Constrained NLS estimator

Miron Ivanov

September 19, 2017

1 Estimators

We applied four different estimators :

1. Sample estimator
2. Ledoit and Wolf estimator(LW)
3. Min-Var K-Fold Cross-Validation Estimator
4. Min-Var Constrained Estimator

Note that the last two of them have a different loss function: they are specifically constructed to estimate optimal shrinkage values for minimum variance portfolio. While LW estimator has a loss function that is designed for max Sharpe ratio portfolio or - as concluded in Ledoit and Wolf (2017a) - for Frobenius norm minimization.

Following the notation in Papanicolaou (2017) we define a data matrix with i.i.d random variables:

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

with T being the number of observations and N being the number of features. Denote the population covariance matrix of X as Σ_N with eigenvalues $\lambda_i : i = 1, ..., N$ lying on the main diagonal of D and a matrix of N eigenvectors V. Thus:

$$\Sigma_N = VDV^\intercal, V = [v_1, ..., v_N], D = Diag(\lambda_1, ..., \lambda_N)$$

Also, consider the following class of estimators :

$$\hat{\Sigma} = U\hat{D}^* U^\intercal$$

where $\hat{\Sigma}$ is estimated population covariance matrix. We would like to find a matrix of optimal shrinkage eigenvalues

$$\hat{D}^* = Diag(d_1^*, ..., d_N^*)$$

keeping the matrix of sample eigenvectors U constant such that the respective loss function $f(\Sigma_N, \hat{\Sigma}_N)$ (different for every class of estimators) is minimized.

1.1 Ledoit and Wolf estimator(LW)

We are using a bona fide nonlinear shrinkage(NLS) estimator introduced by Ledoit and Péché (2011). The equation for optimal estimator is:

$$\hat{d^*} = \delta(\hat{\lambda}_i) = \begin{cases} \frac{\hat{\lambda}_i}{|1-y^{-1}-y^{-1}\hat{\lambda}_i s(\hat{\lambda}_i)|^2} & \hat{\lambda}_i > 0, \\ 0 & \text{otherwise.} \end{cases}$$

where $y = T/N$ and $s(\hat{\lambda}_i)$ denotes Stieltjes transform of the $i-th$ sample eigenvalue $\hat{\lambda}_i : i = 1, ..., N$. As proved by Ledoit and Wolf (2017a) the estimator above is optimal for both Frobenius-norm-based loss function

$$\mathcal{L}(\Sigma_N, \hat{\Sigma}_N) := \frac{1}{N}||\Sigma_N - \hat{\Sigma}_N||_F^2$$

and loss function based on out-of-sample variance

$$\mathcal{L}(\Sigma_N, \hat{\Sigma}_N) := m^\intercal m \times \frac{m^\intercal \hat{\Sigma}_N^{-1} \Sigma_N \hat{\Sigma}_N^{-1} m}{(m^\intercal \hat{\Sigma}_N^{-1} m)^2}$$

with m being expected returns. More specifically, we are using a numerical approximation of $s(\hat{\lambda}_i)$ described in Ledoit and Wolf (2017b) and implemented in R package nlshrink. Note that LW estimator is not optimal for minimum variance portfolio.

1.2 Sample estimator

Sample estimator refers to sample covariance matrix that is taken as is, without transforming the eigenvalues:

$$\hat{\Sigma}_N = S_N = \frac{1}{T} X_N^\intercal X_N = U\hat{D}U^\intercal$$

where $\hat{D} = Diag(\hat{\lambda}_1, ..., \hat{\lambda}_N)$.

1.3 Min-Var K-Fold CV Estimator

K-Fold Cross validation(CV) covariance matrix estimator was first introduced by Bartz (2016) as a fast alternative to LW estimator:

$$\hat{d}_i^* = \frac{1}{K} \sum_{k=1}^{K} u_i^{(k)\intercal} X_k X_k^\intercal u_i^{(k)}$$

here X_k contains the observations from the $k-th$ fold and $u_i^{(k)}$ is the $i-th$ eigenvector of the sample covariance matrix computed on the data with the $k-th$ fold removed. However, the corresponding loss function of this estimator is the same as the one for LW. Which does not fully suit our needs since we require an estimator for minimum variance portfolio. The adapted bona fide version of the estimator was initially derived by Papanicolaou (2017) for the case of the minimum variance portfolio loss function:

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

Following the same notation for k subscripts as above, let

$$\begin{aligned} \alpha^{(k)} &= {U^{(k)}}^T \mathbb{1}, \\ C^{(k)} &= U^{(k)\intercal} X_k X_k^\intercal U^{(k)}, \\ A^{(k)} &= \mathrm{Diag}(\alpha_1^{(k)}, \ldots, \alpha_N^{(k)}), \end{aligned}$$

Find the solution z of the system

$$Pz = -q$$

where

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

then the final estimator of eigenvalues might be expressed as the result of applying the isotonic regression (see Bartz (2016)) to the inverse of z values:

$$\begin{aligned} \bar{d}_i &= 1/z_i, \\ \hat{d}^* &= Isotonic(\bar{d}) \end{aligned}$$

Isotonic regression is helpful here because the values of z are not guaranteed to be monotonically increasing. It is a nice trick to mitigate this problem.

