ to have a Toeplitz structure defined by

$$\Sigma_{ij} = \rho^{|i-j|}, \tag{20}$$

where $\rho \in (0, 1)$ controls the decay rate of correlations between assets. This specification implies that the correlation between assets decays geometrically with their index separation, which in turn reflects the similarity in Sharpe ratios. That is, if assets i and j have similar Sharpe ratios ($\gamma_i \approx \gamma_j$), then $|i - j|$ is small, and thus Σ_{ij} is large.

This structure induces a *pseudo local factor model*: assets that are close in Sharpe ratio space tend to be more correlated, capturing the intuition that assets with similar risk-adjusted returns are likely driven by similar underlying sources of variation. Put differently, this setup reflects the idea that assets with comparable performance metrics should exhibit high co-movement—otherwise, their differences could be systematically exploited. While not a formal no-arbitrage condition, this structure naturally groups similar assets by their exposure patterns.



(a) Identity correlation

(b) Toeplitz correlation

Figure 7: Portfolio Sharpe ratio in Simulations

Note. This figure shows the out-of-sample Sharpe ratio of MSRR-ridge portfolios with simulated assets. In panel (a), the simulated assets are uncorrelated and have population Sharpe ratios according to $\lambda_n = ae^{-bn}$ with $a = 0.2$ and $b = 0.005$. In Panel (b), the simulated assets have a correlation matrix with a Toeplitz structure with $\rho = 0.95$, and their population Sharpe ratios are set to match the empirical Sharpe ratio distributions of assets in the Chen and Dim (2023) dataset. We compare the performance of SRS and Ridge. Simulated portfolio Sharpe ratios are averaged over 100 realizations.

Panel (b) in Figure 7 shows simulation results that closely match the empirical pattern. We set Sharpe ratios of individual asset to match that of Chen and Dim (2023) dataset. The average correlation in the simulation is 6.5% (versus the average absolute correlation in the empirical data-mined strategies being 10%). Despite the small average correlation

between strategies, it generates portfolio performance similar to what we see in the data: Ridge shrinkage produces a pronounced hump shape while SRS exceeds and maintains the peak performance obtained by Ridge shrinkage.

5 Empirical Illustration of Shrinkage Alignment

We illustrate the efficacy of MSRR-SRS by conducting the same asset expansion experiment as in Section 2.1, where we gradually increase the number of data-mined strategies with decreasing Sharpe ratios, to see how the method accommodates assets with strongly heterogeneous profitability.

Panel (a) of Figure 8 reports the out-of-sample Sharpe ratios of MSRR–SRS in this experiment. MSRR–SRS significantly outperforms MSRR–ridge when asset Sharpe ratios are highly heterogeneous. Unlike the standard MSRR–ridge, the out-of-sample performance of MSRR–SRS remains stable as lower–Sharpe-ratio assets are added, thereby preserving the virtue of complexity in this setting.

Panel (b) compares the absolute portfolio weights from MSRR–SRS and MSRR–ridge. MSRR–SRS achieves its superior performance by assigning slightly larger weights to assets with high in-sample Sharpe ratios and slightly smaller weights to those with lower in-sample Sharpe ratios. This behavior reflects the goal of shrinkage alignment: by applying weaker shrinkage to assets that are likely to have high true Sharpe ratios (as indicated by their in-sample performance), the portfolio captures more of their contribution to overall efficiency.[17]

Another way to observe the benefits of shrinkage alignment is to consider the following shrinkage target

$$D(\alpha) = \text{diag}([d_1, d_2, \cdots, d_N]) \tag{21}$$

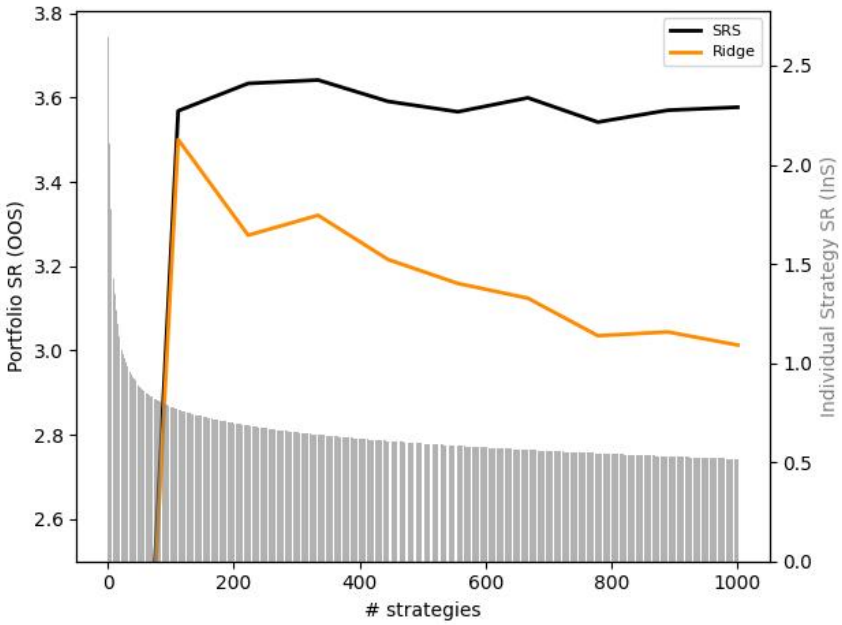
where

$$d_n = \alpha \cdot 1/\widehat{SR}_n^2 + (1 - \alpha) \tag{22}$$

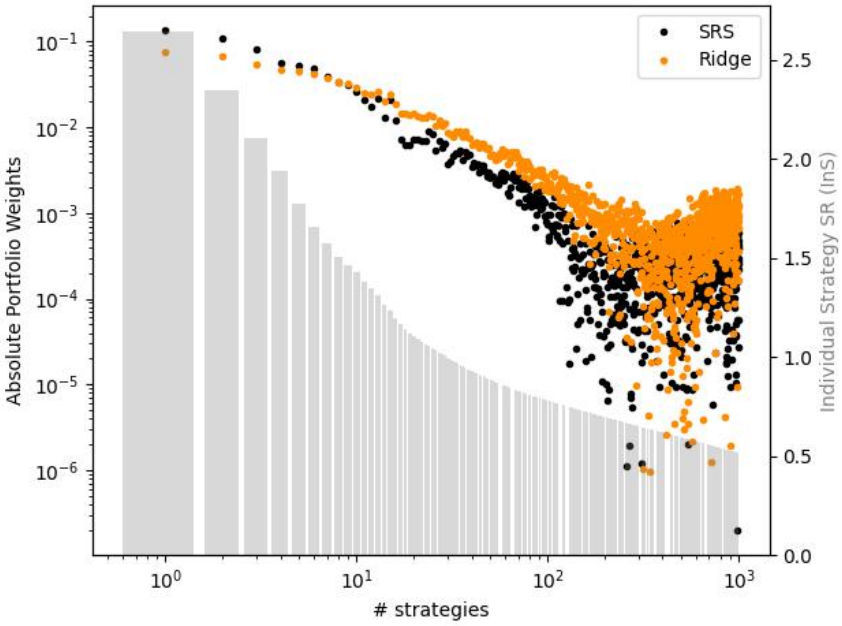
In this formulation, $\alpha \in [0, 1]$ controls the extent to which the shrinkage strength is aligned with each asset's in-sample Sharpe ratio.[18] When $\alpha = 0$, all $d_n = 1$, corresponding to the standard MSRR–ridge case; when $\alpha = 1$, we obtain the MSRR–SRS portfolio.

[17] Figure 17 shows that MSRR–SRS delivers higher out-of-sample Sharpe ratios relative to all benchmark portfolio methods discussed in Section 2. The MSRR–SRS portfolio also produces significant alphas with respect to these benchmarks, as reported in Figure 18.

[18] We scale the L_2-norm of $1/\widehat{SR}_n^2$ such that the L_2-norm of $D(\alpha)$ remains constant across all values of α.



(a) Out-of-sample Sharpe ratios of MSRR-SRS and MSRR-ridge



(b) Absolute Portfolio Weights

Figure 8: Out-of-sample MSRR-SRS vs MSRR-ridge Sharpe Ratios and Weights

Note. This figure shows the out-of-sample Sharpe ratios of MSRR–SRS (yellow) and MSRR–ridge (black) portfolios as the number of strategies included in portfolio construction increases in descending order of in-sample Sharpe ratios. Panel (a) reports results when the penalty parameter z is selected via leave-one-out cross-validation from a grid $z \in 10^{-7}, 10^{-6}, \ldots, 10^{3}$ for each portfolio size. For the MSRR–SRS shrinkage target $D = \text{diag}([1/\widehat{SR}_1^2, 1/\widehat{SR}_2^2, \ldots, 1/\widehat{SR}_N^2])$, D is scaled to have unit L_2 norm. The gray bars (right axis) show the average in-sample Sharpe ratios of individual strategies by rank across rolling windows. Panel (b) reports the average absolute portfolio weight assigned to each strategy by rank. Results are computed using a 360-month rolling window over the out-of-sample period from January 1994 to December 2022.

Figure 9 reports the out-of-sample portfolio Sharpe ratios as functions of α and the overall shrinkage parameter z. In the data-mined strategy setting, out-of-sample performance improves monotonically with α, highlighting that *shrinkage alignment* is critical for achieving superior portfolio performance in high-dimensional environments.

