Random Matrix Theory

Alex Papanicolaou

August 18, 2017

This note outlines a new approach to convariance shrinkage of eigenvalues for the purpose of improved estimation of minimum variance portfolios and their risk. We outline first the basic setup of the high-dimensional estimation problem and explore the fundamentals of Random Matrix Theory and the Marchenko- Pastur theorem. We then motivate the new shrinkage approach by studying the non-linear shrinkage approach of Ledoit and Wolf (2017), especially as it relates to minimizing out-of-sample portfolio variance. Finally we sketch out the preliminaries of how these results may be extended to producing unbiased variance forecasts.

1 Introduction

We first setup the problem and introduce the concepts of Random Matrix Theory and the Marchenko-Pastur theorem

1.1 Setup

To begin, assume we observe i.i.d. random values in a $T \times N$ matrix,

$$X_N = \begin{bmatrix} — & x_1 & — \\ & \vdots & \\ — & x_T & — \end{bmatrix} \tag{1.1}$$

where T is the number of observations and N is the number of variables (ie. $x_i \in \mathbb{R}^N$). We index by N since T and N will be growing to infinity. In particular, we will take $T/N \to y > 1$. The quantity y is known as the concentration ratio. Let Σ_N be the covariance matrix for X_N. Furthermore, define the empirical spectral distribution (e.s.d.) for the eigenvalues of Σ_N as,

$$H_N(\tau) = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{[\tau_i,+\infty)}(\tau) \tag{1.2}$$

where $\tau_1, \dots, \tau_N$ is the system of eigenvalues. It will be implicit that this system depends on N.

A basic assumption is that $H_N(\tau)$ converges to a nonrandom limit everywhere with support on a compact interval[1]. This is controversial in the context of a factor model but for now we refrain from going into the details of why.

An example of a limit law H would be a truncated exponential distribution where,

$$H_\gamma(\tau) = \frac{1 - e^{-\gamma(\tau-1)}}{1 - e^{-\gamma}}, \quad \mathrm{supp}(H) = [1, 2]. \tag{1.3}$$

[1] We are going to assume the limit is continuous and on a compact interval to simplify this since otherwise this statement needs to be more precise.

The formula for the eigenvalues would be,

$$\tau = 1 - \frac{1}{\gamma} \log \left(1 - x(1 - e^{-\gamma})\right), \quad x \in [0, 1], \tag{1.4}$$

or for some finite N,

$$\tau_i = 1 - \frac{1}{\gamma} \log \left(1 - x_i(1 - e^{-\gamma})\right), \quad x_i = \frac{i-1}{N-1},\ i = 1, \ldots, N \tag{1.5}$$

Figure 1.1 shows the spectrum for H_γ as well as the empirical spectrum of the sample covariance matrix for different N, which we define next.

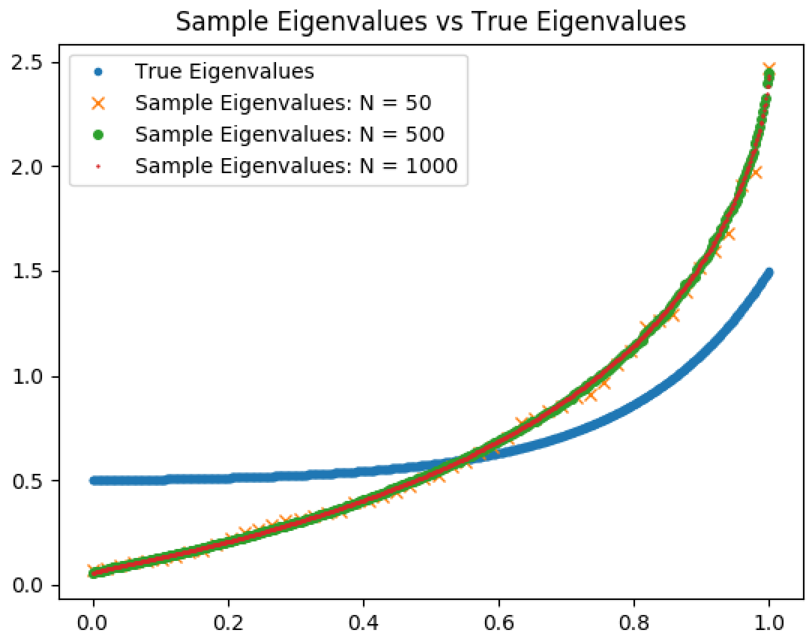

Figure 1.1: Example of sample covariance eigenvalue spectrum for various N and $y = 2$. The true eigenvalue spectrum is given by $H_\gamma(\tau) = \frac{1-e^{-\gamma(\tau-1)}}{1-e^{-\gamma}}$ with $\text{supp}(H) = [1, 2]$. The true spectrum is approximated for some N by $\tau_i = 1 - \frac{1}{\gamma} \log(1 - x_i(1 - e^{-\gamma}))$ where $x_i = \frac{i-1}{N-1}$, $i = 1, \ldots, N$.

2 Marchenko-Pastur Theorem

Define the sample covariance matrix as,

$$S_N = \frac{1}{T} X_N^T X_N \tag{2.1}$$

and the e.s.d. of S_N as

$$F_N(\lambda) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{[\lambda_i, +\infty)}(\lambda) \tag{2.2}$$

where $\lambda_1, \ldots, \lambda_N$ is the system of eigenvalues. It is not assumed that F_N has a limit or what it is. That's where the Marchenko-Pastur theorem comes in.

The main result of Random Matrix Theory and Marchenko-Pastur (MP) is convergence of F_N to a limit law F almost surely. The result is not direct but instead states convergence in terms of the *Stieltjes transform* of F_N and F.

The Stieltjes transform of a nondecreasing function G is,

$$m_G(z) = \int_{-\infty}^{\infty} \frac{dG(x)}{x - z} \tag{2.3}$$

where $z \in \mathbb{C}^+ = \{z \in \mathbb{C} : \mathrm{Im}(z) > 0\}$.

For an empirical distribution like F_N, this is,

$$m_{F_N}(z) = \frac{1}{T}\sum_{i=1}^{N} \frac{1}{\lambda_i - z} = \frac{1}{T}\mathrm{Tr}\left[(S_n - zI)^{-1}\right] \tag{2.4}$$

The Marchenko-Pastur theorem is that $m_{F_N}(z) \to m_F(z)$ almost surely for all $z \in \mathbb{C}^+$ and $m_F(z)$ is defined through the *Marchenko-Pastur equation*,

$$m_F(z) = \int_{\mathrm{supp}(H)} \frac{dH(\tau)}{\tau\left[1 - y^{-1} - y^{-1} z m_F(z)\right] - z}. \tag{2.5}$$

In few cases[2] can the Marchenko-Pastur equation be solved analytically. It can potentially be solved numerically for specified H but this is not particularly effective since it requires a model for H. Moreover, the limit F would come from the inversion of $m_F(z)$ given by,

$$F(b) - F(a) = \lim_{\eta \to 0^+} \frac{1}{\pi} \int_a^b \mathrm{Im}\left[m_F(\xi + i\eta\right] d\xi$$

Figure 1.1 shows the e.s.d. F_N for various values of N. It may be difficult to see but for $N = 1000$, the e.s.d. shows convergence to the limit compared to $N = 50$.

