

De Nard, Gianluca; Kostovic, Damjan

**Working Paper**

## AI shrinkage: A data-driven approach for risk-optimized portfolios

Working Paper, No. 470

**Provided in Cooperation with:**

Department of Economics, University of Zurich

*Suggested Citation:* De Nard, Gianluca; Kostovic, Damjan (2025) : AI shrinkage: A data-driven approach for risk-optimized portfolios, Working Paper, No. 470, University of Zurich, Department of Economics, Zurich,  
<https://doi.org/10.5167/uzh-277803>

This Version is available at:

<https://hdl.handle.net/10419/322286>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



**University of  
Zurich**<sup>UZH</sup>

University of Zurich  
Department of Economics

Working Paper Series  
ISSN 1664-7041 (print)  
ISSN 1664-705X (online)

---

Working Paper No. 470

# **AI Shrinkage: A Data-Driven Approach for Risk-Optimized Portfolios**

Gianluca De Nard and Damjan Kostovic

May 2025

---

# AI Shrinkage: A Data-Driven Approach for Risk-Optimized Portfolios\*

Gianluca De Nard<sup>1, 2, 3</sup> and Damjan Kostovic<sup>3</sup>

<sup>1</sup>Department of Economics, University of Zurich

<sup>2</sup>Stern Volatility and Risk Institute, New York University

<sup>3</sup>OLZ AG, Zurich

May 2025

## Abstract

The paper introduces a new type of shrinkage estimation that is not based on asymptotic optimality but uses artificial intelligence (AI) techniques to shrink the sample eigenvalues. The proposed AI Shrinkage estimator applies to both linear and nonlinear shrinkage, demonstrating improved performance compared to the classic shrinkage estimators. Our results demonstrate that reinforcement learning solutions identify a downward bias in classic shrinkage intensity estimates derived under the i.i.d. assumption and automatically correct for it in response to prevailing market conditions. Additionally, our data-driven approach enables more efficient implementation of risk-optimized portfolios and is well-suited for real-world investment applications including various optimization constraints.

**Keywords:** Covariance matrix estimation; linear and nonlinear shrinkage; portfolio management reinforcement learning; risk optimization.

**JEL Classification:** C13, C58, G11.

---

\*Contact: [gianluca.denard@econ.uzh.ch](mailto:gianluca.denard@econ.uzh.ch)

# 1 Introduction

Financial applications such as portfolio optimization, asset pricing, and risk management rely heavily on the covariance matrix of asset returns. [Markowitz \(1952\)](#) recognizes it as a central risk measure in his seminal work, which has since sparked extensive research on estimating both expected returns and the covariance matrix. While much focus has been placed on return or variance estimation, our emphasis lies on estimating the covariance matrix.

The most common estimator, the sample covariance matrix, often performs poorly out-of-sample, especially in high-dimensional settings. When the number of assets  $N$  approaches or exceeds the number of time observations  $T$ , the sample covariance matrix becomes ill-conditioned or even non-invertible. As the concentration ratio  $c := N/T$  increases, estimation errors grow and portfolio optimization becomes unreliable, leading to poor performance.

[Jobson and Korkie \(1980\)](#) and [Michaud \(1989\)](#) highlight that the sample covariance matrix suffers from large estimation errors, often amplifying extreme and unreliable coefficients. [Ledoit and Wolf \(2017\)](#) reframe this issue as a degrees-of-freedom problem, where estimating  $\mathcal{O}(N^2)$  parameters from limited data is infeasible. Their work and that of others introduced linear and nonlinear shrinkage methods to reduce dimensionality and improve estimation; for an overview see [Ledoit and Wolf \(2022a\)](#).

Linear shrinkage combines the sample covariance matrix with a structured target, reducing extreme values and improving estimation accuracy. Empirical results show it enhances information ratios, diversification, and portfolio stability. In constrained settings, shrinkage can be interpreted as an implied covariance matrix ([Roncalli, 2011](#)), and is widely used by practitioners ([Jagannathan and Ma, 2003](#)).

Nonlinear shrinkage, proposed by [Ledoit and Wolf](#), shrinks individual eigenvalues without a fixed target and is particularly effective in high-dimensional contexts. Both linear and nonlinear shrinkage estimators remain invertible for  $c > 1$ , providing feasible portfolio solutions. However, their derivations assume i.i.d. returns. Although extensions exist (e.g., [Sancetta, 2008](#)), they can introduce high bias in small samples ([Bartz and Müller, 2014](#)).

The main contribution of this paper is to introduce a data-driven approach to compute the shrinkage intensity(ies) without any assumption on the data generating process. We propose a new type of linear and nonlinear shrinkage estimation that is not based on asymptotic optimality, but uses artificial intelligence (AI) techniques to shrink the sample

eigenvalues. The new AI Shrinkage estimators deliver improved estimation accuracy, especially in high-dimensional settings where the sample covariance matrix is ill-conditioned or even non-invertible. The data-driven AI approach is based on reinforcement learning (RL). Thus, the shrinkage intensity(ies) is (are) learned from an agent (artificial portfolio manager) that interacts with its environment (trades based on market observations).

RL is particularly well-suited for applying empirical data analysis methodologies, especially for estimating covariance matrices. It has emerged as a leading field in artificial intelligence, revolutionizing the way we approach complex decision-making problems. At the core of RL is the concept of an agent that learns optimal behaviors by interacting with its environment. As the agent navigates in an often unknown environment, it collects rewards and strives to maximize these rewards over time. In recent years, RL has garnered significant attention for its success in solving non-trivial problems in a variety of tasks. Notable examples include mastering Atari games from raw pixels (Mnih et al., 2015), defeating world champions in the game of Go (Silver et al., 2016), and surpassing professional poker players (Brown et al., 2017; Moravčík et al., 2017). Furthermore, RL techniques are increasingly being applied to real-world challenges, such as autonomous driving (Pan et al., 2017) as well as to portfolio management; e.g., see Cong et al. (2021).

The idea of improving shrinkage estimation of the covariance matrix with RL has been introduced by Matera and Matera (2023) and Lu et al. (2024). Lu et al. (2024) show how RL based on alternative data and textual analysis can improve the optimal linear shrinkage policy for correlation matrices. Their focus is on so-called text-based networks (TBN) introduced by Hoberg and Phillips (2016) to determine the correlation matrix shrinkage target and to use deep reinforcement learning to compute the shrinkage intensity that maximizes the expected out-of-sample utility of an investor in line with the theoretical work by Bodnar et al. (2022). Despite some promising results in their empirical analysis, their approach has some (practical) limitations. First, the TBN based shrinkage target depends on the Hoberg-Phillips Data Library and is therefore only applicable to stocks included in the library and not implementable for real-life trading. Second, it is a shrinkage approach for the correlation matrix, ignoring the large estimation errors coming from the variances. Third, it is limited to linear shrinkage. Fourth, the portfolios are rebalanced, respectively the covariance/correlation matrix is re-estimated, on a yearly basis, which is too infrequent regarding the dynamics

of the market and of the optimal shrinkage intensity.<sup>1</sup> Finally, the data-driven shrinkage intensity computation can be applied to all shrinkage targets and is not just restricted to TBN.

A related study of [Matera and Matera \(2023\)](#) uses Fama-French industry portfolio return data to derive the shrinkage intensity for covariance matrix estimation using RL. For one of their models they report improved out-of-sample Sharpe ratio numbers. However, the RL architecture and empirical analysis of [Matera and Matera \(2023\)](#) is insufficient to evaluate the goodness of a covariance matrix. The review and guide to covariance matrix estimation of [Ledoit and Wolf \(2022a\)](#), as well as the references therein, show that for financial applications, the out-of-sample standard deviation of the global minimum variance (GMV) portfolio should be computed to assess the covariance matrix estimation quality. Estimating the GMV portfolio is a ‘clean’ problem in terms of evaluating the quality of a covariance matrix estimator since it abstracts from having to estimate the vector of expected returns at the same time. Therefore, the approach of [Matera and Matera \(2023\)](#) to implement an RL agent with maximum Sharpe ratio utility function is misleading for covariance matrix estimation. Even though their research design is not suitable to assess the power of RL to improve the shrinkage intensity computation for covariance matrix estimation compared to [Ledoit and Wolf \(2004b\)](#), it is an interesting alternative to [Cong et al. \(2021\)](#) who use RL to directly derive portfolio weights with maximal Sharpe ratio without estimating the expected returns and the covariance matrix. Additionally, the study does not consider daily stock returns but monthly Fama-French industry portfolio returns instead. This results in a practically unimplementable portfolio and low-frequency analysis with a very short out-of-sample period of 3 years, respectively 36 observations, which is far from the literature standard and it is dangerous to claim (statistically significant) outperformance. Finally, the methodology of [Matera and Matera \(2023\)](#) is only applicable to linear shrinkage and has been analyzed only for the most basic (scalar multiple of the) identity matrix shrinkage target, neglecting the off-diagonal entries of the covariance matrix; see [De Nard \(2022\)](#).

As mentioned above, the main contribution of this paper to the literature is to introduce a data-driven approach to learn the optimal shrinkage intensity(ies) with RL. The introduced AI Shrinkage estimator can be applied to linear and nonlinear shrinkage, with improved

---

<sup>1</sup>The shrinkage intensity can spike during periods of financial turmoil where more shrinkage is needed; e.g., see Figure A1. of [De Nard \(2022\)](#).

performance compared to the classic shrinkage estimators discussed in [Ledoit and Wolf \(2022a\)](#). The improvement comes from dropping the i.i.d. assumption when computing the shrinkage intensity(ies) and following a data-driven approach without any assumption on the data generating process. Note that the proposed AI Shrinkage estimators are based only on stock return data and do not need, but can easily be extended with, factor or alternative data signals. We show that an RL agent learns that the i.i.d. derived shrinkage intensity(ies) is (are) biased (too low), and automatically corrects for this bias (increases the shrinkage intensity(ies) depending on the market environment). The learned and adjusted shrinkage intensity(ies) remains stable, intuitive, and correlated with the classic i.i.d. approach. To assess the quality of our covariance matrix estimators we run an extensive empirical study in a high-dimensional setting focusing on the out-of-sample standard deviation of GMV portfolios, where the RL agent has a minimum-variance utility function to learn the shrinkage intensity(ies). Finally, the AI Shrinkage estimators deliver a more efficient implementation of risk-optimized portfolios and should be used for real-life investments. However, our goal is to improve the estimation of covariance matrices in general and our methodology is therefore not restricted to portfolio selection, or even financial, applications.

A further contribution of this paper is the application of RL to constrained portfolio optimization, recognizing its practical relevance and the flexibility of RL to adapt to specific investor preferences. Unlike classic shrinkage methods, which are independent of the optimization step, RL dynamically adjusts shrinkage based on the specific constraints, risk-return objectives, and thus utility functions of the investor. Most classic shrinkage methods tend to shrink too aggressively in constrained optimization problems, as they do not account for the implicit shrinkage impact of constraints, potentially leading to suboptimal portfolio allocations. This adaptability allows RL to incorporate the interaction between shrinkage and portfolio performance in a forward-looking manner, making it particularly well-suited for real-world investment settings where constraints play a crucial role.

The rest of the paper is organized as follows. [Section 2](#) provides a selective overview of the shrinkage estimation and reinforcement learning literature and presents our methodology on how to combine them. [Section 3](#) contains the backtest results and empirical analysis. [Section 4](#) concludes. Appendices contain additional results and auxiliary materials.

## 2 Methodology

### 2.1 Shrinkage Estimation

As the population covariance matrix  $\Sigma$  is not directly observable, it needs to be estimated using historical data. The most straightforward method to estimate  $\Sigma$  is simply to use the sample covariance matrix  $S$ :

$$S := \frac{1}{T-1} \tilde{R}' \tilde{R} , \quad (2.1)$$

where  $\tilde{R}$  denotes the  $T \times N$  demeaned asset return matrix,  $N$  being the number of assets, and  $T$  the number of time periods (observations).

The sample covariance matrix is an intuitive estimator, easy to compute, and unbiased. However, we show that it is unsuitable for portfolio optimization, aligning with and extending the conclusions of [Ledoit and Wolf \(2022a\)](#).

One potential solution to mitigate the impact of estimation error is to impose structure on the covariance matrix, thereby reducing the large proportion of random noise in the sample moments. This approach leads to a more parsimonious parametrization of the covariance matrix, which facilitates the efficient extraction of systematic information from historical correlations. In turn, this reduces estimation error and enhances forecasting accuracy. A significant body of research has explored this approach (e.g., [Sharpe, 1963](#); [Elton and Gruber, 1963](#); [Jobson and Korkie, 1980](#)). In this paper, we focus specifically on the statistical shrinkage technique introduced by [Stein \(1956\)](#). Subsequent work by [Ledoit and Wolf \(2004a, 2017\)](#), [Engle et al. \(2019\)](#), and [De Nard et al. \(2021, 2022, 2024, 2025\)](#) has proposed various linear and nonlinear shrinkage estimators for the covariance matrix to this end. To readers not already familiar with these shrinkage estimators, we refer to the overview paper of [Ledoit and Wolf \(2022a\)](#).

#### 2.1.1 Linear Shrinkage

As discussed above, linear shrinkage combines a highly structured estimator  $F$  and the sample covariance matrix  $S$ . [Ledoit and Wolf \(2003\)](#) suggest a convex linear combination of the form:

$$\tilde{\Sigma}_L := \delta F + (1 - \delta) S , \quad (2.2)$$



where  $\delta$  is the shrinkage intensity, or shrinkage constant, strictly between 0 and 1, and  $F$  denotes the shrinkage target. The sample covariance matrix is therefore shrunk towards the structured estimator with an intensity of  $\delta$ . [Ledoit and Wolf \(2003\)](#) define this shrinkage constant as the weight that is given to the structured measure and that there should be only one optimal shrinkage constant that minimizes the expected distance between the shrinkage estimator  $\tilde{\Sigma}_L$  and the true covariance matrix  $\Sigma$ . Moreover, they propose a loss function  $\mathcal{L}_F$  based on the Frobenius norm, a quadratic measure of distance between the true and estimated covariance matrix, that does not depend on the inverse of the covariance matrix ([Frost and Savarino, 1986](#)) and therefore also works for  $N > T$ . The optimal shrinkage intensity is therefore the constant that minimizes the expected quadratic loss function of the form:

$$\mathbb{E}[\mathcal{L}_F(\delta)] := \mathbb{E} \left[ \|\tilde{\Sigma}_L - \Sigma\|_F^2 \right] = \mathbb{E} \left[ \|\delta F + (1 - \delta)S - \Sigma\|_F^2 \right] , \quad (2.3)$$

where  $\|\cdot\|_F^2$  is the (squared) Frobenius norm:  $\|A\|_F^2 := \langle A, A \rangle := \text{tr}(AA') = \sum_{i=1}^N \sum_{j=1}^N a_{ij}^2$  for a squared matrix  $A$  with entries  $(a_{ij})_{i,j=1,\dots,N}$ .

