On the investigation of a new estimator for the covariance matrix

Pierre-Luc Nadeau

Thesis supervisor: Dimitris Karyampas
Master of Quantitative Finance and Risk Management

Bocconi University

July 24, 2016

Abstract

The quality of the new estimator provided by Abadir et al.(2014)[1] to compute covariance matrices is investigated and compared with the classical way of estimating covariance matrices, as well as with the currently popular shrinkage method proposed by Wolf and Ledoit(2003)[2], via Monte Carlo methods. The conclusion of the comparisons is that the performance of the new estimator is more stable over the space of possible combinations of the number of observations (n) and the number of variables (k). Specifically, the new approach is strongly superior to the two other ones whenever n is smaller than k, and slightly inferior to the shrinkage approach when k is smaller than n. Moreover, a possibly optimal region of m's to use in the computation of the grand average estimator from [1] is found and it is significantly narrower than the one proposed in [1].

1 Introduction

The computation of large covariance matrices is of major importance in finance. It is used to estimate the optimal portfolio weights to obtain a well-diversified portfolio, estimate the covariances and correlations to implement a quantitative trading strategy, to find the right correlations in pricing derivatives on multiple assets and in time series analysis among others. But the issue that everybody encounters is that, as soon as the space of stocks becomes more than a few, the classical approach to compute the covariance matrix becomes fastly unreliable, as we usually don't have enough data to get relevant results from it. A space of 100 stocks would require thousands of daily observations to get acceptable estimates. Of course, years of data can be available, but the point is that computing the covariances and correlations of a set of stocks over years is not relevant, as the importance of past information fades away, while the importance of the new activity is greatly more important when it comes down to predicting the future behavior of prices, volatility and correlations. Hence the need to compute matrices that have more variables than the number of observations, leading to non-positive definite matrices, which are yielding bad estimations.

Multiple methods have been developped over the years to compute covariance matrices of this type. We refer the reader to Abadir et al. (2014)[1] for a summary of the past developments, as this paper will concern only two of those methods: a comparison is made between the new approach of regularizing the eigenvalues from [1] with the shrinkage approach from Wolf and Ledoit (2003)[2].

This paper is organized as follows. Section 2 describes the Monte Carlo methodology and the formulas of the various estimators as well as the methodology used to find some insight into finding the optimal region of values for the parameter m present in [1]. Section 3 shows and comments the results of the Monte Carlo experiments. Section 4 concludes.

2 Methodology

In order to see which estimator produces the best results, Monte Carlo simulations are produced by generating random correlated numbers of k variables, representing the number of stocks, each with a given amount of observations n. By standardizing these variables, the estimation of the covariance matrix is equivalent to estimating the correlation matrix, since when all the variances are 1, $diag[\Sigma] = I$ and the correlation matrix

$$corr[X] = (diag[\Sigma])^{-\frac{1}{2}}\Sigma(diag[\Sigma])^{-\frac{1}{2}}$$

simplifies to

$$corr[X] = I^{-\frac{1}{2}}\Sigma I^{-\frac{1}{2}} = \Sigma.$$

The generation of the matrix of n-by-k random correlated numbers is done via the use of the R package *mvtnorm*. Since the classical approach to compute covariance matrices is ill-conditioned when $n < k$, the input matrix to produce the numbers is coming from the covariance matrix computed under the shrinkage approach (via the use of the R package *tawny*) by using one year of standardized daily observations from k stocks of the S&P 500. Moreover, the market data comes from the CRSP database. Since the data used is standardized, this input matrix is equivalent to the correlation matrix and it represents the true correlation matrix of the numbers generated that way. It will therefore be the benchmark against which the errors of the different estimators will be computed.

For every comparison of estimators, an n-by-k matrix of random correlated numbers is generated for every point of the grid formed by all the combinations of n's and k's, with $n \in \{70, 140, ..., 420\}$ and $k \in \{35, 70, ..., 420\}$. Then, the correlation matrix of every matrix on that grid is computed for each of the estimators we want to compare. Finally, the l^2-*norm* of the half-vectorization of every estimated correlation matrix is computed. The results are shown in section 3. Moreover, the number of stocks and k, as well as the number of days and n will be used interchangeably in the paper.