2.1 A Generalization of the Marchenko-Pastur Theorem

Ledoit and Péché (2011) show a generalization of the Marchenko-Pastur Theorem to functionals of the form,

$$\Omega_N^g(\lambda) = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{[\lambda_i, +\infty)}(\lambda) \sum_{j=1}^{N} |u_i^T v_j|^2 g(\tau_j) \tag{2.6}$$

where the sample covariance and population covariance have eigendecompositions,

$$\begin{aligned} S_N &= U\Lambda U^T, \quad U = \begin{bmatrix} u_1 & \cdots & u_N \end{bmatrix} \quad \Lambda = \mathrm{Diag}(\lambda_1, \ldots, \lambda_N), \\ \Sigma_N &= VTV^T, \quad V = \begin{bmatrix} v_1 & \cdots & v_N \end{bmatrix} \quad T = \mathrm{Diag}(\tau_1, \ldots, \tau_N). \end{aligned}$$

The Stieltjes transform of Ω_N^g is given by,

$$\Theta_N^g(z) = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{\lambda_i - z} \sum_{j=1}^{N} |u_i^T v_j|^2 g(\tau_j) \tag{2.7}$$

$$= \frac{1}{N}\mathrm{Tr}\left[(S_N - zI)^{-1} g(\Sigma_N)\right] \tag{2.8}$$

where in an abuse of notation, $g(\Sigma_N)$ is considered as a spectral function, ie. it applies g to the eigenvalues of Σ_N.

[2]Perhaps only one? This would be the case when all eigenvalues are equal to a common value

The relationship to the Trace operator can be seen through some simple linear algebra:

$$\begin{aligned}
\text{Tr}\left[(S_N - zI)^{-1} g(\Sigma_N)\right] &= \text{Tr}\left[(U\Lambda U^T - zI)^{-1} V g(T) V^T\right] \\
&= \text{Tr}\left[U^T(\Lambda - zI)^{-1} U V g(T) V^T\right] \\
&= \text{Tr}\left[(\Lambda - zI)^{-1}(UV) g(T)(UV)^T\right]
\end{aligned}$$

where expanding out the last term gives $\Theta_N^g(z)$. Furthermore, if we take $g \equiv 1$, we get back F_N and m_{F_N}.

The generalized result for $\Omega_N^g(\lambda)$ and $\Theta_N^g(z)$ is $\Theta_N^g(z)$ converges a.s. to $\Theta^g(z)$ for all $z \in \mathbb{C}^+$ where $\Theta^g(z)$ is given by,

$$\Theta^g(z) = \int_{\text{supp}(H)} \frac{g(\tau) dH(\tau)}{\tau\left[1 - y^{-1} - y^{-1} z m_F(z)\right] - z}. \tag{2.9}$$

Note how the integration kernel in the denominator stays the same and the result only differs in our weighting scheme $g(\tau)$.

3 Nonlinear Shrinkage Estimators

3.1 Rotation Invariant Class

We restrict our consideration to the class of estimators,

$$\hat{\Sigma}_N = UDU^T \tag{3.1}$$

where U are the eigenvectors of the sample covariance matrix S. The choice variable is the diagonal matrix of eigenvalues D.

3.2 Frobenius Loss

We aim to minimize Frobenius error via,

$$\min_D \|UDU^T - \Sigma_N\|_F^2. \tag{3.2}$$

The optimal oracle eigenvalues are

$$D^* = \text{Diag}(\tilde{d}_1, \ldots, \tilde{d}_N) = \text{Diag}(u_1^T \Sigma_N u_1, \ldots, u_N^T \Sigma_N u_N) \tag{3.3}$$

From before, if we take $\Omega_N^g(\lambda)$ and $\Theta_N^g(z)$ for $g(\tau) = \tau$, then if we define,

$$\Delta_N(\lambda) = \frac{1}{N} \sum_{i=1}^N \tilde{d}_i \mathbb{1}_{[\lambda_i, \infty)}(\lambda) = \Omega_N^g(\lambda)$$

then based on previous stated result about the convergence of $\Theta_N^g(z)$, Ledoit and Péché (2011) show through verifying with calculus that $\Delta_N(\lambda) \to \Delta(\lambda)$ almost surely for all $\lambda \neq 0$ and

$$\Delta(\lambda) = \Omega^g(\lambda) = \int_{-\infty}^{\lambda} \delta(x) dF(x)$$

where

$$\delta(\lambda) = \begin{cases} \frac{\lambda}{|1 - y^{-1} - y^{-1}\lambda m_F(\lambda)|^2} & \lambda > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3.4}$$

The summary of this is that while the oracle estimates that are optimal, $\tilde{d}_i = u_i^T \Sigma_N u_i$, are completely infeasible, we use $\delta(\lambda_i)$ as the consistent estimate for $\tilde{d}_i$. This is seen by considering,

$$\tilde{d}_i = \lim_{\epsilon \to 0^+} \frac{\Delta_N(\lambda_i + \epsilon) - \Delta_N(\lambda_i - \epsilon)}{F_N(\lambda_i + \epsilon) - F_N(\lambda_i - \epsilon)} \approx \delta(\lambda_i)$$

where the approximation comes from the asymptotic theory. Now, it remains to actually estimate $m_F(z)$ and thus produce a consistent estimate of the shrinkage function but this has in fact been done in Ledoit and Wolf (2015).

3.3 Out-of-Sample Variance Loss

An alternative goal illustrated in Ledoit and Wolf (2017) is to minimize the out-of-sample variance of a portfolio. The loss function is given by,

$$\mathcal{L}(\hat{\Sigma}_N, \Sigma_N, \hat{w}) := \hat{w}^T \Sigma_N \hat{w} \tag{3.5}$$

for some rotation-equivariant estimator $\hat{\Sigma}_N$ given by (3.1) and some estimated portfolio $\hat{w}$.

