Large Scale Mean–Variance Investing Via Nuclear Hedging

RASMUS LÖNN ALBERTO QUAINI MING YUAN*

First version: January 2025
This version: December 13, 2025

Preliminary work: Please do not circulate

Abstract

We propose a new approach for constructing global minimum–variance and mean–variance efficient portfolios in large asset markets. Instead of estimating the mean vector and precision matrix of returns separately, we express both objects through the coefficients and residual variances of a system of multi–response hedging regressions, in which each asset is optimally hedged using all others. We show that, under general (possibly approximate) factor structures, the matrix of hedging portfolio returns is low-rank: a small number of common hedging portfolios spans most of the systematic comovement in returns. This result motivates a penalized reduced–rank hedging estimator that (i) enforces a low–dimensional hedging structure, (ii) shrinks excessive hedging coefficients promoting diversified portfolios, and (iii) regularizes mean estimates toward economically plausible targets. This approach delivers a dense but structured precision matrix that reflects realistic factor exposures without imposing hard sparsity or orthogonality. In simulations and empirical applications with both $N < T$ and $N > T$, the resulting portfolios are stable, well diversified, and exhibit substantially lower realized volatility and higher out–of–sample Sharpe ratios than leading benchmark methods.

Keywords: Portfolio choice; Mean–variance portfolios; Hedging regressions; Reduced–rank estimation; Nuclear norm regularization; High–dimensional covariance
JEL codes: G11 (Portfolio Choice); C58 (Financial Econometrics); C13 (Estimation); C55 (Large Data Sets); C61 (Optimization Techniques)

*Rasmus Lönn (email: `lonn@ese.eur.nl`) and Alberto Quaini (email: `quaini@ese.eur.nl`) are with Erasmus University of Rotterdam, Erasmus school of Economics, Econometric Institute. Ming Yuan (email: `Ming.Yuan@columbia.edu`) is with Columbia University, Department of Statistics and Data Science Institute.

1 Introduction

Minimum–variance portfolios are widely used as simple yet powerful benchmarks for mean–variance investing and are known to deliver high ex post efficiency in many empirical applications (DeMiguel, Garlappi, and Uppal, 2009). Their construction rests on two key ingredients: the vector of expected excess returns and the precision (inverse covariance) matrix of asset returns. In large asset markets, estimating these objects reliably is inherently difficult. Sample means are noisy, sample covariance matrices become ill-conditioned, and most existing regularization schemes treat the mean and the covariance separately. As a result, small perturbations in the data can translate into large swings in portfolio weights, especially when the number of assets is comparable to or exceeds the length of the return history (Jobson and Korkie, 1981; Best and Grauer, 1991; Kan and Zhou, 2007). In practice, this instability often manifests itself in frequent and sizable rebalancing, amplifying transaction costs and degrading net performance once realistic trading frictions are taken into account (Kirby and Ostdiek, 2012).

This paper proposes a different perspective on the joint estimation of the mean vector and the precision matrix. Rather than treating them as separate moment objects, we exploit an exact representation that links both to a system of *hedging regressions* in the spirit of Stevens (1998). In this system, each asset is regressed on the remaining assets, and the regression coefficients are interpreted as the weights of hedging portfolios that minimize asset–specific risk. A simple linear–algebra identity shows that the coefficients of these hedging regressions and the variances of their residuals jointly determine both the mean vector and the precision matrix of returns. Consequently, all mean–variance efficient portfolios can be recovered from the parameters of a single multi–response regression, without ever inverting a high–dimensional covariance estimate.

The central idea of the paper is to regularize this joint hedging system directly. We develop a *penalized reduced–rank* (PRR) hedging regression estimator that simultaneously: (i) imposes a low–rank structure on the matrix of hedging portfolio returns and (ii) stabilizes the mean vector by shrinking the regression intercepts toward a target such as the cross-sectional grand mean, thereby addressing the main challenges of high-dimensional estimation. The key device is a penalty on the *nuclear norm* of the hedging return matrix, that is, on the sum of its singular values. This encourages the time series of hedging portfolio returns to lie in a low-dimensional subspace, so that a small number of diversified hedging portfolios span the dominant sources of systematic risk in the cross-section. On top of this, a Frobenius penalty on the hedging coefficients shrinks individual hedging positions toward more moderate levels. This directly targets implementation stability: it curbs extreme long–short exposures, mitigates the sensitivity of weights to small fluctuations in the data, and thus reduces the impact of trading frictions on realized performance.

A major contribution of the paper is to show that the main, low-rank regularization term is not merely convenient but also structurally justified. Under a broad class of approximate factor structures—including pervasive and weak factors and allowing for general idiosyncratic spectra—we prove that the matrix of hedging portfolio returns is *nearly reduced–rank*: up to a noise term of moderate size, it can be approximated by a matrix of rank equal to the number of systematic factors. Intuitively, when returns are driven by a limited number of underlying risk factors, only a few "core" hedging portfolios are needed to span the relevant risk space for a mean–variance investor. Additional hedging portfolios mainly load on idiosyncratic variation and contribute little to the shape of the efficient frontier. This near low-rank property provides an economically grounded rationale for penalizing the nuclear norm of the hedging return matrix and differentiates our approach from ad hoc factor or dimension-reduction techniques.

Our methodology also relates to, and departs from, existing approaches that regularize the precision matrix by imposing sparsity. A growing literature builds on Stevens' decomposition to estimate sparse precision matrices or sparse hedging systems using, for instance, the graphical Lasso (Goto and Xu, 2015) or nodewise Lasso regressions (Callot et al., 2021). These methods have appealing features: they are computationally tractable in high dimensions, they produce interpretable conditional-dependence graphs, and they can lead to parsimonious hedging portfolios in which each asset is hedged using only a subset of the remaining assets. From a statistical perspective, sparsity provides a clear mechanism to control complexity and to avoid overfitting.

From a financial perspective, however, exact sparsity of the precision matrix is a strong and often implausible assumption. Zero entries in the precision matrix correspond to exact conditional independence relationships between assets. In realistic factor models, pervasive macroeconomic or sectoral factors induce dense, though structured, conditional dependence in returns, even when idiosyncratic risk is uncorrelated. Enforcing many zero entries in the precision matrix therefore amounts to imposing strong restrictions on the factor structure (for example, many assets with zero factor loadings or artificially segmented markets). Moreover, sparse hedging portfolios may under-utilize available diversification opportunities, relying on a small set of hedging assets and becoming overly sensitive to idiosyncratic noise. Finally, sparse methods are typically built from separate nodewise regressions, which do not explicitly exploit the joint structure of hedging opportunities across assets.

Our PRR approach offers an alternative regularization principle that is more closely aligned with standard asset–pricing intuition. Rather than forcing most conditional correlations to zero, we allow the precision matrix and the hedging coefficients to be dense, but constrain their *effective dimension* through low–rank structure in the hedging returns and shrink their overall magnitude. This matches the idea that many assets may be connected through common factors, yet only a small number of latent hedging directions are needed to capture those connections for portfolio construction. In this sense, our method replaces hard sparsity with an economically motivated low-rank structure that preserves diversification while keeping the model parsimonious.

A related class of alternatives relies on principal components analysis (PCA) to regularize the covariance and precision matrices (e.g. Ledoit and Wolf, 2003). PCA approximates the sample covariance by retaining the leading eigenvectors and adding a diagonal idiosyncratic component, yielding a precision matrix that is "diagonal plus low rank" after inversion. This construction is natural from a variance-explanation standpoint: the leading principal components maximize the proportion of total variance captured. However, minimum-variance portfolios depend on the precision matrix, whose eigenstructure emphasizes low-variance directions that are not necessarily well captured by the leading principal components. The hedging system implied by a PCA-based precision matrix is thus an algebraic byproduct of spectral truncation, not the solution to an economically motivated hedging problem. By targeting the hedging regressions directly, our PRR estimator aligns the regularization with the investor's objective of risk reduction rather than with variance representation per se.

On the implementation side, the PRR estimator is defined by a convex optimization problem that combines a least-squares loss with nuclear, Frobenius and ridge penalties, and a linear no–self-hedging constraint on the hedging coefficients. The problem remains tractable in large panels thanks to a tailored alternating direction method of multipliers (ADMM) algorithm. Each iteration reduces to a small number of simple building blocks: a ridge update for the intercepts (and thus the mean), a singular-value soft-thresholding step for the hedging returns, a linear solve for the hedging coefficients, and a projection that enforces the zero diagonal. We also discuss practical tuning of the three regularization parameters: we propose complementary selection criteria that are statistically oriented (model-selection style), financially oriented (based on portfolio variance and Sharpe ratio), and cross-validated, allowing the practitioner to balance statistical fit and portfolio performance in a transparent way.

Empirically, we show that the proposed estimator delivers stable and well-diversified minimum-variance and mean–variance portfolios across a range of datasets and dimensions. Relative to benchmark approaches based on sample, shrinkage, PCA, or sparse hedging methods, our portfolios achieve competitive or lower volatility and higher realized Sharpe ratios while relying on smoother and more diversified allocations. These gains can be traced back to the joint treatment of mean and precision via the hedging system and to the economically motivated low-rank regularization of the hedging returns.

The remainder of the paper is organized as follows. Section 2 reviews the related literature on precision matrix estimation, hedging regressions, and factor-based portfolio methods. Section 4 develops the hedging regression representation of the mean and precision, establishes the near low-rank structure of hedging returns under general factor models, and introduces our regularized reduced–rank hedging estimator together with its computational algorithm and tuning procedures. Section 6 presents the empirical analysis, benchmarks our approach against competing methods, and documents its performance across small, medium, and large cross-sections. Finally, Section 7 concludes and outlines directions for future research.

2 Literature review

There is a large literature documenting that the mean-variance optimal allocation of Markowitz (1952) obtains low out-of-sample performance. Merton (1980) highlights the empirical challenges associated with estimating expected returns, and Jobson and Korkie (1980) demonstrates the finite sample performance of mean-variance allocations using sample moments as plug-in estimators. Michaud (1989) highlights that the mean-variance portfolios tend to overweight assets with high estimation error. This makes the portfolios unstable and prone to extreme positions, undermining their financial performance. Considering a large set of econometric approaches, DeMiguel et al. (2009) find that it is hard to establish a single approach that outperforms an equally weighted portfolio consistently across datasets. These shortcomings are due to parameter uncertainty, which generates estimation risk in the portfolio weights. Assuming a small cross-section of assets and multivariate normally distributed returns, Kan and Zhou (2007) derives the magnitude of the expected economic losses due to parameter uncertainty. In this paper we propose a novel econometric approach tailored for portfolio formation that provides reliable out-of-sample performance.

There is a multitude of approaches to robust mean-variance optimization explored in the literature.[1] Li (2015) and Yen (2016) cast the portfolio optimization into a linear regression framework, which enables shrinkage on the portfolio weights to promote sparse portfolio allocations. For a dense portfolio, the allocation can be made more robust by combining different allocation rules, for example see Kan and Zhou (2007), DeMiguel et al. (2009), and Tu and Zhou (2011). These portfolio combinations can similarly be understood as a form of shrinkage of the portfolio weights. Alternatively, constraints can be imposed on the uncertainty associated with the portfolios financial performance, thus steering the optimization towards a stable and effective allocation of wealth. (Ban et al., 2018). However, in this paper our focus is instead on the moments of returns, and we impose statistical regularization on the plug-in estimates of the portfolios.[2]

Our methodology emphasizes the estimation of the inverse covariance matrix of returns. There is an extensive literature focused on regularization of the covariance matrix. Ledoit and Wolf (2003) propose estimators which linearly shrink the sample covariance matrix towards a biased, but more stable, target matrix. This approach is extended by Ledoit and Wolf (2012) to enable non-linear shrinkage. Our focus is on cases where the number of assets is allowed to increase beyond the length of the time series. Ledoit and Wolf (2004, 2015) also emphasize this challenge within a statistical context, while the portfolio optimization challenge is specifically highlighted in Ledoit and Wolf (2017).[3] Other works, including Fan et al. (2013) and De Nard et al. (2021), include factor models that use latent and observed factors for covariance estimation and portfolio optimization.

[1] To keep our discussion brief, we focus our attention on the estimators without an explicit focus on time-variation in the moments of returns. For a recent evaluation of state-of-the-art dynamic methods within a portfolio optimization context, see Moura et al. (2020).

[2] DeMiguel et al. (2013) provides a comprehensive analysis of the magnitude of statistical shrinkage for optimal weights and different moments of returns in the context of portfolio optimization.

[3] See Ledoit and Wolf (2022) and references therein for an exhaustive discussion of covariance matrix estimators.

The estimator we propose directly targets the inverse covariance matrix, circumventing the inversion operation, and addresses estimation uncertainty through statistical regularization. The financial interpretation is that we estimate a set of hedging portfolios that span the joint variation across a set of assets. Other works that explore shrinkage estimation of the precision matrix include Kourtis et al. (2012), Goto and Xu (2015), and Callot et al. (2021). The latter two propose intuitive estimators that impose shrinkage and promote sparsity. Sparsity in this context carries important conceptual and practical implications. Conceptually, a zero entry (i, j) in the precision matrix requires that assets i and j be conditionally uncorrelated given all other assets. In a factor-driven market, imposing strict sparsity restricts the presence of pervasive factors—those that load on most assets. Practically, portfolios constructed from sparse hedging portfolios run the risk of being sensitive to idiosyncratic fluctuations in the hedging assets. Our methodology addresses these points by drawing on the insight that effectively hedging a set of assets often requires diversified (i.e., dense) portfolios, but those portfolios can be restricted to a lower-dimensional space. In other words, if there is sufficient multicollinearity or common factor structure in returns, a few distinct hedging portfolios suffice to mimic the majority of the assets' risk.

The methodology that we propose promotes portfolios that can be maintained with low turnover. The degree of turnover controls the costs associated with the portfolio and can significantly undermine out-of-sample performance. The broad empirical study of DeMiguel et al. (2009) reveals that the turnover of mean-variance efficient portfolios is often extraordinarily large relative to simple benchmark strategies. Likewise, they find that the turnover associated with minimum-variance portfolios can also be large. This is further highlighted by Kirby and Ostdiek (2012), who demonstrate the efficacy of straightforward volatility timing without regard for hedging opportunities. An alternative approach is to introduce turnover regularization in the optimization problem. This is explored by Ledoit and Wolf (2025) within the class of Markowitz applications. Hautsch and Voigt (2019) study Markowitz optimization with consideration of transaction costs and show the connection between this optimization problem and shrinkage estimation. The methodology that we propose regulates turnover through the calibration of the hyperparameters. Since our estimation targets the precision matrix directly, we are able to calibrate the hyperparameters using the composition of the portfolio. We use this to regulate deviations from allocations that are known, a priori, to have low turnover.

3 Notation

All random variables and stochastic processes are defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Expectations and variances under $\mathbb{P}$ are denoted by $\mathbb{E}[\cdot]$ and $\mathbb{V}ar(\cdot)$, respectively. For random vectors $\boldsymbol{X} \in \mathbb{R}^N$, we write $\mathbb{V}ar(\boldsymbol{X})$ for the $N \times N$ covariance matrix of $\boldsymbol{X}$ under $\mathbb{P}$.

We use boldface symbols for vectors and matrices, and plain letters for scalars. For a vector $\boldsymbol{v} \in \mathbb{R}^N$, its ith component is v_i. For a matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, its (i, j) entry is A_{ij}. The transpose of a matrix or vector $\boldsymbol{A}$ is denoted by $\boldsymbol{A}^\top$. The i-th column and j-th row of a matrix $\boldsymbol{A}$ are denoted $\boldsymbol{A}^{(i)}$ and $\boldsymbol{A}_j$, respectively, and considered column vectors. We use $\boldsymbol{I}$ to denote the identity matrix of conformable dimension, and we use $\boldsymbol{0}$ ($\boldsymbol{\iota}$) for vectors or matrices of zeros (ones) of conformable dimension.

Given $\boldsymbol{A} \in \mathbb{R}^{N\times N}$, we write $\text{diag}(\boldsymbol{A})$ for the N–dimensional vector collecting the diagonal entries of $\boldsymbol{A}$, and $\text{diag}(\boldsymbol{v})$ for the diagonal matrix with vector $\boldsymbol{v}$ on its main diagonal. The rank of a matrix $\boldsymbol{A}$ is denoted by $\text{rank}(\boldsymbol{A})$, and its trace by $\text{Tr}(\boldsymbol{A})$. For symmetric matrices $\boldsymbol{A} \in \mathbb{R}^{N\times N}$, we denote by $\lambda_{\min}(\boldsymbol{A})$ and $\lambda_{\max}(\boldsymbol{A})$ the smallest and largest eigenvalues of $\boldsymbol{A}$, respectively. We write $\boldsymbol{A} \succ 0$ if $\boldsymbol{A}$ is symmetric positive definite and $\boldsymbol{A} \succeq 0$ if $\boldsymbol{A}$ is symmetric positive semidefinite.