2.1 The definition of the classical approach

The formula to compute the covariance matrix is given by:

$$\boldsymbol{\Sigma} = \mathbb{E}[(\mathbf{X} - \mathbb{E}[\mathbf{X}])(\mathbf{X} - \mathbb{E}[\mathbf{X}])^T]$$

To follow the same notation as in [1], the classical estimator of that matrix can also be rewritten as

$$\hat{\boldsymbol{\Sigma}} := \frac{1}{n}\mathbf{X}^T \boldsymbol{M}_n \mathbf{X}$$

where $\hat{\boldsymbol{\Sigma}}$ is a k-by-k matrix, $\mathbf{X}^T := \begin{bmatrix} \boldsymbol{x}_1 & \dots & \boldsymbol{x}_n \end{bmatrix}$ is a k-by-n matrix and $\boldsymbol{M}_n$ is the demeaning n-by-n matrix defined as

$$\boldsymbol{M}_n := \begin{bmatrix} \frac{n-1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\ -\frac{1}{n} & \frac{n-1}{n} & \cdots & -\frac{1}{n} \\ \vdots & \vdots & \vdots & \vdots \\ -\frac{1}{n} & -\frac{1}{n} & \cdots & \frac{n-1}{n} \end{bmatrix}$$

Moreover, the formula to compute the correlation matrix of a given covariance matrix $\hat{\boldsymbol{\Sigma}}$ is given by:

$$corr[X] = (diag[\hat{\boldsymbol{\Sigma}}])^{-\frac{1}{2}} \hat{\boldsymbol{\Sigma}} (diag[\hat{\boldsymbol{\Sigma}}])^{-\frac{1}{2}}$$

2.2 The definition of the shrinkage approach

The shrinkage approach to estimate a covariance matrix consists in computing a convex linear combination that can be defined as

$$\hat{\boldsymbol{\Sigma}}_{shrink} = \delta\hat{\boldsymbol{\Sigma}} + (1 - \delta)\boldsymbol{F},$$

where $\hat{\boldsymbol{\Sigma}}$ is the classical sample covariance matrix, $\boldsymbol{F}$ is a structured estimator (following a specified model), also called the shrinkage target, and δ is called the shrinkage constant (a number between 0 and 1), that basically represents a weight to be given to $\boldsymbol{F}$. $\boldsymbol{F}$ and δ are chosen such that the estimator $\hat{\boldsymbol{\Sigma}}_{shrink}$ reduces the estimation errors produced by $\hat{\boldsymbol{\Sigma}}$, therefore giving a better estimator for the covariance matrix. The goal of this paper being only to use this estimator to generate simulation results, only a brief intuition was given here and the reader is invited to get more details by consulting the works of Wolf and Ledoit, in particular [2] and [3]. For the computation of $\hat{\boldsymbol{\Sigma}}_{shrink}$, the function *cov.shrink*, coming from the R package *tawny*, is used.

2.3 A new estimator for the covariance matrix

Abadir et al.[1] have found a new way of estimating $\hat{\boldsymbol{\Sigma}}$ producing stunning results, in addition to being design-free (no assumptions have to be made about the data) and always produces positive-definite, non-singular and well-conditioned estimators of $\boldsymbol{\Sigma}$. Let's summarize briefly how to compute the estimator and explain the intuition along the way. The spectral decomposition of the classical estimator $\hat{\boldsymbol{\Sigma}}$,

$$\hat{\boldsymbol{\Sigma}} = \hat{\boldsymbol{P}}\hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{P}}^T, \tag{1}$$

gives us the matrix $\hat{\boldsymbol{P}}$ containing its eigenvectors and the matrix $\hat{\boldsymbol{\Lambda}}$ containing its eigenvalues, which can also be found from

$$\hat{\boldsymbol{\Lambda}} = \hat{\boldsymbol{P}}^T\hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{P}}, \tag{2}$$

where $\hat{\boldsymbol{\Lambda}}$ is a k-by-k diagonal matrix and $\hat{\boldsymbol{P}}$ is a k-by-k orthogonal matrix. The spectral decomposition of a matrix can easily be computed in R by the use of the function *eigen*.