Under the i.i.d assumption, the literature derives the optimal shrinkage intensity (and target), and introduces the optimal feasible and consistent linear shrinkage estimator as

$$\hat{\Sigma}_L := \hat{\delta}F + (1 - \hat{\delta})S , \quad (2.4)$$

$$\hat{\delta} := \min \left\{ \max \left\{ \frac{\hat{\pi} - \hat{\rho}}{\hat{\kappa} \cdot T}, 0 \right\}, 1 \right\} , \quad (2.5)$$

where  $\hat{\pi} := \sum_{i=1}^N \sum_{j=1}^N \frac{1}{T} \sum_{t=1}^T ((r_{it} - \bar{r}_i)(r_{jt} - \bar{r}_j) - s_{ij})^2$  measures the estimation uncertainty in the sample covariance matrix,  $\hat{\rho}$  measures the ('combined') covariance between the shrinkage target and the sample covariance matrix,<sup>2</sup> and  $\hat{\kappa} := \sum_{i=1}^N \sum_{j=1}^N (f_{ij} - s_{ij})^2$  measures how close the (population version of the) shrinkage target is to the population covariance matrix; see [Ledoit and Wolf \(2003, 2004b, 2022a\)](#) and [De Nard \(2022\)](#).

A key advantage of  $\hat{\Sigma}_L$  is that it remains positive definite and thus invertible, even when  $N > T$ , unlike the sample covariance matrix, which becomes rank-deficient and not invertible in such cases. This is due to  $\hat{\Sigma}_L$  being a convex linear combination of two matrices:  $F$ , which should be positive definite, and  $S$ , which is positive semi-definite.

---

<sup>2</sup> $\hat{\rho}$  depends on the choice of shrinkage target and thus requires a case-by-case analysis. For the exact definition of  $\hat{\rho}$ , we therefore reference to the corresponding original paper. Note that the authors show that the impact of  $\hat{\rho}$  on  $\hat{\delta}$  is usually small and can be neglected.

Therefore, the solution (2.4-2.5) of problem (2.3) depends on the choice of the shrinkage target. The question arises: what constitutes an effective shrinkage target? Ideally, it should approximate the true covariance matrix as closely as possible while minimizing the number of parameters. This presents a delicate 'balancing act' between accuracy and parsimony. A good shrinkage target leverages application-specific knowledge, allowing for customization that enhances its relevance and effectiveness in the given context. For an overview of attractive (custom-tailored) shrinkage target see Ledoit and Wolf (2022a). In this paper, we focus on the original proposition of Ledoit and Wolf (2004a) and its generalization of De Nard (2022).

The simplest shrinkage target is to assume that the covariance matrix is a scalar multiple of the identity matrix  $F_I = \hat{\eta}I$ , where  $I$  is the  $N \times N$  identity matrix and  $\hat{\eta}$  is the estimated scalar multiple. Note that  $\eta$  depends on the unobservable parameter  $\Sigma$ , which can be consistently estimated by the sample covariance matrix:  $\hat{\eta} = \frac{\text{tr}(SI)}{N} = \frac{1}{N} \sum_{i=1}^N s_i^2 = \bar{s}^2$ . Consequently,

$$F_I := \bar{s}^2 I . \quad (2.6)$$

Therefore, the linear shrinkage to the identity matrix approach, shrinks all the sample variances towards their grand mean with the estimated (constant) intensity  $\hat{\delta}^I$ . Alternatively, one can reformulate the problem in the eigenvalue space, where all the sample eigenvalues  $\lambda_1, \dots, \lambda_N$  are shrunk towards their grand mean  $\bar{\lambda} := \frac{1}{N} \sum_{i=1}^N \lambda_i$  with the same intensity  $\hat{\delta}^I$ :

$$\lambda_i^I := \hat{\delta}^I \bar{\lambda} + (1 - \hat{\delta}^I) \lambda_i , \quad (2.7)$$

while the sample eigenvectors remain unchanged.<sup>3</sup>

A scalar multiple of the identity matrix is a highly structured estimator that generally exhibits lower estimation error compared to the sample covariance matrix. However, it is limited in that it only considers the diagonal elements of the covariance matrix, effectively disregarding the off-diagonal covariance coefficients by setting them to zero. Given that asset returns typically exhibit positive linear dependence, it is crucial to choose a shrinkage target that captures this relationship. In this regard, De Nard (2022) introduces a generalization of the identity matrix shrinkage estimator, which incorporates knowledge of the non-zero

---

<sup>3</sup>Therefore,  $\hat{\Sigma}_I = \hat{\delta}^I \bar{s}^2 I + (1 - \hat{\delta}^I) S$  is equivalent to  $\hat{\Sigma}_I = U \Lambda_I U'$ , where  $U := [u_1 \dots, u_N]$  denotes an orthogonal matrix whose columns are the sample eigenvectors  $u_i$ , and  $\Lambda_I := \text{Diag}(\lambda_1^I, \dots, \lambda_N^I)$  denotes the diagonal matrix of the  $N$  linearly shrunk eigenvalues.

covariances in asset returns. This is achieved by introducing an additional (off-diagonal) parameter into the shrinkage target, allowing for a more accurate reflection of the covariance matrix structure:

$$F_{\text{CVC}} := \bar{s}^2 I + \bar{s}_{ij} J , \quad (2.8)$$

where  $J := \mathbb{1}\mathbb{1}' - I$  is the off-diagonal matrix,  $\mathbb{1}$  denotes a conformable vector of ones, and  $\bar{s}_{ij} := (N(N-1))^{-1} \sum_{i=1}^N \sum_{j=1, j \neq i}^N s_{ij}$ . Thus, [De Nard \(2022\)](#) proposes not only shrinking the diagonal elements of the sample covariance matrix toward a constant mean variance level, but also shrinking the off-diagonal elements toward a constant mean covariance level. Note that the literature presents even more customized shrinkage targets; for an overview see [Ledoit and Wolf \(2022a\)](#).

### 2.1.2 Nonlinear Shrinkage

The optimal linear combination in Equation (2.4) can be intuitively interpreted as moving each entry of the sample covariance matrix toward the shrinkage target with a common intensity. A natural extension of this approach is to allow for different intensities for different entries of the covariance matrix, or alternatively, shrinking the sample eigenvalues to their grand mean with *individual intensities* while keeping the sample eigenvectors; see Equation (2.7). This latter approach, called nonlinear shrinkage, was introduced and continuously improved by [Ledoit and Wolf \(2012, 2017, 2020, 2022b\)](#).

Nonlinear shrinkage allows sample eigenvalues to be individually moved up or down by an individual amount. Note that this means that the individual shrinkage intensities can be negative, that is, move a sample eigenvalue away from the shrinkage target! Clearly, this more general approach will outperform the use of a common shrinkage intensity, provided that the distinct intensities are selected appropriately. Furthermore, this method will yield a positive-definite estimator, as long as all the transformed eigenvalues remain positive.

The optimization problem for nonlinear shrinkage is as follows:

$$\min_{\tilde{\Lambda}} \|\tilde{\Sigma}_{\text{NL}} - \Sigma\|_F^2 , \quad (2.9)$$

where  $\tilde{\Sigma}_{\text{NL}} := U\tilde{\Lambda}U'$  is the nonlinear shrinkage estimator,  $U := [u_1 \dots, u_N]$  denotes an orthogonal matrix whose columns are the sample eigenvectors  $u_i$ , and  $\tilde{\Lambda} := \text{Diag}(\tilde{\lambda}_1, \dots, \tilde{\lambda}_N)$  denotes the diagonal matrix of the  $N$  shrunk eigenvalues.

For the solution of problem (2.9),  $\widehat{\Sigma}_{\text{NL}} := U\Lambda_{\text{NL}}U'$ , and thus on the actual estimation of the optimal nonlinear shrinkage eigenvalues,  $\Lambda_{\text{NL}} := \text{Diag}(\lambda_1^{\text{NL}}, \dots, \lambda_N^{\text{NL}})$ , we refer to the overview paper of Ledoit and Wolf (2022a) and the references therein.

## 2.2 Reinforcement Learning

In a simple reinforcement learning framework, an agent interacts sequentially with an environment (in our case, a financial market) described by a Markov decision process (MDP). Formally, an MDP is characterized by a tuple  $\mathcal{M} := \langle \mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \zeta \rangle$ , where  $\mathcal{S}$  denotes the set of states (in our case, observable stock prices and other factors),  $\mathcal{A}$  the set of actions (in our case, the shrinkage intensity),  $\mathcal{P}$  the state transition probability matrix (in our case, independent of actions),  $\mathcal{R}$  the reward function (in our case, the portfolio standard deviation), and  $\zeta$  the discount factor (in our case, equal to one for a one-period model).

In RL, actions yield rewards that the agent aims to maximize; in our context, as we are interested in the optimal shrinkage intensity, we minimize risk measured explicitly by the (out-of-sample) portfolio standard deviation of the global minimum variance portfolio; see Section 3.4. The action in our reinforcement learning problem corresponds to choosing a shrinkage intensity; see Section 2.3.

In financial applications, such as portfolio optimization, the MDP structure simplifies considerably. Specifically, the environment's state transitions are independent of the actions taken, i.e., the actions chosen do not influence the underlying (market) state evolution. Formally, this can be described as an MDP with state-transition probabilities independent of actions:

$$P(s_{t+1}|s_t, a_t) = P(s_{t+1}|s_t), \quad \forall \quad a_t \in \mathcal{A} . \quad (2.10)$$

Such a setting naturally reduces the sequential decision-making problem to a series of independent *contextual bandit* problems. In other words, each state can be optimized individually without regard to future states, and the problem naturally reduces to finding the optimal action for each state separately. More formally, for data comprising state  $s_t$ , action  $a_t$ , and the observed portfolio (sample) standard deviation  $\hat{\sigma}(s_t, a_t)$ , our objective is to directly predict the action that minimizes the observed standard deviation for each state:

$$y^*(s_t) = \arg \min_{a_t \in \mathcal{A}} \hat{\sigma}(s_t, a_t) . \quad (2.11)$$

Similar approaches have been applied to portfolio selection problems where sequential decision-making simplifies to context-dependent optimization, demonstrating the suitability of contextual-bandit frameworks for certain classes of financial problems; see, e.g., [Shen et al. \(2015\)](#).

We take advantage of this simplified structure by employing a regression-based policy learning approach, explicitly framing our portfolio optimization as a supervised learning problem. Thus, our final chosen action for a given state  $s$  is obtained by rounding the regression output to the nearest integer action within our discrete action set  $\mathcal{A}$ , where actions are discrete integer actions ranging from 0 to 100. It is important to note that, in this approach, the regression directly delivers an optimal action (action resulting in the minimal out-of-sample portfolio deviation) rather than estimating rewards for each possible action individually. Consequently, we do not explicitly estimate the expected portfolio standard deviation for each action; instead, our loss function directly penalizes deviations from the optimal action observed historically. The optimal policy is given by

$$\pi^*(s) = \arg \min_{\pi \in \Pi} \mathbb{E}_s [(\pi(s) - y^*(s))^2] . \quad (2.12)$$

In our case, the optimal policy is approximated by solving the following regression problem:

$$\hat{\pi}(s) = \arg \min_{f \in \mathcal{F}} \sum_{t=1}^T (f(s_t) - y^*(s_t))^2 , \quad (2.13)$$

where  $f(s_t)$  denotes the predicted continuous action for state  $s_t$ ,  $y^*(s_t)$  is the historically observed optimal discrete action minimizing portfolio standard deviation, and  $\mathcal{F}$  is the regression model function class, which we use to approximate policies from the broader policy space  $\Pi$ . Our regression-based policy  $\hat{\pi}(s)$  provides an explicit approximation to the theoretical optimal policy  $\pi^*(s)$  by minimizing deviations from historically optimal actions  $y^*(s_t)$ .

Although our regression model produces continuous predictions, our action space is discrete (integer-valued actions). Therefore, we transform the continuous model outputs into discrete actions by simply selecting the nearest integer action to the model's predicted value. Although this rounding step introduces minor approximation errors, empirical analyses indicate that the effect is negligible in our context, given the granularity of actions.

### 2.2.1 RL for Linear Shrinkage

There are (at least) two ways how reinforcement learning can help to improve linear shrinkage estimation of the covariance matrix. The first approach, and arguably the more natural way to think about the problem, is to directly learn the optimal linear shrinkage intensity via RL, that is:

$$\widehat{\Sigma}_{\delta^{\text{RL-L}}} := \delta^{\text{RL-L}} F + (1 - \delta^{\text{RL-L}}) S . \quad (2.14)$$

Therefore, if  $\delta^{\text{RL-L}}$  is larger (smaller) than  $\hat{\delta}$ , our data-driven reinforcement learning approach shrinks the sample covariances more (less) towards the shrinkage target compared to the optimal shrinkage intensities from the literature based on the i.i.d assumption. Note that as long as  $\delta^{\text{RL-L}} \in (0, 1]$  and iff  $F$  is positive definite, also  $\widehat{\Sigma}_{\delta^{\text{RL-L}}}$  is positive definite and thus invertible.

One drawback of this approach is that it is not general enough to be applicable also to nonlinear shrinkage. Therefore, let us reformulate the problem not as a convex combination of the sample covariance matrix and a shrinkage target, but as a linear combination of the classic linear or nonlinear shrunk eigenvalues from the literature and the sample eigenvalues in the next section.