1.4 Min-Var Constrained Estimator

Unfortunately, nonmonotonicity is not the only issue with the solutions of the system above. z does not necessarily have to be non-negative even after we apply isotonic regression. In addition, $\hat{d}^*$ appear to be quite noisy as compared to its max Sharpe portfolio counterpart. Reasons for the former are not quite intuitive and are related to the fact that we are targeting the Min-Var portfolio which has the property of error maximization during variance minimization procedure. We could remediate these problems by imposing several constrains on the value of z. Among those are:

1. Non-negativity constraint. We simply solve the linear system above as Non-Negative Least Squares (NNLS) problem, bounding the values of z by zero from below.

2. Trace constraint. We could use the property that the sum of matrix eigenvalues equals to the trace of the matrix and that the eigenvalues of the matrix inverse equal to the inverse of eigenvalues of the original matrix. Thus, with sample covariance matrix being an estimator, we write:

$$\sum_{i=1}^{N} z_i = tr(S_N^{-1})$$

Alternatively, we could set the sum of z directly to the trace of $\hat{\Sigma}^{-1}$:

$$\sum_{i=1}^{N} z_i = tr(U\hat{D}^{-1}U^{\intercal})$$

where $\hat{D}^{-1} = Diag(z_1, ..., z_N)$

3. Monotonicity constraint. Something that is more intuitive is to put the isotonic regression procedure directly into convex optimization by requiring

$$Gz >= 0$$

where $G \in \mathbb{R}^{(N-1)\times N}$ such that $G_{ii} = 1$ and $G_{i,(i+1)} = -1$ and 0 otherwise.

4. Regularization. Optionally, we could also add Tikhonov L_2 regularization to the quadratic program in Min-Var K-Fold CV Estimator.

Finally, we arrive to the modified variant of the Min-Var K-Fold CV Estimator for which the solutions are not exactly optimal(because of constrains), but more stable. We used Python cvxpy package to find solutions for the optimization problem below:

$$\begin{aligned}
&z = \arg\min \left\| C^{(k)} A^{(k)} z - \alpha^{(k)} \right\|_2^2 + \gamma \|z\|_2^2 \\
&\text{subject to:} \\
&z \geqslant 0 \\
&Gz \geqslant 0 \\
&\sum_{i=1}^{N} z_i = tr(S_N^{-1})
\end{aligned}$$

and then as previously

$$\hat{d}^* = 1/z$$

Sometimes the optimizer might still return zero values for z. In this case we perform linear interpolation between its neighbors. For example if $z_i = 0$ then we set

$$z_i = z_{i-1} + (i - (i-1))\frac{z_{i+1} - z_{i-1}}{(i+1) - (i-1)}$$

2 Simulation setup

2.1 Goal

In the simulation we were aimed to prove the derivations in Papanicolaou (2017) empirically by showing that LW estimator and Min-Var estimators are in fact different in terms of shrinked eigenvalues and that Min-Var estimators also differ from sample estimators. Finally, we wanted to observe the error maximization property in Min-Var portfolios by examining the volatility of Min-Var shrinkage estimators compared to others.

2.2 Population covariance and data matrices generation

To generate a population covariance matrix Σ_N we chose a truncated exponential model:

$$H_\gamma(\lambda) = \frac{1 - e^{-\gamma(\lambda-1)}}{1 - e^{-\gamma}}$$

Thus eigenvalues are samples from $H_\gamma(\lambda)$:

$$\lambda_i = a - \frac{1}{\gamma} log(1 - x_i(1 - e^{-\gamma(b-a)}))$$

where $\gamma = 1000, a = 5.6 \cdot 10^{-5}, b = 0.015$ are constant parameters, $x_i \sim U(0, 1), i = 1, ..., N$ and x_i are sorted in descending order. Non-diagonal elements of the population eigenvalue matrix D are set to zero for simplicity and thus

$$D = Diag(\lambda_1, \lambda_2, ..., \lambda_N)$$

and population covariance matrix is

$$\Sigma_N = V D V^\intercal$$

with V being the matrix of eigenvectors. Henceforth, using the obtained covariance matrix and keeping it constant across simulations we compute the normally distributed with zero mean data matrix X_N and sample covariance matrix S_N for m times:

$$X_N^j \sim N(0, \Sigma_N) \qquad S_N^j = \frac{1}{T}(X_N^j)^\intercal X_N^j \qquad j = 1, 2, ..., m$$

Finally, a set of shrinked eigenvalues $\hat{d}^*_{ij}$ for every j is obtained by applying each of the estimators above to $\hat{\lambda}_{ij}$. The resulting values are then compared among themselves.

In this particular study we restrict ourselves to the following combinations of N and T such that the ratio N/T stays constant to approximate the large N limit:

$$N, T = \{(100, 200), (200, 400), (300, 600)\}$$

The number of simulations m is fixed at 100 although we do provide an example when $m = 1$ to illustrate the variability of estimates. Furthermore, across simulations we use only 10-Fold cross validation.

3 Simulation results

Having run simulations, we can see that variability of min-var estimators is indeed much higher than the variability of any other estimators. It is robust to changes in simulation parameters and persists both in large N,m and small N,m cases.

References

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

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

Ledoit, O., & Wolf, M. (2017a). Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets goldilocks. *The Review of Financial Studies*.

Ledoit, O., & Wolf, M. (2017b). Numerical implementation of the quest function. *Computational Statistics & Data Analysis*, *115*, 199–223.

Papanicolaou, A. (2017). Random matrix theory.