In their paper, they start from the fact that orthogonal matrices are never ill-conditioned, meaning that the ill-conditioning of $\hat{\boldsymbol{\Sigma}}$ comes from its eigenvalues. And, basically, what the new estimator does is improving the estimation of the eigenvalues of $\hat{\boldsymbol{\Sigma}}$. By splitting the dataset $\mathbf{X}$ into two subsets,

$$\mathbf{X}^T := \begin{bmatrix}\boldsymbol{x}_1 & \dots & \boldsymbol{x}_n\end{bmatrix} = \begin{bmatrix}\mathbf{X}_1^T & \mathbf{X}_2^T\end{bmatrix},$$

the eigenvectors of the dataset are then estimated using one of the subsets,

$$\hat{\boldsymbol{\Sigma}}_1 := \frac{1}{m}\mathbf{X}_1^T\boldsymbol{M}_m\mathbf{X}_1 \tag{3}$$

$$\hat{\boldsymbol{\Sigma}}_1 = \hat{\boldsymbol{P}}_1\hat{\boldsymbol{\Lambda}}_1\hat{\boldsymbol{P}}_1^T$$

and then computing the estimated $\hat{\boldsymbol{\Sigma}}$ of the second subset,

$$\hat{\boldsymbol{\Sigma}}_2 := \frac{1}{n-m}\mathbf{X}_2^T\boldsymbol{M}_{n-m}\mathbf{X}_2 \tag{4}$$

$$\hat{\boldsymbol{\Sigma}}_2 = \hat{\boldsymbol{P}}_2\hat{\boldsymbol{\Lambda}}_2\hat{\boldsymbol{P}}_2^T$$

one can find the estimator $\tilde{\boldsymbol{\Lambda}}$ of the improved eigenvalues:

$$\tilde{\boldsymbol{\Lambda}} := diag[\hat{\boldsymbol{P}}_1^T\hat{\boldsymbol{\Sigma}}_2\hat{\boldsymbol{P}}_1]. \tag{5}$$

Then, by using the improved eigenvalues, the new estimator of $\hat{\boldsymbol{\Sigma}}$ is given by

$$\tilde{\boldsymbol{\Sigma}} = \hat{\boldsymbol{P}}\tilde{\boldsymbol{\Lambda}}\hat{\boldsymbol{P}}^T. \tag{6}$$

Moreover, by averaging the procedure over different samples (non-repeated random resampling), one then gets a better estimator

$$\tilde{\boldsymbol{\Sigma}}_{m,S} := \frac{1}{S}\sum_{s=1}^{S}\tilde{\boldsymbol{\Sigma}}_{m,s}, \tag{7}$$

called the general estimator. Every resampling simply consists in getting another matrix, with the same dimensions as the original one, by randomly shuffling all the rows of the original data. This means that every resampling will give different subsets $\mathbf{X}_1$ and $\mathbf{X}_2$ for the computation of the estimator. Obviously, the perfomance of this new estimator will depend on the parameter m chosen to split the dataset into two subsets, but here again, by averaging the procedure over different values of this parameter, a better estimator is then obtained

$$\tilde{\boldsymbol{\Sigma}}_{M,S} := \frac{1}{M} \sum_{m \in \mathcal{M}} \tilde{\boldsymbol{\Sigma}}_{m,S} \tag{8}$$

$$where\ \mathcal{M} := m_1, m_2, \dots, m_M,\ 1 < m_1 < m_2 < \cdots < m_M < n-1,$$

called the grand average estimator. (8) is the one used in producing all the results against the classical approach and the shrinkage approach and it will be interchangeably be called the new estimator.

