Optimizing the Ledoit-Wolf Estimator through High-Dimensional Regularization

Youhong Lee (joint with Alex Shkolnik)

Department of Statistics and Applied Probability, UC Santa Barbara

August 2, 2023

Outline

1. The Ledoit-Wolf (LW) Estimator for Covariance Estimation
2. A High-Dimensional Approach to Optimizing the Ledoit-Wolf Estimator
3. Summary

Section 1

The Ledoit-Wolf (LW) Estimator for Covariance Estimation

The Sample Covariance Matrix

The Sample Covariance Matrix

For a $p \times n$ data matrix Y, let the sample covariance matrix be

$$\mathrm{S} = \frac{1}{n}\mathrm{Y}\mathrm{Y}^\top$$

where p is the feature dimension and n is the sample size.

- For our calculation purposes, we make the assumption that the population mean equals zero.
- The sample covariance matrix characterizes the linear relationships among the variables in terms of their variances and covariances.
- The sample covariance matrix is recognized as a consistent estimator for the population covariance matrix, denoted by Σ.

Properties of S in High Dimensions

The Sample Covariance Matrix

For a $p \times n$ data matrix Y, let the sample covariance matrix be

$$\mathrm{S} = \frac{1}{n}\mathrm{Y}\mathrm{Y}^\top$$

where p is the feature dimension and n is the sample size.

- However, in high-dimensional settings where p greatly exceeds n, S displays certain unfavorable characteristics.
- As suggested by Ledoit & Wolf (2004), S lacks invertibility when $p > n$ (this can be a critical issue as the inverse of S may be required for computing the minimum variance portfolio).
- The Marchenko–Pastur law indicates that the spectrum of S deviates considerably from its population counterpart.

The Ledoit-Wolf (LW) Estimator

Ledoit & Wolf (2004)

The Ledoit-Wolf (LW) Estimator is given by

$$\hat{\Sigma} = \alpha \mathrm{S} + (1 - \alpha)\mathrm{F}$$

where S denotes the sample covariance matrix, F stands for a shrinkage target, and $\alpha \in [0, 1]$ represents a shrinkage constant.

- Ledoit & Wolf (2004) proposes a James-Stein type shrinkage estimator for the covariance matrix.
- This estimator is constructed as a linear combination of the sample covariance matrix and a predetermined shrinkage target.
- This approach leads us to two important questions:
  1. What would be the most appropriate shrinkage target?
  2. What degree of shrinkage is optimal?

The Ledoit-Wolf (LW) Estimator

Ledoit & Wolf (2004)

The Ledoit-Wolf (LW) Estimator is given by

$$\hat{\Sigma} = \alpha \mathrm{S} + (1-\alpha)\mathrm{F}$$

where S denotes the sample covariance matrix, F stands for a shrinkage target, and $\alpha \in [0,1]$ represents a shrinkage constant.

- A simply structured deterministic matrix can be used as a shrinkage target.

$$\mathrm{F} = \text{identity}, \text{a factor model}, \text{constant correlation}, \text{etc.}$$

- The optimal shrinkage constant is selected to minimize an error rate.

$$\alpha^* \in \arg\min_{\alpha \in [0,1]} \mathrm{Risk}(\alpha)$$

A Choice of the Optimal Shrinkage Constant

Ledoit & Wolf (2004)

The Ledoit-Wolf (LW) Estimator is given by

$$\hat{\Sigma} = \alpha \mathrm{S} + (1-\alpha)\mathrm{F}$$

where S denotes the sample covariance matrix, F stands for a shrinkage target, and $\alpha \in [0,1]$ represents a shrinkage constant.