For vectors $\boldsymbol{v} \in \mathbb{R}^N$, the ℓ_2 (Euclidean) norm is $\|\boldsymbol{v}\|_2 := (\sum_{i=1}^N v_i^2)^{1/2}$, and the ℓ_1 norm is $\|\boldsymbol{v}\|_1 := \sum_{i=1}^N |v_i|$. For matrices $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, the Frobenius norm is $\|\boldsymbol{A}\|_F := (\sum_{i=1}^m \sum_{j=1}^n A_{ij}^2)^{1/2}$. When $\|\cdot\|_2$ is applied to matrices, it denotes the operator norm induced by the Euclidean norm, i.e. $\|\boldsymbol{A}\|_2 := \sup_{\boldsymbol{v}\neq\boldsymbol{0}} \frac{\|\boldsymbol{A}\boldsymbol{v}\|_2}{\|\boldsymbol{v}\|_2}$, which coincides with the largest singular value of $\boldsymbol{A}$. The nuclear norm of $\boldsymbol{A}$ is $\|\boldsymbol{A}\|_* := \sum_{j=1}^{\mathrm{rank}(\boldsymbol{A})} \sigma_j(\boldsymbol{A})$, where $\sigma_1(\boldsymbol{A}) \geq \cdots \geq \sigma_{\mathrm{rank}(\boldsymbol{A})}(\boldsymbol{A}) > 0$ are the nonzero singular values of $\boldsymbol{A}$.

For a symmetric matrix $\boldsymbol{A}$, the spectral decomposition is written as $\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^\top = \sum_{i=1}^N \lambda_i \boldsymbol{v}^{(i)}\boldsymbol{v}^{(i)\top}$, where $\lambda_1, \ldots, \lambda_N$ are the eigenvalues (ordered as needed) and $\boldsymbol{v}^{(1)}, \ldots, \boldsymbol{v}^{(N)}$ are orthonormal eigenvectors. For a general matrix $\boldsymbol{M} \in \mathbb{R}^{T\times N}$, a singular value decomposition (SVD) is written as $\boldsymbol{M} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^\top$, where $\boldsymbol{U} \in \mathbb{R}^{T\times r}$ and $\boldsymbol{V} \in \mathbb{R}^{N\times r}$ have orthonormal columns, $\boldsymbol{S} = \mathrm{diag}(\sigma_1, \ldots, \sigma_r)$ with $\sigma_1 \geq \cdots \geq \sigma_r > 0$, and $r = \mathrm{rank}(\boldsymbol{M})$. The singular values of $\boldsymbol{M}$ are the eigenvalues of $(\boldsymbol{M}^\top\boldsymbol{M})^{1/2}$ and are denoted by $\sigma_j(\boldsymbol{M})$.

The soft–thresholding operator associated with a threshold $\tau \geq 0$ acts on singular values as follows. Given $\boldsymbol{M} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^\top$ as above, define

$$\boldsymbol{S}_\tau := \mathrm{diag}\big((\sigma_1 - \tau)_+, \ldots, (\sigma_r - \tau)_+\big), \qquad (x)_+ := \max\{x, 0\}.$$

The singular value soft–thresholding of $\boldsymbol{M}$ at level τ is $\mathrm{SVT}(\boldsymbol{M}, \tau) := \boldsymbol{U}\boldsymbol{S}_\tau\boldsymbol{V}^\top$, which shrinks each singular value of $\boldsymbol{M}$ toward zero by τ and sets it to zero whenever $\sigma_j \leq \tau$.

For a scalar-valued function $f : \mathcal{X} \to \mathbb{R}$, we denote by $\mathrm{argmin}_{x\in\mathcal{X}} f(x)$ and $\mathrm{argmax}_{x\in\mathcal{X}} f(x)$ the sets (or selected elements) of minimizers and maximizers, respectively. In algorithmic descriptions, we use the assignment operator "$\leftarrow$" to indicate updates of variables, for instance $\boldsymbol{A} \leftarrow \boldsymbol{A}^\star$ means that $\boldsymbol{A}$ is replaced by the new value $\boldsymbol{A}^\star$.

Asymptotic notation is used in the usual sense. For deterministic sequences $a_n, b_n > 0$, we write $a_n \asymp b_n$ if there exist constants $0 < c \leq C < \infty$ such that $cb_n \leq a_n \leq Cb_n$ for all n sufficiently large. For random sequences X_n and positive deterministic a_n, we write $X_n = O_p(a_n)$ if X_n/a_n is bounded in probability, and $X_n = o_p(a_n)$ if $X_n/a_n \to 0$ in probability, as $n \to \infty$.

Sub-Gaussianity and Orlicz norms are used to quantify tail behavior. For a real-valued random variable X, the sub-Gaussian Orlicz norm is defined as $\|X\|_{\psi_2} :=$

$\inf\{c > 0 : \mathbb{E}[\exp(X^2/c^2)] \leq 2\}$. We say that X is sub-Gaussian if $\|X\|_{\psi_2} < \infty$. For a random vector $\boldsymbol{X}$, we use the induced norm $\|\boldsymbol{X}\|_{\psi_2} := \sup_{\|\boldsymbol{v}\|_2=1} \|\boldsymbol{v}^\top \boldsymbol{X}\|_{\psi_2}$, and say that $\boldsymbol{X}$ is sub-Gaussian if this quantity is finite.

For functions of real variables, the clipping operator, defined for $a \leq b$ by $\text{clip}_{[a,b]}(x) := \min\{\max\{x, a\}, b\}$, truncates x to the interval $[a, b]$. Finally, all norms, eigenvalues, singular values, and asymptotic orders are understood to be taken with respect to the dimension (N, T) indicated by the context; when both N and T diverge, statements such as $N/T = O(1)$ are interpreted in the usual joint asymptotic sense.

4 Methodology

We consider a cross-section of N assets whose excess returns are observed over periods $t = 1, \dots, T$. For each date t, let

$$\boldsymbol{R}_t := [R_{t1}, \dots, R_{tN}]^\top \in \mathbb{R}^N$$

denote the N–dimensional vector of excess returns. We view the sequence $\{\boldsymbol{R}_t\}_{t\geq 1}$ as a strictly stationary and ergodic process of random vectors. The unconditional mean vector and covariance matrix are

$$\boldsymbol{\mu}_N := \mathbb{E}[\boldsymbol{R}_t], \qquad \boldsymbol{\Sigma}_N := \mathbb{V}ar(\boldsymbol{R}_t),$$

with $\boldsymbol{\Sigma}_N$ assumed finite and positive definite. The observed panel of returns is organized into the $T \times N$ data matrix

$$\boldsymbol{R} := [\boldsymbol{R}_1, \dots, \boldsymbol{R}_T]^\top,$$

whose t-th row records the cross-section of excess returns at time t.

The mean–variance investor allocates wealth across the N assets by choosing a vector $\boldsymbol{w} \in \mathbb{R}^N$ satisfying $\boldsymbol{\iota}^\top \boldsymbol{w} = 1$, where $\boldsymbol{\iota}$ is a vector of ones. Since excess returns have mean $\boldsymbol{\mu}_N$ and covariance $\boldsymbol{\Sigma}_N$, the variance of the portfolio return $\boldsymbol{w}^\top \boldsymbol{R}_t$ is $\boldsymbol{w}^\top \boldsymbol{\Sigma}_N \boldsymbol{w}$, and the excess–return Sharpe ratio equals $(\boldsymbol{w}^\top \boldsymbol{\mu}_N)/\sqrt{\boldsymbol{w}^\top \boldsymbol{\Sigma}_N \boldsymbol{w}}$. The *global minimum–variance* (GMV) portfolio solves

$$\boldsymbol{w}_{\text{gmv}} := \underset{\boldsymbol{w}\in\mathbb{R}^N}{\operatorname{argmin}} \left\{ \boldsymbol{w}^\top \boldsymbol{\Sigma}_N \boldsymbol{w} \ : \ \boldsymbol{\iota}^\top \boldsymbol{w} = 1 \right\} = \frac{\boldsymbol{\Theta}_N \boldsymbol{\iota}}{\boldsymbol{\iota}^\top \boldsymbol{\Theta}_N \boldsymbol{\iota}},$$

where $\boldsymbol{\Theta}_N := \boldsymbol{\Sigma}_N^{-1}$ is the *precision matrix*. Instead, the *mean–variance efficient* (MVE) portfolio is

$$\boldsymbol{w}_{\text{mve}} := \underset{\boldsymbol{w}\in\mathbb{R}^N}{\operatorname{argmax}} \left\{ \frac{\boldsymbol{w}^\top \boldsymbol{\mu}_N}{\sqrt{\boldsymbol{w}^\top \boldsymbol{\Sigma}_N \boldsymbol{w}}} \ : \ \boldsymbol{\iota}^\top \boldsymbol{w} = 1 \right\} = \frac{\boldsymbol{\Theta}_N \boldsymbol{\mu}_N}{\boldsymbol{\iota}^\top \boldsymbol{\Theta}_N \boldsymbol{\mu}_N},$$

and has optimal Sharpe ratio of $\sqrt{\boldsymbol{\mu}_N^\top \boldsymbol{\Theta}_N \boldsymbol{\mu}_N}$. These expressions highlight that all mean–variance–optimal portfolios depend only on the two moments $(\boldsymbol{\mu}_N, \boldsymbol{\Sigma}_N)$ through the precision matrix.

When the covariance matrix $\boldsymbol{\Sigma}_N$ is nearly singular—i.e., some eigenvalues are near-zero—or when the cross-sectional dimension N approaches or exceeds the sample size T, the joint estimation of the mean vector $\boldsymbol{\mu}_N$ and the covariance matrix $\boldsymbol{\Sigma}_N$ becomes statistically fragile. Sample means converge slowly in high dimensions, and sample covariance matrices exhibit severe eigenvalue distortions; the subsequent matrix inversion needed to recover the precision matrix $\boldsymbol{\Theta}_N$ amplifies these errors. Even advanced shrinkage, regularization, or factor-based estimators of $(\boldsymbol{\mu}_N, \boldsymbol{\Sigma}_N)$ may deliver inputs that are too unstable for reliable mean–variance portfolio construction.

Our methodology departs from these traditional approaches by exploiting a population reparametrization that links $(\boldsymbol{\mu}_N, \boldsymbol{\Theta}_N)$ to a system of hedging regressions. This representation rewrites the moment estimation problem as a structured multi–response regression, whose coefficients carry an immediate financial interpretation: they describe the optimal hedging portfolios that neutralize asset-specific risk. Estimating this hedging system—rather than $\boldsymbol{\mu}_N$ and $\boldsymbol{\Sigma}_N$ directly—yields regularized, economically meaningful, and statistically stable estimates of the mean and precision matrices, even in challenging high-dimensional environments.

4.1 Mean and precision via hedging regressions

A convenient characterization of the mean vector and the precision matrix arises from viewing excess returns $\boldsymbol{R}_t$ as the sum of a predictable component and an innovation term generated by a system of *hedging regressions*; see Stevens (1998). Consider the decomposition

$$\boldsymbol{R}_t = \boldsymbol{A}_N + \boldsymbol{\Gamma}_N^\top (\boldsymbol{R}_t - \boldsymbol{\mu}_N) + \boldsymbol{E}_t, \tag{1}$$

where $\boldsymbol{E}_t$ collects the innovation terms. The pair $(\boldsymbol{A}_N, \boldsymbol{\Gamma}_N)$ is defined as the population least–squares solution

$$(\boldsymbol{A}_N, \boldsymbol{\Gamma}_N) := \underset{\substack{\boldsymbol{A}\in\mathbb{R}^N \\ \boldsymbol{\Gamma}\in\mathbb{R}^{N\times N}}}{\operatorname{argmin}} \left\{ \mathbb{E}\left[\left\| \boldsymbol{R}_t - \boldsymbol{A} - \boldsymbol{\Gamma}^\top (\boldsymbol{R}_t - \boldsymbol{\mu}_N) \right\|_2^2 \right] \ : \ \operatorname{diag}(\boldsymbol{\Gamma}) = \boldsymbol{0} \right\}.$$

The diagonal restriction excludes self-hedging and ensures that each asset is regressed only on the remaining $N - 1$ assets.

The minimizer of this problem satisfies $\boldsymbol{A}_N = \boldsymbol{\mu}_N$, so the intercepts in the hedging system recover the unconditional mean returns. The matrix $\boldsymbol{\Gamma}_N$ collects the population OLS coefficients from N simultaneous hedging regressions: for asset i, the ith column of $\boldsymbol{\Gamma}_N$ prescribes the positions in the other assets that minimize the conditional variance of $R_{i,t}$ given the cross-section. Large positive entries indicate strong comovement with the corresponding assets that is optimally offset through short positions; small or near-zero entries indicate weak hedging relevance. Viewed collectively, the system (1) encodes the cross-sectional geometry of hedging opportunities. It decomposes each return into a predictable component spanned by other assets and an idiosyncratic innovation, thereby providing a regression-based lens through which systematic and idiosyncratic risk components can be separated.

Let $\boldsymbol{\Psi}_N := \operatorname{diag}(\mathbb{V}ar(E_{1,t}), \dots, \mathbb{V}ar(E_{N,t}))$ collect the marginal innovation variances associated with the hedging system in (1). A classical linear–algebra identity shows that the solution of this system determines both the mean vector and the precision matrix of returns. In particular, whenever $\boldsymbol{\Sigma}_N$ is positive definite, one obtains the decomposition

$$\boldsymbol{\mu}_N = \boldsymbol{A}_N, \qquad \boldsymbol{\Theta}_N := \boldsymbol{\Sigma}_N^{-1} = \boldsymbol{\Psi}_N^{-1}(\boldsymbol{I} - \boldsymbol{\Gamma}_N^\top). \tag{2}$$

Thus the entire precision matrix can be recovered from the hedging system: the matrix $(\boldsymbol{I} - \boldsymbol{\Gamma}_N^\top)$ encodes how each asset loads on the hedging portfolios of the others, while $\boldsymbol{\Psi}_N^{-1}$ scales these exposures by the inverse idiosyncratic variances. From a financial viewpoint, (2) formalizes the intuition that a long position in asset i is optimally hedged by short positions in the other assets with weights $\psi_{N,ii}^{-1}\Gamma_{N,ji}$ for $j \neq i$. Equivalently, the vector $\psi_{N,ii}^{-1}\boldsymbol{\Gamma}_N^{(i)}$ describes precisely the portfolio that removes the asset–specific component of R_{it} while retaining its systematic exposures.

Substituting (2) into the expressions for the GMV and MVE portfolios shows that the entire family of optimal allocations depends solely on $(\boldsymbol{A}_N, \boldsymbol{\Gamma}_N, \boldsymbol{\Psi}_N)$:

$$\boldsymbol{w}_{\text{gmv}} = \frac{\boldsymbol{\Psi}_N^{-1}(\boldsymbol{I} - \boldsymbol{\Gamma}_N^\top)\boldsymbol{\iota}}{\boldsymbol{\iota}^\top \boldsymbol{\Psi}_N^{-1}(\boldsymbol{I} - \boldsymbol{\Gamma}_N^\top)\boldsymbol{\iota}}, \qquad \boldsymbol{w}_{\text{mve}} = \frac{\boldsymbol{\Psi}_N^{-1}(\boldsymbol{I} - \boldsymbol{\Gamma}_N^\top)\boldsymbol{A}_N}{\boldsymbol{\iota}^\top \boldsymbol{\Psi}_N^{-1}(\boldsymbol{I} - \boldsymbol{\Gamma}_N^\top)\boldsymbol{A}_N}.$$

Estimating the hedging system is therefore sufficient to recover both the mean vector and the precision matrix, and hence all mean–variance portfolios, without ever inverting a high-dimensional covariance matrix estimate.

From an estimation perspective, this reparametrization is extremely valuable. When returns exhibit multicollinearity or when N is large relative to T, conventional estimators of $\boldsymbol{\mu}_N$ and $\boldsymbol{\Sigma}_N^{-1}$ become unstable. The hedging formulation, by contrast, lends itself naturally to regularization. For the mean vector, one may stabilize estimation by shrinking $\boldsymbol{A}_N$ toward a plausible target—such as the grand mean of the cross section—thereby preventing the intercepts from reacting excessively to small-sample noise. For the coefficient matrix $\boldsymbol{\Gamma}_N$, however, the appropriate form of regularization is less obvious. A key insight, developed next, is that under general factor structures, the matrix of hedging portfolio returns $\boldsymbol{R}\boldsymbol{\Gamma}_N$ is *nearly reduced-rank*, i.e., some of their singular values are near-zero. Therefore, only a small number of common hedging directions are needed to explain most of the cross-sectional variation. This structural property provides a clear guideline for regularizing $\boldsymbol{\Gamma}_N$: an estimator that encourages a low-rank representation of the hedging portfolios is naturally aligned with economic environments in which a small number of diversified hedging directions span most of the systematic comovement in returns. Such regularization yields stable, interpretable, and cost-efficient portfolio allocations in high-dimensional settings.