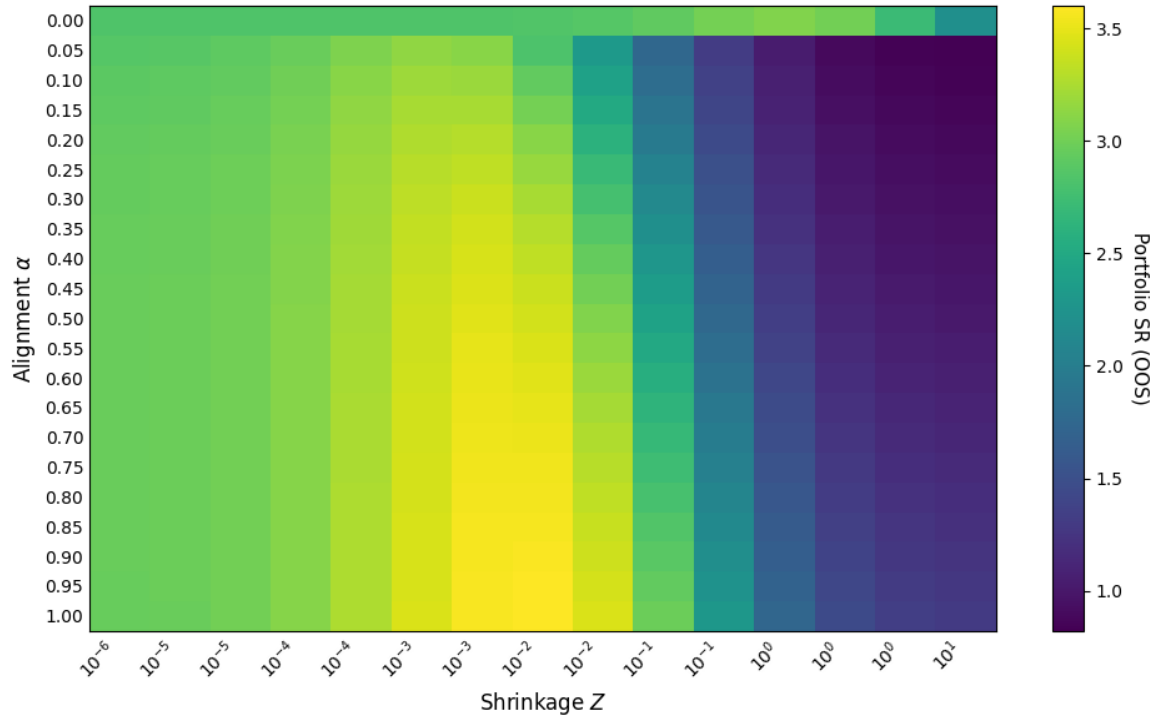


Figure 9: Effect of Shrinkage Alignment for Out-of-sample Portfolio Sharpe Ratios

Note. This plot shows the out-of-sample Sharpe ratio of MSRR-SRS portfolio using shrinkage target $D(\alpha) = \text{diag}([d_1, d_2, \cdots, d_N])$ where $d_n = \alpha \cdot 1/\widehat{SR}_n^2 + (1-\alpha)$. We vary the hyperparameters α and z for different values. Results are based on a 360-month rolling window over the out-of-sample period January 1994 to December 2022.

5.1 Shrinkage Alignment with Principal Components

So far, we considered shrinkage alignment in the setting of high-dimensional assets, where each asset contributes a distinct source of profitability. We also find benefits of shrinkage alignment in the typical APT-style setting (Ross, 1976). In such environments, asset returns typically exhibit a strong factor structure or a few dominant latent components. A common approach is to perform principal component analysis (PCA) and then apply a sparsity-inducing shrinkage such as the lasso. For instance, Kozak (2020) use an elastic net (a combination of lasso and ridge penalties) to select a small subset of principal components (PCs) that yield high out-of-sample Sharpe ratios. However, even when sparsity may appear theoretically justified, discarding information can be suboptimal. It may be preferable instead to retain all components and optimally reweight them. For example, Kelly

et al. (2024b) advocate for nonlinear shrinkage, showing that such non-sparse approaches can achieve high out-of-sample performance even in settings where sparsity might seem desirable.

Our *generalized ridge* approach introduces an alternative, L^2-style non-sparse shrinkage method for constructing efficient portfolios from PCs. An economically motivated investor, equipped with a structural understanding of their environment, would plausibly hold the prior belief that PCs with higher variance explain a greater portion of the cross-sectional variation in returns and, by the principle of no arbitrage, should receive larger portfolio weights.

This idea echoes Kozak et al. (2020), who assume that PC means are normally distributed with variances proportional to the amount of variation they explain. We extend this logic by assuming that the *PC weights themselves* are normally distributed with respect to the variance they explain. Intuitively, components capturing more systematic variation in returns are assigned larger prior variances, reflecting their greater economic relevance.

Formally, we first perform an in-sample eigenvalue decomposition of the factor return covariance matrix:

$$\frac{1}{T}\sum_{\tau=t-T}^{t} F_\tau F_\tau' = U\,\mathrm{diag}(\psi)U',$$

and rotate the space of factor returns using the corresponding in-sample PCs:

$$\tilde{F}_t = U' F_t.$$

We then compute the shrunk portfolio weights of these latent factors as

$$\widehat{w}_t^{\,D(\alpha)}(z) = \left(zD(\alpha) + \mathrm{diag}(\psi)\right)\frac{1}{T}\sum_{\tau=t-T}^{t}\tilde{F}_\tau, \tag{23}$$

where

$$D(\alpha) = \mathrm{diag}\left([d_1, d_2, \ldots, d_K]\right), \tag{24}$$

and

$$d_i = \alpha\psi_i^{-1} + (1-\alpha), \tag{25}$$

with ψ_i[19] denoting the eigenvalue of the ith principal component. The resulting $D(\alpha)$ is a di-

[19]We scale the L_2-norm of ψ_i^{-1} such that the L_2-norm of $D(\alpha)$ remains constant across all values of α.

agonal matrix of size $K \times K$, where $K = \min(T, N)$ is the number of principal components. In this specification, the shrinkage strength is inversely related to each PC's explanatory power: components explaining more variance are assigned lower shrinkage (larger prior variance), while less informative directions are more heavily penalized.

We follow the setup of Kelly et al. (2024b) and estimate portfolios using the 153 value-weighted capped anomaly portfolio returns from Jensen et al. (2022). We then plot out-of-sample Sharpe ratios using a rolling window of $T = 120$ for varying values of $\alpha \in [0, 1]$ in equation (24) and different levels of shrinkage $Z = [10^i : i \in \{-10, -9, \ldots, -1\}]$. The results, reported in Figure 10, reveal a broad optimal region for nonzero α, highlighting the potential benefits of our approach for out-of-sample portfolio formation. We observe a broad plateau of high Sharpe ratios across moderate values of both parameters, indicating that the generalized ridge formulation is robust to hyperparameter choice. Along the α dimension, performance improves as the prior alignment increases, suggesting that portfolios benefit from emphasizing principal components that explain greater variance in returns.

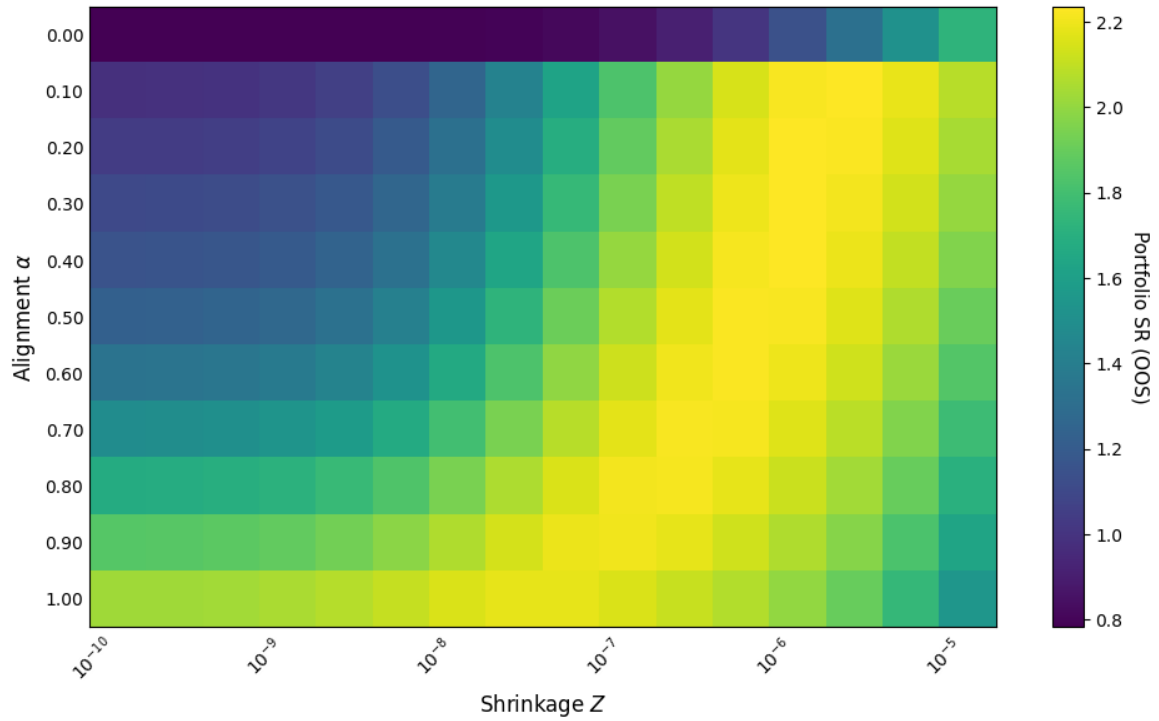


Figure 10: Effect of PC Shrinkage Alignment for Out-of-sample Portfolio Sharpe Ratios

Note. This figure reports the out-of-sample Sharpe ratios of MSRR generalized ridge PC portfolios defined in equation (23), using the shrinkage target $D(\alpha) = \text{diag}([d_{t,1}, d_{t,2}, \ldots, d_{t,K}])$, where $d_{t,i} = \alpha\tilde{\psi}_{t,i}^{-1} + (1 - \alpha)$. The heatmap varies the hyperparameters α and Z across a grid of values. Portfolios are constructed from the 153 value-weighted capped anomaly factors of Jensen et al. (2022), using a rolling window of 120 months. Out-of-sample evaluation period spans December 1981 to December 2023.