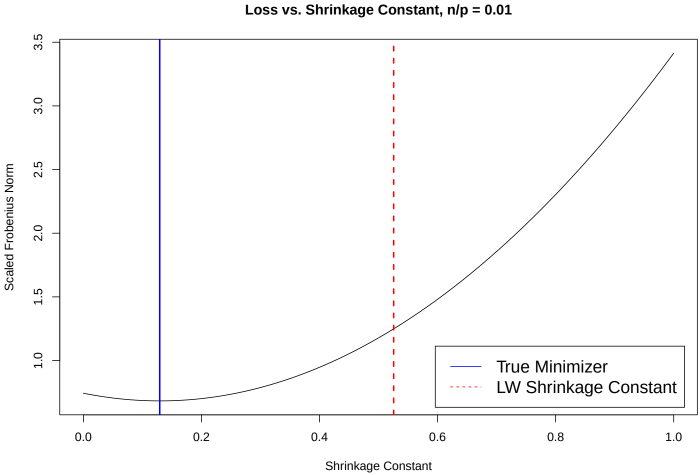
- More specifically, the optimal shrinkage constant is selected to asymptotically minimize the risk function of the Frobenius norm as loss.

$$\lim_{n\to\infty} \hat{\alpha}^* \in \arg\min_{\alpha\in[0,1]} \lim_{n\to\infty} \mathbb{E}\left[\|\hat{\Sigma} - \Sigma\|^2\right]$$

- In other words, $\hat{\alpha}^*$ is a consistent estimator for the minimizer of the risk function.

A Choice of the Optimal Shrinkage Constant

- However, empirical studies indicate that the optimal shrinkage constant may be overestimated in situations where the ratio n/p is relatively small.



Refining the Optimal Shrinkage Constant Selection

- We introduce a new analysis within the high-dimension, low-sample size (HL) regime.
- Our approach involves selecting an optimal shrinkage constant that minimizes its p-asymptotic loss as $p \to \infty$.

$$\lim_{p \to \infty} \alpha^* \in \arg \min_{\alpha \in [0,1]} \lim_{p \to \infty} \mathrm{Error}(\alpha)$$

Section 2

A High-Dimensional Approach to Optimizing the Ledoit-Wolf Estimator

A Loss Function with the Frobenius Norm

Definition 1

Our loss function is defined as the squared Frobenius norm between the true and estimated covariance matrices divided by p,

$$L(\alpha) = \frac{1}{p}\|\Delta(\alpha)\|^2 = \frac{1}{p}\mathrm{tr}\left(\Delta^2(\alpha)\right)$$

where $\Delta = \Sigma - \hat{\Sigma}$ denotes the difference between the true and estimated covariance matrices with $\hat{\Sigma} = \alpha \mathrm{S} + (1-\alpha)\mathrm{F}$.

A Minimizer of the Loss Function

Lemma 1

A minimizer of the the loss function is given by

$$\alpha^* = \frac{\mathrm{tr}\left((\mathrm{F} - \mathrm{S})(\Sigma - \mathrm{S})\right)}{\mathrm{tr}\left((\mathrm{F} - \mathrm{S})^2\right)}$$

- The minimizer is unique almost surely.
- For a convex combination, the minimizer may be truncated to ensure $\alpha \in [0, 1]$.
- Note that the numerator can be expressed as

$$\mathrm{tr}\left((\mathrm{F} - \mathrm{S})(\Sigma - \mathrm{S})\right) = \mathrm{tr}(\mathrm{F}\Sigma) - \mathrm{tr}(\mathrm{S}\Sigma) - \mathrm{tr}\left(\mathrm{S}(\mathrm{F} - \mathrm{S})\right)$$

so we have two unobservable terms, $\mathrm{tr}(\mathrm{F}\Sigma)$ and $\mathrm{tr}(\mathrm{S}\Sigma)$.

The Market Model

- We adopt the assumption that that the population covariance matrix follows the market model

$$\Sigma = p\sigma^2 uu^\top + \delta^2 \mathrm{I}$$

where u is the leading eigenvector with its corresponding eigenvalue $p\sigma^2 + \delta^2$ for some $\sigma^2 > 0$ and $\delta^2 > 0$ and I is the identity matrix.

- Correspondingly, we employ the single factor shrinkage target, represented as

$$\mathrm{F} = p\sigma_F^2 qq^\top + \delta_F^2 \mathrm{I}$$

where q is an educated guess about the leading eigenvector and $\sigma_F^2 > 0$ and $\delta_F^2 > 0$ are constants.