### 2.2.2 RL for Linear and Nonlinear Shrinkage

The more general approach of reinforcement learning that is applicable to both linear and nonlinear shrinkage is to use the spectral decomposition:

$$\widehat{\Sigma}_{\gamma^{\text{RL-*}}} := U \Lambda_{\text{RL-*}} U' , \quad (2.15)$$

where  $\Lambda_{\text{RL-*}} := \text{Diag}(\lambda_1^{\text{RL-*}} \dots, \lambda_N^{\text{RL-*}})$ , and  $*$   $\in \{\text{L}, \text{NL}\}$  denotes if RL is applied to linear (I or CVC) or nonlinear shrunk eigenvalues. To estimate the optimal eigenvalues  $\Lambda_{\text{RL-*}}$  of (2.15), we propose the following convex combination of the sample eigenvalues and the  $*$  shrunk eigenvalues

$$\tilde{\lambda}_i^{\text{RL-*}} := \hat{\gamma} \lambda_i^* + (1 - \hat{\gamma}) \lambda_i , \quad (2.16)$$

where  $\hat{\gamma} \in [0, \infty)$  is learned by the reinforcement learning agent and  $\tilde{\lambda}^{\text{RL-*}} := [\tilde{\lambda}_1^{\text{RL-*}}, \dots, \tilde{\lambda}_N^{\text{RL-*}}]$ .

From Equation (2.16) it follows immediately that:

- the base shrinkage adjustment-factor  $\hat{\gamma}$  is constant;<sup>4</sup>
- we reduce the estimation problem to  $\mathcal{O}(1)$ , while keeping the flexibility of nonlinear transformation, namely, learning the optimal convex combination of the sample eigenvalues and the *shrunk* eigenvalues proposed by the literature;
- for a ‘well-working’ RL estimator, (i)  $\hat{\gamma} = 1$  indicates that the shrinkage estimator from the literature based on the i.i.d assumption is well-suited for the investigated data, (ii)  $\hat{\gamma} > 1$  indicates that the eigenvalues need to be shrunk even more, (iii)  $\hat{\gamma} < 1$  indicates that the eigenvalues need to be shrunk less;<sup>5</sup>
- $\tilde{\lambda}^{\text{RL}*}$  is not guaranteed to be monotonic decreasing in the cases of  $\hat{\gamma} > 1$ ;
- and for too large  $\hat{\gamma}$ ,  $\tilde{\lambda}^{\text{RL}*}$  can have negative entries.

To guarantee that our RL covariance matrix estimator is well defined, that is, to maintain the (monotonic decreasing) order of the (shrunk) eigenvalues as well as their sign (only strictly positive eigenvalues such that the covariance matrix estimator is positive definite and invertible), we apply a custom variant of isotonic regression. We apply a transformation  $\mu$  in the order of  $i = 1, \dots, N$  (from largest to smallest eigenvalue):

$$\lambda_i^{\text{RL}*} = \mu(\tilde{\lambda}_i^{\text{RL}*}, \lambda_i, \tilde{\lambda}^*) := \begin{cases} \lambda_1^{\text{RL}*} := \max[\tilde{\lambda}_1^{\text{RL}*}, \tilde{\lambda}^*] \\ \max\left[\min\left[\tilde{\lambda}_i^{\text{RL}*}, \lambda_{i-1}^{\text{RL}*}\right], \tilde{\lambda}^*\right] & \text{for } \frac{\lambda_i^*}{\lambda_i} < 1 \\ \min\left[\max\left[\tilde{\lambda}_i^{\text{RL}*}, \lambda_{i-1}^{\text{RL}*}\right], \tilde{\lambda}^*\right] & \text{for } \frac{\lambda_i^*}{\lambda_i} > 1 \end{cases} \quad (2.17)$$

where  $\tilde{\lambda}^*$  denotes the crossing point to the 45-degree line (no shrinkage) in Figure 1, where we move from increasing the sample eigenvalues towards actually shrinking the sample eigenvalues. Thus  $\tilde{\lambda}^* := (\lambda_j^* + \lambda_k^*)/2$  where  $j$  is the smallest (shrunk) eigenvalue that fulfills  $\frac{\lambda_j^*}{\lambda_j} < 1$  and  $k$  is the largest (shrunk) eigenvalue that fulfills  $\frac{\lambda_k^*}{\lambda_k} > 1$ .

To visualize the impact of the proposed methodology on actual data, the first two figures plot the eigenvalues of a portfolio with 100 stocks based on daily stock return data, using

---

<sup>4</sup>Our RL approach can be applied also to individual shrinkage adjustments  $\hat{\gamma}_1, \dots, \hat{\gamma}_N$ , but the number of degrees of freedom will increase to  $N$  and the probability to end up with an ill-conditioned estimator increases; e.g., see the last two comments on monotonicity and invertability.

<sup>5</sup>By a ‘well-working’ RL estimator we mean an improved covariance matrix estimator  $\hat{\Sigma}_{\gamma^{\text{RL}*}}$  compared to  $\hat{\Sigma}_*$ , measured by a lower out-of-sample standard deviation of the global minimum variance portfolio as we explain later in Section 3.4.

a lookback window of one trading year. Further details on the data and application are provided in Section 3.

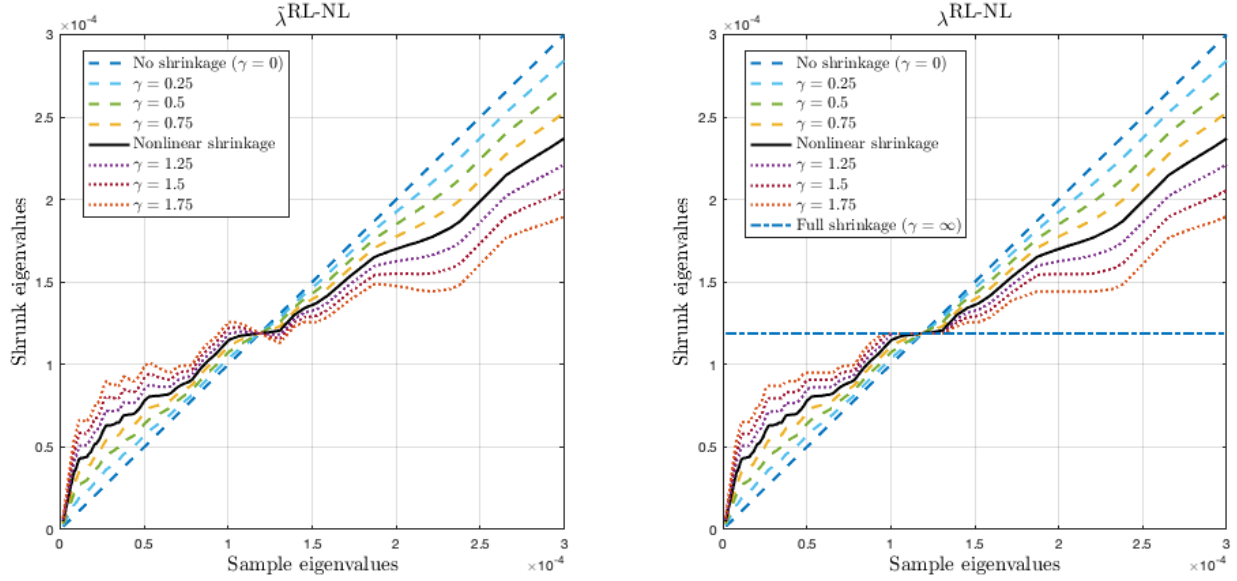


Figure 1: Illustration of the nonlinear shrinkage eigenvalue transformation shown in Equations (2.16) and (2.17) for various  $\gamma$  values. The panel on the left shows the raw transformation, whereas the panel on the right shows the monotonic decreasing shrunk eigenvalues after applying the function  $\mu$ . The illustration is based on a representative sample as of 01/08/2014 for the  $N = 100$  portfolio with  $T = 252$ .

In Figure 1 we visualize the impact of  $\gamma$  on the nonlinear shrunk eigenvalues (in black). A  $\gamma \in [0, 1)$  rotates the original nonlinear shrinkage line counter-clockwise at the rotation point  $(\bar{\lambda}, \bar{\lambda})$ , resulting in less shrinkage. A  $\gamma > 1$  rotates the original nonlinear shrinkage line clockwise at the rotation point  $(\bar{\lambda}, \bar{\lambda})$ , resulting in more shrinkage. The left panel of Figure 1 shows the raw transformations where the eigenvalues are not strictly monotonically decreasing, and the right panel shows the (2.17) adjusted transformation. A  $\gamma = 0$  corresponds to the 45-degree line and therefore no shrinkage is applied. A  $\gamma = \infty$  corresponds to the horizontal line and therefore maximal shrinkage to the grand sample eigenvalue mean, resulting in the identity matrix multiplied by the mean sample variance as covariance matrix estimator.

In Figure (2) we plot various (RL) transformed nonlinear shrinkage eigenvalues divided by their sample eigenvalues. First, more sample eigenvalues are pulled up than pushed down. Second, the smallest sample eigenvalues experience the largest impact, being disproportionately pulled up, whereas the largest sample eigenvalues are only slightly shrunk. Third, sample eigenvalues near their grand mean tend to be shrunk the least. Fourth, the



largest and smallest sample eigenvalues are shrunk less than their ‘neighbors’ closer to the mean.

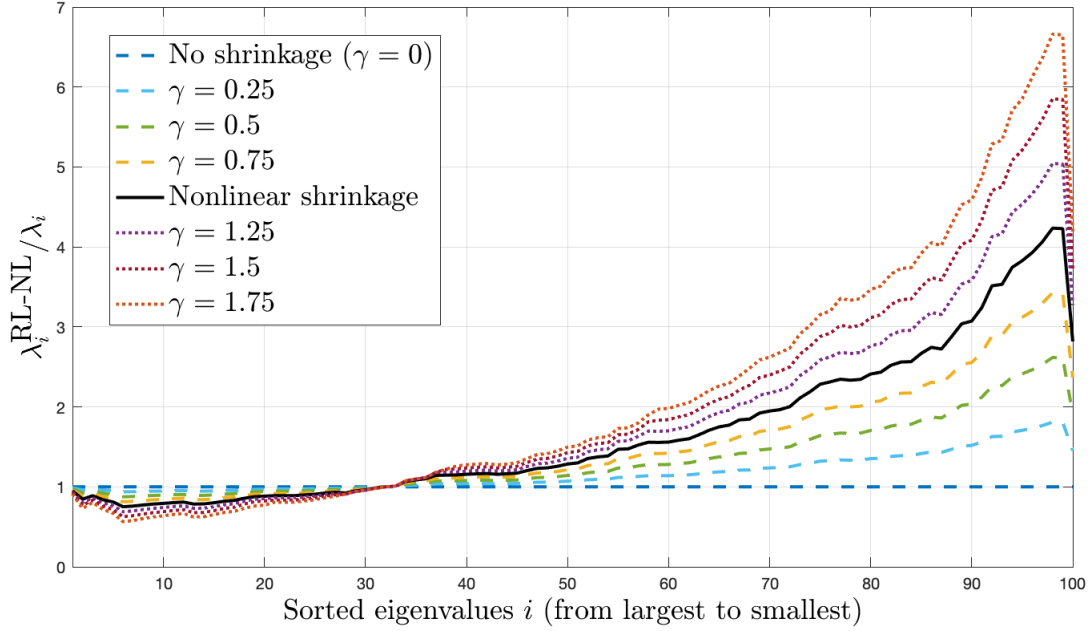


Figure 2: Illustration of the sorted eigenvalue transformations (2.17) for various  $\gamma$  values. The illustration is based on a representative sample as of 01/08/2014 for the  $N = 100$  portfolio with  $T = 252$ .

In Appendix A, we prove that for linear shrinkage towards a scalar multiple of the identity matrix, problem (2.14) is equivalent to problem (2.15–2.17). The intuition is that shrinking the sample covariance matrix towards the identity matrix (scaled by the average sample variance) is equivalent to shrinking the sample eigenvalues towards their grand mean. Consequently, from a mathematical perspective, learning the optimal convex combination  $\hat{\delta}^{\text{RL-I}}$ —which corresponds to the slope of the straight line passing through the point  $(\bar{\lambda}, \bar{\lambda})$ , with the sample eigenvalues on the x-axis and the shrunk eigenvalues on the y-axis, as shown in Figure 1—is equivalent to learning  $\hat{\gamma} \cdot \hat{\delta}^{\text{I}}$ , where  $\hat{\delta}^{\text{I}}$  is the classic slope from linear shrinkage and  $\hat{\gamma}$  is the slope adjustment determined by the reinforcement learning agent.

Note that for other (linear) shrinkage targets, such as CVC, the sample eigenvalues are not shrunk towards their grand mean with the same intensity. As a result, problem (2.14) is not equivalent to problem (2.15–2.17). In unreported results, we find that both approaches outperform classic shrinkage based on the i.i.d. assumption. Moreover, we believe that learning the shrinkage intensity directly is a more natural approach.

Finally, note that the monotonicity transformation in (2.17) ensures that all shrunk eigenvalues remain strictly positive. Consequently, the resulting covariance matrix estimator is positive definite and invertible.

## 2.3 RL Architecture

The action in our reinforcement learning problem corresponds to choosing a shrinkage intensity. For linear shrinkage, we discretize the action space in  $a \in \mathcal{A}_L := \{0, 1, \dots, 100\}$ , where  $\hat{\delta}^{\text{RL-L}} := a/100$ , and thus  $\hat{\delta}^{\text{RL-L}} \in \{0, 0.01, 0.02, \dots, 1\}$ . For nonlinear shrinkage, we discretize the action space in  $a \in \mathcal{A}_{\text{NL}} := \{0, 1, \dots, 30\}$ , where  $\hat{\gamma}^{\text{RL-NL}} := 0.5 + a/20$ , and thus  $\hat{\gamma}^{\text{RL-NL}} \in \{0.5, 0.55, 0.60, \dots, 2\}$ .

In practice, the state space in a reinforcement learning framework can be flexibly specified. However, particularly in financial applications, it is crucial to carefully select state features that balance predictive power and noise. We construct our state space from a set of meaningful universe-specific signals measured at each trading date, summarized in Table 1.

Feature	Description
Average Correlation	Average pairwise sample correlation between stocks.
Average Volatility	Average one-year sample volatility of all individual stocks.
EWMA	Exponentially weighted moving average of previous month returns (decay = 0.1) of the equally-weighted portfolio.
Lagged Optimal Action	Optimal action from previous rebalancing.
Linear Shrinkage	Linear shrinkage intensity of <a href="#">Ledoit and Wolf (2004b)</a> .
Momentum	Fraction of days each stock had positive returns over the previous month, averaged across all stocks.
Rolling Optimal Action	One-year rolling average of optimal action.
Rolling Optimal SD	One-year rolling standard deviation of optimal actions.
Trace	Trace of sample covariance matrix.
Universe Volatility	One-year sample volatility of the equally-weighted universe.