4.2 Low-rank approximations in high dimensions

The hedging representation in (1) suggests that, from the perspective of portfolio construction, the key object is not the entire $T \times N$ return matrix $\boldsymbol{R}$ but rather the matrix of hedging portfolio returns $\boldsymbol{R}\boldsymbol{\Gamma}_N$. Each column of $\boldsymbol{R}\boldsymbol{\Gamma}_N$ is the time series of hedging portfolio return minimizing the residual variance of a given asset, and the cross-sectional geometry of these hedging portfolios encodes how many distinct directions of systematic risk are needed to span the space of return comovements. In high-dimensional environments, it is natural to ask whether this hedging return matrix is effectively low-dimensional, in the sense of being well approximated by a matrix of small rank. Financially, this corresponds to the view that only a few core hedging strategies are needed to span the dominant sources of systematic risk, while the remaining cross-sectional variation is idiosyncratic and therefore economically unimportant.

When excess returns are driven by a small number of latent factors, the answer is affirmative: $\boldsymbol{R}\boldsymbol{\Gamma}_N$ behaves as a nearly reduced-rank matrix, and only a few common hedging portfolios are needed to optimally hedge most of the return variation. To make this precise, we work with centered excess returns

$$\boldsymbol{r}_t := \boldsymbol{R}_t - \boldsymbol{\mu}_N \in \mathbb{R}^N$$

that admit a finite-K factor representation

$$\boldsymbol{r}_t = \boldsymbol{B}_N \boldsymbol{f}_t + \boldsymbol{u}_t, \tag{3}$$

where $\boldsymbol{B}_N \in \mathbb{R}^{N\times K}$ collects the factor loadings, $\boldsymbol{f}_t \in \mathbb{R}^K$ are common factors with mean zero and covariance $\boldsymbol{\Sigma}_f$, and $\boldsymbol{u}_t \in \mathbb{R}^N$ are idiosyncratic components with mean zero and covariance matrix $\boldsymbol{\Sigma}_{u,N}$. The population covariance matrix of returns is

$$\boldsymbol{\Sigma}_N = \boldsymbol{B}_N \boldsymbol{\Sigma}_f \boldsymbol{B}_N^\top + \boldsymbol{\Sigma}_{u,N}.$$

We order the eigenvalues of $\boldsymbol{\Sigma}_N$ as $\lambda_{1,N} \geq \cdots \geq \lambda_{N,N}$ and those of $\boldsymbol{\Omega}_N = \boldsymbol{\Sigma}_N^{-1}$ as $0 < \theta_{1,N} \leq \cdots \leq \theta_{N,N}$.

The factor structure in (3) is coupled with mild restrictions on the idiosyncratic spectrum, the strength of the factor loadings, and the tail behavior of $(\boldsymbol{f}_t, \boldsymbol{u}_t)$. We summarize these conditions formally.

Assumption 1 (Idiosyncratic spectrum). The idiosyncratic covariance matrix $\boldsymbol{\Sigma}_{u,N}$ has eigenvalues uniformly bounded away from zero and infinity: there exist constants $0 < m \leq M < \infty$ such that, for all N large enough,

$$m \leq \lambda_{\min}(\boldsymbol{\Sigma}_{u,N}) \leq \lambda_{\max}(\boldsymbol{\Sigma}_{u,N}) \leq M.$$

Assumption 2 (Spiked factor loadings). Let $s_{1,N} \geq \cdots \geq s_{K,N} > 0$ denote the nonzero eigenvalues of $\boldsymbol{B}_N^\top \boldsymbol{B}_N$. There exist exponents $1 \geq \alpha_1 \geq \cdots \geq \alpha_K > 0$ and constants $0 < c_j \leq C_j < \infty$ such that, for all N large enough,

$$c_j N^{\alpha_j} \leq s_{j,N} \leq C_j N^{\alpha_j}, \qquad j = 1, \ldots, K.$$

Assumption 3 (Sub-Gaussian factors and idiosyncratic terms). The factor and idiosyncratic processes $\{\boldsymbol{f}_t\}_{t\geq 1}$ and $\{\boldsymbol{u}_t\}_{t\geq 1}$ are independent across t and satisfy

$$\mathbb{E}[\boldsymbol{f}_t] = \mathbf{0}, \qquad \mathbb{V}ar(\boldsymbol{f}_t) = \boldsymbol{\Sigma}_f, \qquad \mathbb{E}[\boldsymbol{u}_t] = \mathbf{0}, \qquad \mathbb{V}ar(\boldsymbol{u}_t) = \boldsymbol{\Sigma}_{u,N},$$

where $\boldsymbol{\Sigma}_f$ is positive definite. Moreover, both sequences are sub-Gaussian with uniformly bounded Orlicz norms: there exist finite constants $c_f, c_u < \infty$ such that

$$\sup_t \|\boldsymbol{f}_t\|_{\psi_2} \leq c_f, \qquad \sup_{t,N} \|\boldsymbol{u}_t\|_{\psi_2} \leq c_u.$$

Assumption 1 ensures that idiosyncratic risk remains well behaved as the cross-sectional dimension grows: idiosyncratic variances neither explode nor vanish, and the idiosyncratic component is asymptotically diversifiable but not degenerate. Assumption 2 controls the strength of the factors through the growth of the eigenvalues of $\boldsymbol{B}_N^\top \boldsymbol{B}_N$. The exponents α_j quantify how quickly the j-th factor's contribution grows with N; the case $\alpha_j = 1$ corresponds to a *pervasive* factor, whose influence scales linearly with the size of the cross section, while $0 < \alpha_j < 1$ describes a *weak* factor, whose effect increases more slowly. Assumption 3 is a convenient high-level way to impose concentration of sample moments around their population counterparts, guaranteeing that the factor structure and the idiosyncratic spectrum are visible in finite samples when T is large.

Remark 1 (Factor strength and weak factors). Under Assumptions 1 and 2, the covariance matrix $\boldsymbol{\Sigma}_N$ inherits a spiked-plus-bulk spectrum: the top K eigenvalues grow at rates $\lambda_{j,N} \asymp N^{\alpha_j}$, while the remaining $N-K$ eigenvalues stay of order one. Pervasive factors ($\alpha_j = 1$) generate large spikes and dominate the covariance structure for large N, whereas weak factors ($0 < \alpha_j < 1$) still produce diverging eigenvalues but with a slower growth rate. This distinction is important empirically: weak factors can be difficult to detect with standard principal component methods, yet they still contribute nontrivially to systematic risk and, as we show, to the structure of optimal hedging portfolios.

Remark 2 (Links to classical factor models). The assumptions above are compatible with several classical factor-model frameworks in empirical asset pricing. Strict factor models in the sense of Chamberlain (1983) and Connor and Korajczyk (1986), with diagonal $\boldsymbol{\Sigma}_{u,N}$ and $\frac{1}{N}\boldsymbol{B}_N^\top\boldsymbol{B}_N \to \boldsymbol{Q}$ with $\boldsymbol{Q}$ finite and positive definite, correspond to pervasive factors with $\alpha_j = 1$ and satisfy Assumptions 1–3. Approximate factor models à la Bai and Ng (2002) permit limited cross-sectional dependence in $\boldsymbol{u}_t$ while maintaining a bounded idiosyncratic spectrum and strong factor spikes, which again fit within this framework. Arbitrage pricing theory (APT) models in the spirit of Chamberlain and Rothschild (1983) impose that idiosyncratic risk is asymptotically diversifiable and that a finite number of systematic factors drives the cross-section of returns; under mild conditions, these models also imply a bounded idiosyncratic spectrum and spiked factor eigenvalues. Our setting can therefore be viewed as an abstraction of these familiar structures, accommodating both pervasive and weak factors in a high-dimensional limit.

We now connect these assumptions to the rank properties of the hedging return matrix. Recall that, for centered returns, the population hedging system can be written as

$$\boldsymbol{r}_t = \boldsymbol{\Gamma}_N^\top \boldsymbol{r}_t + \boldsymbol{E}_t,$$

where $\boldsymbol{\Gamma}_N$ is the hollow matrix of population OLS coefficients from regressing each asset on all the others, and $\boldsymbol{E}_t$ collects the associated residuals with diagonal covariance $\boldsymbol{\Psi}_N = \operatorname{diag}(\psi_{1,N}, \ldots, \psi_{N,N})$. Stacking over $t = 1, \ldots, T$ yields the $T \times N$ matrices $\boldsymbol{r} = (\boldsymbol{r}_1, \ldots, \boldsymbol{r}_T)^\top$ and $\boldsymbol{r}\boldsymbol{\Gamma}_N$, the latter collecting the time series of returns on the N hedging portfolios defined by the columns of $\boldsymbol{\Gamma}_N$.

The following proposition formalizes the near low-rank structure of $\boldsymbol{r}\boldsymbol{\Gamma}_N$ in the factor setting just described.

Theorem 4.1 (near low-rank hedging portfolio returns)**.** *Suppose that the factor model* (3) *holds and that Assumptions 1–3 are satisfied. Let $N, T \to \infty$ with $N/T =$*

$O(1)$. Then there exist random matrices $\boldsymbol{M}_N, \boldsymbol{W}_N \in \mathbb{R}^{T\times N}$ such that

$$\boldsymbol{r}\boldsymbol{\Gamma}_N = \boldsymbol{M}_N + \boldsymbol{W}_N, \qquad \mathrm{rank}(\boldsymbol{M}_N) \le K, \qquad \|\boldsymbol{W}_N\|_2 = O_p(\sqrt{T}).$$

Equivalently, the best rank-K approximation error of the hedging return matrix in operator norm satisfies

$$\inf_{\boldsymbol{B}\in\mathbb{R}^{T\times N}} \left\{ \left\|\boldsymbol{r}\boldsymbol{\Gamma}_N - \boldsymbol{B}\right\|_2 \ : \ \mathrm{rank}(\boldsymbol{B}) \le K \right\} = O_p(\sqrt{T}).$$

This result shows that, under a low-dimensional factor structure, the entire panel of hedging portfolio returns is essentially K-dimensional: up to a noise term of size $O_p(\sqrt{T})$ in operator norm, $\boldsymbol{r}\boldsymbol{\Gamma}_N$ lies in a K-dimensional linear subspace. From a portfolio perspective, this means that a small number of common hedging portfolios span nearly all directions of systematic risk that are relevant for mean–variance investors. Additional hedging portfolios mainly load on idiosyncratic components and contribute little to the shape of the efficient frontier. This near low-rank structure is the main theoretical motivation for our nuclear norm–regularized estimator, which explicitly shrinks $\boldsymbol{r}\boldsymbol{\Gamma}_N$ toward low-rank configurations in order to recover stable and economically interpretable hedging portfolios in high-dimensional settings.

Remark 3 (Centered vs. uncentered hedging returns). The result in Proposition 4.1 is stated for the centered returns $\boldsymbol{r}$, but the same conclusion applies to the uncentered hedging return matrix $\boldsymbol{R}\boldsymbol{\Gamma}_N$. Indeed, $\boldsymbol{r} = \boldsymbol{R} - \boldsymbol{\iota}\boldsymbol{\mu}_N^\top$ implies

$$\boldsymbol{R}\boldsymbol{\Gamma}_N = \boldsymbol{r}\boldsymbol{\Gamma}_N + \boldsymbol{\iota}(\boldsymbol{\mu}_N^\top\boldsymbol{\Gamma}_N),$$

where the second term is rank one. Hence the uncentered hedging portfolio returns differ from $\boldsymbol{r}\boldsymbol{\Gamma}_N$ only by a deterministic rank-one shift, which does not affect the near low-rank property nor the $O_p(\sqrt{T})$ approximation bound.

4.3 Penalized reduced-rank hedging regression

The representation in (1) shows that both the mean vector $\boldsymbol{\mu}_N$ and the precision matrix $\boldsymbol{\Theta}_N$ can be recovered from the coefficients $(\boldsymbol{A}_N, \boldsymbol{\Gamma}_N)$ of a system of hedging regressions. Rather than estimating $(\boldsymbol{\mu}_N, \boldsymbol{\Theta}_N)$ directly, our approach regularizes the hedging system itself. Specifically, we (i) regularize $\boldsymbol{\Gamma}_N$ by exploiting the potential near low-rank structure of the hedging return matrix discussed in Section 4.2, and (ii) regularize the intercept vector $\boldsymbol{A}_N$ (and hence the mean vector $\boldsymbol{\mu}_N$) by penalizing deviations from a target mean $\boldsymbol{\mu}^*$, which in our empirical work we take to be the grand mean of excess returns. This strategy yields estimates of $(\boldsymbol{A}_N, \boldsymbol{\Gamma}_N)$ that remain well-behaved in high dimensions and translate directly into stable estimates of $(\boldsymbol{\mu}_N, \boldsymbol{\Theta}_N)$.

We begin by approximating the centered returns $\boldsymbol{r}_t = \boldsymbol{R}_t - \boldsymbol{\mu}_N$ with their sample counterpart. Define

$$\widehat{\boldsymbol{\mu}} := \frac{1}{T}\boldsymbol{R}^\top \boldsymbol{\iota}, \qquad \widetilde{\boldsymbol{R}} := \boldsymbol{R} - \boldsymbol{\iota}\widehat{\boldsymbol{\mu}}^\top,$$

where $\boldsymbol{\iota}$ denotes the T-dimensional vector of ones. We then fit the hedging system by computing the *penalized reduced-rank* (PRR) estimator

$$(\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{\Gamma}}) := \underset{\substack{\boldsymbol{A}\in\mathbb{R}^N \\ \boldsymbol{\Gamma}\in\mathbb{R}^{N\times N}}}{\operatorname{argmin}} \Big\{ \frac{1}{2}\big\|\boldsymbol{R} - \boldsymbol{\iota}\boldsymbol{A}^\top - \widetilde{\boldsymbol{R}}\boldsymbol{\Gamma}\big\|_F^2 + \lambda_*\big\|\widetilde{\boldsymbol{R}}\boldsymbol{\Gamma}\big\|_* + \frac{\lambda_F}{2}\|\boldsymbol{\Gamma}\|_F^2 + \frac{\lambda_r}{2}\|\boldsymbol{A} - \boldsymbol{\mu}^*\|_2^2 \ : \ \operatorname{diag}(\boldsymbol{\Gamma}) = \boldsymbol{0} \Big\}, \tag{4}$$

with tuning parameters $\lambda_*, \lambda_F, \lambda_r \geq 0$.

The first term is the joint least-squares loss across the N hedging regressions: for each asset i, the ith column of $\widehat{\boldsymbol{\Gamma}}$ describes the portfolio of other assets that best predicts $R_{i,t}$, together with intercept $\widehat{A}_i$. Without penalties, (4) reduces to N separate OLS regressions, which is statistically unstable when N is large.

The nuclear norm penalty $\|\widetilde{\boldsymbol{R}}\boldsymbol{\Gamma}\|_*$ directly controls the singular values of the hedging return matrix. In light of Proposition 4.1, this encourages $\widetilde{\boldsymbol{R}}\widehat{\boldsymbol{\Gamma}}$ to concentrate in a low-dimensional subspace spanned by a small number of common hedging portfolios—precisely the structure implied by a low-dimensional factor model. This prevents the estimator from overfitting spurious high-rank components driven by sampling noise.

The Frobenius penalty $\|\mathbf{\Gamma}\|_F^2/2$ shrinks the magnitude of hedging positions. Financially, this produces more diversified and less levered hedging portfolios and limits turnover through smoother coefficient profiles. Finally, the ridge penalty $\|\boldsymbol{A}-\boldsymbol{\mu}^*\|_2^2/2$ stabilizes the estimation of the mean vector, counteracting the well-known erratic behavior of sample means in high dimensions. Shrinking toward $\boldsymbol{\mu}^*$ (taken empirically as the grand mean) yields a disciplined estimate of $\boldsymbol{\mu}_N$ that is both robust and economically interpretable.