Lemmas for Unknown Quantities

- We derive random variables that are asymptotically equivalent to the two unknown quantities, $\mathrm{tr}(\mathrm{F}\Sigma)$ and $\mathrm{tr}(\mathrm{S}\Sigma)$.

Lemma 2

For two sequences of random variables, a_p and b_p, denote $a_p \sim b_p$ if $\lim_{p\to\infty} \frac{a_p}{b_p} = 1$ almos surely. Then, under some assumptions,

(a)

$$\mathrm{tr}(\mathrm{S}\Sigma) \sim p\sigma^2 s^2 \langle u, v\rangle^2$$

where v is the leading eigenvector of S corresponding to the largest eigenvalue s^2 with $\Sigma = p\sigma^2 uu^\top + \delta^2 \mathrm{I}$.

(b)

$$\mathrm{tr}(\mathrm{F}\Sigma) \sim \sigma^2 \sigma_F^2 \langle u, q\rangle^2$$

with $\mathrm{F} = p\sigma_F^2 qq^\top + \delta_F^2 \mathrm{I}$.

An Asymptotic Minimizer of the Loss Function

Theorem 1

Let $\alpha' = \frac{p^2\sigma^2\sigma_F^2\langle u,q\rangle^2 - p\sigma^2 s^2\langle u,v\rangle^2 - \mathrm{tr}(\mathrm{S}(\mathrm{F}-\mathrm{S}))}{\mathrm{tr}((\mathrm{F}-\mathrm{S})^2)}$. Then, under some assumptions, α' asymptotically minimizes the loss function such that

$$\alpha' \sim \alpha^*$$

where α^* is the true minimizer.

- We observe that α' is composed of estimable quantities, specifically σ^2, $\langle u, q\rangle$, and $\langle u, v\rangle$.
- Our analysis replicates the methods used in related literature, such as Jung et al. (2012), Goldberg et al. (2021), and Shkolnik (2022).

Consistent Estimators for Eigenvalues and Eigenvectors

Lemma 3

Let $\hat{\delta}^2 = \frac{(\mathrm{tr(S)} - \mathrm{s}^2) m_p}{n - m_p}$ with $m_p = 1 + n/p$ be an estimator of δ^2, and let $\hat{\sigma}^2 = s^2/p - \hat{\delta}^2/n$ be an estimator of σ^2. Then, under some assumptions,

$$\hat{\sigma}^2 \sim \sigma^2 \qquad \text{and} \qquad \hat{\delta}^2 \sim \delta^2.$$

Moreover, if we define $\psi \geq 0$ by $\psi^2 = \hat{\delta}^2/s^2$, then

$$\psi \sim \langle u, v \rangle \qquad \text{and} \qquad \langle v, q \rangle / \psi \sim \langle u, q \rangle.$$

- Consequently, we derive p-consistent estimators for σ^2, $\langle u, q \rangle$, and $\langle u, v \rangle$.
- These estimators contribute to the construction of a new shrinkage parameter, $\hat{\alpha}'$.