2.4 Towards the optimal m

The parameter m to use in (6), (7) or (9) (coming in this subsection) must be optimimized, that is it must minimize the errors that the estimator will produce using it. In their paper, they define the minimization problem as follows. The bootstrapped (repeated random sampling) classical estimator $\hat{\boldsymbol{\Sigma}}$ is defined as

$$\hat{\boldsymbol{\Sigma}}_B := \frac{n}{(n-1)B} \sum_{b=1}^{B} \hat{\boldsymbol{\Sigma}}_b \tag{9}$$

and the bootstrapped general estimator is given by

$$\tilde{\boldsymbol{\Sigma}}_{m,B} := \frac{1}{B} \sum_{b=1}^{B} \tilde{\boldsymbol{\Sigma}}_{m,b}. \tag{10}$$

Each bootstrap simply consists in rebuilding a matrix, with the same dimensions as the data to be bootstrapped, by randomly sampling rows from the original data and by allowing repeated sampling of those rows. Then, the optimal value of the parameter m is given by minimizing the errors between (9) and (10) for each of the B resamplings:

$$\frac{1}{B} \sum_{b=1}^{B} \|\text{vech}[\tilde{\boldsymbol{\Sigma}}_{m,b} - \tilde{\boldsymbol{\Sigma}}_{m,B}]\|_1^1. \tag{11}$$

Or, equivalently, the optimal value of m is given by

$$m_B := \underset{m \in \mathcal{M}}{\text{argmin}} \frac{1}{B} \sum_{b=1}^{B} \|\text{vech}[\tilde{\boldsymbol{\Sigma}}_{m,b} - \tilde{\boldsymbol{\Sigma}}_{m,B}]\|_1^1. \tag{12}$$

The $l2$-*norm* could also be used equivalently. In the computations that have been done on m_B, of which the results are shown in 3.2, the $l1$-*norm* was used. The results shown there consist of the average of 30 samplings of (11) for every point of the grid formed by all the combinations of n's and k's, with $n \in \{50, 115, ..., 505\}$, $k \in \{75, 125, ..., 425\}$ and $B = 30$. That is, an average of 30 complete grids, each done with 30 bootstraps. The results will give insights about the optimal region to be chosen when one is using the grand average estimator (8).

3 Results

As it has been said in Wolf and Ledoit (2004)[3]:

> "...nobody should be using the sample covariance matrix for the purpose of portfolio optimization. It contains estimation error of the kind most likely to perturb a mean-variance optimizer."

We will shortly get a glimpse at how bad this estimation error can get, and how much better are the possible alternatives. In the following pages, the simulation results are mostly showned 2 contour plots side-by-side in order to compare between estimators. Both plots are standardized using the same range of values for their color scale, and the blue region represents the bottom 10% of the range when the values plotted are all positive. Moreover, the red regions consist of regions where the values are negative.

3.1 The classical approach and the shrinkage approach

In order to see the amplitude of the estimation error under the usual approach, a comparison is made with the currently widely popular shrinkage approach. The results of the Monte Carlo simulations for the classical and the shrinkage approaches are shown in **Figure 1**.