Recovering mean, precision, and portfolio weights. Given $(\widehat{\boldsymbol{A}}, \widehat{\mathbf{\Gamma}})$, we recover the mean vector and precision matrix using the population identities in (2). Let the residuals be

$$\widehat{\boldsymbol{E}} := \boldsymbol{R} - \boldsymbol{\iota}\widehat{\boldsymbol{A}}^\top - \widetilde{\boldsymbol{R}}\widehat{\mathbf{\Gamma}}.$$

Define the diagonal matrix of innovation variances

$$\widehat{\mathbf{\Psi}} := \operatorname{diag}\Big(\widehat{\mathbb{V}ar}[\widehat{\boldsymbol{E}}_1], \ldots, \widehat{\mathbb{V}ar}[\widehat{\boldsymbol{E}}_N]\Big),$$

where, for any $\boldsymbol{v} \in \mathbb{R}^T$,

$$\widehat{\mathbb{V}ar}[\boldsymbol{v}] := \frac{1}{T}\sum_{t=1}^{T}(v_t - \bar{v})^2, \qquad \bar{v} := \frac{1}{T}\sum_{t=1}^{T} v_t.$$

Using the identities $\boldsymbol{\mu}_N = \boldsymbol{A}_N$ and $\mathbf{\Theta}_N = \mathbf{\Psi}_N^{-1}(\boldsymbol{I} - \mathbf{\Gamma}_N^\top)$, we estimate

$$\widehat{\boldsymbol{\mu}} := \widehat{\boldsymbol{A}}, \qquad \widehat{\mathbf{\Theta}} := \widehat{\mathbf{\Psi}}^{-1}(\boldsymbol{I} - \widehat{\mathbf{\Gamma}}^\top).$$

The global minimum-variance and maximum-Sharpe portfolios are therefore estimated by

$$\widehat{\boldsymbol{w}}_{\mathrm{gmv}} = \frac{\widehat{\boldsymbol{\Theta}}\boldsymbol{\iota}}{\boldsymbol{\iota}^\top\widehat{\boldsymbol{\Theta}}\boldsymbol{\iota}}, \qquad \widehat{\boldsymbol{w}}_{\mathrm{mve}} = \frac{\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\mu}}}{\boldsymbol{\iota}^\top\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\mu}}}.$$

These estimated objects provide the fully regularized inputs for our empirical and simulation analyses in the remainder of the paper.

4.4 Comparison to sparse approaches to hedging regressions

A number of recent contributions have revisited Stevens' (1998) decomposition and proposed sparse estimators of the precision matrix $\boldsymbol{\Theta}_N = \boldsymbol{\Sigma}_N^{-1}$ as a means of constructing hedging portfolios, especially in high-dimensional settings. These methods effectively impose sparsity on the hedging coefficient matrix $\boldsymbol{\Gamma}_N$, and hence, via Stevens' decomposition, on the precision matrix $\boldsymbol{\Theta}_N$ itself. In this parametrization, sparsity in $\boldsymbol{\Theta}_N$ is exactly equivalent to assuming that each asset can be hedged using only a small subset of the remaining assets. Because our primary objective is the construction of stable and well-diversified minimum-variance portfolios, it is important to understand both the appeal and the limitations of such sparsity-based approaches.

Graphical Lasso and Nodewise Lasso Estimators. One prominent approach is the *Graphical Lasso* (Goto and Xu, 2015), which estimates a sparse precision matrix by solving the convex program

$$\widehat{\boldsymbol{\Theta}}_N := \underset{\boldsymbol{\Theta}\in\mathbb{R}^{N\times N}}{\operatorname{argmin}} \left\{ \underbrace{\operatorname{tr}(\widehat{\boldsymbol{\Sigma}}_N\boldsymbol{\Theta}) - \log\det(\boldsymbol{\Theta})}_{\text{Gaussian log-likelihood}} + \underbrace{\lambda\|\boldsymbol{\Theta} - \operatorname{diag}(\boldsymbol{\Theta})\|_1}_{\ell_1 \text{ penalty}} \ : \ \boldsymbol{\Theta} \succeq 0 \right\}, \qquad (5)$$

where $\widehat{\boldsymbol{\Sigma}}_N$ denotes the sample covariance matrix and $\lambda > 0$ is a tuning parameter. The ℓ_1 penalty induces sparsity in the off-diagonal entries of $\boldsymbol{\Theta}_N$, thereby producing a system of hedging portfolios in which many assets receive zero weight when hedging one another.

A second influential method is the *nodewise Lasso* (Callot et al., 2021). For each asset i, one estimates the nonzero population regression coefficients $\boldsymbol{\Gamma}_{N,i}$ in

$$\boldsymbol{R}^{(i)} = \boldsymbol{R}\boldsymbol{\Gamma}_N^{(i)} + \boldsymbol{E}_i,$$

by solving

$$\widehat{\boldsymbol{\Gamma}}_N^{(i)} := \operatorname*{argmin}_{\boldsymbol{\gamma} \in \mathbb{R}^N} \left\{ \frac{1}{2} \|\boldsymbol{R}^{(i)} - \boldsymbol{R}\boldsymbol{\gamma}\|_2^2 + \lambda_i \|\boldsymbol{\gamma}\|_1 \ : \ \gamma_i = 0 \right\}, \qquad i = 1, \ldots, N, \tag{6}$$

where $\lambda_i > 0$ are asset-specific tuning parameters. Stacking (horizontally) the $\widehat{\boldsymbol{\Gamma}}_N^{(i)}$ yields a sparse estimator of the hedging matrix $\boldsymbol{\Gamma}_N$ and, via Stevens' identity, a sparse approximation to the precision matrix.

Both (5) and (6) produce N sparse hedging portfolios: each asset is hedged using only a subset of the remaining assets. This structural simplicity can be computationally convenient and may facilitate interpretability in settings where only a few assets are believed to hedge one another.

Financial Limitations of Sparsity. Despite their statistical appeal, sparse estimators of $\boldsymbol{\Theta}_N$ or of the hedging matrix $\boldsymbol{\Gamma}_N$ lack a strong financial justification in most realistic return-generating environments.

First, sparsity in $\boldsymbol{\Theta}_N$ encodes conditional uncorrelation: a zero off-diagonal entry $\Theta_{ij} = 0$ implies that assets i and j are conditionally independent given all other assets. This is a stringent assumption. In typical financial markets, even assets with weak marginal correlations may share exposures to common latent factors, which makes exact conditional independence implausible.

Second, sparsity in the hedging coefficients mechanically restricts the set of assets available for diversification. Because the GMV portfolio relies entirely on covariance structure, excluding assets from the hedging set forces the model to rely excessively on a small subset of assets, which can lead to unstable and poorly diversified portfolios. Sparse hedges are often too "sharp" relative to the smoothness of true factor exposures, amplifying estimation noise rather than controlling it.

A third limitation becomes apparent in the presence of *approximate factor models.* Suppose excess returns satisfy

$$\boldsymbol{R}_t = \boldsymbol{\alpha}_N + \boldsymbol{B}_N \boldsymbol{f}_t + \boldsymbol{u}_t, \qquad \boldsymbol{f}_t \sim (0, \boldsymbol{\Sigma}_f), \quad \boldsymbol{u}_t \sim (0, \boldsymbol{\Sigma}_{u,N}), \tag{7}$$

with $\boldsymbol{\Sigma}_{u,N}$ diagonal for simplicity. By the Woodbury identity,

$$\boldsymbol{\Theta}_N = \boldsymbol{\Sigma}_{u,N}^{-1} - \boldsymbol{\Sigma}_{u,N}^{-1} \boldsymbol{B}_N \left(\boldsymbol{B}_N^\top \boldsymbol{\Sigma}_{u,N}^{-1} \boldsymbol{B}_N + \boldsymbol{\Sigma}_f^{-1} \right)^{-1} \boldsymbol{B}_N^\top \boldsymbol{\Sigma}_{u,N}^{-1}. \tag{8}$$

If $\boldsymbol{B}_N$ is dense, as is typical when many assets load on common macroeconomic factors, then the matrix in (8) is also dense, even when $\boldsymbol{\Sigma}_{u,N}$ is diagonal. Thus, sparsity in $\boldsymbol{\Theta}_N$ requires imposing strong and financially restrictive conditions on the factor structure (e.g. many zero loadings or block-separable sectors). Such assumptions rarely hold in practice.

These arguments highlight that sparsity of the precision matrix or of the hedging coefficients is often more a statistical convenience than a reflection of financial structure. Realistic factor models generate dense conditional dependence patterns, and effective risk reduction typically requires combining information from many assets, not a small subset.

The shortcomings of sparsity motivate alternative forms of regularization that better reflect economic fundamentals. In particular, as shown in Section 4.2, a wide class of factor structures implies that the hedging return matrix $\boldsymbol{R}\boldsymbol{\Gamma}_N$ is nearly reduced-rank. Only a few synthetic hedging portfolios are needed to span the systematic risks in the cross-section. This low-dimensional structure provides a financially sound basis for regularizing $\boldsymbol{\Gamma}_N$ and stands in contrast to sparsity, which lacks comparable economic grounding.

Regularizing via low-rankness therefore strikes a balance between flexibility and parsimony: it preserves the ability to diversify across many assets while capturing the essential factor-driven structure of returns. Our proposed penalized reduced-rank hedging regression builds explicitly on this insight.

4.5 Comparison to PCA–based precision estimators

Principal Components Analysis (PCA) is a widely used dimension–reduction tool in multivariate statistics and has been proposed as a means of regularizing high–dimensional covariance and precision matrices for portfolio allocation (e.g., Ledoit and Wolf, 2003). Because PCA delivers a low–rank approximation of the sample covariance matrix, it appears at first glance to provide a natural alternative to our reduced–rank hedging regression framework. In this subsection, we describe the PCA–based precision estimator, discuss the hedging system it induces, and contrast it with our penalized PRR approach.

PCA–based covariance and precision matrices. Let $\widehat{\boldsymbol{\Sigma}}_N$ denote the sample covariance matrix of excess returns, with spectral decomposition

$$\widehat{\boldsymbol{\Sigma}}_N = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^\top = \sum_{i=1}^{N} \lambda_i \boldsymbol{V}^{(i)}\boldsymbol{V}^{(i)\top},$$

where $\lambda_1 \geq \cdots \geq \lambda_N > 0$ are the sample eigenvalues and $\boldsymbol{V}^{(i)}$ denotes the ith column of $\boldsymbol{V}$, that is, the ith orthonormal eigenvector.

A standard PCA–based covariance model retains the first k principal components and adds an idiosyncratic noise level $\sigma_\varepsilon^2 > 0$:

$$\widehat{\boldsymbol{\Sigma}}_{\text{PCA}}^{(k)} := \boldsymbol{V}_k\boldsymbol{\Lambda}_k\boldsymbol{V}_k^\top + \sigma_\varepsilon^2\boldsymbol{I}_N,$$

where $\boldsymbol{V}_k := [\boldsymbol{V}^{(1)}, \ldots, \boldsymbol{V}^{(k)}] \in \mathbb{R}^{N\times k}$ and $\boldsymbol{\Lambda}_k := \text{diag}(\lambda_1, \ldots, \lambda_k) \in \mathbb{R}^{k\times k}$. In this model, the principal–component subspace $\text{span}(\boldsymbol{V}^{(1)}, \ldots, \boldsymbol{V}^{(k)})$ carries the systematic variation, while $\sigma_\varepsilon^2\boldsymbol{I}_N$ captures residual idiosyncratic risk.

The eigenvalues of $\widehat{\boldsymbol{\Sigma}}_{\text{PCA}}^{(k)}$ are $\lambda_i + \sigma_\varepsilon^2$ for $i \leq k$ and σ_ε^2 for $i > k$, so its inverse admits the closed–form representation

$$
\begin{aligned}
\widehat{\mathbf{\Theta}}_{\text{PCA}}^{(k)} &:= \widehat{\mathbf{\Sigma}}_{\text{PCA}}^{(k)-1} \\
&= \boldsymbol{V}_k(\mathbf{\Lambda}_k + \sigma_\varepsilon^2 \boldsymbol{I}_k)^{-1}\boldsymbol{V}_k^\top + \sigma_\varepsilon^{-2}\big(\boldsymbol{I}_N - \boldsymbol{V}_k\boldsymbol{V}_k^\top\big) \\
&= \sigma_\varepsilon^{-2}\boldsymbol{I}_N - \sigma_\varepsilon^{-2}\boldsymbol{V}_k \operatorname{diag}\Big(\tfrac{\lambda_1}{\lambda_1+\sigma_\varepsilon^2}, \dots, \tfrac{\lambda_k}{\lambda_k+\sigma_\varepsilon^2}\Big)\boldsymbol{V}_k^\top. \qquad (9)
\end{aligned}
$$

Thus the PCA–based precision matrix is the sum of a diagonal component $\sigma_\varepsilon^{-2}\boldsymbol{I}_N$ and a low–rank correction within the principal–component subspace. This structure stabilizes the inverse by shrinking small eigenvalues of $\widehat{\mathbf{\Sigma}}_N$ away from zero; however, the criterion used to select $\boldsymbol{V}_k$ and $\mathbf{\Lambda}_k$ is purely variance–based.

PCA and hedgeability. PCA chooses the first k eigenvectors to maximize the proportion of total variance explained,

$$
\mathrm{EV}_k(\widehat{\mathbf{\Sigma}}_N) := \frac{\sum_{i=1}^k \lambda_i}{\sum_{j=1}^N \lambda_j},
$$

so that $\widehat{\mathbf{\Sigma}}_{\text{PCA}}^{(k)}$ approximates $\widehat{\mathbf{\Sigma}}_N$ in a mean–square sense. Global minimum–variance optimization, by contrast, depends on the precision matrix $\widehat{\mathbf{\Theta}}_N = \widehat{\mathbf{\Sigma}}_N^{-1}$, whose spectral decomposition is

$$
\widehat{\mathbf{\Theta}}_N = \boldsymbol{V}\mathbf{\Lambda}^{-1}\boldsymbol{V}^\top = \sum_{i=1}^N \lambda_i^{-1}\boldsymbol{V}^{(i)}\boldsymbol{V}^{(i)\top}.
$$

The directions that matter most for hedging and partial covariances are those associated with large λ_i^{-1}, that is, with small–variance eigenvectors of $\widehat{\mathbf{\Sigma}}_N$. The fraction of the precision matrix captured by the top k principal components of $\widehat{\mathbf{\Sigma}}_N$,

$$
\mathrm{EV}_k(\widehat{\mathbf{\Theta}}_N) := \frac{\sum_{i=1}^k \lambda_i^{-1}}{\sum_{j=1}^N \lambda_j^{-1}},
$$

is typically small in high–dimensional financial panels: the largest–variance eigenvectors of $\widehat{\mathbf{\Sigma}}_N$ often correspond to the least important directions for $\widehat{\mathbf{\Theta}}_N$ and for GMV hedging. PCA therefore regularizes the covariance matrix in a way that is not explicitly tailored to portfolio risk minimization.

The hedging representation introduced in Section 4 links the precision matrix to a system of population hedging regressions. In particular, if $\mathbf{\Gamma}_N$ collects the population OLS coefficients from regressing each asset on all others, and $\mathbf{\Psi}_N := \text{diag}(\mathbb{V}ar(E_{1,t}), \ldots, \mathbb{V}ar(E_{N,t}))$ collects the associated innovation variances, then the exact identity

$$\mathbf{\Theta}_N = \mathbf{\Psi}_N^{-1}\big(\boldsymbol{I}_N - \mathbf{\Gamma}_N^\top\big)$$

in (2) holds whenever $\mathbf{\Sigma}_N$ is positive definite; see (2). Conversely, given any positive definite precision matrix $\mathbf{\Theta}_N$, there exists a unique pair $(\mathbf{\Gamma}_N, \mathbf{\Psi}_N)$ with $\text{diag}(\mathbf{\Gamma}_N) = \mathbf{0}$ satisfying (2). Elementwise, this mapping can be written as

$$(\mathbf{\Psi}_N)_{ii}^{-1} = (\mathbf{\Theta}_N)_{ii}, \qquad \Gamma_{ji} = -\frac{(\mathbf{\Theta}_N)_{ij}}{(\mathbf{\Theta}_N)_{ii}}, \quad i \neq j,$$

so that for each asset i the ith column $\mathbf{\Gamma}_N^{(i)}$ is recovered from the ith column of $\mathbf{\Theta}_N$.

In particular, the PCA precision matrix $\widehat{\mathbf{\Theta}}_{\text{PCA}}^{(k)}$ in (9) induces a unique hedging system $(\widetilde{\mathbf{\Gamma}}_N, \widetilde{\mathbf{\Psi}}_N)$ satisfying

$$\widehat{\mathbf{\Theta}}_{\text{PCA}}^{(k)} = \widetilde{\mathbf{\Psi}}_N^{-1}\big(\boldsymbol{I}_N - \widetilde{\mathbf{\Gamma}}_N^\top\big),$$

obtained by applying the formulas above with $\mathbf{\Theta}_N = \widehat{\mathbf{\Theta}}_{\text{PCA}}^{(k)}$. These $(\widetilde{\mathbf{\Gamma}}_N, \widetilde{\mathbf{\Psi}}_N)$ describe the hedging coefficients and innovation variances that would arise if the PCA covariance $\widehat{\mathbf{\Sigma}}_{\text{PCA}}^{(k)}$ were the true population covariance. They do not, however, solve a hedging regression fitted directly to the observed returns, nor do they correspond to any standard penalized regression (e.g., ridge or lasso) on the original data. Instead, their structure is entirely dictated by the spectral truncation underlying $\widehat{\mathbf{\Sigma}}_{\text{PCA}}^{(k)}$.