An Asymptotic Minimizer of the Loss Function

Theorem 2

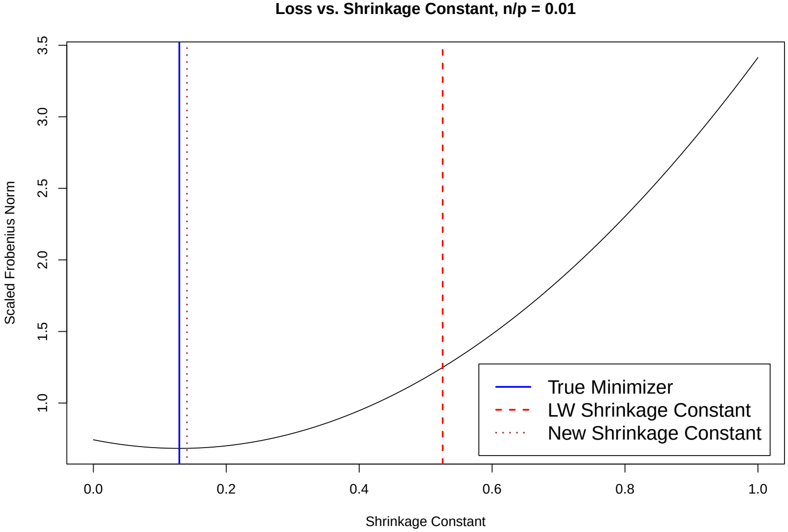
Let $\hat{\alpha}' = \frac{p^2 \hat{\sigma}^2 \sigma_F^2 \langle v, q \rangle^2 / \psi^2 - p \hat{\sigma}^2 s^2 \psi^2 - \mathrm{tr}(\mathrm{S}(\mathrm{F}-\mathrm{S}))}{\mathrm{tr}((\mathrm{F}-\mathrm{S})^2)}$ be an estimator of α^*. Then, under some assumptions, $\hat{\alpha}'$ asymptotically minimizes the loss function such that

$$\hat{\alpha}' \sim \alpha^*$$

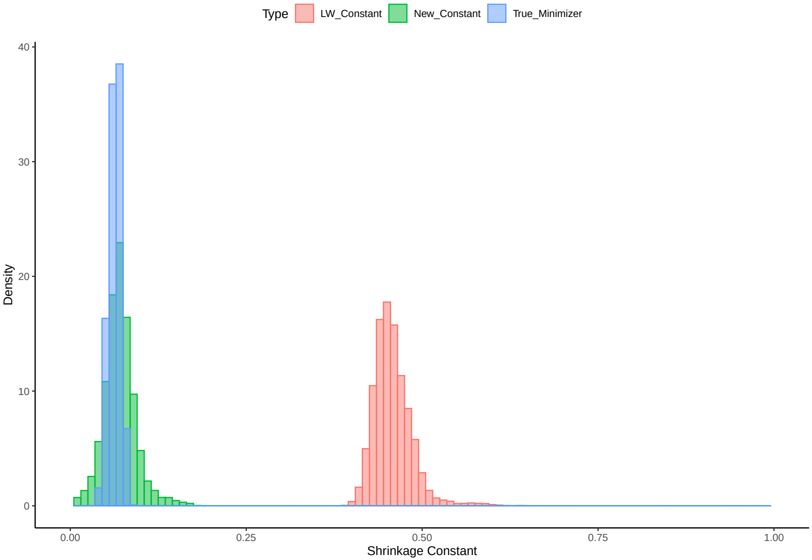
where α^* is the true minimizer.

- We note that $\hat{\alpha}'$ consists solely of sample quantities.
- Thus, $\hat{\alpha}'$ offers a practical formula for determining the shrinkage constant, based on a p-asymptotic analysis.

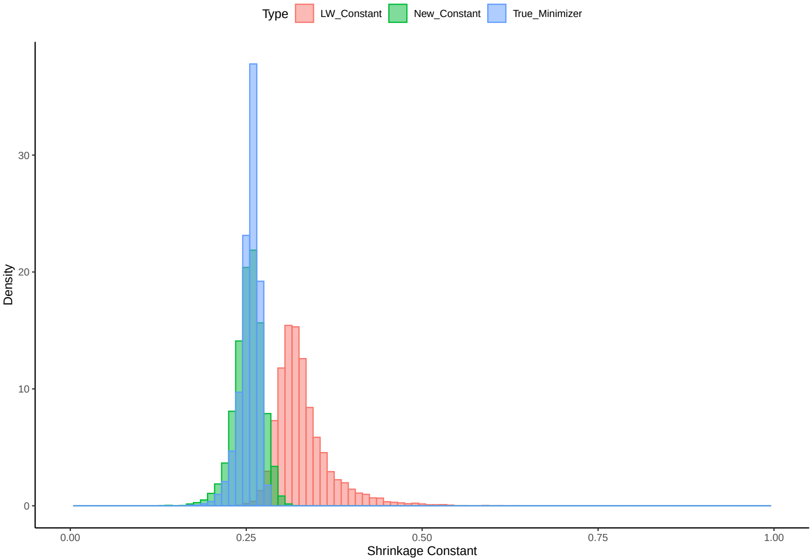
A Comparison of the Optimal Shrinkage Constants