Table 1: Out-of-sample Sharpe Ratios

This table reports the out-of-sample Sharpe ratios for portfolios estimated using the Generalized UPSA, UPSA, Ridge, and KNS approaches. Portfolios are constructed from the 153 value-weighted capped anomaly factors of Jensen et al. (2022). The out-of-sample evaluation period spans December 1981 to December 2023.

	Generalized UPSA	UPSA	Ridge	KNS
JKP VW Cap	2.12	1.91	1.61	1.60
Mega	1.40	1.33	1.32	1.12
Large	1.53	1.35	1.21	1.09
Small	2.37	2.26	2.00	1.99

Table 2: Alphas (in %) and t-Statistics

Annual percentage α from the regression of Generalized UPSA on benchmark portfolios using heteroscedasticity-adjusted standard errors (5 lags). All portfolios are scaled to an annual volatility of 10%. The full sample covers the period from December 1981 to December 2024. t-statistics are reported in parentheses. Stars denote significance levels: *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$.

	UPSA	Ridge	KNS
JKP VW Cap	9.32^{***}	14.15^{***}	14.28^{***}
	(3.42)	(3.70)	(3.68)
Mega	4.52^{**}	5.31^{**}	8.43^{***}
	(3.11)	(3.17)	(4.76)
Large	7.16^{***}	9.29^{***}	10.92^{***}
	(4.47)	(4.66)	(5.49)
Small	7.16^{**}	12.61^{***}	12.79^{***}
	(2.65)	(3.69)	(3.84)

For our full out-of-sample analysis, we begin by setting $\alpha = 1$. We then apply the UPSA non-linear shrinkage method of Kelly et al. (2024b) to adaptively tune the shrinkage of our portfolio estimator, which we refer to as Generalized UPSA. In addition to the 153 factors available on the Jensen et al. (2022) website, we construct additional factors within size groups by forming capped value-weighted portfolios. At each period, stocks are sorted into three size categories using NYSE breakpoints—Mega (top 20%), Large (80th–50th percentile), and Small (50th–20th percentile)—excluding micro and nano stocks for liquidity considerations.

Table 1 reports out-of-sample Sharpe ratios for portfolios constructed using the Generalized UPSA, UPSA, Ridge, and KNS estimators. The Generalized UPSA consistently achieves the highest Sharpe ratios across all size groups, with performance gains that are both economically and statistically meaningful—on average, Sharpe ratios increase by approximately 10–20% relative to the standard UPSA and Ridge benchmarks. Table 2 complements these findings by reporting annualized alphas and corresponding t-statistics from regressions of the Generalized UPSA portfolio on the benchmark portfolios. The Generalized UPSA delivers positive and statistically significant alphas relative to all benchmarks.

5.2 Shrinkage Alignment in Regressions

Finally, we also show that shrinkage alignment is important for achieving good predictive performance in high-dimensional regressions. To show this, we consider a regression problem

$$y_t = \beta' X_t + \epsilon_t \tag{26}$$

and the generalized ridge shrinkage estimator

$$\beta = \left(zD + \sum_{t=1}^{T} X_t X_t' \right)^{-1} \sum_{t=1}^{T} X_t y_t. \tag{27}$$

Instead of standard ridge regression, which sets $D = I$, we consider

$$D = \alpha \text{Diag}([\rho_1^{-1}, \rho_2^{-1}, \cdots \rho_N^{-1}]) + (1-\alpha) I \tag{28}$$

where $\rho_n = Corr(y_t, X_{n,t})$ is the correlation between n'th predictor and y_t. Intuitively, the univariate correlation between each predictor and the target provides rough estimates of the predictor's individual predictability. By aligning shrinkage with the inverse of the individual correlations, the estimator penalizes potentially strong estimators less.

We simulate $N = 1000$ predictors whose β_n is drawn from distribution $\beta_n \sim N(0, b_n)$ and b_n is drawn from a Weibull distribution with shape parameter a.[20] A lower a means the ex-ante magnitude of β_n's are more evenly distributed, while a larger a means the ex-ante magnitudes of β_n's are more dispersed, which simulates a setting of strong heterogeneity in predictability. In Figure 19, we plot the out-of-sample R^2 as α and z vary across settings with different values of a. We find that when a is large, the highest out-of-sample R^2 is achieved when the shrinkage structure aligns with the inverse of individual predictor correlations. This result highlights that shrinkage alignment can also enhance performance in regression settings, particularly when predictors exhibit strong heterogeneity in their degrees of predictability.

6 Conclusion

In this paper, we highlight the critical role of shrinkage alignment in constructing efficient portfolios in high-dimensional settings. We show, both empirically and theoretically, that the conventional "one-size-fits-all" shrinkage employed in standard portfolio methods fails to accommodate heterogeneity in asset profitability, leading to out-of-sample inefficiency. By aligning shrinkage strength with empirical estimates of each asset's risk-adjusted return,

[20] The pdf of a Weibull random variable is given by $f(x;a) = ax^{a-1}e^{-x^a}$ for $x \geq 0$.

investors can mitigate this misalignment, enhance out-of-sample performance, and restore the virtue of complexity.

Our paper provides practical guidance for constructing high-dimensional portfolios. Incorporating the cross-sectional distribution of asset profitability into shrinkage design is essential for efficiency. We identify the conditions under which the virtue of complexity arises: complex models outperform simpler ones only when the marginal benefit of additional complexity exceeds the cost of estimation noise. Proper shrinkage alignment is therefore critical. More broadly, portfolio performance depends on how model design interacts with the empirical environment. While Sharpe Ratio Shrinkage mitigates misalignment, further gains may be achieved by embedding richer asset characteristics—such as covariance structures of profitability—into the shrinkage framework. This remains a promising direction for future research.

References

Ao, M., Yingying, L., and Zheng, X. (2019). Approaching mean-variance efficiency for large portfolios. *The Review of Financial Studies*, 32(7):2890–2919.

Britten-Jones, M. (1999). The sampling error in estimates of mean-variance efficient portfolio weights. *The Journal of Finance*, 54(2):655–671.

Bryzgalova, S., Pelger, M., and Zhu, J. (2024). Forest through the trees: Building cross-sections of stock returns. *Journal of Finance, Forthcoming*.

Cartea, Á., Jin, Q., and Shi, Y. (2025). The limited virtue of complexity in a noisy world. *Available at SSRN*.

Chen, A. Y. and Dim, C. (2023). High-throughput asset pricing. *arXiv preprint arXiv:2311.10685*.

Chen, A. Y., Lopez-Lira, A., and Zimmermann, T. (2022). Peer-reviewed theory does not help predict the cross-section of stock returns. *arXiv preprint arXiv:2212.10317*.

Chen, A. Y. and Zimmermann, T. (2021). Open source cross-sectional asset pricing. *Critical Finance Review, Forthcoming*.

Chen, L., Pelger, M., and Zhu, J. (2024). Deep learning in asset pricing. *Management Science*, 70(2):714–750.

Cong, L. W., Feng, G., He, J., and He, X. (2025). Growing the efficient frontier on panel trees. *Journal of Financial Economics*, 167:104024.

Da, R., Nagel, S., and Xiu, D. (2024). The statistical limit of arbitrage. Technical report, National Bureau of Economic Research.

DeMiguel, V., Garlappi, L., and Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *The review of Financial studies*, 22(5):1915–1953.

Diaconis, P. and Freedman, D. (1986). On the consistency of bayes estimates. *The Annals of Statistics*, pages 1–26.

Didisheim, A., Ke, S. B., Kelly, B. T., and Malamud, S. (2024). Apt or "aipt"? the surprising dominance of large factor models. Technical report, National Bureau of Economic Research.

Dobriban, E. and Wager, S. (2018). High-dimensional asymptotics of prediction: Ridge regression and classification. *The Annals of Statistics*, 46(1):247–279.

Giglio, S., Xiu, D., and Zhang, D. (2025). Test assets and weak factors. *The Journal of Finance*, 80(1):259–319.

Ingersoll Jr, J. E. (1984). Some results in the theory of arbitrage pricing. *The Journal of Finance*, 39(4):1021–1039.

Jensen, T. I., Kelly, B., and Pedersen, L. H. (2022). Is there a replication crisis in finance? *The Journal of Finance*.

Kan, R., Wang, X., and Zheng, X. (2024). In-sample and out-of-sample sharpe ratios of multi-factor asset pricing models. *Journal of Financial Economics*, 155:103837.

Kan, R. and Zhou, G. (2007). Optimal portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis*, 42(3):621–656.

Ke, S. (2023). The double-edged sword of data mining: Implications on asset pricing and information efficiency. *Available at SSRN 4633293*.

Kelly, B., Malamud, S., and Pourmohamadi, M. (2025). Complex rational expectations equilibria (cree). *Working Paper*.