Ledoit and Wolf (2017) consider $\hat{w}$ as a scaled form of the long-short maximum Sharpe ratio portfolio,

$$\hat{w} = \frac{\sqrt{m^T m}}{m^T \hat{\Sigma}_N^{-1} m} \times \hat{\Sigma}_N^{-1} m. \tag{3.6}$$

where m is the *return predictive signal* aka expected return and is distributed independently of S_N and its distribution is rotation invariant. The scaling $\sqrt{m^T m}$ was chosen so that the portfolio weights are invariant to m. Under this portfolio, the loss function can be expressed as,

$$\mathcal{L}(\hat{\Sigma}_N, \Sigma_N, m) = m^T m \times \frac{m^T \hat{\Sigma}_N^{-1} \Sigma_N \hat{\Sigma}_N^{-1} m}{\left(m^T \hat{\Sigma}_N^{-1} m\right)^2} \tag{3.7}$$

From Lemma 1 in Ledoit and Péché (2011) based on the properties of m,

$$\frac{1}{N} m^T \hat{\Sigma}_N^{-1} m - \frac{1}{N} \mathrm{Tr}(\hat{\Sigma}_N^{-1}) \to 0,$$

almost surely so that they converge together to,

$$\int \frac{1}{\hat{\rho}(\lambda)} dF(\lambda).$$

A similar line of reasoning shows that

$$\frac{1}{N} m^T \hat{\Sigma}_N^{-1} \Sigma_N \hat{\Sigma}_N^{-1} m - \frac{1}{N} \mathrm{Tr}(\hat{\Sigma}_N^{-1} \Sigma_N \hat{\Sigma}_N^{-1}) \to 0$$

almost surely so that they converge together. The limit becomes evident by looking at the theorems for the generalization of the Marchenko-Pastur theorem. Namely,

$$\frac{1}{N} \mathrm{Tr}(\hat{\Sigma}_N^{-1} \Sigma_N \hat{\Sigma}_N^{-1}) = \frac{1}{N} \mathrm{Tr}(U^T \Sigma_N U D^{-2}) = \frac{1}{N} \sum_{i=1}^{N} \frac{u_i^T \Sigma_N u_i}{\hat{\rho}(\lambda_i)^2}$$

But this is basically the same as Δ_N from the previous section, so it turns out that the limit is,

$$\int \frac{\delta(\lambda)}{\hat{\rho}(\lambda)^2} dF(\lambda).$$

These two limit results show that,

$$\mathcal{L}(\hat{\Sigma}_N, \Sigma_N, m) \to \frac{\int \frac{\delta(\lambda)}{\hat{\rho}(\lambda)^2} dF(\lambda)}{\left(\int \frac{1}{\hat{\rho}(\lambda)} dF(\lambda)\right)^2}.$$

Differentiation with respect to $\hat{\rho}(\lambda)$[3] yields the result that the first order condition is satified if and only if,

$$\frac{\hat{\rho}(\lambda)}{\delta(\lambda)} = c$$

where c is an arbitrary constant. Choosing the constant equal to 1 satisfies some consistency results with regards to Traces of the estimator and the population covariance.

The upshot here is that even for this new loss function, the shrinkage is actually the same as the one for the Frobenius norm.

3.4 Cross-Validation for Optimal Nonlinear Shrinkage

In Remark 5.2 of Ledoit and Wolf (2012), a cross-validated estimator is proposed at the suggestion of a referee. We illustrate this estimator here and show how it is improved in Bartz (2016).

3.4.1 Leave-One-Out CV

Let $S^{(k)}$ denote the sample covariance matrix computed from all the observed data, except for the k-th observation x_k. Further let $\lambda_1^{(k)}, \dots, \lambda_N^{(k)}$ and $u_1^{(k)}, \dots, u_N^{(k)}$ denote the system of eigenvalues and eigenvectors of $S^{(k)}$, respectively.

From Sections 3.2 and 3.3, the quantity to compute is $\tilde{d}_i = u_i^T \Sigma_N u_i$, which leads to the cross-validation approximation,

$$\tilde{d}_i \approx \rho^{cv}(\lambda_i) \equiv \frac{1}{T} \sum_{k=1}^{T} ({u_i^{(k)}}^T x_k)^2 \tag{3.8}$$

The motivation comes from the fact that,

$$({u_i^{(k)}}^T x_k)^2 = {u_i^{(k)}}^T x_k x_k^T u_i^{(k)}$$

where x_k is independent of $u_i^{(k)}$ and $\mathbb{E}\left[x_k x_k^T\right] = \Sigma_N$. Therefore,

$$\mathbb{E}\left[\rho^{cv}(\lambda_i)\right] = \mathbb{E}\left[{u_i^{(k)}}^T x_k x_k^T u_i^{(k)}\right] = \mathbb{E}\left[{u_i^{(k)}}^T \Sigma_N u_i^{(k)}\right]$$

The full algorithm is given by Algorithm 1 in the appendix.

3.4.2 K-Fold CV

The standard LOO cross-validation estimator in the previous section is known to have poor properties, as discussed in Bartz (2016). One possible fix is to use K-fold sampling of the data to produce a stabilized estimate.

[3] I tried replicating this result in the Goldilocks paper but I can't get it right

Now we let $S^{(k)}$ denote the sample covariance matrix computed from all the observed data, except for the k-th fold set of observations $X_k \in \mathbb{R}^{T/K \times N}$. Further let $\lambda_1^{(k)}, \ldots, \lambda_N^{(k)}$ and $u_1^{(k)}, \ldots, u_N^{(k)}$ denote the system of eigenvalues and eigenvectors of $S^{(k)}$, respectively.

The K-fold cross-validation estimator is given by,

$$\rho^{kcv}(\lambda_i) \equiv \frac{1}{K} \sum_{k=1}^{K} {u_i^{(k)}}^T S^{(k)} u_i^{(k)}. \tag{3.9}$$

The full algorithm is given by Algorithm 2 in the appendix.

3.4.3 Isotonic Regression and Cross-Validation Estimators

Another easy and well-behaved fix is to apply isotonic regression to produce an improved estimator $\rho^{icv}(\lambda_i)$. This is given by the solution to the optimization problem,

$$\begin{aligned}
\underset{a_1,\ldots,a_N}{\text{minimize}} \quad & \sum_{i=1}^{N} (a_i - \rho^{cv}(\lambda_i))^2 \\
\text{subject to} \quad & a_i > a_{i+1}, \\
& \sum_{i=1}^{N} a_i = \sum_{i=1}^{N} \rho^{cv}(\lambda_i).
\end{aligned}$$

The isotonic regression can also be applied to the K-fold CV estimator that will be called $\rho^{ikcv}(\lambda_i)$.

3.4.4 Performance

Bartz (2016) shows that $\rho^{cv}(\lambda_i)$ is a poor performing estimator. The main issue is instability in the eigenvalues & eigenvectors when dropping a single observation. This is most simply mitigated by using $\rho^{kcv}(\lambda_i)$, which shows huge improvement in Figure 3.1. Applying isotonic regression, whether to the $\rho^{cv}(\lambda_i)$ or $\rho^{kcv}(\lambda_i)$ gives the best results.

4 Optimal Nonlinear Shrinkage for Unbiased Variance Ratios

There are a few problems with the above methodologies that we have run into (and need to solidly verify numerically):

1. The above method produces a covariance matrix that will minimize out-of- sample variance for any portfolio that is colinear with that tangency portfolio. In that sense, it is optimal. But what about another portfolio? What about the minimum variance (global or long-only) portfolio? Is it still the correct shrinkage if you are interested in computing the minimum variance portfolio?
2. Numerical results (need to verify again) indicate the variance ratios for minimum variance portfolio computed from the optimal non-linear shrunk covariance matrix are suboptimal. That is, they are not unbiased around 1. As we will see, the optimal non-linear shrunk covariance matrix should produce good variance ratios for the tangency portfolios (need to verify numerically).