Comparison to penalized reduced–rank hedging regression. Our penalized reduced–rank regression (PRR) approach differs from the PCA strategy in both objective and construction. PCA regularizes the covariance matrix by retaining directions of high total variance and shrinking the remaining eigenvalues toward a common idiosyncratic level σ_ε^2. The resulting precision matrix $\widehat{\boldsymbol{\Theta}}_{\text{PCA}}^{(k)}$ is diagonal plus low rank, and the implied hedging system is an algebraic byproduct of this spectral approximation. PCA thus optimizes variance representation, not hedgeability: it prioritizes explaining $\widehat{\boldsymbol{\Sigma}}_N$, whereas mean–variance portfolios are governed by the conditional risk structure encoded in $\boldsymbol{\Gamma}_N$ and $\boldsymbol{\Psi}_N$.

By contrast, our PRR estimator targets the hedging system directly. It estimates $(\boldsymbol{A}_N, \boldsymbol{\Gamma}_N, \boldsymbol{\Psi}_N)$ from the regression (1) and exploits the exact link (2) to recover the precision matrix. Low–rank structure is imposed on the matrix of hedging portfolio returns, not on the covariance matrix itself, and the Frobenius penalty shrinks hedging coefficients toward diversified, low–turnover portfolios. This design accommodates dense and economically structured dependence patterns without imposing orthogonality or hard sparsity on the factor structure.

From a practical standpoint, PCA requires the discrete choice of the number of principal components k, which can lead to instability and high portfolio turnover when eigenvalues cluster and eigenvectors swap order out of sample. Penalized PRR instead uses continuous tuning parameters that control the effective rank of the hedging return matrix and the overall size of hedging positions, yielding smoother tuning behavior and more stable out–of–sample GMV allocations. Moreover, because PRR is built around the hedging regressions themselves, the resulting portfolios have a direct interpretation as regression–based risk–reduction portfolios, rather than as exposures to statistical principal components that may be less closely aligned with diversification and hedging objectives.

5 Implementation

Having introduced the population formulation and the penalized reduced–rank hedging estimator in the previous section, we now turn to its empirical implementation. This section describes how we tune the regularization parameters governing the nuclear, Frobenius, and ridge penalties, and how we solve the resulting convex program efficiently in large panels. We first propose statistically and financially motivated tuning criteria, complemented by portfolio–based cross–validation, and then detail an ADMM algorithm that exploits the structure of the objective and the hollow constraint to deliver stable solutions for large (N, T).

5.1 Hyper-parameter tuning

The PRR estimator in (4) involves three regularization parameters (λ_*, λ_F, and λ_r) governing, respectively, the nuclear norm penalty on the hedging returns, the Frobenius shrinkage of hedging coefficients, and the ridge penalty applied to the mean vector. All three components have distinct economic interpretations, and their magnitudes must be chosen in a disciplined, data-driven manner.

In comparison with approaches that tune a separate penalty parameter for each asset or each regression coefficient, our framework imposes only three scalar hyper-parameters. This dramatically reduces the dimensionality of the tuning problem, while preserving the flexibility needed to regularize both the mean and the hedging structure in high dimensions. Nevertheless, classical cross-validation becomes computationally burdensome when N and T are large, and it may not align well with the financial objectives of the problem. We therefore introduce a pair of tuning criteria, one statistically motivated and one financially motivated, and complement them with cross-validation schemes that evaluate out-of-sample portfolio performance.

Statistical tuning criterion. We consider a *generalized cross validation* criterion (Wahba, 1990) that evaluates the statistical quality of the PRR fit. Given the residual matrix

$$\widehat{\boldsymbol{E}} = \boldsymbol{R} - \boldsymbol{\iota}\widehat{\boldsymbol{A}}^\top - \widetilde{\boldsymbol{R}}\widehat{\boldsymbol{\Gamma}},$$

define a model selection score

$$\mathrm{TC}(\lambda_*, \lambda_F, \lambda_r) = \log\left(\frac{\|\widehat{\boldsymbol{E}}\|_F^2}{TN}\right) - 2\log\left(1 - \mathrm{clip}_{[0,0.95]}\left(\frac{\mathrm{df}_{\mathrm{eff}}}{TN}\right)\right), \tag{10}$$

where the *effective degrees of freedom* are

$$\mathrm{df}_{\mathrm{eff}} := \Big[w\mathrm{df}_* + (1-w)\mathrm{df}_F\Big] + \mathrm{df}_r, \qquad w := \frac{\lambda_*}{\lambda_* + \lambda_F \bar{s}}, \qquad \bar{s} := \frac{1}{N}\mathrm{Tr}(\widetilde{\boldsymbol{R}}^\top \widetilde{\boldsymbol{R}}).$$

The three components quantify the model's complexity:

$$\mathrm{df}_* := r_*(2N - r_*) - N, \qquad r_* := \exp\Big(-\sum_j p_j \log p_j\Big), \qquad p_j := \frac{\sigma_j(\widetilde{\boldsymbol{R}}\widehat{\boldsymbol{\Gamma}})}{\sum_\ell \sigma_\ell(\widetilde{\boldsymbol{R}}\widehat{\boldsymbol{\Gamma}})},$$

encodes the effective rank of the hedging returns,

$$\mathrm{df}_F := (N-1)\mathrm{Tr}\Big(\widetilde{\boldsymbol{R}}(\widetilde{\boldsymbol{R}}^\top \widetilde{\boldsymbol{R}} + \lambda_F \boldsymbol{I})^{-1}\widetilde{\boldsymbol{R}}^\top\Big),$$

captures Frobenius shrinkage, and

$$\mathrm{df}_r := N\frac{T}{T + \lambda_r},$$

captures the ridge penalty applied to the mean vector.

The statistical criterion therefore balances goodness-of-fit with model complexity, blending nuclear norm rank reduction, Frobenius shrinkage, and mean-regularization effects into a single degrees-of-freedom measure.

Financial tuning criterion for the MVE allocation. The first financial tuning approach evaluates how well the estimated parameters $(\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{\Gamma}}, \widehat{\boldsymbol{\Psi}})$ support the construction of a high Sharpe–ratio mean–variance efficient portfolio while penalizing unnecessary financial complexity. Define

$$\widehat{\boldsymbol{\Theta}} := \widehat{\boldsymbol{\Psi}}^{-1}(\boldsymbol{I} - \widehat{\boldsymbol{\Gamma}}^\top), \qquad \widehat{\boldsymbol{\mu}} := \widehat{\boldsymbol{A}}, \qquad \widehat{\boldsymbol{\Sigma}}_N := \frac{1}{T}\boldsymbol{R}\boldsymbol{R}^\top.$$

The corresponding estimated MVE portfolio is $\widehat{\boldsymbol{w}} := \frac{\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\mu}}}{\boldsymbol{\iota}^\top \widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\mu}}}$.

To assess the quality of this estimator, we minimize the *financial tuning criterion*

$$\mathrm{TC}(\lambda_*, \lambda_F, \lambda_r) = \underbrace{\Big[\log\big(\widehat{\boldsymbol{w}}^\top \widehat{\boldsymbol{\Sigma}}_N \widehat{\boldsymbol{w}}\big) - 2\log\big(\max\{\varepsilon, \widehat{\boldsymbol{\mu}}^\top \widehat{\boldsymbol{w}}\}\big)\Big]}_{-\log(\mathrm{Sharpe}(\widehat{\boldsymbol{w}})^2)} + \kappa\sqrt{\frac{N}{T}}\big\|\widehat{\boldsymbol{w}} - \boldsymbol{\iota}/N\big\|_2^2, \quad (11)$$

where $\varepsilon > 0$ is a numerical safeguard and κ is fixed to 1 (or chosen from a small grid around 1).

The first term in (11) equals the negative log of the squared Sharpe ratio of $\widehat{\boldsymbol{w}}$, so minimizing TC rewards portfolios with a high expected excess return per unit of risk. The second term penalizes departures from the naive diversified portfolio $\boldsymbol{\iota}/N$, thereby controlling financial complexity by discouraging excessive leverage, concentration, and long–short exposures. Its weight decreases at rate $\sqrt{N/T}$, so that for large samples the data dominate the penalty, while for N of the same order as T the regularizer provides much-needed stabilization.

When a turnover-adjusted performance measure is desired, the numerator $\widehat{\boldsymbol{\mu}}^\top \widehat{\boldsymbol{w}}$ can be replaced by

$$\widehat{\boldsymbol{\mu}}^\top \widehat{\boldsymbol{w}} - c \cdot \mathrm{TO},$$

where TO denotes the estimated portfolio turnover and $c > 0$ represents per-unit trading costs.

Financial tuning criterion for the GMV allocation. For the global minimum–variance portfolio, expected returns do not enter the optimization problem, so the financial objective focuses solely on producing a stable and low-variance allocation. Given the same fitted objects $(\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{\Gamma}}, \widehat{\boldsymbol{\Psi}})$, the GMV estimator is

$$\widehat{\boldsymbol{w}}_{\mathrm{gmv}} := \frac{\widehat{\boldsymbol{\Theta}}\boldsymbol{\iota}}{\boldsymbol{\iota}^\top \widehat{\boldsymbol{\Theta}}\boldsymbol{\iota}}, \qquad \widehat{\boldsymbol{\Theta}} := \widehat{\boldsymbol{\Psi}}^{-1}(\boldsymbol{I} - \widehat{\boldsymbol{\Gamma}}^\top).$$

The tuning criterion specializes to

$$\mathrm{TC}_{\mathrm{gmv}}(\lambda_*, \lambda_F, \lambda_r) = \log\big(\widehat{\boldsymbol{w}}_{\mathrm{gmv}}^\top \widehat{\boldsymbol{\Sigma}}_N \widehat{\boldsymbol{w}}_{\mathrm{gmv}}\big) + \kappa\sqrt{\frac{N}{T}}\big\|\widehat{\boldsymbol{w}}_{\mathrm{gmv}} - \boldsymbol{\iota}/N\big\|_2^2. \qquad (12)$$

The first term now measures only the (log) portfolio variance, consistent with the defining objective of minimum-variance investing. The second term again discourages excessive financial complexity, particularly when N is large relative to T, ensuring that the selected hyper-parameters emphasize diversification and stability while still allowing the data to drive the solution as the sample grows.

Cross-validation criteria. In addition to the in-sample criteria above, it is natural to assess $(\lambda_*, \lambda_F, \lambda_r)$ using out-of-sample performance via cross-validation. Let $\{\mathcal{I}_\ell^{\mathrm{val}}\}_{\ell=1}^K$ denote a collection of validation blocks, either K non-overlapping folds or a sequence of rolling out-of-sample windows, and let $\mathcal{I}_\ell^{\mathrm{train}}$ denote the corresponding training indices used to fit the model for block ℓ.

For a given triple $(\lambda_*, \lambda_F, \lambda_r)$ and block ℓ, estimate $(\widehat{\boldsymbol{A}}^{(-\ell)}, \widehat{\boldsymbol{\Gamma}}^{(-\ell)}, \widehat{\boldsymbol{\Psi}}^{(-\ell)})$ using only observations in $\mathcal{I}_\ell^{\mathrm{train}}$. Construct the corresponding precision matrix

$$\widehat{\boldsymbol{\Theta}}^{(-\ell)} := \widehat{\boldsymbol{\Psi}}^{(-\ell)-1}\big(\boldsymbol{I} - \widehat{\boldsymbol{\Gamma}}^{(-\ell)}\big).$$

For the MVE allocation, form the out-of-sample portfolio

$$\widehat{\boldsymbol{w}}_{\mathrm{mve}}^{(-\ell)} := \frac{\widehat{\boldsymbol{\Theta}}^{(-\ell)}\widehat{\boldsymbol{\mu}}^{(-\ell)}}{\boldsymbol{\iota}^\top\widehat{\boldsymbol{\Theta}}^{(-\ell)}\widehat{\boldsymbol{\mu}}^{(-\ell)}}, \qquad \widehat{\boldsymbol{\mu}}^{(-\ell)} := \widehat{\boldsymbol{A}}^{(-\ell)},$$

and compute on the validation block $\mathcal{I}_\ell^{\mathrm{val}}$ the sample mean and covariance

$$\widehat{\boldsymbol{\mu}}^{(\ell)} := \frac{1}{|\mathcal{I}_\ell^{\mathrm{val}}|}\sum_{t\in\mathcal{I}_\ell^{\mathrm{val}}}\boldsymbol{R}_t, \qquad \widehat{\boldsymbol{\Sigma}}^{(\ell)} := \frac{1}{|\mathcal{I}_\ell^{\mathrm{val}}|}\sum_{t\in\mathcal{I}_\ell^{\mathrm{val}}}\big(\boldsymbol{R}_t - \widehat{\boldsymbol{\mu}}^{(\ell)}\big)\big(\boldsymbol{R}_t - \widehat{\boldsymbol{\mu}}^{(\ell)}\big)^\top.$$

An out-of-sample Sharpe-based cross-validation score is then

$$\mathrm{CV}_{\mathrm{mve}}(\lambda_*, \lambda_F, \lambda_r) := \frac{1}{K}\sum_{\ell=1}^{K}\Big[\log\big(\widehat{\boldsymbol{w}}_{\mathrm{mve}}^{(-\ell)\top}\widehat{\boldsymbol{\Sigma}}^{(\ell)}\widehat{\boldsymbol{w}}_{\mathrm{mve}}^{(-\ell)}\big) - 2\log\big(\max\{\varepsilon, \widehat{\boldsymbol{\mu}}^{(\ell)\top}\widehat{\boldsymbol{w}}_{\mathrm{mve}}^{(-\ell)}\}\big)\Big],$$

which mirrors the financial criterion (11), but is evaluated out-of-sample on the validation blocks.

For the GMV allocation, define the out-of-sample GMV weights for block ℓ as

$$\widehat{\boldsymbol{w}}_{\mathrm{gmv}}^{(-\ell)} := \frac{\widehat{\boldsymbol{\Theta}}^{(-\ell)}\boldsymbol{\iota}}{\boldsymbol{\iota}^\top\widehat{\boldsymbol{\Theta}}^{(-\ell)}\boldsymbol{\iota}},$$

and set

$$\mathrm{CV}_{\mathrm{gmv}}(\lambda_*, \lambda_F, \lambda_r) := \frac{1}{K}\sum_{\ell=1}^{K}\log\big(\widehat{\boldsymbol{w}}_{\mathrm{gmv}}^{(-\ell)\top}\widehat{\boldsymbol{\Sigma}}^{(\ell)}\widehat{\boldsymbol{w}}_{\mathrm{gmv}}^{(-\ell)}\big),$$

which averages the log out-of-sample portfolio variance across folds or rolling windows. Minimizing $\mathrm{CV}_{\mathrm{gmv}}$ and $\mathrm{CV}_{\mathrm{mve}}$ over $(\lambda_*, \lambda_F, \lambda_r)$ yields tuning choices explicitly geared toward out-of-sample risk and Sharpe performance.

Practical implementation. In empirical work, the statistical criterion (10), the financial criteria (11)–(12), and, when computationally feasible, the cross-validation scores $\mathrm{CV}_{\mathrm{gmv}}$ and $\mathrm{CV}_{\mathrm{mve}}$ can be evaluated on a grid over $(\lambda_*, \lambda_F, \lambda_r)$. The financial criteria are particularly attractive when portfolio performance is the primary objective, whereas the statistical criterion provides a complementary measure of fit and effective complexity. In challenging high-dimensional applications, careful design of the tuning grid (for example, using logarithmic scales and adapting the range of λ_F and λ_* to the spectrum of $\widetilde{\boldsymbol{R}}$) is important to balance numerical tractability and coverage of economically relevant regimes. In our experience, the Frobenius penalty λ_F becomes especially useful in very high-dimensional panels, where it improves conditioning and stabilizes the ADMM updates, while smaller or even negligible values of λ_F can suffice in more moderate dimensions. Across our empirical and Monte Carlo applications, these tuning strategies remain computationally tractable for large

(N, T) and deliver stable, economically interpretable portfolio allocations.

5.2 Solving algorithm

We now describe the algorithm used to compute the PRR estimator in (4). The optimization problem is convex but combines a non-smooth nuclear norm with a Frobenius penalty, a ridge penalty on the mean vector, and the linear constraint $\mathrm{diag}(\boldsymbol{\Gamma}) = \mathbf{0}$. We exploit the structure of the objective and constraints and solve the problem with an alternating direction method of multipliers (ADMM) scheme (Boyd, Parikh, Chu, Peleato, Eckstein et al., 2011) tailored to our hedging setting.