Kelly, B., Malamud, S., and Zhou, K. (2024a). The virtue of complexity in return prediction. *The Journal of Finance*, 79(1):459–503.

Kelly, B. T. and Malamud, S. (2025). Understanding the virtue of complexity. *Available at SSRN 5346842*.

Kelly, B. T., Malamud, S., Pourmohammadi, M., and Trojani, F. (2024b). Universal portfolio shrinkage. Technical report, National Bureau of Economic Research.

Kelly, B. T., Malamud, S., and Zhou, K. (2022). The virtue of complexity everywhere. Swiss Finance Institute Research Paper 22-57, Swiss Finance Institute.

Kelly, B. T., Pruitt, S., and Su, Y. (2019). Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics*, 134(3):501–524.

Kelly, B. T. and Xiu, D. (2023). Financial machine learning. *Available at SSRN*.

Kim, S., Korajczyk, R. A., and Neuhierl, A. (2021). Arbitrage portfolios. *The Review of Financial Studies*, 34(6):2813–2856.

Kozak, S. (2020). Kernel trick for the cross-section. *Available at SSRN 3307895*.

Kozak, S., Nagel, S., and Santosh, S. (2018). Interpreting factor models. *The Journal of Finance*, 73(3):1183–1223.

Kozak, S., Nagel, S., and Santosh, S. (2020). Shrinking the cross-section. *Journal of Financial Economics*, 135(2):271–292.

Ledoit, O. and Wolf, M. (2004). Honey, i shrunk the sample covariance matrix. *The Journal of Portfolio Management*, 30(4):110–119.

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

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

Liao, Y., Ma, X., Neuhierl, A., and Shi, Z. (2023). Does noise hurt economic forecasts? *Available at SSRN 4659309*.

MacKay, D. J. C. (1994). Bayesian methods for backpropagation networks. *Models of Neural Networks*, 6:211–254.

Martin, I. W. and Nagel, S. (2022). Market efficiency in the age of big data. *Journal of financial economics*, 145(1):154–177.

Moreira, A. and Muir, T. (2017). Volatility-managed portfolios. *The Journal of Finance*, 72(4):1611–1644.

Neal, R. M. (1996). *Bayesian Learning for Neural Networks*, volume 118 of *Lecture Notes in Statistics*. Springer.

Ritov, Y., Bickel, P. J., Gamst, A. C., and Kleijn, B. J. K. (2014). The bayesian analysis of complex, high-dimensional models: Can it be coda? *Statistical Science*, 29(4):619–639.

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13(3):341–360.

Shen, Z. and Xiu, D. (2025). Can machines learn weak signals? Technical report, National Bureau of Economic Research.

Tu, J. and Zhou, G. (2011). Markowitz meets talmud: A combination of sophisticated and naive diversification strategies. *Journal of Financial Economics*, 99(1):204–215.

Wu, D. and Xu, J. (2020). On the optimal weighted l2 regularization in overparameterized linear regression. *Advances in Neural Information Processing Systems*, 33:10112–10123.

Yuan, M. and Zhou, G. (2024). Why naive diversification is not so naive, and how to beat it? *Journal of Financial and Quantitative Analysis*, 59(8):3601–3632.

A Additional Empirical Results for Section 2

A.1 Different Portfolio Construction Methodologies

In this section, we describe the alternative portfolio construction methods analyzed in Section 2.2. For each method[21], hyperparameters are selected either through leave-one-out cross-validation when computationally efficient implementations are available, or through 3-fold cross-validation when such methods are not feasible.

- **UPSA**: We follow Kelly et al. (2024b) and implement non-linear spectral shrinkage that maximizes expected quadratic utility:

$$\max_f \bar{E}\big[R_t^{\bar{\pi}(f)}\big] \; - \; \tfrac{1}{2}\,\bar{E}\big[(R_t^{\bar{\pi}(f)})^2\big] \quad \text{s.t.} \quad f(\lambda) = \sum_{i=1}^{L} w_i f_{z_i}(\lambda) = \sum_{i=1}^{L} w_i (z_i + \lambda)^{-1}, \tag{29}$$

  where $z_i \in \{10^{-7}, 10^{-6}, \ldots, 10^{3}\}$ and the weights w_i are chosen using leave-one-out cross-validation.

- **LW**: Following Ledoit and Wolf (2020), we implement non-linear spectral shrinkage that minimizes the Frobenius norm of the deviation between the empirical and population covariance matrices:

$$\min_f \; \|\bar{U}\, f(\bar{\psi})\, \bar{U}' - \Sigma\|_F.$$

  where $\bar{\Sigma} = \bar{U}\bar{\psi}\bar{U}'$ is the eigenvalue decomposition for the empirical covariance matrix.

- **LASSO**: Following Ao et al. (2019), we add an ℓ_1 penalty to the portfolio optimization problem. Specifically, the objective function is

$$\hat{w}_T = \arg\max_{w \in \mathbb{R}^N} \Big\{ w'\,\bar{E}[F_t] \; - \; \tfrac{1}{2}\, w'\,\bar{E}[F_t F_t']\, w \Big\} \quad \text{s.t.} \quad \|w\|_1 \le \alpha, \tag{30}$$

  where $\alpha \in \{10^{-7}, 10^{-6}, \ldots, 10^{3}\}$ is tuned using 3-fold cross-validation.

- **PCA**: Following Kozak et al. (2018), we construct the efficient portfolio using the principal components of the data-mined strategies. We first extract the top K principal components, and then solve the mean–variance problem in PC space:

$$\hat{w}^{PC} = \arg\max_{w \in \mathbb{R}^K} \Big\{ w'\,\bar{E}[F_t^{PC}] \; - \; \tfrac{1}{2}\, w'\,\bar{E}[F_t^{PC} F_t^{PC\prime}]\, w \Big\}, \tag{31}$$

[21] We define $\bar{E}$ to be the in-sample mean using the rolling window of size T. Given the heterogeneity in shrinkage methods, we normalize portfolio weights to ensure comparable out-of-sample return volatility across samples. Specifically, following Ledoit and Wolf (2020), we scale weights so that the implied in-sample variance after shrinkage equals the in-sample variance of the empirical eigenvalues.

where K is selected based on 3-fold cross-validation.

- **KNS**: Following Kozak et al. (2020), we implement Elastic Net estimation in the principal component space. We extract all $min(T, N)$ principal components in each rolling window, where T is the size of the rolling window and N is the number of assets. We then estimate:

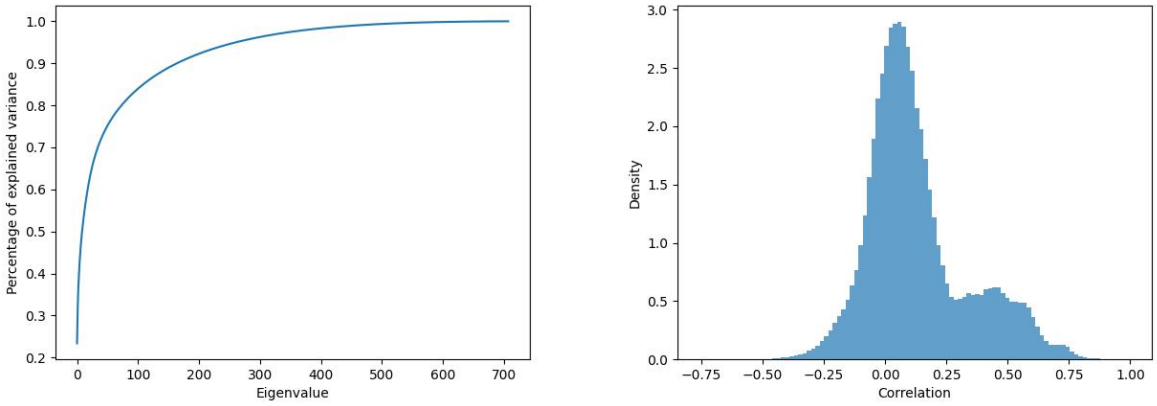
$$\hat{w}^{KNS} = \arg\max_{w \in \mathbb{R}^T} \Big\{ w' \, \bar{E}[F_t^{PC}] \; - \; \tfrac{1}{2} \, w' \, \bar{E}[F_t^{PC} F_t^{PC\prime}] \, w \; - \; z\|w\|_2 - \alpha\|w\|_1 \Big\}, \quad (32)$$

where $z, \alpha \in \{10^{-7}, 10^{-6}, \ldots, 10^{3}\}$ are selected using 3-fold cross-validation.

- **1/N**: Following DeMiguel et al. (2009), we construct the equal-weighted portfolio as

$$\hat{w} = \tfrac{1}{N} \mathbf{1}_N.$$

A.2 Additional Figures



(a) Variance Explained by Number of PCs **(b)** Off Diagonal Correlations

Figure 11: Weak Factor Structure and Correlation Evidence

Note. The left panel shows the cumulative variance explained by the principal components (PCs) of the strategy returns, ordered from largest to smallest eigenvalue and estimated over the full sample. The right panel displays the distribution of the off-diagonal elements of the covariance matrix.

