Bias reduction in covariance matrices and its effects on high-dimensional portfolios.

Benito Rodriguez Camejo, benito.rodriguez@cimat.mx
Advisor: Dr. Andrés García Medina, andres.garcia@cimat.mx
CIMAT Monterrey

The problem in question

The current world keeps evolving, and in our era the challenge of "too much data" keeps popping up, in particular in the world of portfolio optimization, therein lies the opportunity of providing more accurate results through the use of the relatively new tools developed in the random matrix theory literature to reduce the bias in the sample covariance matrix of some financial data (real or simulated). In this work we will present a couple of these tools and a brief summary of the results that can be achieved with them. We must also extend our gratitude to the works of [1] and [2] for being the major inspirations.

Suppose that the inputs of a large $n \times p$ matrix X are independent and identically distributed random variables with distribution $N(0,1)$, and that $\{\lambda_1, \ldots, \lambda_p\}$ are the eigenvalues in descending order of the covariance matrix $E = n^{-1}XX'$. And if $\lim_{n,p\to\infty} \frac{p}{n} = \gamma$ where $\gamma \in (0,1]$ then the eigenvalues density behave converges to the Marcenko-Pastur distribution.

Clipping method

First developed by [4], the clipping method consists of using the Marcenko-Pastur law to create a λ_+ an upper bound of the eigenvalues of the empirical covariance matrix $E = n^{-1}XX' = \sum_{i=1}^{N} \xi_i \mathbf{u}_i \mathbf{u}_i'$ (where $\mathbf{u}_i$ are the eigenvectors and ξ_i the eigenvalues). We consider that any eigenvalue greater than or equal to the upper bound $\lambda_+ = (1+\sqrt{q})^2$ it is interpreted as a signal and the rest as pure noise.

Then, with this information our new diagonal matrix is built, where the signals remain with their original value, but the values called noise are replaced by a constant value $\bar{\lambda}$ which only restriction is that this constant must ensure that the trace of the new resulting covariance matrix must be equal to the trace of the original $Tr\Xi^{\text{clip.}} = TrE$. The method uses the mean of the noise values as a constant.

$$\Xi^{\text{clip.}} := \sum_{i=1}^{N} \xi_i^{\text{clip.}} \mathbf{u}_i \mathbf{u}_i', \quad \xi_i^{\text{clip.}} = \begin{cases} \lambda_i & \text{if} \lambda_i \geqslant (1+\sqrt{q})^2 \\ \bar{\lambda} & \text{otherwise} \end{cases}$$

Thus, we have created a new diagonal matrix, and by substituting with the original in the eigendecomposition we have completed the clipping method and obtained a new sample covariance matrix reducing bias to the underlying "true" covariance matrix.

Tracy-Widom method

The Tracy-Widom method or Tracy-Widom test is very similar to the clipping method with the only difference being the way to identify the signals between the eigenvalues. This estimator uses the Tracy-Widom distribution to build a statistical test and compare normalized data against a probability density, this method is done according to the proofs of [3].

$$\mu_{np} = (\sqrt{n-1} + \sqrt{p})^2, \ \sigma_{np} = \sqrt{n-1} + \sqrt{p}\left(\frac{1}{\sqrt{n-1}} + \frac{1}{\sqrt{p}}\right)^{\frac{1}{3}}$$

$$\frac{\lambda_1 - \mu_{np}}{\sigma_{np}} \xrightarrow{dist} W_1 \sim F_1$$

We find the corresponding percentile to the statistic and then compare it against our confidence level (in our case $\alpha = 0.01$), if the percentile is bigger we identify the eigenvalue as a signal.

Linear shrinkage method

The linear shrinkage method proposed by [6] determines an asymptotic optimal formula to estimate α_s directly from the data and is defined as:

$$\beta := \frac{1}{p} Tr\left[(E - I_p)(E - I_p)^*\right]$$

$$\gamma := max\left(\beta, \frac{1}{n^2}\sum_{k=1}^{n} \frac{1}{p} Tr\left[(y_k y_k^* - E)(y_k y_k^* - E)^*\right]\right)$$

$$\alpha_s := 1 - \frac{\beta}{\gamma} \quad \Xi^{\text{lin}} = \sum_{i=1}^{p} \xi^{\text{lin}} u_i u_i^*, \quad \xi^{\text{lin}} = 1 + \alpha_s(\lambda_i - 1)$$

Non-linear shrinkage method

In [7] the authors define the function QuEST $Q_{p,n}$ as a multivariate deterministic function that maps $[0,\infty)^p$ on itself, in which given a set of population eigenvalues $\boldsymbol{L} = \{l_1, \ldots, l_p\}$ as input, the function returns a deterministic equivalent of the sample eigenvalues $Q_{p,n}(\boldsymbol{L}) = (q_{p,n}^1(\boldsymbol{L}), q_{p,n}^2(\boldsymbol{L}), \ldots, q_{p,n}^p(\boldsymbol{L}))$. Therefore, the eigenvalues of the population can be estimated by numerically inverting the QuEST function:

$$\tilde{\boldsymbol{\tau}} := \underset{\mathbf{L} \in [0,\infty)^p}{\text{argmin}} \frac{1}{p} \sum_{i=1}^{n} \left[q_{p,n}^i(\boldsymbol{L}) - \lambda_i\right]^2. \qquad (1)$$