Consider the loss function,

$$\mathcal{R}(\hat{\Sigma}_N, \Sigma_N, \hat{w}_N) := \left(1 - \frac{\hat{w}_N^T \hat{\Sigma}_N \hat{w}_N}{\hat{w}_N^T \Sigma_N \hat{w}_N}\right)^2$$

where $\hat{w}_N$ is some portfolio rule dependent on the covariance matrix and $\hat{\Sigma}_N$ is an estimator.

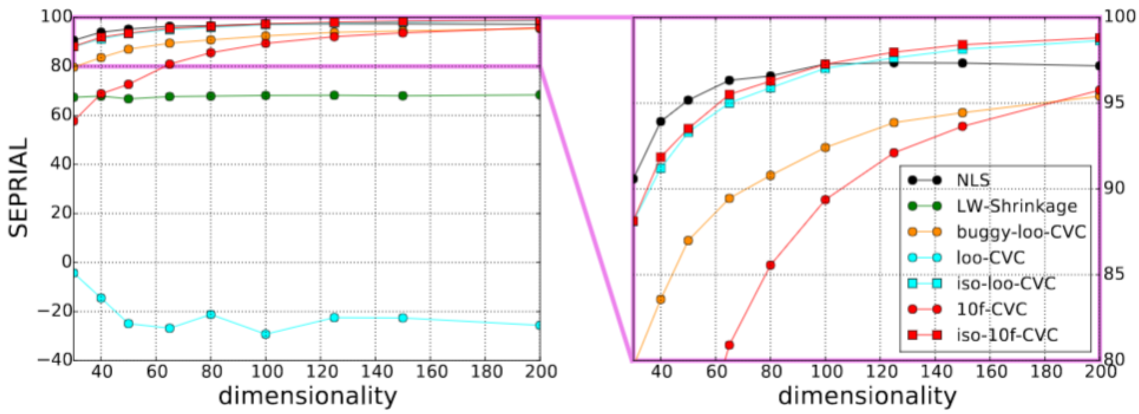

Figure 3.1: From Bartz (2016). Estimation error of cross- validated covariance estimators from Bartz (2016). Averaged over 50 repetitions. *NLS* refers to the LW optimal shrinkage. *LW-Shrinkage* refers to the linear shrinkage of Ledoit and Wolf (2004). *buggy-loo-CVC* refers to an incorrectly implemented leave-one-out cross-validated estimator that appears in Ledoit and Wolf (2012). *loo-CVC* refers to the correctly implemented leave-one-out cross-validated estimator proposed in Ledoit and Wolf (2012) and described in Section 3.4.1. *iso-loo-CVC* refers to the isotonic version of the leave-one-out cross- validated estimator described in 3.4.3. *10f-CVC* refers to the 10-fold cross-validated estimator described in 3.4.2. And *iso-10f-CVC* refers to the 10-fold, isotonic cross-validated estimator described in 3.4.3.

4.1 Maximum Sharpe Portfolio

Theorem 4.1 (Asymptotic Shrinkage for Variance Ratio of Maximum Sharpe Ratio Portfolio)**.** *Under the same setup as Theorem 4.1 of Ledoit and Wolf (2017), the asymptotic nonlinear shrinkage estimator $\delta(\lambda)$ given in* (3.4) *that that minimizes Frobenius loss* (3.2) *and out-of-sample variance in* (3.7) *also minimizes the almost sure limit of the loss function $\mathcal{R}(\hat{\Sigma}_N, \Sigma_N, \hat{w}_N)$ where $\hat{w}_N$ is the maximum Sharpe ratio portfolio given by* (3.6)*.*

Sketch:

If we consider the same maximum Sharpe ratio portfolios as before, then,

$$\mathcal{R}(\hat{\Sigma}_N, \Sigma_N, \hat{w}_N) := \left(1 - \frac{m^T \hat{\Sigma}_N^{-1} \Sigma_N \hat{\Sigma}_N^{-1} m}{m^T \hat{\Sigma}_N^{-1} m}\right)^2.$$

From the same theory as above,

$$\mathcal{R}(\hat{\Sigma}_N, \Sigma_N, \hat{w}_N) \to \left(1 - \frac{\int \frac{\delta(\lambda)}{\hat{\rho}(\lambda)^2} dF(\lambda)}{\int \frac{1}{\hat{\rho}(\lambda)} dF(\lambda)}\right)^2,$$

which is minimized and equal to 0 if $\hat{\rho}(\lambda) = \delta(\lambda)$. Moreover, this indicates the optimal non-linear shrinkage estimator should provide good variance forecasting for the maximum Sharpe ratio portfolio.

4.2 Global Minimum Variance Portfolio

Now we consider the global minimum variance portfolio,

$$\hat{w}_N = \frac{\hat{\Sigma}_N^{-1} \mathbb{1}}{\mathbb{1}^T \hat{\Sigma}_N^{-1} \mathbb{1}}.$$

The loss function is given by,

$$\mathcal{R}(\hat{\Sigma}_N, \Sigma_N, \hat{w}_N) = \left(1 - \frac{\mathbb{1}^T \hat{\Sigma}_N^{-1} \Sigma_N \hat{\Sigma}_N^{-1} \mathbb{1}}{\mathbb{1}^T \hat{\Sigma}_N^{-1} \mathbb{1}}\right)^2.$$

Letting $\alpha = U^T \mathbb{1}$, the above loss can be written as,

$$\mathcal{R}(\hat{\Sigma}_N, \Sigma_N, \hat{w}_N) = \left(1 - \frac{\alpha^T D^{-1} U^T \Sigma_N U D^{-1} \alpha}{\alpha^T D^{-1} \alpha}\right)^2.$$

4.2.1 Oracle Optimal Shrinkage

Theorem 4.2 (Oracle Shrinkage for Variance Ratio of Minimum Variance Portfolio)**.** *Let $X \in \mathbb{R}^{T \times N}$ denote T i.i.d. draws from an N-dimensional distribution with covariance Σ_N. Define the sample covariance matrix as $S = \frac{1}{T} X^T X$ with eigendecomposition $S = U \Lambda U^T$.*

Let the class of rotation invariant covariance estimators be defined by,

$$\hat{\Sigma}_N = U D U^T$$

where $D = \mathrm{Diag}(d_1, \ldots, d_N)$ are the chosen eigenvalues for the estimator.

Finally, define the plug-in estimator for global minimum variance portfolio as

$$\hat{w}_N = \frac{\hat{\Sigma}_N^{-1} \mathbb{1}}{\mathbb{1}^T \hat{\Sigma}_N^{-1} \mathbb{1}}.$$

The optimal finite sample oracle shrinkage estimator $\hat{\Sigma}_N^* = U D^* U^T$ with respect to the loss function $\mathcal{R}(\hat{\Sigma}_N, \Sigma_N, \hat{w}_N)$ is given by $D_{ii}^* = d_i^* = 1/z_i$ where $z \in \mathbb{R}^N$ is the solution to the linear system,

$$CAz = \alpha.$$

for $C = U^T \Sigma_N U$, $\alpha = U^T \mathbb{1}$, and $A = \mathrm{Diag}(\alpha_1, \ldots, \alpha_N)$.