Table 1: Summary of market state features, computed over one-year estimation windows based on the eligible stock universe.

To approximate the optimal policy, we choose the elastic net (ENet) as a regularized regression method over more complex deep learning methods due to the relatively small number of actions and the moderate dimensionality of the state space. ENet provides

robustness to noise, interpretability, and inherent regularization, which effectively mitigates overfitting; a common problem of financial backtests. Additionally, its simplicity helps to avoid the computational complexity and instability often associated with deep learning methods when applied to limited and noisy financial datasets.

Our model is trained on approximately 40 years of historical financial data, with roughly half of this dataset (approximately 20 years) used for initial training. For each trading day, we evaluate all available actions (101 in the L-RL model and 31 in the NL-RL model) by calculating the 21-days a-head portfolio standard deviation. The optimal action, defined as the one that minimizes this realized risk measure, becomes our target variable.

We conduct in-sample hyperparameter tuning over the grid summarized in Table 2, noting that predictive performance remains robust and largely unchanged across this grid. The optimized ENet hyperparameters are defined as follows:

- $\alpha$ : Controls overall regularization strength. Higher  $\alpha$  implies stronger regularization.
- **l1\_ratio**: Controls the balance between  $L_1$  and  $L_2$  penalties. Specifically,  $l1\_ratio = 0$  results in pure  $L_2$  regularization,  $l1\_ratio = 1$  results in pure  $L_1$  regularization, and intermediate values provide a balance.
- **max\_iter**: Maximum iterations allowed for optimization convergence.
- **tol**: Convergence tolerance; optimization stops once parameter updates become smaller than this value.

Hyperparameter	Values
$\alpha$	0.5, 1.0, 1.5, 2.0, 5.0
l1_ratio	0.1, 0.25, 0.5, 0.75, 0.9
max_iter	500, 1000, 1500, 2000
tol	$10^{-3}$ , $10^{-4}$ , $10^{-5}$

Table 2: Hyperparameter grid for ENet regression.

Additionally, we employ a rolling-window retraining scheme, updating our model every 21 trading days, corresponding to each rebalancing period, by incorporating the most recent 21 observations and discarding the oldest 21 observations. This rolling-window approach ensures that our model adapts continually to the latest market conditions, providing predictions closely aligned with current market regimes. This is another benefit of employing a simpler, less computationally intensive model.

**Remark 2.1** (Adjustments for Long-Only Portfolios). As we point out later, long-only portfolios need less shrinkage. This is why in the RL-NL long-only case we extend the action space by  $a \in \mathcal{A}_{\text{NL-long-only}} := \{0, 1, \dots, 40\}$ , where  $\hat{\gamma}^{\text{RL-NL-long-only}} := a/20$ , and thus  $\hat{\gamma}^{\text{RL-NL-long-only}} \in \{0, 0.05, 0.10, \dots, 2\}$ .

In our long-only portfolio context, we additionally employ a cost-sensitive learning approach, biasing the model toward predicting lower actions and thereby less shrinkage. In our case, this is an inverse logarithmic weighting  $\frac{1}{\log(1+a)}$  of the actions in the training set, weighting lower actions, and therefore smaller shrinkage intensities, more. This is based on our in-sample findings indicating that lower actions generally yield lower standard deviations. This empirical evidence aligns with prior research, which indicates that long-only portfolios generally require substantially less regularization compared to their unconstrained (long-short) counterparts; see, e.g., [Jagannathan and Ma \(2003\)](#) and [Dom et al. \(2025\)](#). ■

## 3 Empirical Analysis

### 3.1 Data and General Portfolio-Construction Rules

We download daily stock return data from the Center for Research in Security Prices (CRSP) starting on January 1, 1980 through January 31, 2022. We restrict attention to stocks from the NYSE, AMEX, and NASDAQ stock exchanges.

Roughly half of the data is allocated for training our AI Shrinkage estimators, with the remaining half used for evaluation. Hence, our out-of-sample period spans from 12/18/2000 to 01/31/2022. This results in a total of 253 months or 5,313 trading days for evaluation. All portfolios are updated monthly.<sup>6</sup> For simplicity, and in line with much of the literature, we adopt the common convention that 21 consecutive trading days constitute one (trading) ‘month’. This implies that the covariance matrix must be estimated every 21 trading days, i.e., every rebalancing date  $h = 1, \dots, 253$ . At any investment date  $h$ , a covariance matrix is estimated based on the most recent 252 daily returns, which roughly corresponds to using 1 year of past data.

We consider the following portfolio sizes:  $N \in \{30, 50, 100, 225, 500\}$ . For a given

---

<sup>6</sup>Monthly updating is common practice to avoid an unreasonable amount of turnover and thus transaction costs. During a month, from one day to the next, we hold number of shares fixed rather than portfolio weights; in this way, there are no transactions during a month.

combination  $(h, N)$ , the investment universe is obtained as follows. We find the set of stocks that have an almost complete return history over the most recent  $T = 252$  days as well as a complete return ‘future’ over the next 21 days.<sup>7</sup>

From the remaining set of stocks, we then pick the largest  $N$  stocks (as measured by their market capitalization on investment date  $h$ ) as our investment universe. In this way, the investment universe changes relatively slowly from one investment date to the next.

There is a great advantage in having a well-defined rule that does not involve drawing stocks at random, as such a scheme would have to be replicated many times and averaged over to give stable results. As far as rules go, the one we have chosen seems the most reasonable because it avoids so-called ‘penny stocks’ whose behavior is often erratic; also, high-market-cap stocks tend to have the lowest bid-ask spreads and the highest depth in the order book, which allows large investment funds to invest in them without breaching standard safety guidelines.

## 3.2 Competing Covariance Matrix Estimators

We now detail the various covariance matrix estimators included in our empirical analysis.

- **L:** the linear shrinkage estimator of [Ledoit and Wolf \(2004a\)](#); that is, we use the scalar multiple of the identity matrix (I) shrinkage target.
- **NL:** the nonlinear shrinkage estimator of [Ledoit and Wolf \(2022b\)](#), that is, we use the quadratic-inverse shrinkage (QIS) estimator.
- **RL-L:** the linear shrinkage estimator of [Ledoit and Wolf \(2004a\)](#), but with data-driven shrinkage intensity computed via reinforcement learning as described in [Section 2.2.1](#).
- **RL-NL:** the nonlinear shrinkage estimator of [Ledoit and Wolf \(2022b\)](#), but with data-driven shrinkage intensities computed via reinforcement learning as described in [Section 2.2.2](#).

It is of interest to include the sample covariance matrix and other linear and nonlinear shrinkage estimators from the literature to compare their performance. However, as summarized in [Ledoit and Wolf \(2022a\)](#), nonlinear shrinkage estimators tend to perform

---

<sup>7</sup>The first restriction allows for up to 2.5% of missing returns over the most recent 252 days, and replaces missing values by zero. The latter, ‘forward-looking’ restriction is not feasible in practice but is commonly used in the literature. Although it might affect (in a minor way) absolute performance due to survivorship bias, it does not systematically affect relative performance of various methods.

similarly, whereas the sample covariance matrix is consistently outperformed by both linear and nonlinear shrinkage estimators. We confirm these findings and, to save space, do not report (all) results for the sample covariance matrix, instead focusing on the QIS nonlinear shrinkage estimator. For linear shrinkage, we present results only for the identity-based shrinkage target (I), as it is the only case where both reinforcement learning approaches yield identical results: (i) learning the optimal convex combination  $\delta$  of the sample covariance matrix and the shrinkage target, as detailed in Equation (2.14), or (ii) learning the optimal convex combination  $\gamma$  of the sample eigenvalues and the shrunk eigenvalues, as described in Equation (2.16); see Appendix A. Nonetheless, in unreported results, we find that for more sophisticated linear shrinkage targets—such as the constant-variance-covariance shrinkage target of De Nard (2022) in Equation (2.8)—the results discussed in Section 3.5 hold as well and generally show improved performance for both approaches.

### 3.3 Performance Measures

- **AV:** We compute the **average** of the 5,313 out-of-sample returns as geometric mean, and then multiply by  $\sqrt{252}$  to annualize.
- **SD:** We compute the **standard deviation** of the 5,313 out-of-sample returns, and then multiply by  $\sqrt{252}$  to annualize.
- **SR:** We compute the annualized **Sharpe ratio** as the ratio AV/SD.
- **TO:** We compute average (monthly) **turnover** as  $\frac{1}{252} \sum_{h=1}^{252} \|\hat{w}_{h+1} - \hat{w}_h^{\text{hold}}\|_1$ , where  $\|\cdot\|_1$  denotes the  $L_1$  norm and  $\hat{w}_h^{\text{hold}}$  denotes the vector of the ‘hold’ portfolio weights at the end of month  $h$ .<sup>8</sup>
- **GL:** We compute average (monthly) excess **gross leverage** as  $\frac{1}{253} \sum_{k=1}^{253} \|\hat{w}_k\|_1 - 1$ .
- **PL:** We compute average (monthly) **proportion of leverage** as  $\frac{1}{253 \times N} \sum_{h=1}^{253} \sum_{i=1}^N \mathbb{1}_{\{\hat{w}_{i,h} < 0\}}$ .

---

<sup>8</sup>The vector  $\hat{w}_h^{\text{hold}}$  is determined by the initial vector of portfolio weights,  $\hat{w}_h$ , together with the evolution of the various prices of the  $N$  stocks in the portfolio during month  $h$ .

### 3.4 Global Minimum Variance Portfolio

We consider the problem of estimating the global minimum variance (GMV) portfolio in the absence of short-sales constraints. The problem is formulated as

$$\arg \min_w w' \Sigma w \quad (3.1)$$

$$\text{subject to } w' \mathbb{1} = 1, \quad (3.2)$$

where  $\mathbb{1}$  denotes a vector of ones of dimension  $N \times 1$ . It has the analytical solution

$$w = \frac{\Sigma^{-1} \mathbb{1}}{\mathbb{1}' \Sigma^{-1} \mathbb{1}}. \quad (3.3)$$

The natural strategy in practice is to replace the unknown covariance matrix  $\Sigma$  by an estimator  $\hat{\Sigma}$  in formula (3.3), yielding a feasible portfolio

$$\hat{w} := \frac{\hat{\Sigma}^{-1} \mathbb{1}}{\mathbb{1}' \hat{\Sigma}^{-1} \mathbb{1}}. \quad (3.4)$$

Estimating the GMV portfolio is a ‘clean’ problem in terms of evaluating the quality of a covariance matrix estimator, as it abstracts from having to estimate the vector of expected returns at the same time. In addition, researchers have established that estimated GMV portfolios have desirable out-of-sample properties not only in terms of risk but also in terms of reward-to-risk, that is, in terms of Sharpe and information ratios; for example, see [Haugen and Baker \(1991\)](#), [Jagannathan and Ma \(2003\)](#), and [Nielsen and Aylursubramanian \(2008\)](#).

Following the advice of [Dom et al. \(2025\)](#) and focusing on more practically relevant portfolios, we also include an analysis of long-only minimum variance portfolios for investors who cannot take short positions. Notably, although the covariance matrix estimator in classic shrinkage is independent of the specific investment problem, for RL shrinkage, it plays a crucial role. The RL agent learns the optimal shrinkage intensity based on the investment environment and portfolio constraints, which means that the intensity for long-short and long-only portfolios can differ. As we will demonstrate, they indeed do.

In the long-only case, the weights of the global minimum variance portfolio need to be derived numerically. This is a constrained quadratic optimization problem, which we solve with the python package *CVXPY*.

In addition to Markowitz portfolios based on formula (3.4), we also include as a simple-minded benchmark the equal-weighted portfolio promoted by DeMiguel et al. (2009), among others, as it has been claimed to be difficult to outperform. We denote the equal-weighted portfolio by **EW**. We also include the value-weighted portfolio, denoted by **VW**, which is sometimes favored over the equal-weighted portfolio, for example because it incurs less turnover and ‘controls’ for small-cap biases.

Our stance is that in the context of the GMV portfolio, the most important performance measure is the out-of-sample standard deviation, SD. The true (but unfeasible) GMV portfolio is given by (3.3). It is designed to minimize the variance (and thus the standard deviation) rather than to maximize the expected return or the information ratio. Therefore, any portfolio that implements the GMV portfolio should be primarily evaluated by how successfully it achieves this goal. A high out-of-sample average return, AV, and a high out-of-sample Sharpe ratio, SR, are naturally also desirable, but should be considered of secondary importance from the point of view of evaluating the quality of a covariance matrix estimator.

We also consider the question of whether one estimation model delivers a lower out-of-sample standard deviation than another estimation model. To avoid a multiple testing problem and as a major goal of this paper is to show that using a data-driven reinforcement learning approach to obtain the shrinkage intensities improves the estimation of (large-dimensional) covariances matrices, we restrict attention to two comparisons: (i) L with RL-L, and (ii) NL with RL-NL. For a given universe size, a two-sided  $p$ -value for the null hypothesis of equal standard deviations is obtained by the prewhitened  $HAC_{PW}$  method described in Ledoit and Wolf (2011, Section 3.1).<sup>9</sup>

## 3.5 Empirical Results

### 3.5.1 Main Results

The main results are presented in Table 3 and can be summarized as follows. Unless stated otherwise, the findings are with respect to SD as the performance measure.

- All models outperform EW and VW by a wide margin. This also holds for Sharpe ratio

---

<sup>9</sup>As the out-of-sample size is very large at 5,313, there is no need to use the computationally more involved bootstrap method described in Ledoit and Wolf (2011, Section 3.2), which is preferred for small sample sizes.



comparisons.

- Each RL-L and RL-NL model consistently—and often markedly—outperforms its respective classic L and NL base model. Moreover, the outperformance of RL-L over L, and RL-NL over NL, is both statistically significant and economically meaningful.
- There is a consistent ranking across smaller portfolio sizes  $N \in \{30, 50, 100\}$  (from best to worst): RL-L, RL-NL, NL, L. For larger portfolio sizes  $N \in \{225, 500\}$ , the ranking shifts to (from best to worst): RL-NL, NL, RL-L, L, VW, EW.
- In terms of SR, the results for linear shrinkage are mixed. However, for nonlinear shrinkage, RL-NL consistently outperforms NL. With the exception of  $N = 500$ , RL-NL achieves the highest SR.