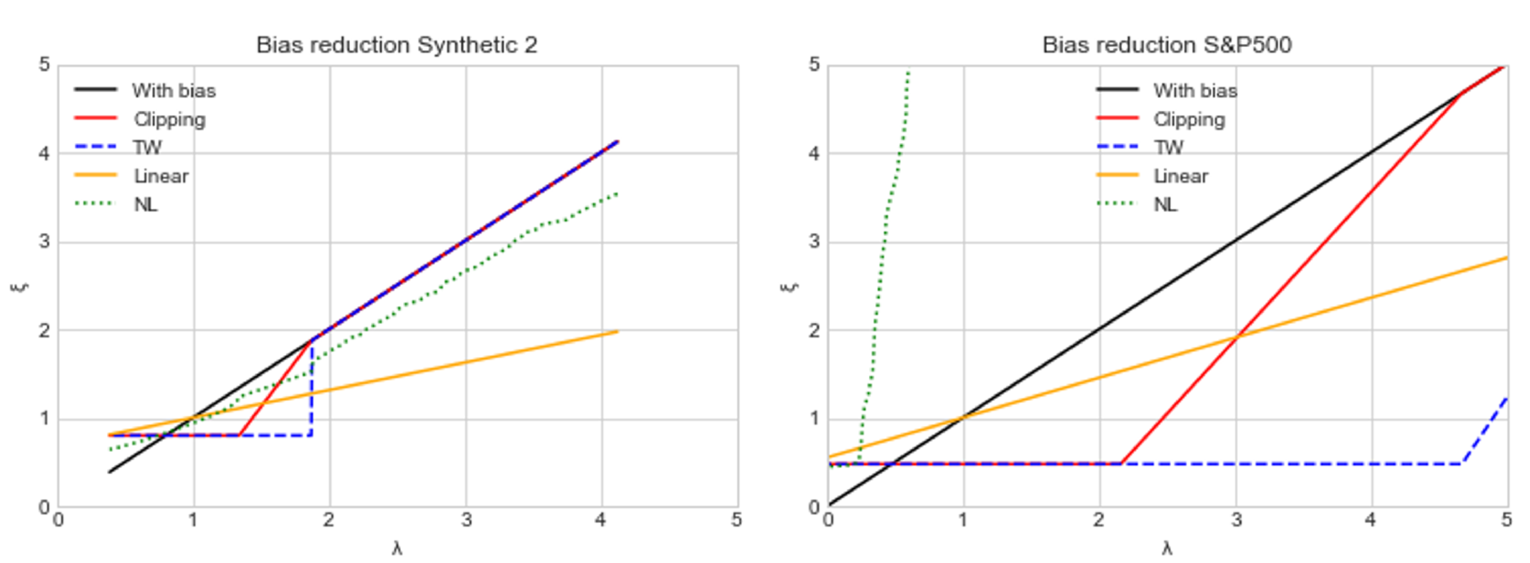
Where $\tilde{\boldsymbol{\tau}} = (\tilde{\tau}_1, \tilde{\tau}_2, \ldots, \tilde{\tau}_p))$ represents the estimated population eigenvalues. Then making use of the work done by [5] we can find an asymptotically optimal formula, which gives us an estimator $\tilde{\lambda}(\tilde{\boldsymbol{\tau}})$ consistent for the variance out of sample under large-dimension asymptotic situations.

About the data

We have 2 cases:

- Synthetic gaussian $N(0,\Sigma)$ data with $p = 500$ variables and $n = 5000$ observations. Then, we modify the covariance matrix to add 100 [8].
- Real data from $p = 96$ S&P500's stock profits from (January 01, 2020 to May 05, 2021) for a total of $n = 338$ daily observations downloaded from Yahoo finance.

Reduction results



Figures 1-2. Modified eigenvalues comparison for synthetic and real data, respectively.