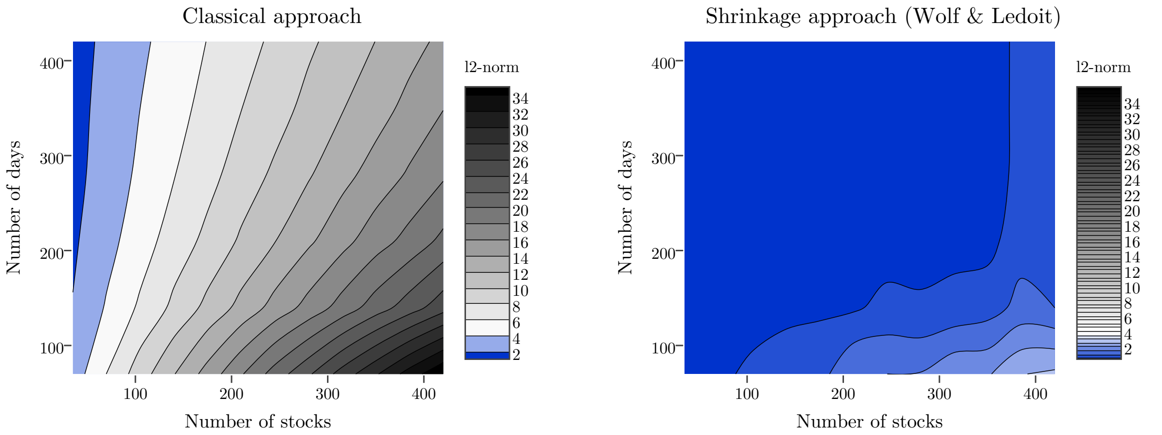

Figure 1: The classical approach versus the shrinkage approach

The fact that the range of errors under the usual approach, $[1.16, 35.82]$, is incredibly worst than the range under the shrinkage one, which is only $[0.01, 2.6]$, is well summarized in the above figure. The range of errors under the shrinkage approach is entirely within the blue region. Moreover, it is easy to see that the performance of the usual approach is clearly dependent on the number of days and on the number of stocks used to compute $\hat{\boldsymbol{\Sigma}}$.

Since it is difficult to distinguish if this is also the case for the shrinkage approach when plotted relatively to the usual approach, **Figure 2** shows the plot for the shrinkage approach alone. There is also a similar dependence, but a somewhat better one. In fact, it is performing uniformly very well above the line $k < n$, represented by the dotted line, whereas the classical approach performs badly very well above it. The drawback of the shrinkage approach, while immensely superior to the classical one, is that it starts to perform more and more poorly as the ratio $\frac{k}{n}$ gets bigger and bigger.

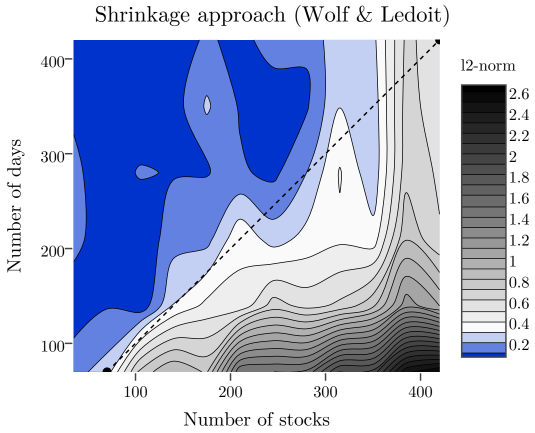

Figure 2: The shrinkage approach alone

3.2 Towards the optimal m

Before moving forward to compare the shrinkage approach with the new one, one must get an idea of which range of m values should be used to compute the grand average estimator **(8)**. The grid used produced 8 rows (the n parameter) and 8 columns (the k parameter). For every row, a contour plot of the surface of its values is showned. A graph containing all the different slices taken at each k of these contour plots is produced as well, for each contour plot. Only a few of them are showned in the present section to depict the summary of this set of results. To get a more complete picture, the reader is invited to consult the appendix (see 5), where the whole set of plots is available.

The perfomance of the new approach will vary depending on which value of the parameter m is used. This m should be optimized when the other estimators, **(6)**, **(7)** or **(9)** are used instead of **(8)**, the grand average one. The purpose of this section is not to give the value of m_B (the optimal m) for every combination of n and k, but rather to give the best range of values to use when computing **(8)**. In **Figure 3**, the slices of the errors from row $n = 50$ and row $n = 505$ are showned. One should notice that the curves are u-shaped. Interestingly, they all converge (see 5) towards the same curve as n gets bigger and the region containing the minimum values is well defined in all of them, pointing towards $m \in [0.2, 0.4]$.