Global minimum variance portfolio						
	L	RL-L	NL	RL-NL	EW	VW
SD						
$N = 30$	13.50	<b>13.41</b>	13.45	13.42	18.72	19.13
$N = 50$	13.64	13.28*	13.32	<b>13.25*</b>	18.86	19.04
$N = 100$	13.26	<b>12.57*</b>	12.68	12.62*	19.51	19.20
$N = 225$	12.64	11.45*	11.42	<b>11.35*</b>	19.87	19.32
$N = 500$	11.21	10.45*	10.38	<b>10.36</b>	20.62	19.52
SR						
$N = 30$	0.84	0.79	0.82	<b>0.83</b>	0.52	0.52
$N = 50$	0.80	0.79	0.80	<b>0.80</b>	0.52	0.53
$N = 100$	0.56	0.68	0.65	<b>0.68</b>	0.51	0.52
$N = 225$	0.78	0.83	0.86	<b>0.86</b>	0.55	0.54
$N = 500$	<b>0.95</b>	0.88	0.91	0.92	0.58	0.55

Table 3: Annualized out-of-sample performance measures for the competing estimators of the GMV portfolio, in percentage. All measures are based on 5,313 daily out-of-sample returns from 12/18/2000 until 01/31/2022. For any performance measure (SD and SR), the best estimator is indicated in **bold blue** font. In the columns RL-L, respectively RL-NL, significance at the 0.01 level of outperformance in SD over L, respectively NL, is denoted by an asterisk \*.

To sum up, the proposed reinforcement learning approach for computing shrinkage intensity(ies) outperforms classic linear and nonlinear shrinkage estimators based on the

i.i.d. assumption.

We also report a comparison over time of the classic linear shrinkage intensity estimates and those learned by RL in Figure 3 for  $N = 500$ . For other portfolio sizes, see Figure B.1. For L, the shrinkage intensity is very low and stable, with an average shrinkage intensity of 0.07. This is due to the fact that the scalar multiple of the identity matrix as a shrinkage target has very high bias but very low estimation error.<sup>10</sup> In contrast, the purely data-driven RL-L approach yields much higher and more adaptive shrinkage intensities, with an average shrinkage intensity of 0.43. Comparing these shrinkage intensities with the (one-year moving average of the) Oracle—defined as the best ex-post monthly shrinkage intensity that minimizes monthly out-of-sample SD—we observe that RL-L successfully learns a higher level of shrinkage, closely aligning with the smoothed Oracle (average shrinkage intensity of 0.42). Overall, we find that classic linear shrinkage based on the i.i.d. assumption results in a substantially lower shrinkage intensity, which can be consistently improved by RL-L.

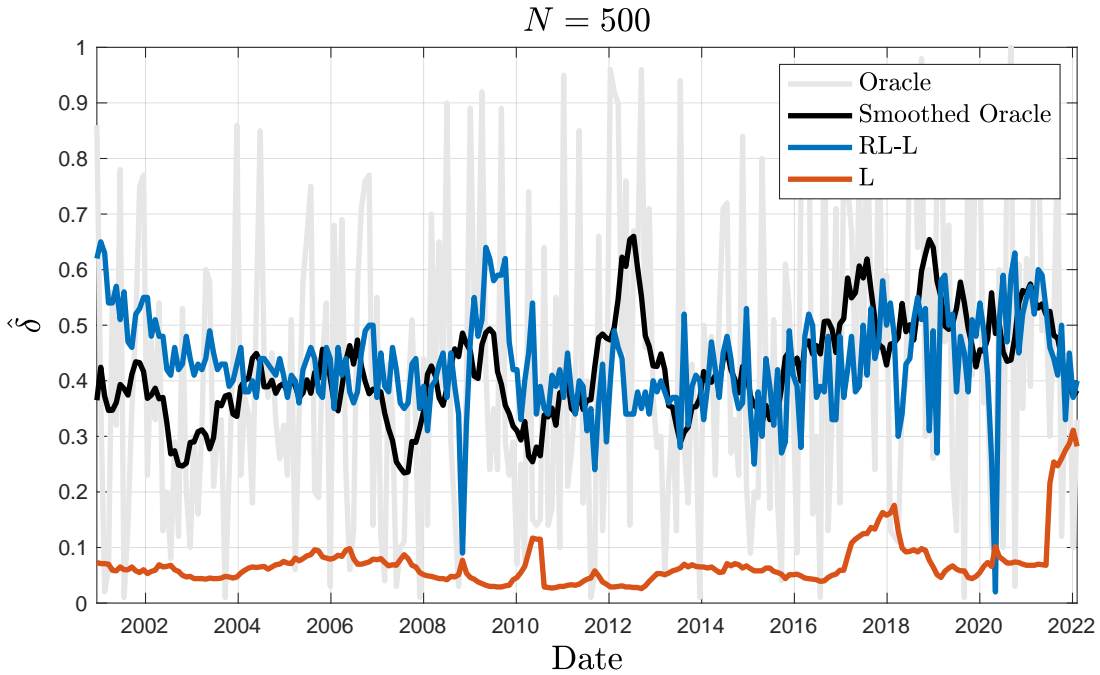


Figure 3: Comparison of the monthly Oracle, RL-L and L estimated linear shrinkage intensities over time for  $N = 500$ . For the smoothed Oracle we use a one-year moving average smoothing.

<sup>10</sup>See De Nard (2022). Note that for other linear shrinkage estimators, bias can be reduced by incorporating a more realistic shrinkage target, which increases the optimal shrinkage intensity. However, more realistic shrinkage targets require estimating additional (and more complex) parameters, leading to higher estimation errors. As a result, the estimated optimal shrinkage intensities become more volatile.

A similar but less extreme pattern can also be observed for nonlinear shrinkage. In Figure 4, we plot the adjustment of nonlinear shrinkage intensities over time for  $N = 500$ ; for other portfolio sizes, see Figure B.2. Since  $\hat{\gamma} = 1$  corresponds to classic nonlinear shrinkage, we find that RL-NL consistently exhibits substantially higher shrinkage intensities. This suggests that sample eigenvalues require even greater adjustment.

This observation is further supported by the Oracle nonlinear shrinkage intensities, which are generally well above 1. In particular, their smoothed version (with an average adjustment of 1.26) closely aligns with the RL-NL values (average adjustment of 1.28).

Notably, RL shrinkage intensities spike during periods of financial turmoil—such as the dot-com bubble, the global financial crisis, and the COVID-19 pandemic—indicating that the need for shrinkage is highest when market volatility is elevated.

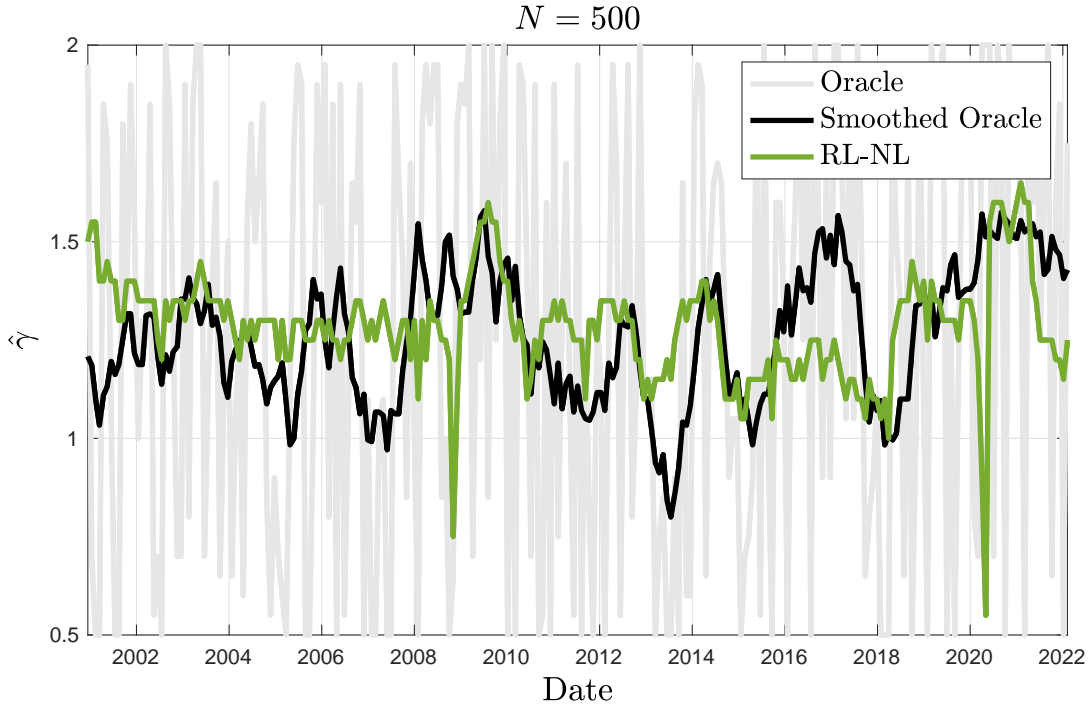


Figure 4: Comparison of the Oracle and RL-NL estimated nonlinear shrinkage intensities adjustment over time for  $N = 500$ . For the smoothed Oracle we use a one-year moving average smoothing.

To visualize the monthly out-of-sample shrinkage intensity estimation errors, we present the mean, mean absolute, and mean squared deviations of the oracle intensities minus the estimated intensities in Table 4 and Figure 5. The estimation error results can be summarized as follows:

- For any portfolio size and deviation metric, RL shrinkage intensities exhibit consistently and significantly lower estimation errors compared to classic shrinkage intensities.
- The positive mean error values for L and NL indicate a substantial negative bias in shrinkage intensities due to the i.i.d. assumption.
- In contrast, the mean errors of the data-driven RL shrinkage intensities are close to zero.

Monthly shrinkage intensity estimation errors of GMV portfolios												
	Mean error				Mean absolute error				Mean squared error			
	L	RL-L	NL	RL-NL	L	RL-L	NL	RL-NL	L	RL-L	NL	RL-NL
$N = 30$	0.21	-0.01	0.35	0.04	0.25	0.24	0.65	0.58	0.13	0.08	0.52	0.40
$N = 50$	0.25	0.00	0.37	0.00	0.28	0.24	0.63	0.56	0.15	0.08	0.51	0.37
$N = 100$	0.29	0.02	0.41	0.06	0.30	0.23	0.67	0.57	0.16	0.08	0.54	0.39
$N = 225$	0.36	0.02	0.43	0.07	0.37	0.23	0.60	0.49	0.20	0.08	0.47	0.30
$N = 500$	0.35	-0.01	0.26	-0.02	0.36	0.23	0.53	0.49	0.20	0.08	0.37	0.30

Table 4: Monthly shrinkage-intensity estimation errors of the GMV portfolio. All numbers are based on 5'313 daily out-of-sample returns, respectively 253 monthly predicted shrinkage intensities vs. the oracle intensities, from 12/18/2000 until 01/31/2022.

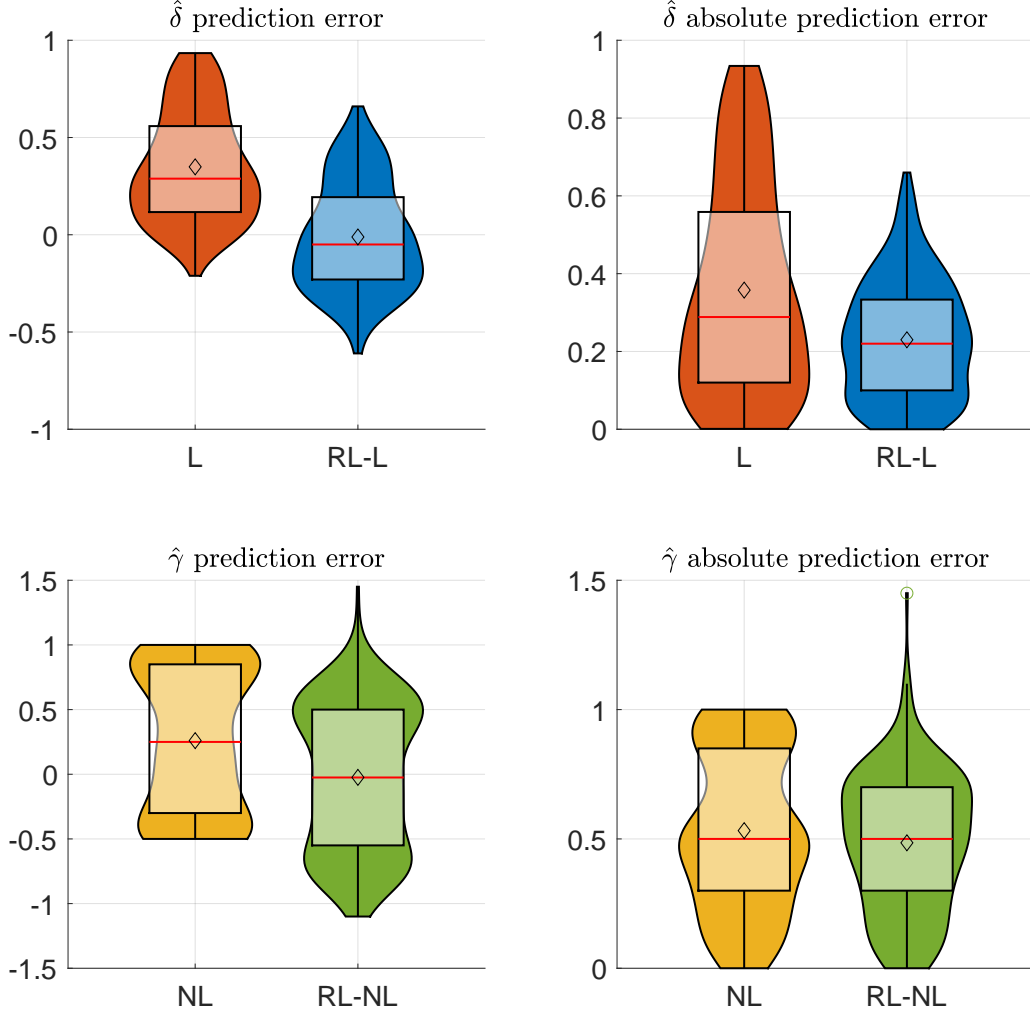


Figure 5: Violin plots of monthly Oracle vs. estimated optimal shrinkage intensity differences (panels on the left) and absolute differences (panels on the right) for the  $N = 500$  GMV portfolio. The upper panels show the differences for linear shrinkage and the lower panels for nonlinear shrinkage. A violin plot visually summarizes the distribution of a data set by combining the corresponding box plot with a (rotated) kernel density estimator; note that within the box of the box plot the sample median is indicated by a horizontal line whereas the sample average is indicated by a diamond.