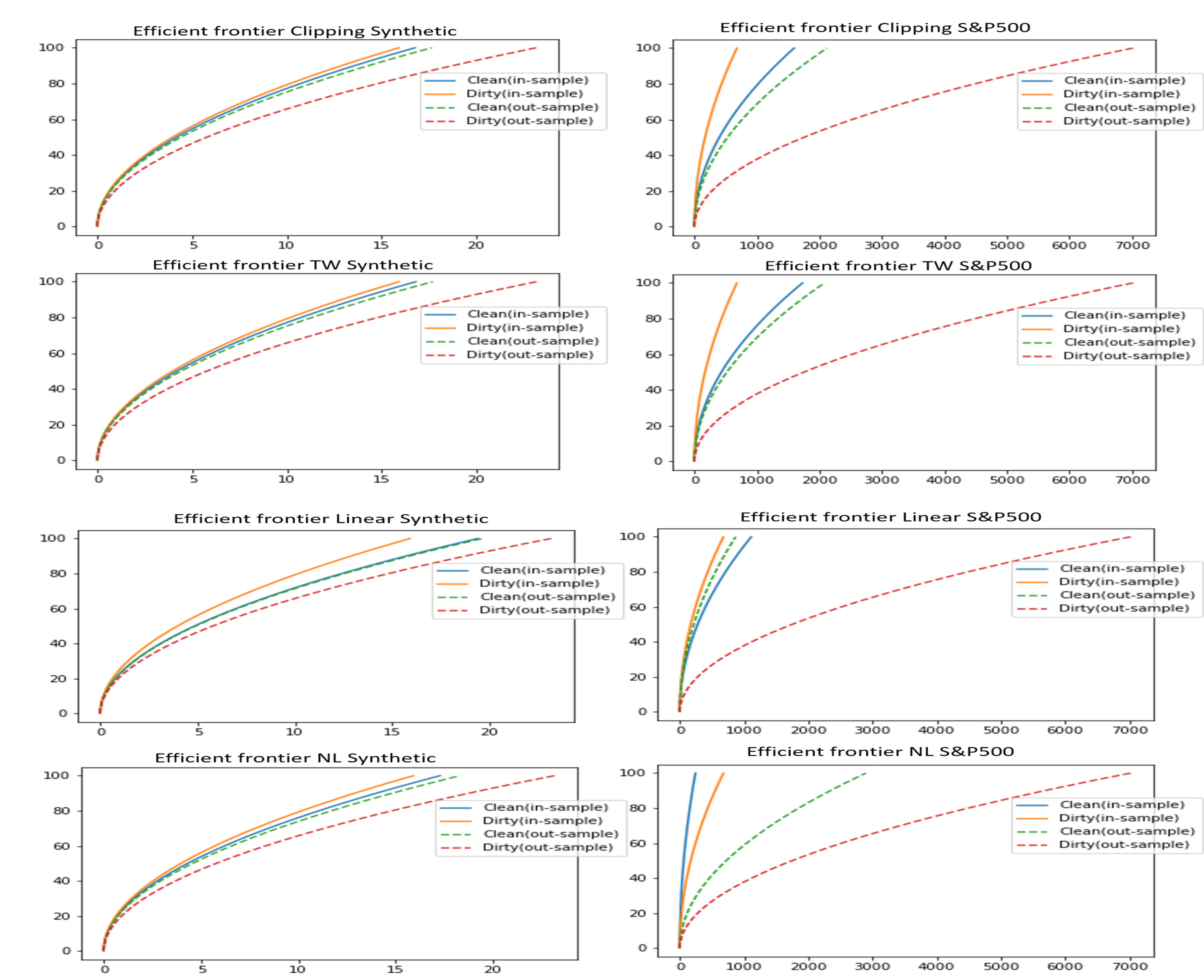
Markowitz model

The portfolio model used is the Markowitz model with only 1 restriction. Markowitz's basic portfolio finds an allocation vector w that minimizes risk while maximizing expected return:

$$max_w \mu^T w - \frac{1}{2} w^T \Sigma w, \text{ subject to } \Sigma w_i = 1$$

where μ represents the expected profits and Σ is the covariance matrix. We consider the ideal case of $\mu = (1, \ldots, 1)$.

Final Results



Figures 3-6. Efficient frontiers for Synthetic (On the left). Efficient frontiers for real data. (On the right)

We can see in the figures 3-6 that all methods have yield smaller efficient frontiers than the non-modified covariance matrix, while the figures 7-10 show that real data gives a similar result with the exception on NL shrinkage.

References

[1] Joël Bun, Jean-Philippe Bouchaud, and Marc Potters. "Cleaning large correlation matrices:tools from random matrix theory". In:PR666 (2017), pp. 1–109.

[2] Robert F Engle, Olivier Ledoit, and Michael Wolf. "Large dynamic covariance matrices". In:JBES37.2 (2019), pp. 363–375.

[3] Iain M Johnstone. "On the distribution of the largest eigenvalue in principal components anal-ysis". In:AS(2001), pp. 295–327.

[4] Laurent Laloux et al. "Noise dressing of financial correlation matrices". In:PRL83.7 (1999),p. 1467.

[5] Olivier Ledoit and Sandrine Péché. "Eigenvectors of some large sample covariance matrixensembles". In:PTRF151.1 (2011), pp. 233–264.

[6] Olivier Ledoit and Michael Wolf. "A well-conditioned estimator for large-dimensional covariancematrices". In:JMA88.2 (2004), pp. 365–411.

[7] Olivier Ledoit and Michael Wolf. "Spectrum estimation: A unified framework for covariancematrix estimation and PCA in large dimensions". In:JMA139 (2015), pp. 360–384.

[8] Marcos M López de Prado.Machine learning for asset managers. 2020.