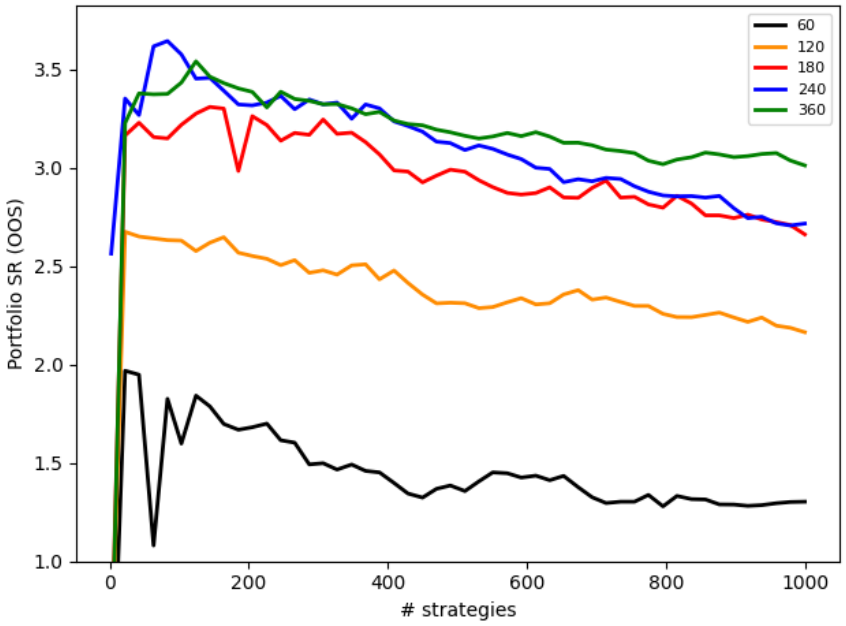


Figure 12: MSRR-ridge Portfolio with Different Rolling Windows

Note. This figure replicates Figure 1 using alternative rolling window lengths to examine the robustness of the results to the choice of estimation horizon. As in Figure 1, strategies are added in descending order of their in-sample Sharpe ratios. The ridge penalty is selected via leave-one-out cross-validation from a grid of shrinkage values $z \in \{10^{-7}, 10^{-6}, \ldots, 10^{3}\}$. All results are based on rolling windows of 60, 120, 180 240, and 360, updated monthly, over the out-of-sample period January 1994 to December 2022.

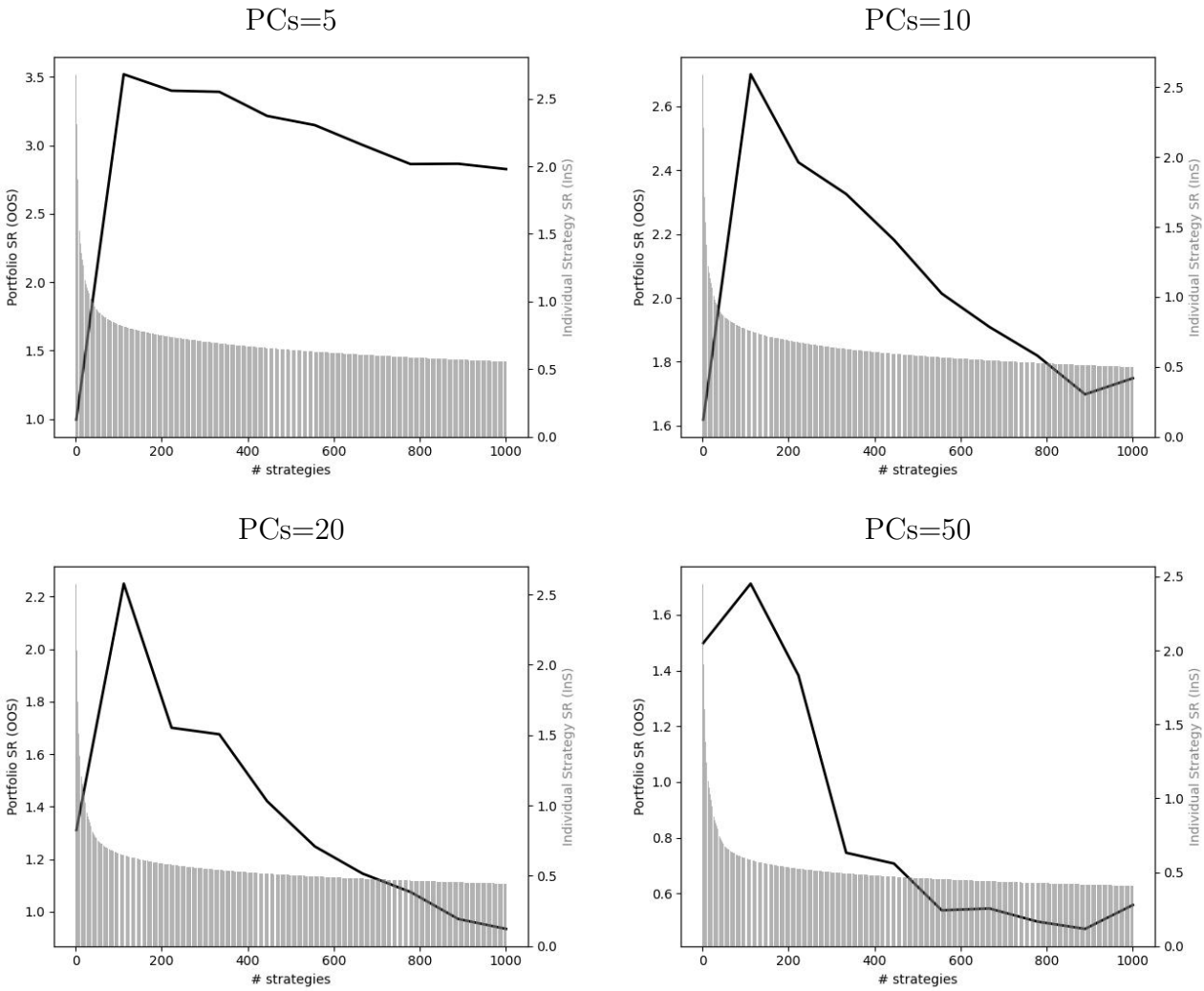


Figure 13: Adding Strategies Based on Sharpe ratios: Removing PCs

Note. This figure replicates Figure 1 using residualized strategy returns, obtained after removing the top principal components from the full-sample strategy set. Residualized strategy are added in order of their in-sample Sharpe ratios. The gray bars display the average in-sample Sharpe ratios of the residualized strategies (right axis), computed for each rank across rolling windows. The black line plots the out-of-sample Sharpe ratio of the MSRR–ridge portfolio constructed with varying numbers of residualized strategies. The ridge penalty is selected via leave-one-out cross-validation from a grid of shrinkage values $z \in \{10^{-7}, 10^{-6}, \ldots, 10^{3}\}$. Results are computed using a 360-month rolling window, updated monthly, over the out-of-sample period January 1994 to December 2022.

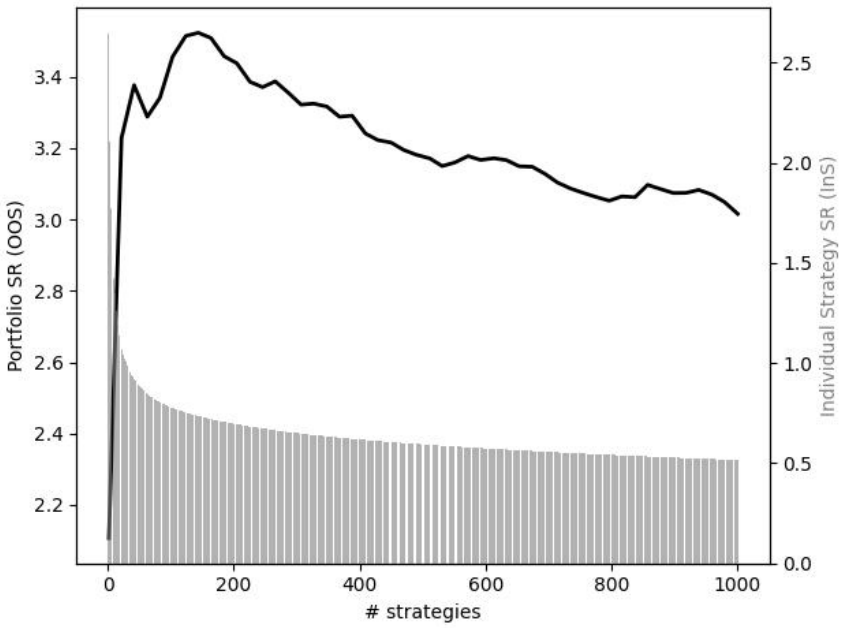


Figure 14: Adding Strategies Based on Sharpe Ratios: Ex-post Optimal Shrinkage

Note. This figure shows the out-of-sample Sharpe ratio of the MSRR–ridge portfolio as the number of strategies used in portfolio construction increases in order of their in-sample Sharpe ratio. The gray bars display the average in-sample Sharpe ratios of individual strategies (right axis), computed for each rank across rolling windows. The black line plots the out-of-sample Sharpe ratio of the portfolio for different numbers of strategies, where the ex-post optimal level of ridge shrinkage is selected from a grid of shrinkage values $z \in \{10^{-7}, 10^{-6}, \ldots, 10^{3}\}$. Results are computed using a 360-month rolling window, updated monthly, over the out-of-sample period January 1994 to December 2022.