In Figure 6, we plot the correlation of linear shrinkage intensities across all models and find that they are positively correlated. Notably, we observe two distinct clusters of models with high correlations: (i) The classic L models across  $N$ , and (ii) the new RL-L models across  $N$ . The correlations between L and RL-L are substantially lower, indicating that the data-driven approach not only increases the level of shrinkage but also alters its dynamics. This is further evident in Figures 3 and B.1.

In unreported results, we find that the added value of RL extends beyond merely reducing the negative bias of the estimated shrinkage intensities (i.e., increasing the shrinkage level). It also enhances the shrinkage dynamics. To test this, we adjust the linear shrinkage intensities by adding the bias term (first L column of Table 4).

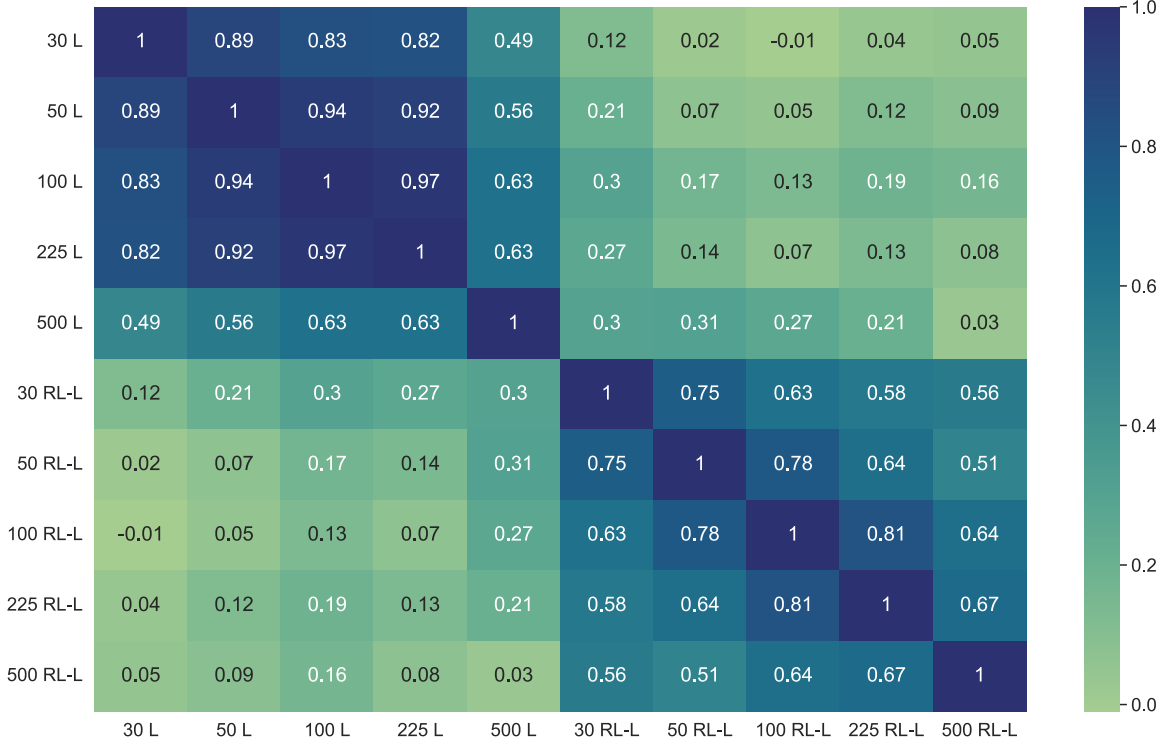


Figure 6: Correlation matrix of L and RL monthly shrinkage intensities across GMV portfolio sizes. All numbers are based on 253 monthly shrinkage intensities from 12/18/2000 until 01/31/2022.

To illustrate the impact of the shrinkage level, we also report the standard deviation of the GMV portfolio for constant shrinkage intensities in Table 5:

- All classic and RL shrinkage models consistently outperform the sample covariance matrix  $S$  (i.e.,  $\delta = \gamma = 0$ ). The SD of  $S$  increases with  $N$  due to greater estimation error in higher dimensions, whereas shrinkage models reduce SD by improving estimation accuracy and enhancing diversification in higher dimensions.
- Total shrinkage results in the identity matrix as the covariance matrix estimator, which is equivalent to the EW portfolio ( $\delta = 1$  or  $\gamma = \infty$ ).
- Consequently, both excessively small and excessively large shrinkage levels are

suboptimal. We find that the optimal constant linear shrinkage intensity increases with  $N$  due to the curse of dimensionality.

- For linear shrinkage toward the identity matrix, the optimal constant shrinkage intensity consistently outperforms L and is comparable to RL-L. This suggests that L can be improved by increasing its shrinkage intensity. However, since the optimal constant shrinkage intensity is determined ex post, it is not directly applicable, whereas RL-L is.
- For nonlinear shrinkage, we observe that the shrunk eigenvalues require even stronger shrinkage, as optimal  $\gamma$  values are greater than one. However, the optimal constant  $\gamma$  adjustment decreases with  $N$ .
- The SDs of the ex post evaluated optimal constant  $\gamma$  adjustments are comparable to those of RL-NL.

Linear shrinkage											
$\delta$	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$N = 30$	13.64	<b>13.35</b>	13.36	13.49	13.70	13.99	14.37	14.88	15.59	16.68	18.72
$N = 50$	13.96	13.35	<b>13.24</b>	13.29	13.42	13.64	13.94	14.37	15.04	16.19	18.86
$N = 100$	14.27	12.80	<b>12.60</b>	12.62	12.73	12.91	13.19	13.59	14.23	15.49	19.51
$N = 225$	26.50	11.75	11.41	<b>11.39</b>	11.48	11.67	11.93	12.32	12.91	14.08	19.87
$N = 500$	NA	10.73	10.46	<b>10.43</b>	10.50	10.65	10.89	11.24	11.79	12.87	20.62
Nonlinear shrinkage											
$\gamma$	0.5	0.6	0.7	0.8	0.9	1.0	1.2	1.4	1.6	1.8	2.0
$N = 30$	13.53	13.50	13.49	13.47	13.46	13.45	13.42	13.42	13.38	<b>13.38</b>	13.39
$N = 50$	13.52	13.46	13.41	13.38	13.37	13.32	13.27	13.24	13.20	<b>13.19</b>	13.18
$N = 100$	13.00	12.90	12.83	12.76	12.72	12.68	12.64	<b>12.61</b>	12.62	12.63	12.65
$N = 225$	11.88	11.72	11.61	11.52	11.46	11.42	11.36	<b>11.34</b>	11.34	11.36	11.42
$N = 500$	10.62	10.53	10.47	10.43	10.40	10.38	<b>10.37</b>	10.39	10.46	10.54	10.62

Table 5: Annualized out-of-sample standard deviations of the GMV portfolio for the linear (I) and nonlinear (QIS) shrinkage estimators with fixed  $\delta$  respectively  $\gamma$ . All measures are based on 5,313 daily out-of-sample returns from 12/18/2000 until 01/31/2022. For both estimators and for any portfolio size the best fixed  $\delta$ , respectively  $\gamma$ , is indicated in **bold blue** font.

Additionally, we report results on average turnover and leverage. The results, presented in Table 6, can be summarized as follows. Note that our optimization does not impose constraints on leverage or turnover.

- For linear shrinkage, RL consistently and significantly reduces turnover. Overall, RL-L exhibits the lowest turnover across all models.
- For linear shrinkage, RL also results in lower leverage for larger portfolio sizes.
- Turnover and leverage values are comparable between NL and RL-NL.

Global minimum variance portfolio						
	L	RL-L	NL	RL-NL	EW	VW
	TO					
$N = 30$	0.61	<b>0.09</b>	0.80	0.80	0.06	0.07
$N = 50$	0.94	<b>0.16</b>	0.86	0.86	0.06	0.07
$N = 100$	1.60	<b>0.29</b>	1.02	1.11	0.06	0.07
$N = 225$	3.03	<b>0.54</b>	1.90	1.79	0.05	0.06
$N = 500$	3.10	<b>1.07</b>	3.55	3.59	0.05	0.06
	GL					
$N = 30$	2.05	3.82	2.20	<b>2.15</b>	1.00	1.00
$N = 50$	2.61	3.86	2.36	<b>2.30</b>	1.00	1.00
$N = 100$	3.62	3.62	2.67	<b>2.60</b>	1.00	1.00
$N = 225$	5.59	<b>3.25</b>	4.02	3.97	1.00	1.00
$N = 500$	5.67	<b>2.96</b>	4.99	5.09	1.00	1.00
	PL					
$N = 30$	0.35	0.41	0.33	<b>0.33</b>	0.00	0.00
$N = 50$	0.38	0.41	0.35	<b>0.34</b>	0.00	0.00
$N = 100$	0.43	0.41	0.38	<b>0.37</b>	0.00	0.00
$N = 225$	0.45	<b>0.39</b>	0.42	0.42	0.00	0.00
$N = 500$	0.44	<b>0.38</b>	0.43	0.43	0.00	0.00

Table 6: Monthly average portfolio metrics for the the competing estimators of the GMV portfolio. All measures are based on 253 monthly out-of-sample weight vectors from 12/18/2000 until 01/31/2022. The best estimator for any measure (TO, GL and PL) is indicated in **bold blue** font. The EW and VW portfolio are not considered in this comparison.



### 3.5.2 Robustness Checks

Table 3 presents the ‘single’ results over the entire out-of-sample period from 12/18/2000 to 01/31/2022. A natural question arises: is the relative performance of the various models stable throughout this period, or does it change during certain subperiods, such as periods of ‘boom’ or ‘bust’ compared to normal conditions? To investigate this, we conduct a rolling-window analysis based on shorter out-of-sample periods.

The results for out-of-sample standard deviations using a one-year rolling window are displayed in Figure 7. The figure shows results for linear shrinkage GMV portfolios with a universe size of  $N = 500$ . The relative performance appears remarkably stable over time, with RL-L consistently performing best across all subperiods. Specifically, RL-L outperforms L in 95% of trading days based on one-year rolling SDs. Moreover, RL-L remains very close to the Oracle performance for most of the period, whereas L consistently performs markedly worse than the theoretical optimum.

Interestingly, the advantage of RL over classic shrinkage intensities remains largely time-invariant: we do not observe systematic differences in RL’s outperformance during periods of ‘boom’ or ‘bust’ compared to normal periods. This finding reinforces the robustness and stability of RL’s improvements.

To save space and avoid redundancy, we do not report results for nonlinear shrinkage. However, the same pattern of robust outperformance over time is observed for RL-NL versus NL. That said, the relative improvement of RL is smaller for nonlinear shrinkage, as also shown in Table 3.

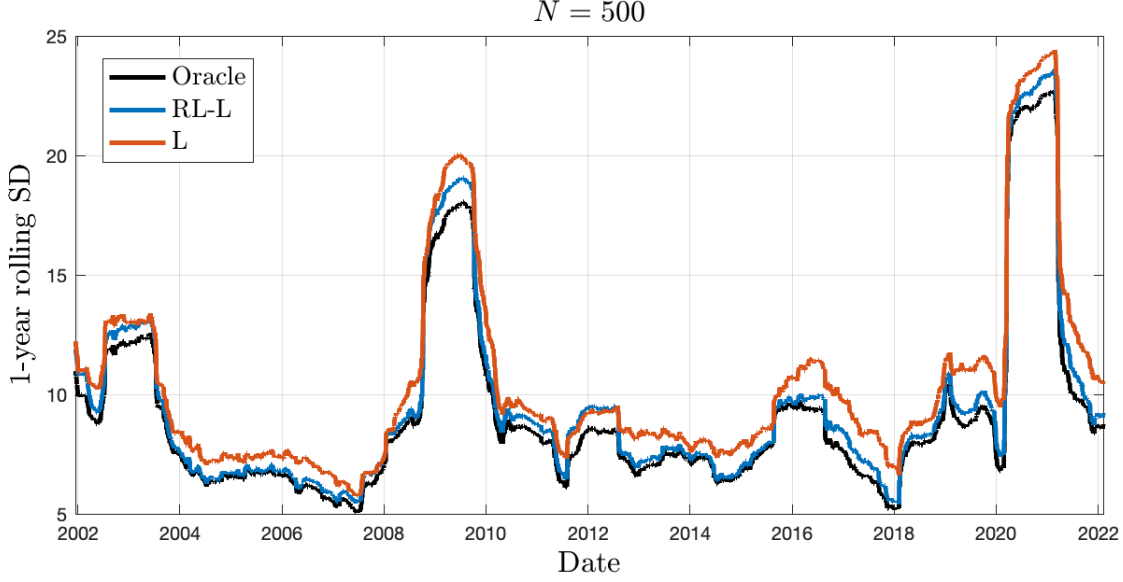


Figure 7: One-year rolling-window of out-of-sample standard deviations (SD) for linear shrinkage models based on the  $N = 500$  GMV portfolio.

For linear shrinkage, we also examine alternative shrinkage targets proposed in the literature, such as CVC. In unreported results, we find that the RL version consistently improves upon classic linear shrinkage across various shrinkage targets. Although the reduction in SD is statistically significant, the relative improvement tends to be smaller for more sophisticated shrinkage targets. This is because their shrinkage intensity is often substantially larger due to their enhanced accuracy; see [De Nard \(2022\)](#).<sup>11</sup>

Additionally, we conduct various robustness checks on the number of stocks in the portfolio ( $N$ ), the number of observations in the estimation window ( $T$ ), and, consequently, different concentration ratios. The results confirm that our findings are robust and not driven by specific parameter choices.

### 3.5.3 Long-Only Minimum Variance Portfolios

In this section, we present the results for long-only minimum variance portfolios. While classic shrinkage covariance matrix estimators are independent of the portfolio optimization step, the ex post optimal shrinkage intensity (Oracle) differs, which also affects the learned shrinkage intensity of the RL agent. The long-only results are summarized in [Table 7](#) and visualized in [Figures 8 and 9](#).

<sup>11</sup>The results are available upon request.