For notational simplicity, we present the algorithm in a slightly simplified form where we shrink $\boldsymbol{A}$ toward the origin, that is, with $\lambda_r \|\boldsymbol{A}\|_2^2/2$ instead of $\lambda_r \|\boldsymbol{A} - \boldsymbol{\mu}^*\|_2^2/2$ in (4). The extension to a general target $\boldsymbol{\mu}^*$ is trivial and implemented in our code by reparametrization.

Recall the PRR objective (4):

$$\min_{\substack{\boldsymbol{A}\in\mathbb{R}^N \\ \boldsymbol{\Gamma}\in\mathbb{R}^{N\times N}}} \left\{ \frac{1}{2}\left\|\boldsymbol{R} - \boldsymbol{\iota}\boldsymbol{A}^\top - \widetilde{\boldsymbol{R}}\boldsymbol{\Gamma}\right\|_F^2 + \lambda_* \left\|\widetilde{\boldsymbol{R}}\boldsymbol{\Gamma}\right\|_* + \frac{\lambda_F}{2}\|\boldsymbol{\Gamma}\|_F^2 + \frac{\lambda_r}{2}\|\boldsymbol{A} - \boldsymbol{\mu}^*\|_2^2 \ : \ \mathrm{diag}(\boldsymbol{\Gamma}) = \mathbf{0} \right\},$$

To apply ADMM, we introduce two auxiliary variables:

$$\boldsymbol{W} := \widetilde{\boldsymbol{R}}\boldsymbol{\Gamma}, \qquad \boldsymbol{H} := \boldsymbol{\Gamma}.$$

The variable $\boldsymbol{W}$ decouples the nuclear norm from the least-squares term, while $\boldsymbol{H}$ carries the hollow constraint. The problem can then be written as

$$\min_{\boldsymbol{A},\boldsymbol{\Gamma},\boldsymbol{W},\boldsymbol{H}} \frac{1}{2}\left\|\boldsymbol{R} - \boldsymbol{\iota}\boldsymbol{A}^\top - \boldsymbol{W}\right\|_F^2 + \lambda_*\|\boldsymbol{W}\|_* + \frac{\lambda_F}{2}\|\boldsymbol{\Gamma}\|_F^2 + \frac{\lambda_r}{2}\|\boldsymbol{A}\|_2^2$$

subject to

$$\boldsymbol{W} = \widetilde{\boldsymbol{R}}\boldsymbol{\Gamma}, \qquad \boldsymbol{H} = \boldsymbol{\Gamma}, \qquad \mathrm{diag}(\boldsymbol{H}) = \mathbf{0}.$$

The last two constraints ensure that $\boldsymbol{\Gamma}$ and $\boldsymbol{H}$ coincide while enforcing the no–self-hedging restriction on the diagonal of $\boldsymbol{\Gamma}$.

Augmented Lagrangian and ADMM updates. Let $\boldsymbol{U}$ and $\boldsymbol{V}$ denote the dual variables associated with the constraints $\boldsymbol{W} = \widetilde{\boldsymbol{R}}\boldsymbol{\Gamma}$ and $\boldsymbol{H} = \boldsymbol{\Gamma}$, respectively, and let $\rho_1, \rho_2 > 0$ be penalty parameters. The augmented Lagrangian is

$$\begin{aligned}\mathcal{L}(\boldsymbol{A}, \boldsymbol{\Gamma}, \boldsymbol{W}, \boldsymbol{H}, \boldsymbol{U}, \boldsymbol{V}) = \frac{1}{2}\big\|\boldsymbol{R} - \boldsymbol{\iota}\boldsymbol{A}^\top - \boldsymbol{W}\big\|_F^2 + \lambda_*\|\boldsymbol{W}\|_* + \frac{\lambda_F}{2}\|\boldsymbol{\Gamma}\|_F^2 + \frac{\lambda_r}{2}\|\boldsymbol{A}\|_2^2 \\ + \frac{\rho_1}{2}\big\|\boldsymbol{W} - \widetilde{\boldsymbol{R}}\boldsymbol{\Gamma} + \boldsymbol{U}\big\|_F^2 + \frac{\rho_2}{2}\big\|\boldsymbol{\Gamma} - \boldsymbol{H} + \boldsymbol{V}\big\|_F^2,\end{aligned}$$

with the understanding that $\boldsymbol{H}$ is always projected onto the set of hollow matrices, $\{\boldsymbol{H} \in \mathbb{R}^{N\times N} : \operatorname{diag}(\boldsymbol{H}) = \boldsymbol{0}\}$.

ADMM proceeds by minimizing $\mathcal{L}$ blockwise in $(\boldsymbol{A}, \boldsymbol{W}, \boldsymbol{\Gamma}, \boldsymbol{H})$ while updating the dual variables by gradient ascent. One iteration consists of the following steps.

1. Update of $\boldsymbol{A}$ (ridge intercepts). Conditioning on $(\boldsymbol{W}, \boldsymbol{\Gamma}, \boldsymbol{H}, \boldsymbol{U}, \boldsymbol{V})$, the augmented Lagrangian is quadratic in $\boldsymbol{A}$:

$$\min_{\boldsymbol{A}\in\mathbb{R}^N} \frac{1}{2}\big\|\boldsymbol{R} - \boldsymbol{\iota}\boldsymbol{A}^\top - \boldsymbol{W}\big\|_F^2 + \frac{\lambda_r}{2}\|\boldsymbol{A}\|_2^2.$$

This is a ridge regression problem with closed-form solution

$$\boldsymbol{A} \leftarrow \frac{(\boldsymbol{R} - \boldsymbol{W})^\top \boldsymbol{\iota}}{T + \lambda_r}.$$

2. Update of $\boldsymbol{W}$ (nuclear norm proximal step). The $\boldsymbol{W}$-update solves

$$\min_{\boldsymbol{W}} \frac{1}{2}\big\|\boldsymbol{R} - \boldsymbol{\iota}\boldsymbol{A}^\top - \boldsymbol{W}\big\|_F^2 + \lambda_*\|\boldsymbol{W}\|_* + \frac{\rho_1}{2}\big\|\boldsymbol{W} - \widetilde{\boldsymbol{R}}\boldsymbol{\Gamma} + \boldsymbol{U}\big\|_F^2.$$

This is a proximal problem for the nuclear norm. Defining

$$\boldsymbol{B} := \boldsymbol{R} - \boldsymbol{\iota}\boldsymbol{A}^\top, \qquad \boldsymbol{C} := \widetilde{\boldsymbol{R}}\boldsymbol{\Gamma} - \boldsymbol{U}, \qquad \boldsymbol{M} := \frac{\boldsymbol{B} + \rho_1\boldsymbol{C}}{1 + \rho_1},$$

The solution is given by singular value soft–thresholding. Let

$$\boldsymbol{M} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^\top$$

be a singular value decomposition (SVD) of $\boldsymbol{M}$, where $\boldsymbol{U}$ and $\boldsymbol{V}$ are orthonormal and $\boldsymbol{S} = \operatorname{diag}(\sigma_1, \dots, \sigma_r)$ collects the singular values $\sigma_1 \geq \cdots \geq \sigma_r \geq 0$. Define the thresholded diagonal matrix

$$\boldsymbol{S}_\tau := \operatorname{diag}\big((\sigma_1 - \tau)_+, \dots, (\sigma_r - \tau)_+\big), \qquad (x)_+ := \max\{x, 0\},$$

with $\tau := \lambda_*/(1+\rho_1)$. Then the $\boldsymbol{W}$–update is

$$\boldsymbol{W} \leftarrow \operatorname{SVT}(\boldsymbol{M}, \tau) := \boldsymbol{U}\boldsymbol{S}_\tau\boldsymbol{V}^\top,$$

that is, we shrink each singular value of $\boldsymbol{M}$ toward zero by τ, setting it to zero whenever $\sigma_j \leq \tau$.

3. Update of $\boldsymbol{\Gamma}$ *(linear system).* The $\boldsymbol{\Gamma}$-update solves

$$\min_{\boldsymbol{\Gamma}} \frac{\lambda_F}{2}\|\boldsymbol{\Gamma}\|_F^2 + \frac{\rho_1}{2}\big\|\boldsymbol{W} - \widetilde{\boldsymbol{R}}\boldsymbol{\Gamma} + \boldsymbol{U}\big\|_F^2 + \frac{\rho_2}{2}\big\|\boldsymbol{\Gamma} - \boldsymbol{H} + \boldsymbol{V}\big\|_F^2.$$

This is a strictly convex quadratic problem with normal equations

$$\big(\rho_1\widetilde{\boldsymbol{R}}^\top\widetilde{\boldsymbol{R}} + (\lambda_F + \rho_2)\boldsymbol{I}\big)\boldsymbol{\Gamma} = \rho_1\widetilde{\boldsymbol{R}}^\top(\boldsymbol{W} + \boldsymbol{U}) + \rho_2(\boldsymbol{H} - \boldsymbol{V}),$$

which we solve using a pre-factorization of $\rho_1\widetilde{\boldsymbol{R}}^\top\widetilde{\boldsymbol{R}} + (\lambda_F + \rho_2)\boldsymbol{I}$.

4. Update of $\boldsymbol{H}$ *(projection onto hollow matrices).* The variable $\boldsymbol{H}$ enforces the zero-diagonal constraint through

$$\boldsymbol{H} \leftarrow \boldsymbol{\Gamma} + \boldsymbol{V}, \qquad \operatorname{diag}(\boldsymbol{H}) := \boldsymbol{0}.$$

Thus the $\boldsymbol{H}$-update is simply a projection step that zeroes the diagonal of $\boldsymbol{\Gamma} + \boldsymbol{V}$.

5. Dual variable updates. Finally, the dual variables are updated by

$$\boldsymbol{U} \leftarrow \boldsymbol{U} + (\boldsymbol{W} - \widetilde{\boldsymbol{R}}\boldsymbol{\Gamma}), \qquad \boldsymbol{V} \leftarrow \boldsymbol{V} + (\boldsymbol{\Gamma} - \boldsymbol{H}).$$

Convergence, initialization, and warm starts. Under convexity and mild regularity conditions, ADMM with fixed $\rho_1, \rho_2 > 0$ converges at an $\mathcal{O}(1/k)$ rate in terms of the primal–dual gap, where k is the iteration index; see Boyd et al. (2011) for rate results and Eckstein and Bertsekas (1992) for the underlying operator-splitting convergence theory. In practice, we monitor both the primal residuals $\boldsymbol{W} - \widetilde{\boldsymbol{R}}\boldsymbol{\Gamma}$ and $\boldsymbol{\Gamma} - \boldsymbol{H}$ and the dual residuals implied by the updates of $\boldsymbol{U}$ and $\boldsymbol{V}$, and we stop once they fall below standard absolute and relative tolerances.

We initialize $(\boldsymbol{A}, \boldsymbol{\Gamma}, \boldsymbol{W}, \boldsymbol{U}, \boldsymbol{V})$ either at zero or using a warm start from a nearby point in the tuning grid. In particular, when computing a path over decreasing values of λ_* (for fixed λ_F, λ_r), we use the solution at the previous grid point as initialization for the next one. This warm-start strategy substantially reduces computation time and stabilizes the algorithm in high-dimensional panels.

Overall, the ADMM solver exploits the special structure of the PRR problem: ridge regression for the intercepts, a simple SVD-based proximal step for the nuclear norm, a linear system for the hedging coefficients, and a trivial projection for the hollow constraint. This yields a numerically stable and scalable algorithm that can handle large cross sections and long time series while faithfully implementing the nuclear-regularized hedging framework.

6 Empirical analysis

This section examines the out-of-sample performance of our proposed method, with particular emphasis on net Sharpe ratios as the primary evaluation metric. To provide additional insight, we also analyze volatility, expected returns, and key characteristics of the resulting portfolio positions, including turnover and the extent of short selling.

Our analysis covers three datasets of managed portfolios, representing low-, medium-, and high-dimensional cross-sections, to assess the robustness of the method across different data sizes. We implement our methodology to construct minimum-variance portfolios, calibrating the shrinkage intensity via 5-fold cross-validation.

6.1 Data

We collect three sets of daily returns on managed portfolios from Kenneth French's data library. The data spans January 1967 to December 2019. The first set includes 49 industry sorts; the second set includes 100 independent portfolio sorts based on size and investment characteristics; and the third set is the union of both industry and managed portfolios. After removing assets with missing values at any time between January 1965 and January 2020, the resulting balanced panels contain 47, 100, and 147, respectively.

To reduce computational costs, we update estimates of the portfolio weights once per quarter. On the first day of every new quarter, the allocations are estimated using the prior 90 days of returns ($M = 90$). The portfolios are then rebalanced daily, such that the optimal wealth allocation is maintained until the next estimation date. We assess the financial performance in excess of the 1-month T-bill obtained from Kenneth French's data library.

6.2 Performance evaluation

We evaluate the performance of our approach relative to several prominent alternatives in the literature. As a fundamental benchmark, we include the minimum-variance portfolio constructed using the sample covariance matrix. DeMiguel et al. (2009) show that an equally weighted portfolio can deliver competitive mean-variance performance out of sample, particularly once transaction costs are taken into account. Accordingly, we consider an equally weighted portfolio, as well as the optimal combination of minimum-variance and equally weighted allocations derived in their study under the assumption of multivariate normality.

To assess the economic value of hedging relationships between assets, we also consider the volatility-timing portfolios proposed by Kirby and Ostdiek (2012). In this framework, portfolio weights for asset i are given by

$$w_i = \frac{(1/\sigma_i^2)^\eta}{\sum_{i=1}^{N}(1/\sigma_i^2)^\eta},$$

where $\eta \geq 0$ governs the aggressiveness with which portfolio weights adjust to changes in asset volatility. This approach provides a useful benchmark because it does not explicitly account for hedging opportunities across assets. Moreover, volatility-timing portfolios avoid the need for covariance matrix inversion.

Table 1: Minimum variance allocations considered

Abbreviation	Description	Reference
RRR	Nuclear and Frobenius regularization	
Sample	Sample covariance matrix	
Combination	Combines the $1/N$ and sample gmv portfolio	DeMiguel et al. (2009)
GLasso	Graphical lasso estimator of the precision matrix	Goto and Xu (2015)
Nodewise	Nodewise estimator of the precision matrix	Callot et al. (2021)
Vol-timing-1	Volatility timing portfolio setting $\eta = 1$	Kirby and Ostdiek (2012)
Vol-timing-4	Volatility timing portfolio setting $\eta = 4$	Kirby and Ostdiek (2012)
Lin-Corr	Linear shrinkage towards equal correlation	Ledoit and Wolf (2003)
Lin-Diag	Linear shrinkage towards diagonal matrix	Ledoit and Wolf (2003)
Lin-Iden	Linear shrinkage towards scaled identity	Ledoit and Wolf (2004)
Non-Lin	Non-linear shrinkage of covariance matrix	Ledoit and Wolf (2015)

Alternative methods that directly target the precision matrix in our empirical evaluation include the Graphical Lasso (Goto and Xu, 2015) and the nodewise regression approach proposed by Callot et al. (2021). We further include covariance matrix estimators well-suited for high-dimensional estimation challenges, such as the linear shrinkage estimators proposed by Ledoit and Wolf (2004) and the non-linear shrinkage estimators proposed by Ledoit and Wolf (2015). Ledoit and Wolf (2022) provide an extensive discussion about many shrinkage estimators and their use in portfolio optimization, emphasizing the efficiency of the non-linear shrinkage estimator. This is not an exhaustive evaluation of all relevant estimators proposed in the literature. We restrict our attention to these methods due to their clear methodological contrast to the methodology that we propose. We summarize the methods in Table 1.

For each of the k portfolio allocations we compute the daily out-of-sample return $R_t^{(k)} = \boldsymbol{R}_t' \boldsymbol{w}_{t-1}^{(k)}$, along with excess returns over the 1-month T-bill, and transaction costs:

$$tc_t^{(k)} = \sum_{i=1}^{N} c_{i,t} \left| w_{i,t} - w_{i,t-1} \frac{1 + R_{i,t}}{1 + R_t^{(k)}} \right| .$$

We compute the out-of-sample volatility and the net Sharpe ratio of the respective portfolio allocations as follows:

$$\text{Vol}^{(k)} = \sqrt{\mathbb{V}ar\left[R_t^{(k),excess}\right]}, \tag{13}$$

$$\text{SR}^{(k),net} = \frac{\mathbb{E}\left[R_t^{(k),excess} - tc_t^{(k)}\right]}{\sqrt{\mathbb{V}ar\left[R_t^{(k),excess}\right]}}. \tag{14}$$

The population quantities are estimated by their sample counterparts over the out-of-sample data starting at $M + 1$ and ending at T. Since we obtain a large sample of out-of-sample portfolio returns, we perform inference on the performance differences using the HAC procedures proposed by Ledoit and Wolf (2008) and Ledoit and Wolf (2011) respectively.