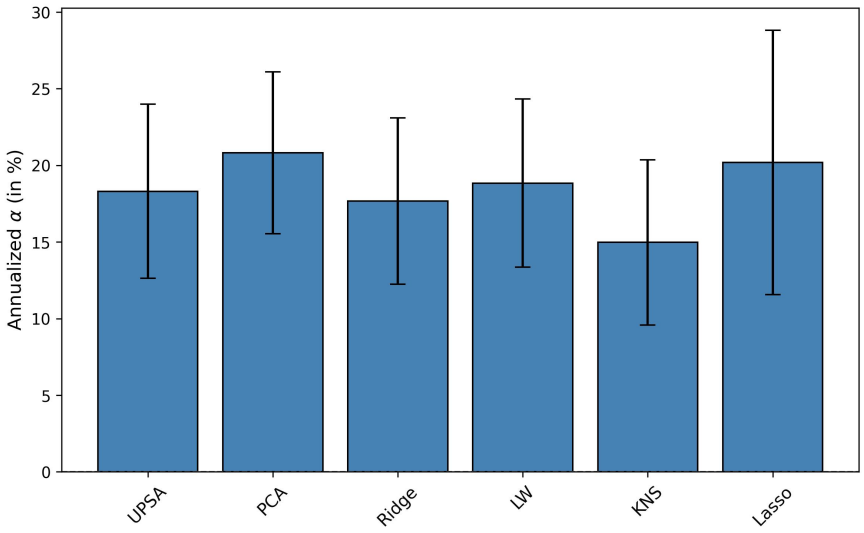
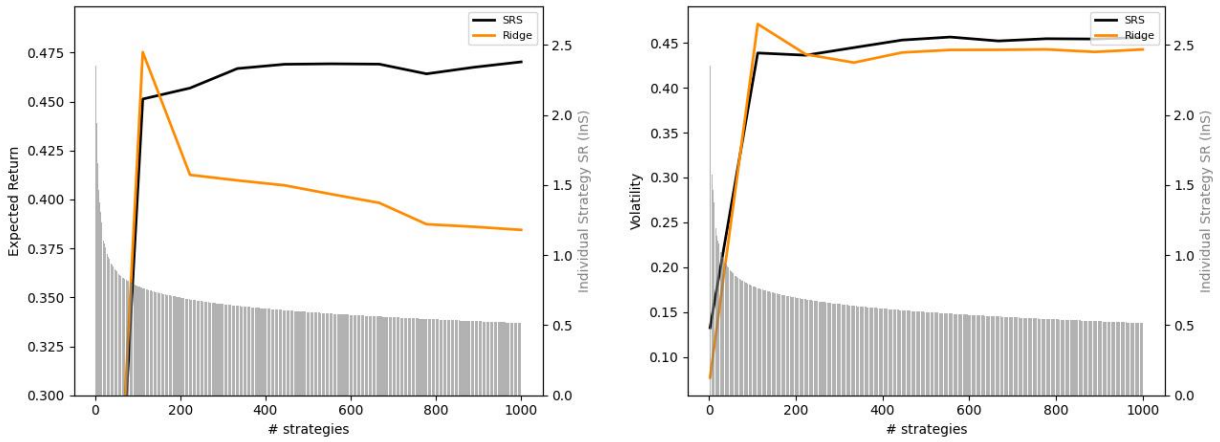


Figure 15: Negative Alpha of Portfolio with $N = 1000$ on Portfolio with $N = 100$

Note. This plot shows the annualized α when regressing the out-of-sample portfolio return with $N = 1000$ strategies on the out-of-sample portfolio return with $N = 100$ strategies. We use top N strategies based on their in-sample Sharpe ratio in each rolling window to construct the portfolios. Hyperparameters are chosen with leave-one-out cross-validation. The markers report the HAC-adjusted standard errors for five lags. Results are computed using a 360-month rolling window, updated monthly, over the out-of-sample period January 1994 to December 2022.



(a) Expected return **(b)** Volatility

Figure 16: Out-of-sample SRS vs MSRR-ridge Portfolio Expected Return and Volatility

Note. This figure shows compares the out-of-sample SRS vs MSRR-ridge portfolio expected return in panel (a) and volatility in panel (b) when strategies are added in order of their in-sample Sharpe ratios. The gray bars show the Sharpe ratio of individual strategies (right axis). The gray bars show the Sharpe ratio of individual strategies (right axis). Results are computed using a 360-month rolling window, updated monthly, over the out-of-sample period January 1994 to December 2022.

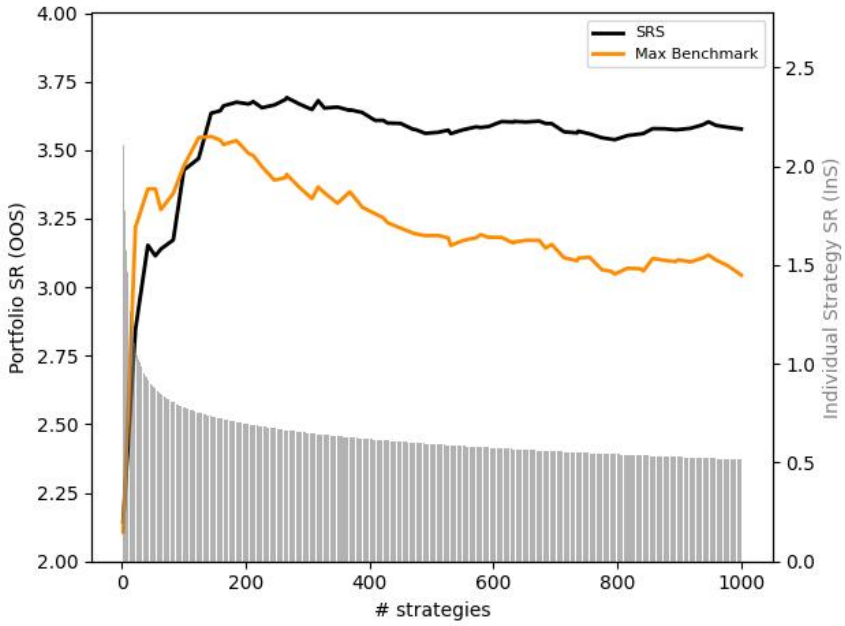


Figure 17: Out-of-sample MSRR-SRS vs Standard portfolio Methods Sharpe Ratio

Note. This figure shows the out-of-sample Sharpe ratio of MSRR-SRS portfolios and the ex-post optimal portfolios built using all methods in Section 2.1 as the number of data-mined strategies used in portfolio construction increases. Hyperparameters are is selected by leave-one-out cross-validation. For the MSRR-SRS shrinkage target $D = \text{diag}([1/\widehat{SR}_1^2, 1/\widehat{SR}_2^2, \cdots, 1/\widehat{SR}_N^2])$ we scale it to have unit L_2 norm. The gray bar shows the in-sample Sharpe ratio of individual strategies (right axis), which is averaged for each ranking across rolling windows. Results are computed using a 360-month rolling window, updated monthly, over the out-of-sample period January 1994 to December 2022.

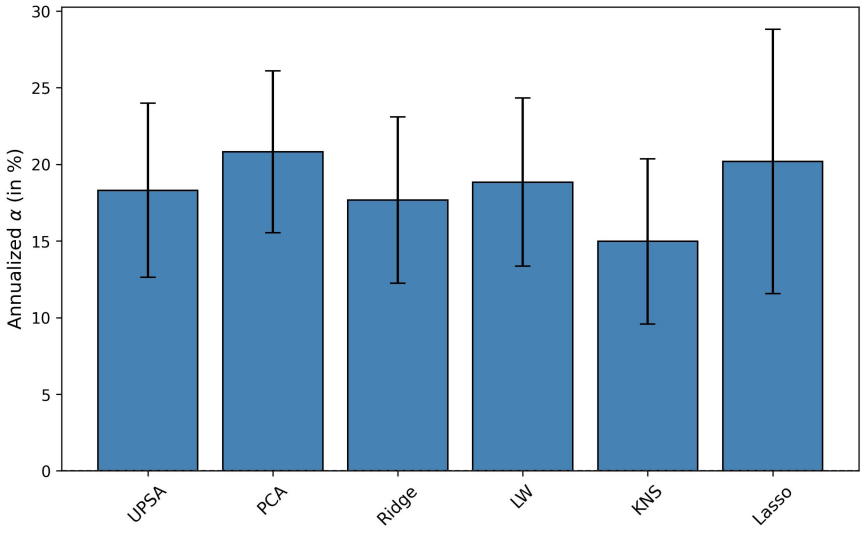
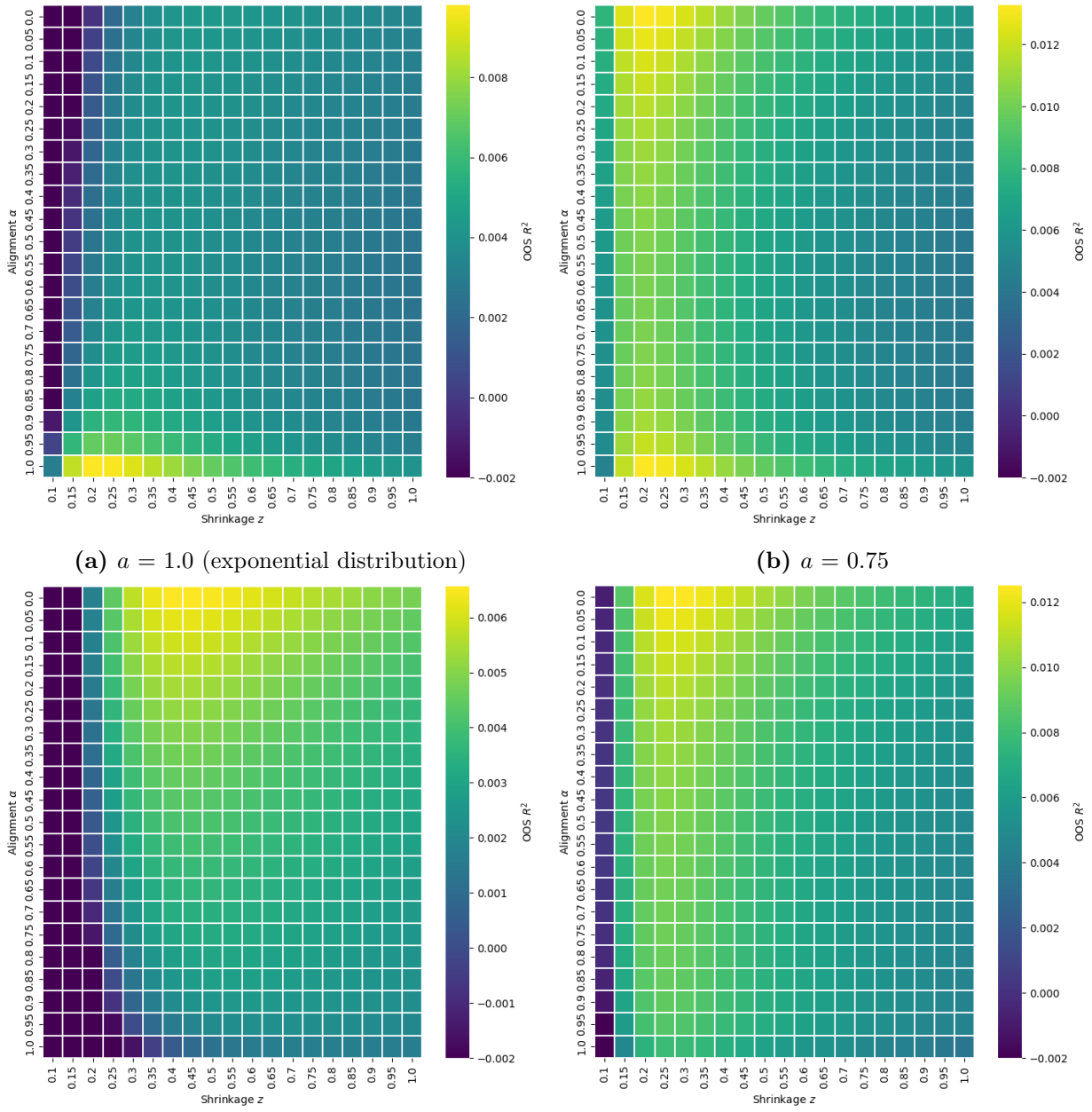