This estimator also minimizes out-of-sample variance,

$$\mathcal{L}(\hat{\Sigma}_N, \Sigma_N, \hat{w}_N) = \hat{w}_N^T \Sigma_N \hat{w}_N = \frac{\mathbb{1}^T \hat{\Sigma}_N^{-1} \Sigma_N \hat{\Sigma}_N^{-1} \mathbb{1}}{\left(\mathbb{1}^T \hat{\Sigma}_N^{-1} \mathbb{1}\right)^2}$$

Sketch:

From the preceding section, we can write,

$$\alpha^T D^{-1} \alpha = \mathrm{Tr}(D^{-1} \alpha \alpha^T) = \sum_{i=1}^{N} \frac{\alpha_i^2}{d_i}$$

and,

$$\alpha^T D^{-1} U^T \Sigma_N U D^{-1} \alpha = \mathrm{Tr}(D^{-1} U^T \Sigma_N U D^{-1} \alpha \alpha^T) = \sum_{i=1}^{N} \frac{\alpha_i}{d_i} \sum_{j=1}^{N} \frac{C_{ij} \alpha_j}{d_j}$$

where $C = U^T \Sigma_N U$ and $C_{ij} = u_i^T \Sigma_N u_j$.

If we can choose d_i to equate α_i and $\frac{C_{ji} \alpha_j}{d_j}$, then the loss function $\mathcal{R}$ will be minimized and equal to 0. This can be achieved by solving a linear system.

Let,

$$A = \mathrm{Diag}(\alpha_1, \ldots, \alpha_N)$$

Then $d_i = 1/z_i$ where $z \in \mathbb{R}^N$ is the solution to,

$$CAz = \alpha.$$

This gives

$$z = A^{-1}U^T\Sigma_N^{-1}U\alpha = A^{-1}U^T\Sigma_N^{-1}\mathbb{1}$$

It turns out this estimator also minimizes out-of-sample variance for the global minimum variance portfolio.

The out-of-sample variance simplifies to,

$$\mathcal{L}(\hat{\Sigma}_N, \Sigma_N, \hat{w}_N) = \frac{\alpha^T D^{-1} U^T \Sigma_N U D^{-1}\alpha}{(\alpha^T D^{-1}\alpha)^2}$$

Expanding terms as before and differentiating with respect to the i-th entry in D gives,

$$\frac{\partial \mathcal{L}}{\partial d_i} = -2\frac{\frac{\alpha_i}{d_i^2}\sum_{j=1}^N \frac{C_{ij}\alpha_j}{d_j}}{Q_2^2} + 2\frac{Q_1}{Q_2^3}\frac{\alpha_i^2}{d_i^2}$$

where $Q_1 = \alpha^T D^{-1}U^T\Sigma_N U D^{-1}\alpha$ and $Q_2 = \alpha^T D^{-1}\alpha$.

The first-order condition $0 = \frac{\partial \mathcal{L}}{\partial d_i}$ reduces to,

$$Q_2 \sum_{j=1}^N \frac{C_{ij}\alpha_j}{d_j} = Q_1\alpha_i.$$

If d_i derived from the solution to $CAz = \alpha$, then $Q_2 = Q_1$ and the first-order condition is satisfied.

To see this is a minimum, we rely on the chain-rule. Let $F(w) = w^T\Sigma_N w$ where w is the plug-in estimator for global minimum variance portfolio $\hat{w}_N$ and is therefore a function of D.

By the chain rule,

$$\nabla_D^2 F(w) = \nabla_D^T w \cdot \nabla_w^2 F(w) \cdot (\nabla_D^T w)^T + \nabla_D^2 w \cdot \nabla_w F(w)$$

where,

$$\begin{aligned}
(\nabla_D^2 F(w))_{ij} &= \frac{\partial^2 F(w)}{\partial d_i \partial d_j}, \\
(\nabla_w F(w))_i &= \frac{\partial F(w)}{\partial w_i}, \\
(\nabla_w^2 F(w))_{ij} &= \frac{\partial^2 F(w)}{\partial w_i \partial w_j}, \\
(\nabla_D^T w)_{ij} &= \frac{\partial w_i}{\partial d_j}, \\
(\nabla_D^2 w)_{ijk} &= \frac{\partial^2 w_i}{\partial d_j \partial d_k}.
\end{aligned}$$

The first-order condition means $\nabla_w F(w) = 0$ and thus since $\nabla_w^2 F(w) = 2\Sigma_N \succ 0$, $\nabla_D^2 F(w) \succ 0$. This verifies that the oracle estimator for D^* minimizes out-of-sample variance.

4.2.2 Asymptotically Optimal Shrinkage

From the above, it may be possible to put $\mathcal{R}(\hat{\Sigma}_N, \Sigma_N, \hat{w}_N)$ and the optimal oracle shrinkage in the RMT asymptotic framework and derive the limiting results that would indicate how one should choose D. One possible work that may have related results is Mestre (2008) which potentially derive the asymptotic results for these kinds of functionals. More needs to be done to study that work and check if it fits the framework.

4.2.3 Proposed Cross-Validation Estimators

The solution z to the linear system,

$$CAz = \alpha,$$

has a few issues with it. First, despite being an oracle estimator like the optimal values in (3.3), the resulting values are very noisy. The optimal values in (3.3) are also noisy so this is not a problem exclusive to the new values. Second, there are no guarantees on positivity or monotonicty of the values z or d.

The following are proposed estimators that seek to remedy these problems and serve as *bona fide* estimators.

1. *MinVar Leave-One-Out Cross-Validation with Isotonic Regression*

   See Algorithm 3 in the appendix for more detailed implementation.

   The first proposed estimator is a simple cross-validation estimator using Leave- One-Out cross-validation to generate multiple solutions to the above linear system to stabilize estimates of the eigenvalues.

   First we make the following definitions:

$$\alpha^{(k)} = {U^{(k)}}^T \mathbb{1},$$
$$C^{(k)} = {U^{(k)}}^T x_k x_k^T U^{(k)},$$
$$A^{(k)} = \text{Diag}(\alpha_1^{(k)}, \ldots, \alpha_N^{(k)}),$$

   where $\cdot^{(k)}$ denotes a quantity computed when leaving out the k-th observation, x_k.

   Then compute a set of intermediate values

$$\bar{d}_i = \frac{1}{T} \sum_{k=1}^{T} \frac{1}{z_i^{(k)}}$$

   where $z^{(k)}$ is the solution to,

$$C^{(k)} A^{(k)} z^{(k)} = \alpha^{(k)}.$$

   Finally, the final estimator values are obtained following applying isotonic regression such that,

$$\hat{d}^* = \text{Isotonic}(\bar{d}).$$

   Variants of this method:

   - Use a median computation instead of an averaging for computing $\bar{d}$
   - Non-negativity constraint for $z^{(k)}$, ie. solve the linear system as a non-negative least squares problem.
   - L^2 regularization on $z^{(k)}$.