- All models significantly outperform EW and VW benchmarks, including in terms of the Sharpe ratio.
- Each RL-L and RL-NL model consistently outperforms its respective classic L and NL base model. For NL and higher-dimensional settings, RL-NL’s outperformance is both statistically significant and economically meaningful.
- A consistent ranking emerges for smaller portfolio sizes  $N \in \{30, 50, 100\}$ , ordered from best to worst as: RL-L, L, RL-NL, NL. For larger portfolio sizes  $N \in \{225, 500\}$ , the ranking shifts to: RL-NL, NL, RL-L, L, VW, EW.

Long-only minimum variance portfolio						
	L	RL-L	NL	RL-NL	EW	VW
	SD					
$N = 30$	13.82	<b>13.81</b>	13.87	13.87	18.72	19.13
$N = 50$	13.86	<b>13.85</b>	13.91	13.90	18.86	19.04
$N = 100$	13.49	<b>13.47</b>	13.52	13.51	19.51	19.20
$N = 225$	12.77	12.76	12.82	<b>12.75*</b>	19.87	19.32
$N = 500$	12.14	12.13	12.24	<b>12.10*</b>	20.62	19.52

Table 7: Annualized out-of-sample standard deviations for the competing estimators of the long-only minimum variance portfolio, in percentage. All measures are based on 5,313 daily out-of-sample returns from 12/18/2000 until 01/31/2022. The best estimator with the lowest SD is indicated in **bold blue** font. In the columns RL-L, respectively RL-NL, significance at the 0.01 level of outperformance in SD over L, respectively NL, is denoted by an asterisk \*.

The outperformance of RL-L over L is again driven by an increased average shrinkage intensity; see Figure 8. For the  $N = 500$  long-only GMV portfolio, the average shrinkage intensity of RL-L is 0.2, which is very close to the Oracle’s average shrinkage intensity of 0.21, especially in comparison to L, which has an average shrinkage intensity of 0.07.

However, the outperformance of RL-L is smaller in the long-only setting due to the upper bounds imposed on portfolio weights, which act as constraints in the optimization. As a result, less shrinkage is required due to the implicit shrinkage effect described in Jagannathan and Ma (2003). This reduced need for shrinkage is evident in the lower Oracle and RL-L shrinkage intensities compared to the unconstrained long-short setting, where shrinkage intensity is approximately halved; see Figure 3 for comparison.

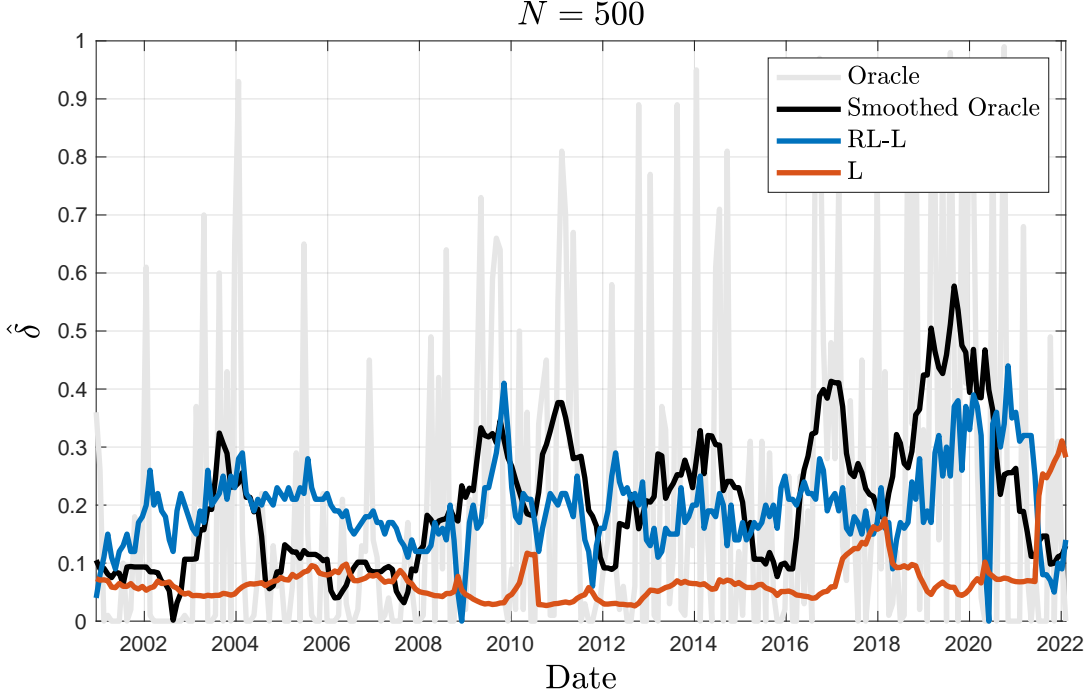


Figure 8: Comparison of the monthly Oracle, RL-L and L estimated linear shrinkage intensities over time for the  $N = 500$  long-only GMV portfolio. For the smoothed Oracle we use a one-year moving average smoothing.

The reduced need for shrinkage due to the long-only constraint can also be observed for nonlinear shrinkage; see Figure 9. While in the unconstrained setting, a higher shrinkage intensity was required, with an average Oracle adjustment of 1.26, the long-only case requires significantly less shrinkage, with an average Oracle adjustment of 0.68.

RL offers an advantage over classic shrinkage methods by dynamically adjusting shrinkage intensity based on the specific optimization problem and performance metric of interest. Unlike classic approaches, which derive a theoretical optimal shrinkage intensity under the i.i.d. assumption independent of the portfolio optimization step, RL learns to optimize shrinkage in a forward-looking manner, directly incorporating the interaction between shrinkage, constraints and portfolio performance. This adaptive approach allows RL to tailor shrinkage intensity to different portfolio constraints, asset universes, and risk-return objectives, ultimately leading to more robust and economically meaningful improvements in out-of-sample performance.

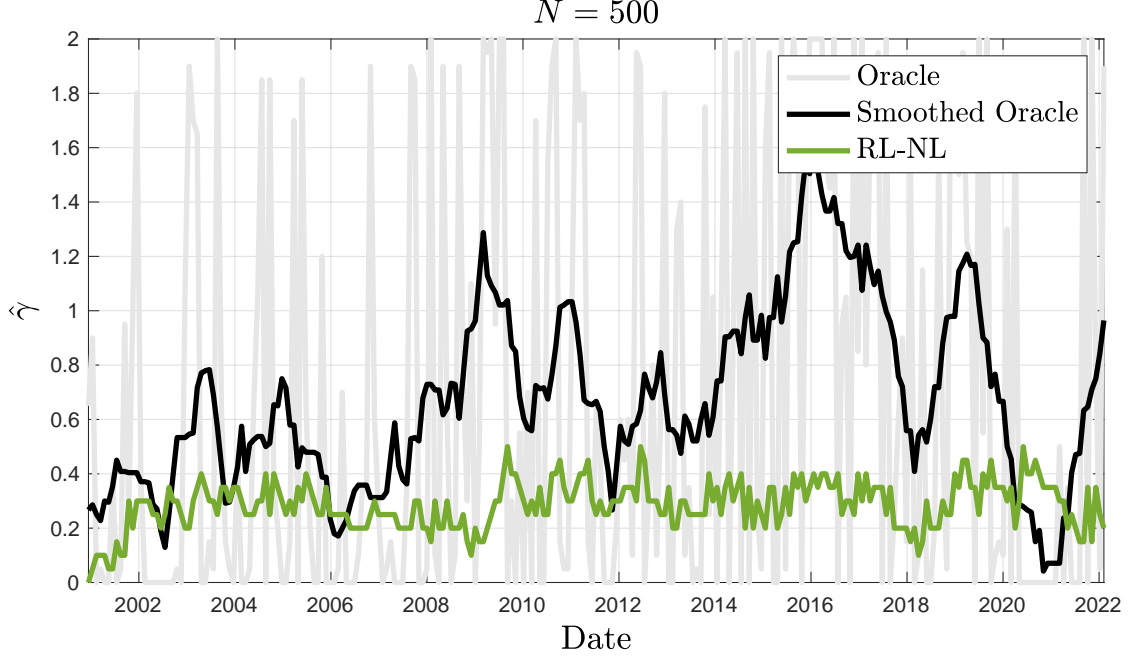


Figure 9: Comparison of the Oracle and RL-NL estimated nonlinear shrinkage intensities adjustment over time for the  $N = 500$  long-only GMV portfolio. For the smoothed Oracle we use a one-year moving average smoothing.

### 3.6 Future Research

In this paper, we focus on the added value of RL for the computation of shrinkage intensity(ies) without making any assumptions about the data-generating process. However, RL-L and RL-NL remain static covariance matrix estimators. It would therefore be of interest to combine our approach with the dynamic conditional correlation (DCC) model of [Engle et al. \(2019\)](#), specifically integrating the DCC multivariate GARCH model with RL-L or RL-NL for correlation targeting.

Another potential enhancement to improve performance is incorporating higher-frequency data (see, e.g., [Callot et al., 2017](#); [De Nard et al., 2022](#)) or factor data (see, e.g., [De Nard et al., 2021](#); [Alves et al., 2023](#); [De Nard and Zhao, 2022, 2023](#)). An alternative approach is to apply RL not to shrinkage estimators but to other dimensionality reduction techniques such as principal component analysis (see, e.g., [Fan et al., 2013](#)).

Finally, RL could be a powerful tool for learning the optimal time-varying combination of different covariance matrix estimators. For example, RL could be used to determine the shrinkage intensities for multi-target shrinkage estimation (see, e.g., [Lancewicki and Aladjem,](#)

2014) or, more generally, to derive the optimal weighting scheme for ensemble methods that aggregate multiple covariance matrix estimators.

## 4 Conclusion

The main contribution of this paper is to introduce a data-driven approach to learn the optimal shrinkage intensity(ies) using reinforcement learning . The proposed AI Shrinkage estimator applies to both linear and nonlinear shrinkage, demonstrating improved performance compared to classic shrinkage estimators. The improvement stems from dropping the i.i.d. assumption when computing the shrinkage intensity(ies) and instead leveraging a data-driven approach that does not impose any assumptions on the data-generating process. Notably, the AI Shrinkage estimators rely solely on stock return data but can be easily extended to incorporate factor or alternative data signals.

Our results show that an RL agent recognizes that the shrinkage intensity(ies) derived under the i.i.d. assumption are biased downward and automatically adjusts for this bias by increasing shrinkage depending on the market environment. The learned shrinkage intensity(ies) remain stable, intuitive, and correlated with the classic i.i.d. approach. To assess the quality of our covariance matrix estimators, we conduct an extensive empirical study in a high-dimensional setting, focusing on the out-of-sample standard deviation of GMV portfolios, where the RL agent optimizes shrinkage under a minimum-variance utility function.

A further contribution of this paper is the application of RL to constrained portfolio optimization, recognizing its practical relevance and the flexibility of RL to adapt to specific investor preferences. Unlike classic shrinkage methods, which are independent of the optimization step, RL dynamically adjusts shrinkage based on the specific constraints, risk-return objectives, and utility functions of the investor. Most classic shrinkage methods tend to shrink too aggressively in constrained optimization problems, as they do not account for the implicit shrinkage impact of constraints, potentially leading to suboptimal portfolio allocations. This adaptability allows RL to incorporate the interaction between shrinkage and portfolio performance in a forward-looking manner, making it particularly well-suited for real-world investment settings where constraints play a crucial role.

Finally, the AI Shrinkage estimators enable more efficient implementation of risk-optimized

portfolios and are well-suited for real-world investment applications. However, our approach is not limited to portfolio selection or even financial applications; it provides a general framework for improving covariance matrix estimation across various domains.

## References

- Alves, R. P., de Brito, D. S., Medeiros, M. C., and Ribeiro, R. M. (2023). Forecasting large realized covariance matrices: The benefits of factor models and shrinkage. *Journal of Financial Econometrics*, 22(3):696–742.
- Bartz, D. and Müller, K. (2014). Covariance shrinkage for autocorrelated data. NIPS Proceedings 1592–1600, Advances in Neural Information Processing Systems 27.
- Bodnar, T., Okhrin, Y., and Parolya, N. (2022). Optimal shrinkage-based portfolio selection in high dimensions. *Journal of Business & Economic Statistics*, 41(1):140–156.
- Brown, N., Sandholm, T., and Machine, S. (2017). Libratus: The superhuman AI for no-limit poker. In *IJCAI*, pages 5226–5228.
- Callot, L. A. F., Kock, A. B., and Medeiros, M. C. (2017). Modeling and forecasting large realized covariance matrices and portfolio choice. *Journal of Applied Econometrics*, 32(1):140–158.
- Cong, L., Tang, K., Wang, J., and Zhang, Y. (2021). Alphaportfolio: Direct construction through deep reinforcement learning and interpretable AI. Working paper, Cornell University.
- De Nard, G. (2022). Oops! I shrunk the sample covariance matrix again: Blockbuster meets shrinkage. *Journal of Financial Econometrics*, 20(4):569–611.
- De Nard, G., Engle, R. F., and Kelly, B. (2024). Factor-mimicking portfolios for climate risk. *Financial Analysts Journal*, 80(3):37–58.
- De Nard, G., Engle, R. F., Ledoit, O., and Wolf, M. (2022). Large dynamic covariance matrices: Enhancements based on intraday data. *Journal of Banking and Finance*, 138:106426.

- De Nard, G., Ledoit, O., and Wolf, M. (2021). Factor models for portfolio selection in large dimensions: The good, the better and the ugly. *Journal of Financial Econometrics*, 19(2):236–257.
- De Nard, G., Ledoit, O., and Wolf, M. (2025). Improved tracking-error management for active and passive investing. *The Journal of Portfolio Management*, 51(4):40–62.
- De Nard, G. and Zhao, Z. (2022). A large-dimensional test for cross-sectional anomalies: Efficient sorting revisited. *International Review of Economics and Finance*, 80:654–676.
- De Nard, G. and Zhao, Z. (2023). Using, taming or avoiding the factor zoo? A double-shrinkage estimator for covariance matrices. *Journal of Empirical Finance*, 72:23–35.
- DeMiguel, V., Garlappi, L., and Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the  $1/N$  portfolio strategy? *Review of Financial Studies*, 22:1915–1953.
- Dom, M. S., Howard, C., Jansen, M., and Lohre, H. (2025). Beyond GMV: the relevance of covariance matrix estimation for risk-based portfolio construction. *Quantitative Finance*, Forthcoming.
- Elton, E. J. and Gruber, M. J. (1973). Estimating the dependence structure of share prices ? implication for portfolio selection. *Journal of Finance*, 28:1203–1232.
- Engle, R. F., Ledoit, O., and Wolf, M. (2019). Large dynamic covariance matrices. *Journal of Business & Economic Statistics*, 37(2):363–375.
- Fan, J., Liao, Y., and Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements (with discussion). *Journal of the Royal Statistical Society, Series B*, 75(4):603–680.
- Frost, P. A. and Savarino, J. E. (1986). An empirical Bayes approach to portfolio selection. *Journal of Financial and Quantitative Analysis*, 21:293–305.
- Haugen, R. A. and Baker, N. L. (1991). The efficient market inefficiency of capitalization-weighted stock portfolios. *Journal of Portfolio Management*, 17(3):35–40.
- Hoberg, G. and Phillips, G. (2016). Text-based network industries and endogenous product differentiation. *Journal of Political Economy*, 124(5):1423–1465.