Figure 18: Alpha of MSRR-SRS Portfolios

Note. This plot shows the annualized α of MSRR-SRS Portfolios on different portfolio methods considered in Section 2.1. The markers report the HAC-adjusted standard errors with 5 lags. Results are computed using a 360-month rolling window, updated monthly, over the out-of-sample period January 1994 to December 2022.



(a) $a = 1.0$ (exponential distribution)

(b) $a = 0.75$

(c) $a = 0.5$

(d) $a = 0.25$

Figure 19: Out-of-sample R^2 in Regression Simulation

Note. This figure shows the out-of-sample R^2 of regression prediction $\widehat{\beta}' X_t$ where $\widehat{\beta}$ is estimated using 27. We simulate $N = 1000$ predictors whose β_n is drawn from distribution $\beta_n \sim N(0, b_n)$ and b_b is drawn from a Weibull distribution with shape parameter a. We simulate $X_{n,t} \sim N(0, \frac{1}{16000})$ and $\epsilon_t \sim N(0, 1)$ so the population R^2 is 6%. We estimate $\widehat{\beta}$ using an training sample size of 360 observations and compute 360 out-of-sample predictions. We conduct 1000 simulations and compute average R^2.

A.3 Random Fourier Feature Factors

Following Didisheim et al. (2024), we construct P random Fourier factors as

$$F_{t+1} = S_t' R_{t+1} \tag{33}$$

where R_{t+1} is the vector of individual stock return, and S_t is the matrix denoting random Fourier features constructed from underlying stock characteristics

$$[S_{2p-1,t}, S_{2p,t}] \in \mathbb{R}^{N_t \times 2} = [\sin(\upsilon Z_t \omega_p), \cos(\upsilon Z_t \omega_p)]', \quad \omega_p \sim \text{i.i.d } N(0, I), \quad p = 1, ..., P/2 \tag{34}$$

Each RFF basis function forms a random linear combination ω_p of the raw characteristics, which we take as the 153 characteristics constructed in Jensen et al. (2022). For each $\upsilon \in \{0.01, 0.02, \cdots 0.09, 0.1, 0.2, \cdots 1\}$, we make 20 random draws of 1000 RFF factors, and perform our main exercise described in Section 2 by gradually increase the number of factors used in the efficient portfolio in the decreasing order of their in-sample Sharpe ratio. We compute whether the out-of-sample Sharpe ratio always exhibit an increasing pattern, and if not, what is the shortfall (percentage differences in Sharpe ratio) between the interior optimal and the final portfolio using all 1000 factors.

Table 3 reports the result. We conduct the exercise separately for different size groups. We find that about 80% of the time, expanding factors by sorting on their in-sample Sharpe ratios lead to an interior optimal, suggesting that the performance is higher when only a subset of factors are used. The expected shortfall is small on average around 4%, but in some realizations the shortfall can reach to more than 28%.

	Mega	Large	Small	Micro
Exist Interior Optimum	83.42%	82.10%	79.25%	75.52%
Expected Shortfall	4.49%	4.21%	2.39%	1.48%
Maximum Shortfall	28.02%	11.24%	8.52%	7.48%

Table 3: Experiment with Random Fourier Feature Factors

Note. In this table, we sort randomly generated Fourier feature factors by their in-sample Sharpe ratios in the descending orders and increase the number of factors used in building the efficient portfolio.

B Proofs and Additional Theoretical Results

B.1 Theoretical Moments of the Feasible Ridge Portfolio

Proposition 3 (Large-N, T Sharpe of Feasible Ridge Portfolio)
Let $F_t^N \in \mathbb{R}^N$ and λ^N be as in Assumption 1, and set $c = N/T$. Define the feasible ridge return

$$R_{t+1}^{\text{feas}}(z,c) = \hat{w}_T(z,c)' F_{t+1}^N,$$

with $\hat{w}_T(z,c)$ from (5). In the limit $N,T \to \infty$ with $N/T \to c$, its out-of-sample moments satisfy

$$\lim_{N,T} \mathbb{E}\big[R_{t+1}^{\text{feas}}(z,c)\big] = \mathcal{E}\big(Z^*;\, \gamma(c)\big)$$

$$\lim_{N,T} \text{Var}\big[R_{t+1}^{\text{feas}}(z,c)\big] = \mathcal{V}\big(Z^*;\, \gamma(c)\big) + G\big(z;c\big)\left[\big(1 - \mathcal{E}\big(Z^*;\, \gamma(c)\big)\big)^2 + \mathcal{V}\big(Z^*;\, \gamma(c)\big)\right],$$

$$\lim_{N,T} \frac{\text{Var}\big[R_{t+1}^{\text{feas}}(z,c)\big]}{\mathbb{E}\big[R_{t+1}^{\text{feas}}(z,c)\big]^2} \equiv \frac{1}{\text{SR}(R_{t+1}^{\text{feas}}(z,c))^2} = \big(1+G(z;c)\big)\,\frac{1}{\gamma(c)} + G(z;c)\left(\frac{1-\mathcal{E}\big(Z^*;\,\gamma(c)\big)}{\mathcal{E}\big(Z^*;\,\gamma(c)\big)}\right)^2.$$

where $\mathcal{E}$ and $\mathcal{V}$ are from Proposition 1, and

$$Z^* = z\,(1+\xi(z;c)), \quad G(z;c) = \frac{\partial}{\partial z}(z\xi(z;c)), \quad \xi(z;c) = cm(-z;c),$$

with the Marčenko–Pastur Stieltjes transform in closed form

$$m(-z;c) = \frac{-((1-c)+z) + \sqrt{((1-c)+z)^2 + 4cz}}{2cz}$$

Proposition 3 is a special case of Theorem 3 in Didisheim et al. (2024). It characterizes the feasible portfolio behavior as a function of ridge shrinkage parameter z, complexity c, and the norm of expected return of the strategies γ. The formula for $\mathbb{E}\big[R_{t+1}^{\text{feas}}(z,c)\big]$ shows how the feasible portfolio expected return is related to that of the infeasible portfolio through Z^*, which is the "implicit shrinkage" of the feasible estimator.[22] Holding z fixed, a rise in complexity to $c > 0$ induces additional bias in the estimator and drives down the expected return of the portfolio. In other words, the challenge of learning in a complex setting is equivalent to knowing the true moments but being forced to use an unduly large shrinkage.

The formula for $\text{Var}\big[R_{t+1}^{\text{feas}}(z,c)\big]$ has two terms. The first term is related to the implicit shrinkage effect in the same way as in the expected return. The second term represents a

[22] See Didisheim et al. (2024) for more discussion on implicit shrinkage.

different phenomenon that captures sampling variation that exists even in the large T limit. It is a pure variance effect governed by the function $G(z;c)$, independent of the expected returns of strategies, and only depends on the eigenvalue distribution of the strategies. For a simple estimation problem with $c = 0$, there are infinitely more observations than parameters, so the portfolio weight estimator converges to a non-random limit, thus $G(z;0) = 0$. However, when $c > 0$, sampling variation survives because the number of parameters is too large to be accurately informed from the data, which means $G(z;c) > 0$ for $c > 0$. Using large shrinkage z will stabilize the variance of the portfolio, and this term will go to zero.