2. *MinVar K-Fold Cross-Validation with Isotonic Regression*

   A variant to the Leave-One-Out estimator using K-Fold cross-validation instead. See Algorithm 4 in the appendix.

3. *MinVar Leave-One-Out Joint Cross-Validation with Isotonic Regression*

   See Algorithm 5 in the appendix for a detailed implementation.

   Averaging of the individual cross-validation values serves to denoise the raw values to produce better performance before applying Isotonic regression. The solution z could be obtained from solving a larger least squares problem.

   With the definitions as the basic Leave-One-Out method,

$$\alpha^{(k)} = {U^{(k)}}^T \mathbb{1},$$
$$C^{(k)} = {U^{(k)}}^T x_k x_k^T U^{(k)},$$
$$A^{(k)} = \mathrm{Diag}(\alpha_1^{(k)}, \ldots, \alpha_N^{(k)}),$$

   the least squares problem is defined as,

$$\text{minimize} \quad \textstyle\sum_{k=1}^T \|C^{(k)} A^{(k)} z - \alpha^{(k)}\|_2^2, \tag{4.1}$$

   so that $\bar{d}_i = \frac{1}{z_i}$. The final estimator is obtained via Isotonic regression,

$$\hat{d}^* = \mathrm{Isotonic}(\bar{d}).$$

   The solution to (4.1) can also be expressed as the solution to a quadratic program given by,

$$\text{minimize} \quad (1/2) z^T P z + q^T z, \tag{4.2}$$

   where

$$P = \sum_{k=1}^T A^{(k)} {C^{(k)}}^T C^{(k)} A^{(k)}, \quad q = -\sum_{k=1}^T A^{(k)} {C^{(k)}}^T \alpha^{(k)},$$

   which reduces to solving the system $Pz = -q$.

   Variants of this method:

   - Non-negativity constraint for $z^{(k)}$, ie. solve the least squares problem instead as a non-negative least squares problem.
   - Instead of using Isotonic Regression, the monotonicity constraint could be directly applied when solving z.
   - L^2 regularization on $z^{(k)}$.
   - Use a robust least squares computation (this is too advanced to consider for now).

4. *MinVar K-Fold Joint Cross-Validation with Isotonic Regression*

   A variant to the Leave-One-Out Joint Cross-Validation estimator using K-Fold cross-validation instead. See Algorithm 6 in the appendix.

5 To Address

5.1 Metrics

- Out-of-Sample Variance
- Variance forecast
- Bias statistics: $r_w/(w\Sigma w)^{1/2}$

5.2 Portfolios

- Global minimum variance
- Long-only minimum variance
- Equally weighted
- Maximum Sharpe Ratio
- Concentrated

5.3 Models

- Covariances: simple factor, broad-narrow factor, identity, known H with random U or identity U
- Returns: equivariant

5.4 Questions to Address

- What is the baseline performance of the classical sample covariance estimator?
- Does long-only outperform out-of-sample var vs global min var?
- When substituting in true eigenvectors and/or true eigenvalues, which is most important in fixing issues of estimation? What does that tell us about covariance estimation?
- How does SLR improve when model is big factor model? Where are its deficiencies? With respect to which metrics or portfolios?
- How does non-linear shrinkage do? Does the numerical verify the theory? How is the var ratio for the Max Sharpe and min var portfolios? Compare out of sample variance to portfolios along mean-variance frontier (true and estimated)
- Can we replicate the success/failure of non-linear shrinkage on the global min var and the long-only min var?
- For the oracle estimator for best var ratio performance: what portfolio does that give? Where does it fall on mean-variance frontier (true or estimated)? Does it give unbiased var ratio? Does it give good var performance too despite being designed for var ratio?
- Does the cross-validated optimal var ratio estimator work?

6 Appendix

6.1 Algorithms

References

Bartz, D. (2016). Cross-validation based nonlinear shrinkage. *arXiv preprint arXiv:1611.00798*.

Ledoit, O. and Péché, S. (2011). Eigenvectors of some large sample covariance matrix ensembles. *Probability Theory and Related Fields*, 151(1):233–264.

Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2):365–411.

Algorithm 1 Leave-One-Out Cross Validation for Non-Linear Shrinkage

Require: $X = [x_1, \ldots, x_T]^T \in \mathbb{R}^{T\times N}$
Require: $\Sigma_N = \mathrm{Cov}(X)$
procedure LOO-CV(X)
  $S \leftarrow \frac{1}{T} X^T X$
  $U \leftarrow [u_1, \ldots, u_N] \in \mathbb{R}^{N\times N}, \quad \Lambda \leftarrow \mathrm{Diag}(\lambda_1, \ldots, \lambda_N), \quad \text{s.t.} \quad S = U\Lambda U^T$
  $\tilde{d} \leftarrow [0, \ldots, 0]^T \in \mathbb{R}^N$
  while $k \leq T$ **do**
    $X_{-k} \leftarrow [x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_T]^T$
    $S^{(k)} \leftarrow \frac{1}{T-1} X_{-k}^T X_{-k}$
    $U^{(k)} \leftarrow [u_1^{(k)}, \ldots, u_N^{(k)}] \in \mathbb{R}^{N\times N}, \quad \Lambda^{(k)} \leftarrow \mathrm{Diag}(\lambda_1^{(k)}, \ldots, \lambda_N^{(k)}), \quad \text{s.t.} \quad S^{(k)} = U^{(k)}\Lambda^{(k)}{U^{(k)}}^T$
    $\tilde{d}_i \leftarrow \tilde{d}_i + \frac{1}{T}({U^{(k)}}^T x_k)_i^2, \quad i = 1, \ldots, N$
    $k \leftarrow k + 1$
  end while
  $\hat{\Sigma}_N \leftarrow UDU^T, \quad \text{where} \quad D = \mathrm{Diag}(\tilde{d}_1, \ldots, \tilde{d}_N)$
  return $\hat{\Sigma}_N$ ▷ The LOO-CV non-linear shrinkage estimator for Σ_N.
end procedure