- Jagannathan, R. and Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *Journal of Finance*, 54(4):1651–1684.
- Jobson, J. D. and Korkie, B. M. (1980). Estimation for Markowitz efficient portfolios. *Journal of the American Statistical Association*, 75:544–554.
- Lancewicki, T. and Aladjem, M. (2014). Multi-target shrinkage estimation for covariance matrices. *IEEE Transactions on Signal Processing*, 62(24):6380–6390.
- Ledoit, O. and Wolf, M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance*, 10(5):603–621.
- Ledoit, O. and Wolf, M. (2004a). Honey, I shrunk the sample covariance matrix. *Journal of Portfolio Management*, 30(4):110–119.
- Ledoit, O. and Wolf, M. (2004b). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2):365–411.
- Ledoit, O. and Wolf, M. (2011). Robust performance hypothesis testing with the variance. *Wilmott Magazine*, September:86–89.
- Ledoit, O. and Wolf, M. (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. *Annals of Statistics*, 40(2):1024–1060.
- Ledoit, O. and Wolf, M. (2017). Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets Goldilocks. *Review of Financial Studies*, 30(12):4349–4388.
- Ledoit, O. and Wolf, M. (2020). Analytical nonlinear shrinkage of large-dimensional covariance matrices”. *Annals of Statistics*, 48:3043–3065.
- Ledoit, O. and Wolf, M. (2022a). The power of (non-)linear shrinking: A review and guide to covariance matrix estimation. *Journal of Financial Econometrics*, 20(1):187–218.
- Ledoit, O. and Wolf, M. (2022b). Quadratic shrinkage for large covariance matrices. *Bernoulli*, 28(3):1519–1547.
- Lu, C., Ndiaye, P. M., and Simaan, M. (2024). Improved estimation of the correlation matrix using reinforcement learning and text-based networks. *International Review of Financial Analysis*, 96:73–105.

- Markowitz, H. (1952). Portfolio selection. *Journal of Finance*, 7:77–91.
- Matera, G. and Matera, R. (2023). Shrinkage estimation with reinforcement learning of large variance matrices for portfolio selection. *Intelligent Systems with Applications*, 17:73–105.
- Michaud, R. (1989). The Markowitz optimization enigma: Is optimized optimal? *Financial Analysts Journal*, 45:31–42.
- Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., et al. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533.
- Moravčik, M., Schmid, M., Burch, N., Lisỳ, V., Morrill, D., Bard, N., Davis, T., Waugh, K., Johanson, M., and Bowling, M. (2017). Deepstack: Expert-level artificial intelligence in heads-up no-limit poker. *Science*, 356(6337):508–513.
- Nielsen, F. and Aylursubramanian, R. (2008). Far from the madding crowd — Volatility efficient indices. Research insights, MSCI Barra.
- Pan, X., You, Y., Wang, Z., and Lu, C. (2017). Virtual to real reinforcement learning for autonomous driving. *arXiv preprint arXiv:1704.03952*.
- Roncalli, T. (2011). Understanding the impact of weights constraints in portfolio theory. Working Paper 36753, MPRA.
- Sancetta, A. (2008). Sample covariance shrinkage for high dimensional dependent data. *Journal of Multivariate Analysis*, 99:949–967.
- Sharpe, W. F. (1963). A simplified model for portfolio analysis. *Management Science*, 9(1):277–293.
- Shen, W., Wang, J., Jiang, Y.-G., and Zha, H. (2015). Portfolio choices with orthogonal bandit learning. In *IJCAI*, volume 15, pages 974–980.
- Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., et al. (2016). Mastering the game of Go with deep neural networks and tree search. *nature*, 529(7587):484–489.

Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, pages 197–206. University of California Press.

## A Proof of Equivalence for Shrinkage Towards the Identity Matrix

By substituting the definition of  $\lambda_i^{\text{I}}$  from Equation (2.7) into Equation (2.16) and carrying out some algebraic manipulations, we obtain:

$$\tilde{\lambda}_i^{\text{RL-I}} := \hat{\gamma} \lambda_i^{\text{I}} + (1 - \hat{\gamma}) \lambda_i \quad (\text{A.1})$$

$$= \hat{\gamma} (\hat{\delta}^{\text{I}} \bar{\lambda} + (1 - \hat{\delta}^{\text{I}}) \lambda_i) + (1 - \hat{\gamma}) \lambda_i \quad (\text{A.2})$$

$$= \hat{\gamma} \hat{\delta}^{\text{I}} \bar{\lambda} + \hat{\gamma} (1 - \hat{\delta}^{\text{I}}) \lambda_i + (1 - \hat{\gamma}) \lambda_i \quad (\text{A.3})$$

$$= \hat{\gamma} \hat{\delta}^{\text{I}} \bar{\lambda} + (\hat{\gamma} (1 - \hat{\delta}^{\text{I}}) + 1 - \hat{\gamma}) \lambda_i \quad (\text{A.4})$$

$$= \hat{\gamma} \hat{\delta}^{\text{I}} \bar{\lambda} + (1 - \hat{\gamma} \hat{\delta}^{\text{I}}) \lambda_i . \quad (\text{A.5})$$

As  $\hat{\gamma} \hat{\delta}^{\text{I}} \geq 0$ , and for  $\hat{\gamma} \leq 1/\hat{\delta}^{\text{I}}$ , we can write

$$\lambda_i^{\text{RL-I}} = \hat{\delta}^{\text{RL-I}} \bar{\lambda} + (1 - \hat{\delta}^{\text{RL-I}}) \lambda_i , \quad (\text{A.6})$$

where  $\hat{\delta}^{\text{RL-I}} := \hat{\gamma} \hat{\delta}^{\text{I}}$  is the same as in the direct approach (2.14).

## B Additional Figures

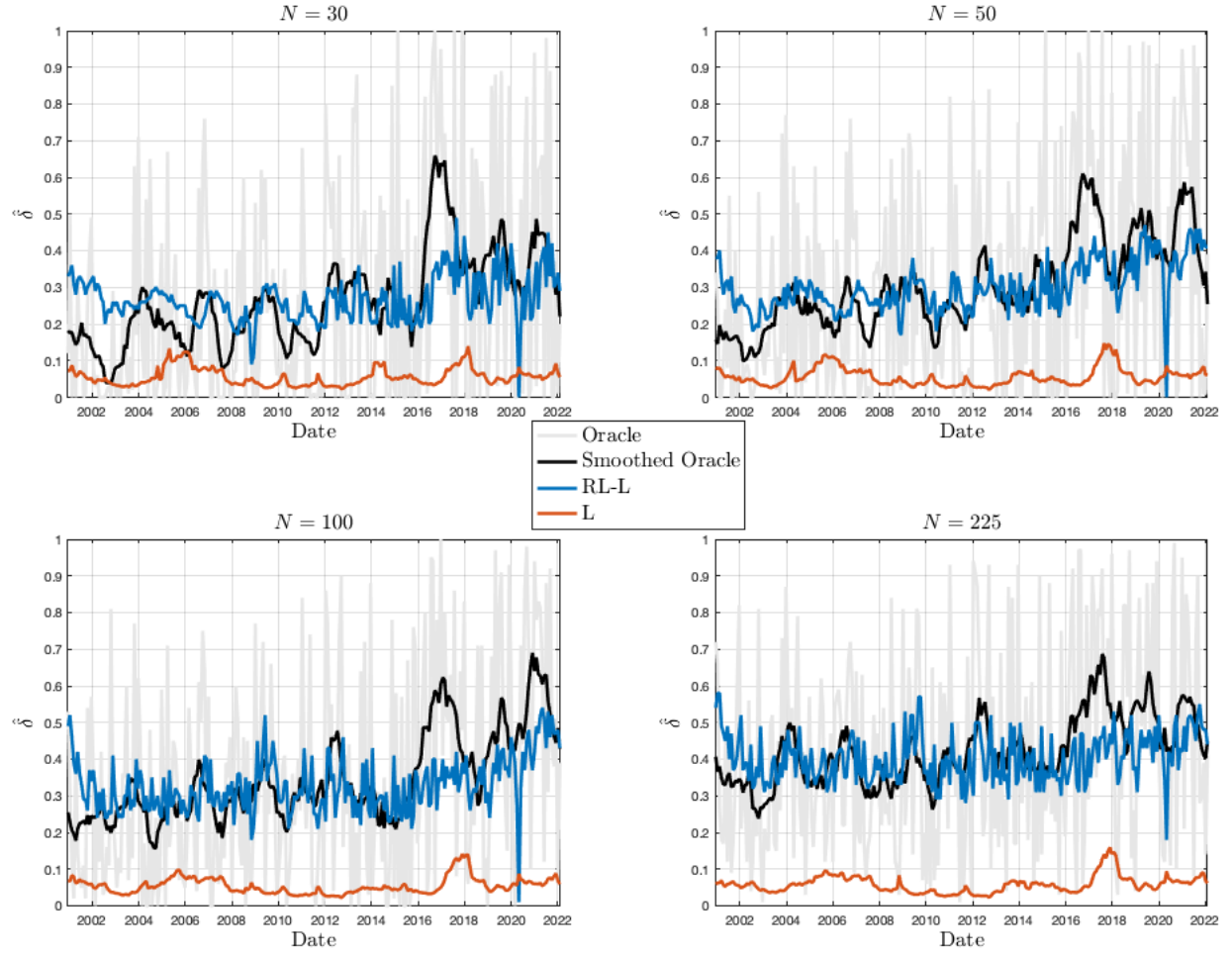


Figure B.1: Comparison of the monthly Oracle, RL-L and L estimated linear shrinkage intensities over time for various portfolio sizes. For the smoothed Oracle we use a one-year moving average smoothing.

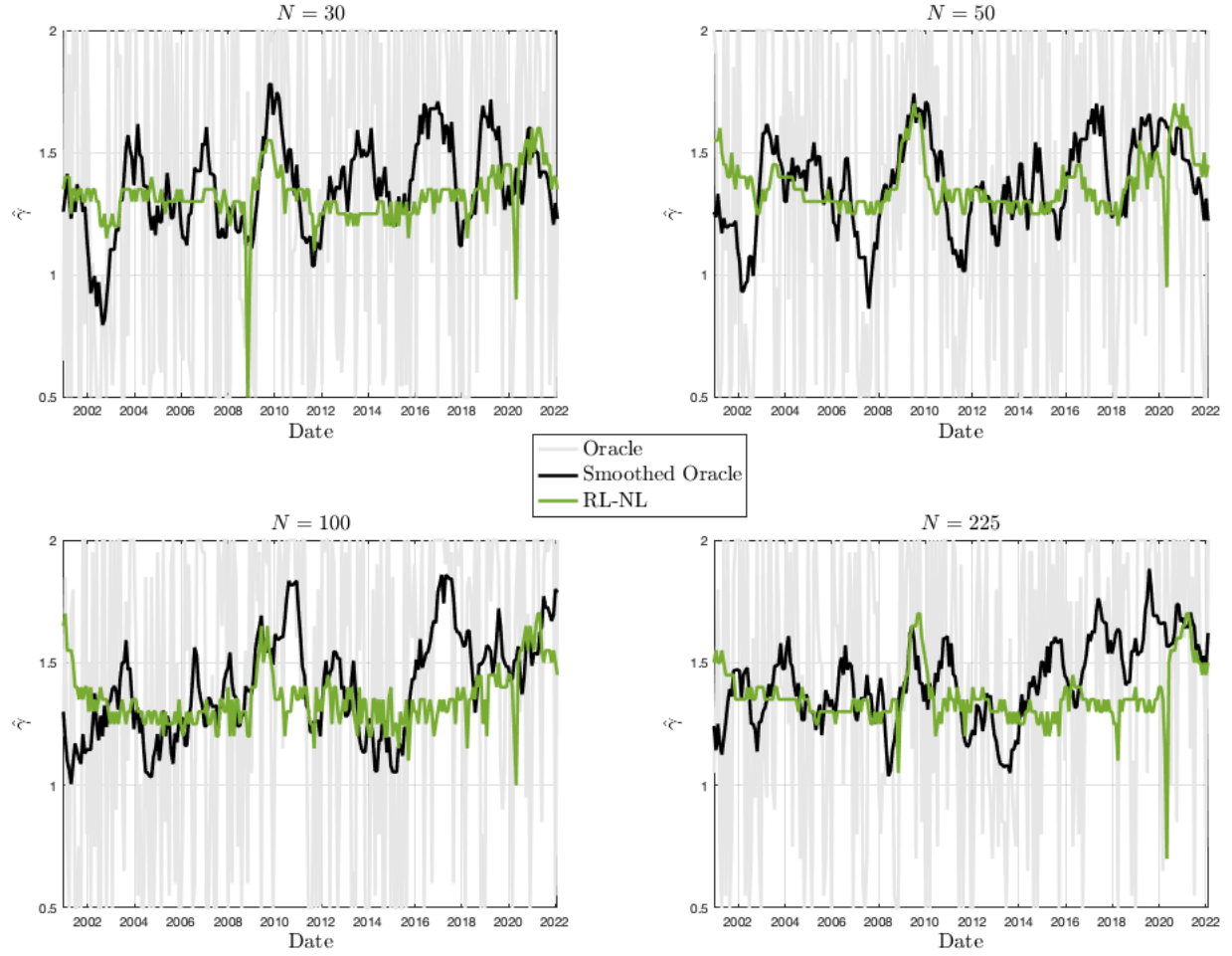


Figure B.2: Comparison of the Oracle and RL-NL estimated nonlinear shrinkage intensities adjustment over time for various portfolio sizes. For the smoothed Oracle we use a one-year moving average smoothing.