6.3 Volatilities and Sharpe ratios

The out-of-sample portfolio volatilities and Sharpe ratios obtained using the respective methods are summarized in Table 2. Overall, we find that the estimator that we propose delivers low portfolio variance and high net Sharpe ratios once transaction costs are taken into account. We discuss the performance of the different classes of estimators in turn.

The performance of the shrinkage estimators is reported in Panel D of Table 2. These methods are designed to mitigate estimation error when the cross-section of

Table 2: Out-of-sample Volatility and Sharpe ratio performance

	Industries $N = 47$		ME/BE $N = 96$		Ind. + ME/BE $N = 143$	
	$\hat{\sigma}$	SR^{net}	$\hat{\sigma}$	SR^{net}	$\hat{\sigma}$	SR^{net}
	Panel A: Nuclear and Frobenius regularization					
RRR	10.40	0.26	9.51	0.71	8.85	0.62
	Panel B: Unconstrained moments					
Sample	12.29***	−0.26	—	—	—	—
Combination	12.04***	0.16	—	—	—	—
	Panel C: Moment restrictions					
$1/N$	15.68***	0.44	16.02***	0.51	15.68***	0.51
Vol-timing-1	14.19***	0.47	14.90***	0.56	14.41***	0.56
Vol-timing-4	11.62***	0.44	12.67***	0.71	11.55***	0.66
	Panel D: Shrinkage estimators					
Lin-Corr	10.38	0.07	8.98***	0.61	8.41***	0.34
Lin-Diag	10.39	0.07	9.18***	0.32	8.55***	0.22
Lin-Iden	10.45	0.18	9.38	0.34	8.63***	0.29
Non-Lin	10.44	0.19	9.08***	0.57	8.42***	0.53
Nodewise	12.04***	0.33	12.53***	0.46	12.28***	0.49
GLasso	15.68***	0.44	16.02***	0.51	15.67***	0.51

The table reports out-of-sample portfolio performance during the period January 4 1967 - December 31 2019. The length of the estimation window is $M = 90$. Annualized Sharpe ratio is denoted SR. We test the null hypothesis of equal Sharpe ratios using the HAC procedure proposed by Ledoit and Wolf (2008). The tests evaluate differences in performance between the RRR approach and the competing estimators in a pairwise fashion. We indicate significant differences at 1 percent, 5 percent and 10 percent significance levels by ***, ** and * respectively.

assets is large relative to the available time-series observations. Focusing first on the 47 industry portfolios, we find that both linear and non-linear shrinkage estimators achieve volatility levels that are comparable to those obtained under the reduced-rank regularization (RRR) approach. However, their net Sharpe ratios are consistently lower. In particular, linear shrinkage estimators targeting the covariance matrix deliver net Sharpe ratios that are close to zero, reflecting a pronounced sensitivity to transaction costs.

As the cross-section expands to include portfolios managed on size and value, volatility reductions achieved by shrinkage estimators become statistically significant relative to RRR. However, these reductions do not translate into superior net performance. The non-linear shrinkage estimator improves upon its linear counterparts but still falls short of the RRR benchmark in terms of SR^{net}. The Nodewise estimator and the Graphical Lasso are less adversely affected by transaction costs and achieve intermediate performance, this is likely because of aggressive shrinkage of off-diagonal elements in the precision matrix. Nevertheless, even these methods do not obtain higher net Sharpe ratios than the reduced-rank approach when the cross-section of assets is large.

Panel C presents the performance of portfolios constructed under moment restrictions. The volatility-timing strategies do not exploit any hedging opportunities in the set of assets. In the small cross-section of industry portfolios, the losses due to omitting hedging are small. The portfolio variance obtained by emphasizing allocation of wealth to assets with low variance (*i.e.* Vol-timing-4) is higher than if hedging opportunities are used, but both the Vol-timing-1 and Vol-timing-4 strategies deliver higher net Sharpe ratios than most competing estimators.

However, omitting hedging relations becomes costly in terms of volatility reduction when the portfolio includes assets managed on financial characteristics that exhibit a strong factor structure. Within the set of size and value managed portfolios there are substantial hedging opportunities, which the RRR approach exploits efficiently to achieve improved diversification. Along with the greater reduction in portfolio volatility, the reduced-rank approach also delivers high net performance that is well in excess of the equally weighted portfolio allocation, even in the presence of transaction costs.

Turning to the performance of the sample-based minimum-variance portfolio, and the combination of the minimum-variance allocation and the equally weighed portfolio, we find that the portfolio variances are high and the net performance is comparably low. This result underscores the importance of regularization, in particular when the cross-section of assets is large.

6.4 Expected returns, gross Sharpe ratios, and Turnover

We decompose net Sharpe ratio performance by reporting annualized gross Sharpe ratios, mean returns, and average daily turnover, all presented in Table 3. The table highlights that our methodology also delivers high gross performance and keeps trading intensity at moderate levels relative to competing approaches that exploit hedging opportunities.

The results show that omitting regularization and forming the minimum variance portfolio using the sample covariance matrix yields portfolio weights that are substantially over-fitted. This is reflected in low gross performance and very high turnover. Introducing additional bias through regularization on the nuclear and Frobenius norms improves performance and reduces trading. Similarly, combining the minimum variance allocation with the equally weighted portfolio reduces turnover sharply relative to the pure sample based minimum variance portfolio, while delivering higher mean returns and a higher Sharpe ratio. Nevertheless, the combined portfolio still falls short of the RRR approach in terms of gross Sharpe ratio, and its turnover remains at least as high as that required to maintain the RRR portfolio in the set of industry portfolios.

All shrinkage estimators reported in Panel D deliver minimum variance weights that incorporate hedging opportunities across assets. In the smaller cross section of industry portfolios, RRR performs strongly relative to the shrinkage alternatives. The RRR approach attains one of the highest gross Sharpe ratios in Panel D while maintaining lower turnover than the linear shrinkage estimators. Linear shrinkage methods targeting the covariance matrix produce Sharpe ratios that are substantially below RRR, and their turnover is materially higher. The non linear shrinkage estimator improves gross performance relative to the linear methods and comes closer to RRR in terms of Sharpe ratio, but it still requires higher turnover than RRR in the industry universe.

As the cross section expands to include portfolios managed on size and value, sev-

Table 3: Out-of-sample Expected returns, Sharpe ratios, and Turnover

	Industries $N = 47$			ME/BE $N = 96$			Ind + ME/BE $N = 143$		
	$\hat{\mu}$	SR	TO	$\hat{\mu}$	SR	TO	$\hat{\mu}$	SR	TO
	Panel A: Nuclear and Frobenius regularization								
RRR	5.39	0.52	5.47	9.71	1.02	5.96	9.07	1.03	7.19
	Panel B: Unconstrained moments								
Sample	3.92	0.32	14.26						
Combination	4.76	0.40	5.69						
	Panel C: Moment restrictions								
1/N	7.26	0.46	0.63	8.43	0.53	0.46	8.20	0.52	0.51
Vol-timing-1	7.01	0.49	0.83	8.77	0.59	0.72	8.43	0.59	0.75
Vol-timing-4	5.72	0.49	1.32	9.69	0.76	1.41	8.34	0.72	1.41
	Panel D: Shrinkage estimators								
Lin-Corr	4.37	0.42	7.22	9.37	1.04	7.86	9.04	1.07	12.30
Lin-Diag	4.62	0.44	7.77	9.84	1.07	13.76	9.04	1.06	14.24
Lin-Iden	5.38	0.52	7.09	9.80	1.04	13.14	9.30	1.08	13.56
Non-Lin	5.09	0.49	6.10	9.39	1.03	8.35	8.73	1.04	8.56
Nodewise	5.23	0.43	2.49	7.45	0.59	3.40	7.43	0.60	2.79
GLasso	7.26	0.46	0.63	8.43	0.53	0.46	8.20	0.52	0.51

The table reports out-of-sample portfolio performance during the period January 4, 1967–December 31, 2019. The length of the estimation window is $M = 90$. Annualized mean returns are denoted $\hat{\mu}$ (%) and SR denotes the annualized Sharpe ratio. The average daily turnover is denoted TO (%). Hypothesis tests on Sharpe ratios are omitted.

eral shrinkage estimators achieve Sharpe ratios that are similar to, and in some cases slightly above, the RRR benchmark. However, these gains are accompanied by much higher trading intensity. In the largest cross section, the linear and non linear shrinkage estimators deliver Sharpe ratios around one, but their turnover is substantially larger than that of RRR. In contrast, the RRR approach achieves a Sharpe ratio close to the top performing shrinkage methods while requiring meaningfully less turnover, indicating that it is better positioned to preserve performance once transaction costs are introduced.

Estimators that target the precision matrix directly require less turnover than covariance based shrinkage methods. The Nodewise regression approach stands out as requiring very low turnover across all three universes. The drawback is that its mean returns and Sharpe ratios are materially lower than those obtained under RRR in the medium and large cross sections. The Graphical Lasso results coincide with the equally weighted allocation throughout, implying very low turnover but also a level of performance that does not improve with the expansion of the asset universe.

Panel C reports the performance of portfolios constructed under moment restrictions. These strategies do not exploit cross asset hedging relations. The equally weighted portfolio yields low turnover and a Sharpe ratio that is stable across the three universes, but it remains well below the RRR benchmark in the medium and large cross sections. The volatility timing strategies deliver higher Sharpe ratios than the equally weighted portfolio while still requiring very low turnover. Their improvements are most pronounced when size and value managed portfolios are included. Even so, the RRR approach remains competitive in terms of risk adjusted performance, and it does so while explicitly exploiting hedging opportunities across assets. Taken together, the results in Table 3 indicate that the reduced rank approach effectively balances the benefits of exploiting hedging relations against the trading required to maintain those positions, leading to strong overall performance.

6.5 Portfolio weights

Lastly, we turn to the composition of weights in the portfolios. Table 4 summarizes the time-series averages of the fraction of negative weights, the sum of squared weights, and the smallest and largest weights obtained across the respective methods. Overall, we find that the RRR estimator produces portfolios that are well diversified and avoid excessively extreme positions relative to most alternative approaches that rely on estimating the full covariance matrix of returns.

We assess the degree of diversification across the portfolios through the time-series average of the sum of squared weights. Comparing our methodology with estimators that target the full covariance matrix shows that, across all portfolio universes, the RRR methodology consistently produces portfolios with relatively low concentration. In the largest dataset, the sum of squared weights obtained under RRR is comparable to that of the portfolios formed by aggressive volatility timing and closely matches the values produced by precision matrix estimators. While portfolios that completely omit hedging relations achieve the lowest concentration, among methods that exploit hedging opportunities we find that RRR delivers a high degree of diversification that is similar to that of the GLasso and non-linear shrinkage estimators.

Similarly, when considering the magnitude of extreme positions in portfolios that incorporate hedging opportunities, the RRR methodology generally yields less extreme long positions than linear shrinkage estimators and the sample-based approach. At the same time, the average largest weight under RRR remains more pronounced than those obtained using the Nodewise regression approach, particularly in larger cross sections. With respect to short positions, our methodology performs favorably. The smallest weights under RRR are substantially less extreme than those implied by most shrinkage estimators and are comparable to those produced by GLasso across all datasets.

In practice, forming portfolios with extensive short-selling positions is often costly and difficult to implement. The methodology that we propose does not rely on short-selling to a greater extent than most competing approaches. On average, the RRR methodology implies that approximately 41 to 44 percent of assets carry negative weights, depending on the size of the investment universe. Among the methods that permit short-selling, only the Nodewise regression approach consistently produces portfolios that rely on a markedly smaller fraction of short positions.

Table 4: Portfolio weights

	Industries $N = 47$				ME/BE $N = 96$				Industries + ME/BE $N = 143$			
	$\overline{(\boldsymbol{w} < 0)/N}$	$\overline{\Vert\boldsymbol{w}\Vert_2^2}$	$\overline{\min(\boldsymbol{w})}$	$\overline{\max(\boldsymbol{w})}$	$\overline{(\boldsymbol{w} < 0)/N}$	$\overline{\Vert\boldsymbol{w}\Vert_2^2}$	$\overline{\min(\boldsymbol{w})}$	$\overline{\max(\boldsymbol{w})}$	$\overline{(\boldsymbol{w} < 0)/N}$	$\overline{\Vert\boldsymbol{w}\Vert_2^2}$	$\overline{\min(\boldsymbol{w})}$	$\overline{\max(\boldsymbol{w})}$
	Panel A: Nuclear and Frobenius regularization											
RRR	0.41	0.38	−0.13	0.35	0.44	0.22	−0.08	0.20	0.43	0.17	−0.08	0.16
	Panel B: Unconstrained moments											
Sample	0.46	1.36	−0.34	0.59								
Combination	0.34	0.30	−0.13	0.27								
	Panel C: Moment restrictions											
$1/N$	0.00	0.02	0.02	0.02	0.00	0.01	0.01	0.01	0.00	0.01	0.01	0.01
Vol-timing-1	0.00	0.03	0.00	0.07	0.00	0.01	0.00	0.03	0.00	0.01	0.00	0.02
Vol-timing-4	0.00	0.28	0.00	0.42	0.00	0.08	0.00	0.19	0.00	0.11	0.00	0.22
	Panel D: Shrinkage estimators											
Lin-Corr	0.49	0.60	−0.15	0.47	0.52	0.37	−0.08	0.27	0.51	0.40	−0.09	0.25
Lin-Diag	0.46	0.60	−0.17	0.45	0.49	0.68	−0.15	0.33	0.49	0.46	−0.11	0.25
Lin-Iden	0.42	0.41	−0.16	0.29	0.46	0.56	−0.16	0.23	0.44	0.36	−0.12	0.15
Non-Lin	0.40	0.33	−0.14	0.26	0.42	0.29	−0.12	0.16	0.41	0.17	−0.08	0.10
Nodewise	0.19	0.12	−0.03	0.23	0.21	0.09	−0.03	0.15	0.16	0.05	−0.01	0.12
GLasso	0.41	0.38	−0.13	0.35	0.44	0.22	−0.08	0.20	0.43	0.17	−0.08	0.16

The table reports out-of-sample portfolio performance during the period January 4 1967 - December 31 2019. The length of the estimation window is $M = 90$. The columns summarize the time-series average of the: mean number of short-selling positions $\left(\overline{(\boldsymbol{w} < 0)/N}\right)$, sum of squared weights $\left(\overline{\vert\vert\boldsymbol{w}\vert\vert_2^2}\right)$, smallest weight $\left(\overline{\min(\boldsymbol{w})}\right)$ and largest weight $\left(\overline{\max(\boldsymbol{w})}\right)$.

7 Conclusions

In this paper we approach modern portfolio optimization through a novel estimator of the precision matrix of returns. Our method controls both the effective dimensionality of hedging portfolio returns and the size of the hedging positions, thereby mitigating idiosyncratic risk and reducing transaction costs. We achieve this through Nuclear norm and Frobenius norm penalties. The approach that we promote remains viable when the number of assets exceeds the length of the time-series, and provides a financially intuitive departure from many existing methods. In particular, we circumvent the need for restrictive sparsity assumptions on the precision matrix of returns. This makes our approach attractive in cases where returns are driven by a strong factor structure.

In an empirical experiment across datasets of various dimensions, we find that our method delivers portfolio allocations with low turnover, and high net Sharpe ratios relative to competing methods, including sparse precision estimators, shrinkage covariance estimators, and volatility-timing strategies. Our results highlight the advantages of low-rank specifications in capturing the latent factor structure of returns without imposing strict sparsity. Empirically, we demonstrate that such regularization not only stabilizes estimates when the cross-section of assets grows large but also yields practical benefits such as lower leverage and fewer extreme positions. Moreover, by simultaneously penalizing the hedging coefficients and their overall span, the method effectively balances systematic hedging opportunities against the need to constrain trading costs.

A number of open avenues remain for further research. Interesting refinements to our methodology include the possibility of informing hedging positions in accordance with asset pricing models with observable factors. There are also open questions remaining about the most effective manner of introducing cost mitigation within the portfolio allocation challenge that we explore.

References

Bai, J. and S. Ng (2002): "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70, 191–221.

Ban, G.-Y., N. El Karoui, and A. E. Lim (2018): "Machine learning and portfolio optimization," *Management Science*, 64, 1136–1154.

Best, M. J. and R. R. Grauer (1991): "On the sensitivity of mean-variance-efficient portfolios to changes in asset means: some analytical and computational results," *The review of financial studies*, 4, 315–342.

Boyd, S., N. Parikh, E. Chu, B. Peleato, J. Eckstein, et al. (2011): "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine learning*, 3, 1–122.