B.2 Proofs

Proof of Proposition 1 The infeasible portfolio weight is obtained by plugging in population moments from Assumption 1 into the ridge portfolio solution (5)

$$\begin{aligned} w^{infeas}(z,c) &= \left(E[F_t^N F_t^{N\prime}] + zI_N\right)^{-1} E[F_t^N] \\ &= (\lambda^N \lambda^{N\prime} + I_N + zI_N)^{-1}\lambda_N \end{aligned} \tag{35}$$

Therefore, the expected return for the infeasible portfolio is given by

$$\begin{aligned} E[R_{t+1}^{infeas}(z;c)] &= E[w^{infeas}(z,c)' F_{t+1}] \\ &= \lambda_N'(\lambda_N\lambda_N' + I_N + zI_N)^{-1}\lambda_N \\ &= \frac{||\lambda_N||^2}{1+z+||\lambda_N||^2} \\ &= \frac{\gamma(c)}{1+z+\gamma(c)} \equiv \mathcal{E}(z;\gamma(c)) \end{aligned} \tag{36}$$

The variance of the infeasible portfolio is

$$\begin{aligned} Var(R_{t+1}^{infeas}(z;c)) &= E[R_{t+1}^{infeas}(z;c)^2] - \mathcal{E}(z;\gamma(c))^2 \\ &= E[w^{infeas}(z,c)' F_{t+1}F_{t+1}' w^{infeas}(z,c)] - \mathcal{E}(z;\gamma(c))^2 \\ &= w^{infeas}(z,c)' E[F_{t+1}F_{t+1}'] w^{infeas}(z,c) - \mathcal{E}(z;\gamma(c))^2 \\ &= w^{infeas}(z,c)'(\lambda^N\lambda^{N\prime} + I_N) w^{infeas}(z,c) - \mathcal{E}(z;\gamma(c))^2 \\ &= \frac{||\lambda_N||^4 + ||\lambda_N||^2}{(1+z+||\lambda_N||^2)^2} - \mathcal{E}(z;\gamma(c))^2 \\ &= \frac{\gamma(c)}{(1+z+\gamma(c))^2} \equiv \mathcal{V}(z;c) \end{aligned} \tag{37}$$

Therefore, the squared Sharpe ratio is given by

$$SR(R_{t+1}^{infeas}(z;c))^2 = \frac{E[R_{t+1}^{infeas}(z;c)]^2}{Var[R_{t+1}^{infeas}(z;c)]} = \gamma(c) \tag{38}$$

Proof of Lemma 1 Given

$$m(-z;c) = \frac{-((1-c)+z) + \sqrt{((1-c)+z)^2 + 4cz}}{2cz}$$

We have

$$\xi(z;c) = \frac{c(1-zm(-z;c))}{1-c(1-zm(-z;c))} = \frac{-((1-c)+z) + \sqrt{((1-c)+z)^2 + 4cz}}{2z} \tag{39}$$

and therefore

$$Z^* = z(1+\xi(z;c)) = \frac{-((1-c)-z) + \sqrt{((1-c)+z)^2 + 4cz}}{2} \tag{40}$$

We also have

$$G(z;c) = \frac{\partial}{\partial z}(z\xi(z;c)) = -\frac{1}{2} + \frac{1+c+z}{2\sqrt{(1-c+z)^2+4cz}} \tag{41}$$

Now we compute $1/SR(R_{t+1}^{feas}(z,c))^2$. First, notice that

$$\left(\frac{1-\mathcal{E}(Z^*;\gamma(c))}{\mathcal{E}(Z^*;\gamma(c))}\right)^2 = \left(\frac{1+Z^*}{\gamma(c)}\right)^2 \tag{42}$$

and thus

$$\frac{1}{SR(R_{t+1}^{feas}(z,c))^2} \propto G(z;c) + \left(\frac{(1+Z^*)^2}{\gamma(c)}\right) G(z;c) \tag{43}$$

The second term is

$$\left(\frac{(1+Z^*)^2}{\gamma(c)}\right) G(z;c) = \frac{c(1+c+z+\sqrt{((1-c)+z)^2+4cz})}{4\gamma(c)\sqrt{((1-c)+z)^2+4cz}} \tag{44}$$

and thus

$$\begin{aligned}\frac{1}{SR(R_{t+1}^{feas}(z,c))^2} &\propto G(z;c) + \left(\frac{(1+Z^*)^2}{\gamma(c)}\right) G(z;c) \\ &= \frac{2\gamma(c)(1+c+z) + c(1+c+z+\sqrt{((1-c)+z)^2+4cz})}{\sqrt{((1-c)+z)^2+4cz}} \\ &\propto \frac{1+c+z}{\sqrt{((1-c)+z)^2+4cz}} \equiv f(z;c)\end{aligned} \tag{45}$$

Take partial derivative of $f(z;c)$ with respect to z, we find that

$$\frac{\partial f(z;c)}{\partial z} = -\frac{4c}{\sqrt{((1-c)+z)^2+4cz}^3} < 0 \tag{46}$$

for all z, which means $SR(R_{t+1}^{feas}(z,c))$ is strictly increasing in z. Therefore, the optimal shrinkage is $z \to \infty$. □

Proof of Proposition 2 From Proposition 3 and the proof of Lemma 1, we know

$$\begin{aligned}\frac{1}{\mathrm{SR}(R_{t+1}^{\mathrm{feas}}(z,c))^2} &= \big(1+G(z;c)\big)\,\frac{1}{\gamma(c)} \;+\; G(z;c)\left(\frac{1-\mathcal{E}\big(Z^*;\,\gamma(c)\big)}{\mathcal{E}\big(Z^*;\,\gamma(c)\big)}\right)^2 \\ &= \frac{1}{\gamma(c)} + \frac{G(z;c)}{\gamma(c)} + \frac{1}{\gamma(c)^2}G(z;c)(1+Z^*)^2\end{aligned} \tag{47}$$

From (41) we conclude that $G(z;c) \to 0$ as $z \to \infty$. Intuitively, as shrinkage goes to infinity, the implicit shrinkage also goes to infinity, therefore the variance contribution goes to zero. For the second term, from direct calculation of $G(z;c)(1+Z^*)^2$ using (41) and (40) we have

$$G(z;c)(1+Z^*)^2 \to c \tag{48}$$

as $z \to \infty$. Putting everything together, we conclude that

$$\lim_{z\to\infty} \mathrm{SR}(R_{t+1}^{feas}(z;c)^2) = \frac{\gamma(c)^2}{\gamma(c)+c} \tag{49}$$

Now for the derivative we have

$$\begin{aligned}\frac{\partial \mathrm{SR}(R^{\mathrm{feas}}_{t+1}(z,c))^2}{\partial c} &= \frac{\partial \frac{\gamma^2}{\gamma+c}}{\partial c} = \frac{2\gamma'\gamma(\gamma+c)-(\gamma'+1)\gamma^2}{(\gamma+c)^2} \\ &= \frac{\gamma(\gamma'(\gamma+2c)-\gamma)}{(\gamma+c)^2} \\ &\equiv \gamma'(\gamma+2c)-\gamma \end{aligned} \tag{50}$$

where we have defined $\frac{\partial \gamma}{\partial c} = \gamma'$ for notational convenience. Therefore Sharpe declines if

$$\gamma' < \frac{\gamma}{\gamma+2c} \tag{51}$$

Second derivative. From Equation (50) we have the first derivative, and we proceed to take the second derivative: Define $N(c) = \gamma(\gamma'(\gamma+2c)-\gamma)$. Therefore

$$\begin{aligned} N'(c) &= \gamma'(\gamma'(\gamma+2c)-\gamma) + \gamma(\frac{\partial}{\partial c}(\gamma'(\gamma+2c)-\gamma)) \\ &= \gamma'(\gamma'(\gamma+2c)-\gamma) + \gamma(\gamma''(\gamma+2c)+\gamma'(\gamma'+2)-\gamma') \\ &= 2(\gamma')^2(\gamma+c)+\gamma''\gamma(\gamma+2c) \end{aligned}$$

$$\begin{aligned}\frac{\partial^2 \mathrm{SR}(R^{\mathrm{feas}}_{t+1}(z,c))^2}{\partial c^2} &= \frac{N'(c)(\gamma+c)^2 - 2N(c)(\gamma+c)(\gamma'+1)}{(\gamma+c)^4} \\ &\equiv N'(c)(\gamma+c) - 2N(c)(\gamma'+1) \\ &= (2(\gamma')^2(\gamma+c)+\gamma''\gamma(\gamma+2c))(\gamma+c) - 2\gamma(\gamma'(\gamma+2c)-\gamma)(\gamma'+1) \\ &= \gamma\,\gamma''\,(\gamma+c)(\gamma+2c) + 2(\gamma - c\,\gamma')^2 + 2(\gamma - c\,\gamma')^2 \end{aligned}$$

For the function to be concave we would need

$$\gamma'' < \frac{2(\gamma - c\,\gamma')^2}{\gamma\,(\gamma+c)(\gamma+2c)} \tag{52}$$

□

Proof of Lemma 2 The second part of Lemma 2 is directly implied in the proof of Proposition 2. For the first part, since $\frac{\partial \gamma}{\partial c} = T\bar{\gamma}^2$, we have

$$\bar{\gamma}^2 > \frac{\gamma(c)}{\gamma(c)+2c} = \frac{T\bar{\gamma}^2 c}{T\bar{\gamma}^2 c + 2c} = \frac{T\bar{\gamma}^2}{T\bar{\gamma}^2+2}. \tag{53}$$

which is always satisfied.