Algorithm 2 K-Fold Cross Validation for Non-Linear Shrinkage

Require: $X = [x_1, \ldots, x_T]^T \in \mathbb{R}^{T\times N}$
Require: $\Sigma_N = \mathrm{Cov}(X)$
Require: $K \in \mathbb{Z}, \quad K > 1$
procedure K-FOLD-CV(X)
  $S \leftarrow \frac{1}{T} X^T X$
  $U \leftarrow [u_1, \ldots, u_N] \in \mathbb{R}^{N\times N}, \quad \Lambda \leftarrow \mathrm{Diag}(\lambda_1, \ldots, \lambda_N), \quad \text{s.t.} \quad S = U\Lambda U^T$
  $m \leftarrow T/K$
  $\mathcal{K} \leftarrow \{(1, \ldots m), \ldots, (T-m+1, \ldots, T)\}$
  $\tilde{d} \leftarrow [0, \ldots, 0]^T \in \mathbb{R}^N$
  $k \leftarrow 1$
  while $k \leq K$ **do**
    $X_{-k} \leftarrow [x_{j_1}, \ldots, x_{j_{(T-m)}}]^T \in \mathbb{R}^{(T-m)\times N}, \quad j_i \notin \mathcal{K}_k$
    $X_k \leftarrow [x_{\ell_1}, \ldots, x_{\ell_m}]^T \in \mathbb{R}^{m\times N}, \quad \ell_i \in \mathcal{K}_k$
    $S^{(k)} \leftarrow \frac{1}{T-m} X_{-k}^T X_{-k}$
    $U^{(k)} \leftarrow [u_1^{(k)}, \ldots, u_N^{(k)}] \in \mathbb{R}^{N\times N}, \quad \Lambda^{(k)} \leftarrow \mathrm{Diag}(\lambda_1^{(k)}, \ldots, \lambda_N^{(k)}), \quad \text{s.t.} \quad S^{(k)} = U^{(k)}\Lambda^{(k)}{U^{(k)}}^T$
    $\tilde{d}_i \leftarrow \tilde{d}_i + \frac{1}{T}{u_i^{(k)}}^T(\frac{1}{m}X_k^T X_k)u_i^{(k)} \quad i = 1, \ldots, N$
    $k \leftarrow k + 1$
  end while
  $\hat{\Sigma}_N \leftarrow UDU^T, \quad \text{where} \quad D = \mathrm{Diag}(\tilde{d}_1, \ldots, \tilde{d}_N)$
  return $\hat{\Sigma}_N$ ▷ The K-Fold-CV non-linear shrinkage estimator for Σ_N.
end procedure

Algorithm 3 MinVar Leave-One-Out Cross-Validation with Isotonic Regression

Require: $X = [x_1, \dots, x_T]^T \in \mathbb{R}^{T \times N}$
Require: $\Sigma_N = \mathrm{Cov}(X)$
procedure MinVar-LOO-CV(X)
  $S \leftarrow \frac{1}{T} X^T X$
  $U \leftarrow [u_1, \dots, u_N] \in \mathbb{R}^{N \times N}, \quad \Lambda \leftarrow \mathrm{Diag}(\lambda_1, \dots, \lambda_N), \quad \text{s.t.} \quad S = U \Lambda U^T$
  $\bar{d} \leftarrow [0, \dots, 0]^T \in \mathbb{R}^N$
  while $k \leq T$ **do**
    $X_{-k} \leftarrow [x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_T]^T$
    $S^{(k)} \leftarrow \frac{1}{T-1} X_{-k}^T X_{-k}$
    $U^{(k)} \leftarrow [u_1^{(k)}, \dots, u_N^{(k)}] \in \mathbb{R}^{N \times N}, \quad \Lambda^{(k)} \leftarrow \mathrm{Diag}(\lambda_1^{(k)}, \dots, \lambda_N^{(k)}), \quad \text{s.t.} \quad S^{(k)} = U^{(k)} \Lambda^{(k)} {U^{(k)}}^T$
    $\alpha^{(k)} \leftarrow {U^{(k)}}^T \mathbb{1}$
    $C^{(k)} \leftarrow {U^{(k)}}^T x_k x_k^T U^{(k)}$
    $A^{(k)} \leftarrow \mathrm{Diag}(\alpha_1^{(k)}, \dots, \alpha_N^{(k)})$
    $\bar{d}_i \leftarrow \bar{d}_i + \frac{1}{T} \frac{1}{z_i^{(k)}}, \quad i = 1, \dots, N, \quad C^{(k)} A^{(k)} z^{(k)} = \alpha^{(k)}.$
    $k \leftarrow k + 1$
  end while
  $\hat{d}^* \leftarrow \mathrm{Isotonic}(\bar{d}).$ ▷ Isotonic Regression applied to $\bar{d}$
  $\hat{\Sigma}_N \leftarrow U D U^T, \quad \text{where} \quad D = \mathrm{Diag}(\hat{d}_1^*, \dots, \hat{d}_N^*)$
  return $\hat{\Sigma}_N$ ▷ The MinVar LOO-CV non-linear shrinkage estimator for Σ_N.
end procedure

Algorithm 4 MinVar K-Fold Cross-Validation with Isotonic Regression

Require: $X = [x_1, \dots, x_T]^T \in \mathbb{R}^{T \times N}$
Require: $\Sigma_N = \mathrm{Cov}(X)$
Require: $K \in \mathbb{Z}, \quad K > 1$
procedure MinVar-K-Fold-CV(X)
  $S \leftarrow \frac{1}{T} X^T X$
  $U \leftarrow [u_1, \dots, u_N] \in \mathbb{R}^{N \times N}, \quad \Lambda \leftarrow \mathrm{Diag}(\lambda_1, \dots, \lambda_N), \quad \text{s.t.} \quad S = U \Lambda U^T$
  $m \leftarrow T/K$
  $\mathcal{K} \leftarrow \{(1, \dots m), \dots, (T - m + 1, \dots, T)\}$
  $\bar{d} \leftarrow [0, \dots, 0]^T \in \mathbb{R}^N$
  $k \leftarrow 1$
  while $k \leq K$ **do**
    $X_{-k} \leftarrow [x_{j_1}, \dots, x_{j_{(T-m)}}]^T \in \mathbb{R}^{(T-m) \times N}, \quad j_i \notin \mathcal{K}_k$
    $X_k \leftarrow [x_{\ell_1}, \dots, x_{\ell_m}]^T \in \mathbb{R}^{m \times N}, \quad \ell_i \in \mathcal{K}_k$
    $S^{(k)} \leftarrow \frac{1}{T-m} X_{-k}^T X_{-k}$
    $U^{(k)} \leftarrow [u_1^{(k)}, \dots, u_N^{(k)}] \in \mathbb{R}^{N \times N}, \quad \Lambda^{(k)} \leftarrow \mathrm{Diag}(\lambda_1^{(k)}, \dots, \lambda_N^{(k)}), \quad \text{s.t.} \quad S^{(k)} = U^{(k)} \Lambda^{(k)} {U^{(k)}}^T$
    $\alpha^{(k)} \leftarrow {U^{(k)}}^T \mathbb{1}$
    $C^{(k)} \leftarrow {U^{(k)}}^T (\frac{1}{m} X_k^T X_k) U^{(k)}$
    $A^{(k)} \leftarrow \mathrm{Diag}(\alpha_1^{(k)}, \dots, \alpha_N^{(k)})$
    $\bar{d}_i \leftarrow \bar{d}_i + \frac{1}{K} \frac{1}{z_i^{(k)}}, \quad i = 1, \dots, N, \quad C^{(k)} A^{(k)} z^{(k)} = \alpha^{(k)}.$
    $k \leftarrow k + 1$
  end while
  $\hat{d}^* \leftarrow \mathrm{Isotonic}(\bar{d}).$ ▷ Isotonic Regression applied to $\bar{d}$
  $\hat{\Sigma}_N \leftarrow U D U^T, \quad \text{where} \quad D = \mathrm{Diag}(\hat{d}_1^*, \dots, \hat{d}_N^*)$
  return $\hat{\Sigma}_N$ ▷ The MinVar K-Fold-CV non-linear shrinkage estimator for Σ_N.
end procedure