Callot, L., M. Caner, A. Ö. Önder, and E. Ulaşan (2021): "A nodewise regression approach to estimating large portfolios," *Journal of Business & Economic Statistics*, 39, 520–531.

Chamberlain, G. (1983): "Funds, factors, and diversification in arbitrage pricing models," *Econometrica: Journal of the Econometric Society*, 1305–1323.

Chamberlain, G. and M. Rothschild (1983): "Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets," *Econometrica*, 51, 1281–304.

Connor, G. and R. A. Korajczyk (1986): "Performance measurement with the arbitrage pricing theory: A new framework for analysis," *Journal of financial economics*, 15, 373–394.

De Nard, G., O. Ledoit, and M. Wolf (2021): "Factor models for portfolio selection in large dimensions: The good, the better and the ugly," *Journal of Financial Econometrics*, 19, 236–257.

DeMiguel, V., L. Garlappi, and R. Uppal (2009): "Optimal versus Naive Diversification: How Inefficient Is the 1/N Portfolio Strategy?" *Review of Financial Studies*, 22, 1915–1953.

DeMiguel, V., A. Martin-Utrera, and F. J. Nogales (2013): "Size matters: Optimal calibration of shrinkage estimators for portfolio selection," *Journal of Banking & Finance*, 37, 3018–3034.

Eckstein, J. and D. P. Bertsekas (1992): "On the Douglas—Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Mathematical programming*, 55, 293–318.

Fan, J., Y. Liao, and M. Mincheva (2013): "Large covariance estimation by thresholding principal orthogonal complements," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75, 603–680.

Goto, S. and Y. Xu (2015): "Improving mean variance optimization through sparse hedging restrictions," *Journal of Financial and Quantitative Analysis*, 50, 1415–1441.

Hautsch, N. and S. Voigt (2019): "Large-scale portfolio allocation under transaction costs and model uncertainty," *Journal of Econometrics*, 212, 221–240.

Jobson, J. and B. Korkie (1980): "Estimation for Markowitz efficient portfolios," *Journal of the American Statistical Society*, 371, 544–554.

Jobson, J. D. and R. M. Korkie (1981): "Putting Markowitz theory to work," *The Journal of Portfolio Management*, 7, 70–74.

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

Kirby, C. and B. Ostdiek (2012): "It's all in the timing: simple active portfolio strategies that outperform naive diversification," *Journal of financial and quantitative analysis*, 47, 437–467.

Kourtis, A., G. Dotsis, and R. N. Markellos (2012): "Parameter uncertainty in portfolio selection: Shrinking the inverse covariance matrix," *Journal of Banking & Finance*, 36, 2522–2531.

Ledoit, O. and M. Wolf (2003): "Improved Estimation of the Covariane Matrix of Stock Returns with an Application to Portfolio Selection," *Journal of Empirical Finance*, 10, 603–621.

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

——— (2008): "Robust performance hypothesis testing with the Sharpe ratio," *Journal of Empirical Finance*, 15, 850–859.

——— (2011): "Robust performances hypothesis testing with the variance," *Wilmott*, 2011, 86–89.

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

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

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

——— (2022): "The power of (non-) linear shrinking: A review and guide to covariance matrix estimation," *Journal of Financial Econometrics*, 20, 187–218.

——— (2025): "Markowitz portfolios under transaction costs," *The Quarterly Review of Economics and Finance*, 101962.

Li, J. (2015): "Sparse and stable portfolio selection with parameter uncertainty," *Journal of Business & Economic Statistics*, 33, 381–392.

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

Merton, R. C. (1980): "On estimating the expected return on the market: An exploratory investigation," *Journal of Financial Economics*, 8, 323–361.

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

Moura, G. V., A. A. Santos, and E. Ruiz (2020): "Comparing high-dimensional conditional covariance matrices: Implications for portfolio selection," *Journal of Banking & Finance*, 118, 105882.

Stevens, G. V. (1998): "On the inverse of the covariance matrix in portfolio analysis," *The Journal of Finance*, 53, 1821–1827.

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

Wahba, G. (1990): *Spline models for observational data*, SIAM.

Yen, Y.-M. (2016): "Sparse weighted-norm minimum variance portfolios," *Review of Finance*, 20, 1259–1287.

Appendix

A Monte Carlo analysis

In this section, we examine the finite-sample properties of our proposed method in a controlled setting. Specifically, we generate asset returns from a linear latent multi-factor model of the form:

$$\boldsymbol{R}_t \;=\; \boldsymbol{\beta}(\boldsymbol{V}_t + \boldsymbol{\gamma}) \;+\; \boldsymbol{u}_t,$$

where $\boldsymbol{V}_t \sim N(\boldsymbol{0}, \boldsymbol{I})$ captures L latent factors, $\boldsymbol{\beta} \in \mathbb{R}^{N \times L}$ is a matrix of asset-factor exposures, and $\boldsymbol{\gamma} \in \mathbb{R}^{L}$ represents factor risk prices. The idiosyncratic noise $\boldsymbol{u}_t$ is independent of $\boldsymbol{V}_t$ and it follows a normal distribution with zero mean and a diagonal covariance matrix $\boldsymbol{H} \in \mathbb{R}^{N \times N}$.

To calibrate parameters $(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{H})$, we use a set of 349 asset returns comprising 49 industry portfolios, 100 size and book-to-market portfolios, 100 size and investment portfolios, and 100 size and operating profitability portfolios. We extract $L = 5$ principal components from these returns as the latent factors $\boldsymbol{V}_t$. With this data, we obtain parameter estimates and construct the expectation vector $\boldsymbol{\mu} = \boldsymbol{\beta}\boldsymbol{\gamma}$ and covariance matrix $\boldsymbol{\Sigma} = \boldsymbol{\beta}\boldsymbol{\beta}' + \boldsymbol{H}$, as implied by the model.

Next, we perform $M = 1000$ Monte Carlo simulations, each drawing $N \in \{50, 150, 300\}$ asset returns over $T = 250$ time periods from a multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. We then estimate optimal portfolio weights under various methods, including our proposed approach (denoted RRR), the sample minimum-variance allocation (sample), the nonlinear shrinkage estimator of Ledoit and Wolf (2012) (Non-Lin), the volatility-timing strategy of Kirby and Ostdiek (2012) (Vol-timing-1), and the nodewise-regression technique of Callot, Caner, Önder, and Ulaşan (2021) (Nodewise).

To assess out-of-sample performance, we compute the Sharpe ratio and volatility of the estimated portfolio weights $\hat{w}$ against the true (population) moments:

$$SR(\hat{w}) = \frac{\hat{w}'\,\mu}{\sqrt{\hat{w}'\,\Sigma\,\hat{w}}}, \qquad Vol(\hat{w}) = \sqrt{\hat{w}'\,\Sigma\,\hat{w}}.$$

We compare these statistics to the population-level performance:

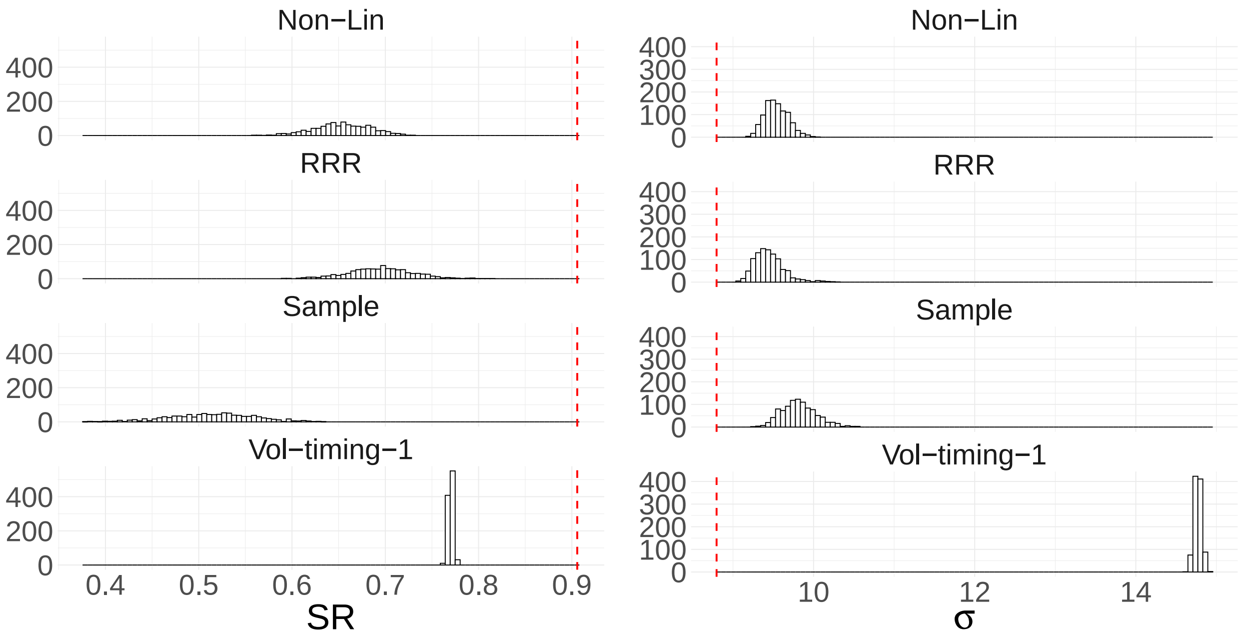
$$SR(w_{tan}) = \sqrt{\mu'\,\Sigma^{-1}\,\mu}, \qquad Vol(w_{gmv}) = \left(\sqrt{\iota'\,\Sigma^{-1}\,\iota}\right)^{-1},$$

where w_{tan} and w_{gmv} respectively denote the population-tangency and global minimum-variance portfolios, and ι is a vector of ones of appropriate dimension. Our analysis focuses on the discrepancy between estimated and population-optimal Sharpe ratios and volatilities, offering insight into how effectively each method performs under various dimensional regimes.

Table A1: Bias-Variance decomposition

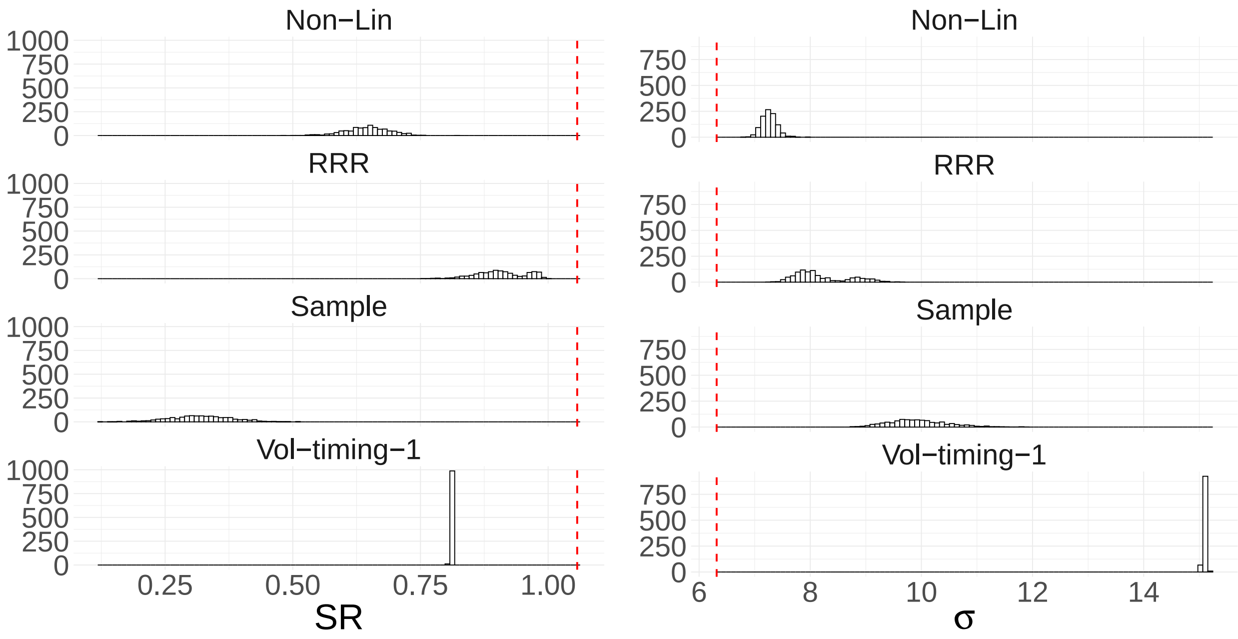
		Sample	RRR	Non-Lin	Vol-timing-1	Nodewise
$N = 50$	Sd	38.65	18.93	24.07	1.08	8.66
	Bias	12.22	7.95	9.40	21.40	18.06
	MSE	1643.28	421.75	667.67	459.09	401.04
	Leverage	9.94	4.11	5.40	1.00	1.12
$N = 150$	Sd	83.65	16.69	24.40	0.56	4.52
	Bias	26.45	9.54	10.02	19.68	18.44
	MSE	7697.37	369.68	696.10	387.70	360.60
	Leverage	4.00	2.84	3.24	1.00	1.20
$N = 300$	Sd	—	0.82	1.15	0.15	0.40
	Bias	—	10.78	10.58	18.15	17.45
	MSE	—	116.94	113.28	329.55	304.74
	Leverage	—	4.30	6.65	1.00	1.07

The table reports the bias-variance decomposition of the minimum-variance weights, along with average portfolio leverage.



(a) Sharpe ratios (b) Volatility

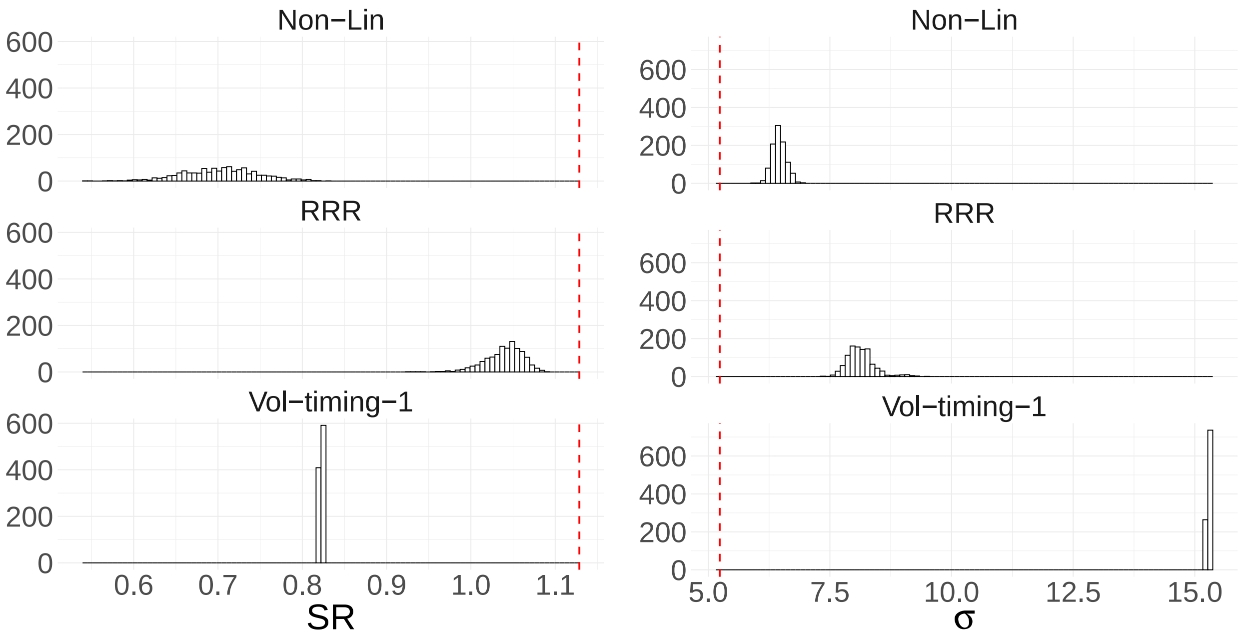
Figure A1: Sampling distribution of out-of-sample performance measures. The dashed red line denotes the optimal out-of-sample performance, *i.e.* the Sharpe ratio obtained using the tangency portfolio formed with population moments, and the variance of the minimum-variance portfolio formed with population moments.



(a) Sharpe ratios

(b) Volatility

Figure A2: Sampling distribution of out-of-sample performance measures. The dashed red line denotes the optimal out-of-sample performance, *i.e.* the Sharpe ratio obtained using the tangency portfolio formed with population moments, and the variance of the minimum-variance portfolio formed with population moments.



(a) Sharpe ratios

(b) Volatility

Figure A3: Sampling distribution of out-of-sample performance measures. The dashed red line denotes the optimal out-of-sample performance, *i.e.* the Sharpe ratio obtained using the tangency portfolio formed with population moments, and the variance of the minimum-variance portfolio formed with population moments.