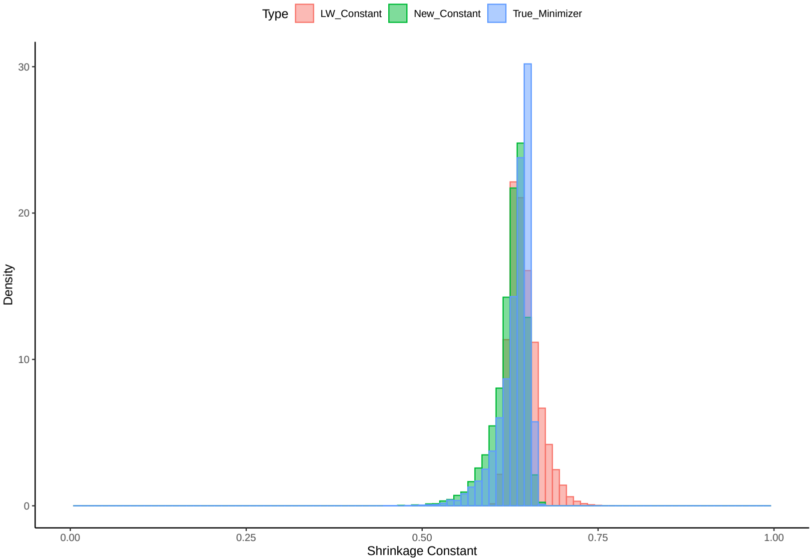
Numerics: Distribution of Shrinkage Constants with $n/p = 0.01$



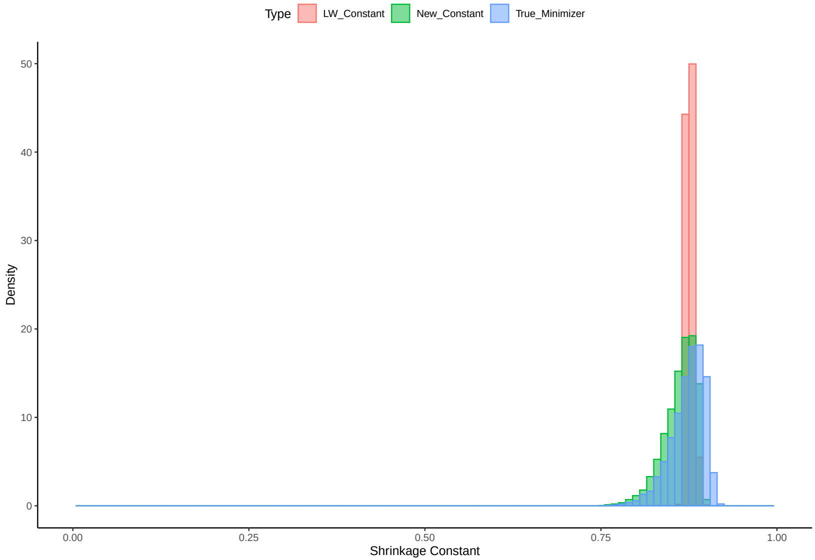
Numerics: Distribution of Shrinkage Constants with $n/p = 0.05$



Numerics: Distribution of Shrinkage Constants with $n/p = 0.25$



Numerics: Distribution of Shrinkage Constants with $n/p = 1$



Section 3

Summary

Summary

1. The Ledoit-Wolf estimator is a widely adopted method for covariance estimation.
2. We present a novel strategy for determining an optimal shrinkage constant using p-asymptotic analysis.
3. Both theoretical and numerical results demonstrate that our new optimal constant delivers decreased error rates in situations where n/p is small.

Thank You!