Algorithm 5 MinVar Leave-One-Out Joint Cross-Validation with Isotonic Regression

Require: $X = [x_1, \ldots, x_T]^T \in \mathbb{R}^{T \times N}$
Require: $\Sigma_N = \mathrm{Cov}(X)$
procedure MinVar-LOO-Joint-CV(X)
  $S \leftarrow \frac{1}{T} X^T X$
  $U \leftarrow [u_1, \ldots, u_N] \in \mathbb{R}^{N \times N}, \quad \Lambda \leftarrow \mathrm{Diag}(\lambda_1, \ldots, \lambda_N), \quad \text{s.t.} \quad S = U \Lambda U^T$
  $P \leftarrow 0 \in \mathbb{R}^{N \times N}$
  $q \leftarrow 0 \in \mathbb{R}^N$
  while $k \leq T$ **do**
    $X_{-k} \leftarrow [x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_T]^T$
    $S^{(k)} \leftarrow \frac{1}{T-1} X_{-k}^T X_{-k}$
    $U^{(k)} \leftarrow [u_1^{(k)}, \ldots, u_N^{(k)}] \in \mathbb{R}^{N \times N}, \quad \Lambda^{(k)} \leftarrow \mathrm{Diag}(\lambda_1^{(k)}, \ldots, \lambda_N^{(k)}), \quad \text{s.t.} \quad S^{(k)} = U^{(k)} \Lambda^{(k)} {U^{(k)}}^T$
    $\alpha^{(k)} \leftarrow {U^{(k)}}^T \mathbb{1}$
    $C^{(k)} \leftarrow {U^{(k)}}^T x_k x_k^T U^{(k)}$
    $A^{(k)} \leftarrow \mathrm{Diag}(\alpha_1^{(k)}, \ldots, \alpha_N^{(k)})$
    $P \leftarrow P + A^{(k)} {C^{(k)}}^T C^{(k)} A^{(k)}$
    $q \leftarrow q + A^{(k)} {C^{(k)}}^T \alpha^{(k)}$
    $k \leftarrow k + 1$
  end while
  $\bar{d}_i \leftarrow \frac{1}{z_i}, \quad \text{s.t.} \quad Pz = -q$
  $\hat{d}^* \leftarrow \mathrm{Isotonic}(\bar{d})$. ▷ Isotonic Regression applied to $\bar{d}$
  $\hat{\Sigma}_N \leftarrow UDU^T, \quad \text{where} \quad D = \mathrm{Diag}(\hat{d}_1^*, \ldots, \hat{d}_N^*)$
  return $\hat{\Sigma}_N$ ▷ The MinVar LOO-Joint-CV non-linear shrinkage estimator for Σ_N.
end procedure

Ledoit, O. and Wolf, M. (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. *The Annals of Statistics*, 40(2):1024–1060.

Ledoit, O. and Wolf, M. (2015). Spectrum estimation: A unified framework for covariance matrix estimation and pca in large dimensions. *Journal of Multivariate Analysis*, 139:360–384.

Ledoit, O. and Wolf, M. (2017). Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets goldilocks.

Mestre, X. (2008). Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates. *IEEE Transactions on Information Theory*, 54(11):5113–5129.

Algorithm 6 MinVar K-Fold Joint Cross-Validation with Isotonic Regression

Require: $X = [x_1, \dots, x_T]^T \in \mathbb{R}^{T \times N}$
Require: $\Sigma_N = \mathrm{Cov}(X)$
Require: $K \in \mathbb{Z}, \quad K > 1$

procedure MinVar-K-Fold-CV(X)
  $S \leftarrow \frac{1}{T} X^T X$
  $U \leftarrow [u_1, \dots, u_N] \in \mathbb{R}^{N \times N}, \quad \Lambda \leftarrow \mathrm{Diag}(\lambda_1, \dots, \lambda_N), \quad \text{s.t.} \quad S = U \Lambda U^T$
  $m \leftarrow T/K$
  $\mathcal{K} \leftarrow \{(1, \dots m), \dots, (T - m + 1, \dots, T)\}$
  $P \leftarrow 0 \in \mathbb{R}^{N \times N}$
  $q \leftarrow 0 \in \mathbb{R}^N$
  $k \leftarrow 1$
  while $k \leq K$ **do**
    $X_{-k} \leftarrow [x_{j_1}, \dots, x_{j_{(T-m)}}]^T \in \mathbb{R}^{(T-m) \times N}, \quad j_i \notin \mathcal{K}_k$
    $X_k \leftarrow [x_{\ell_1}, \dots, x_{\ell_m}]^T \in \mathbb{R}^{m \times N}, \quad \ell_i \in \mathcal{K}_k$
    $S^{(k)} \leftarrow \frac{1}{T-m} X_{-k}^T X_{-k}$
    $U^{(k)} \leftarrow [u_1^{(k)}, \dots, u_N^{(k)}] \in \mathbb{R}^{N \times N}, \quad \Lambda^{(k)} \leftarrow \mathrm{Diag}(\lambda_1^{(k)}, \dots, \lambda_N^{(k)}), \quad \text{s.t.} \quad S^{(k)} = U^{(k)} \Lambda^{(k)} {U^{(k)}}^T$
    $\alpha^{(k)} \leftarrow {U^{(k)}}^T \mathbb{1}$
    $C^{(k)} \leftarrow {U^{(k)}}^T (\frac{1}{m} X_k^T X_k) U^{(k)}$
    $A^{(k)} \leftarrow \mathrm{Diag}(\alpha_1^{(k)}, \dots, \alpha_N^{(k)})$
    $P \leftarrow P + A^{(k)} {C^{(k)}}^T C^{(k)} A^{(k)}$
    $q \leftarrow q + A^{(k)} {C^{(k)}}^T \alpha^{(k)}$
    $k \leftarrow k + 1$
  end while
  $\bar{d}_i \leftarrow \frac{1}{z_i}, \quad \text{s.t.} \quad Pz = -q$
  $\hat{d}^* \leftarrow \mathrm{Isotonic}(\bar{d})$. ▷ Isotonic Regression applied to $\bar{d}$
  $\hat{\Sigma}_N \leftarrow U D U^T, \quad \text{where} \quad D = \mathrm{Diag}(\hat{d}_1^*, \dots, \hat{d}_N^*)$
  return $\hat{\Sigma}_N$ ▷ The MinVar K-Fold-Joint-CV non-linear shrinkage estimator for Σ_N.
end procedure