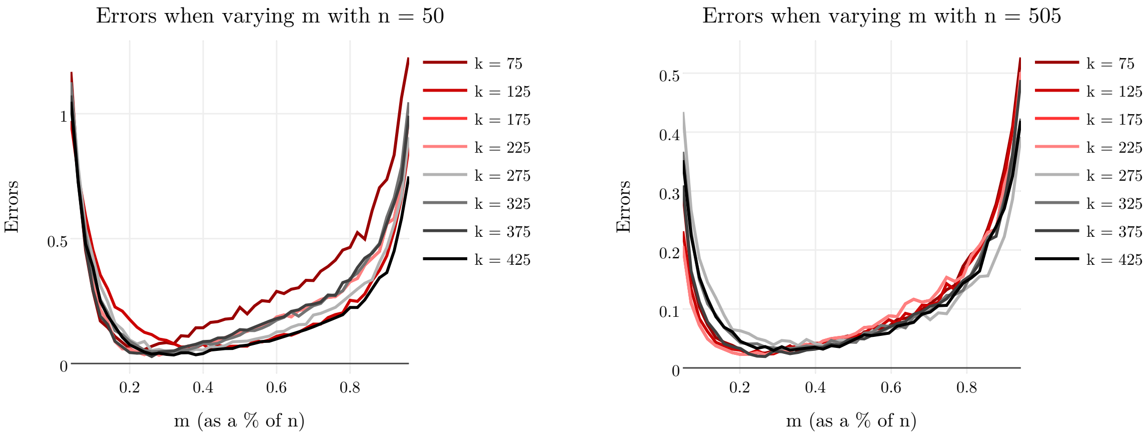

Figure 3: The distribution of the errors when varying m has a u-shape

In **Figure 4**, the contour plots of the errors from row $n = 50$ and row $n = 505$ are showned. The darker blue region contains the bottom 5% of the range of errors and the lighter blue region the bottom 10%. Here again, there is an indication that the region that must contain the optimal m_B should be within $m \in [0.2, 0.4]$, as were indicating the slices from the previous figure. There is a positive relationship between the wideness of the optimal region and n. When considering the whole set of contour plots, one can see the blue region growing (shrinking) as n becomes bigger (smaller). Also, from the results shown in this paper, the optimal region seems to be independent of k. When n is small, the optimal region seems to be around $m \in [0.3, 0.4]$ and when n is big it seems to be $m \in [0.2, 0.4]$. The notation $\mathcal{M}$ will be used to refer to this region as the set of m's used when computing (8).

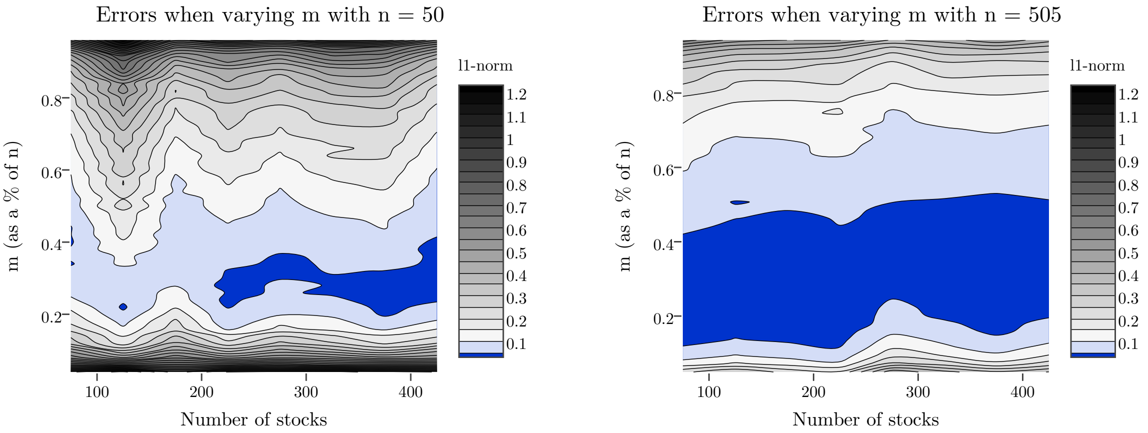

Figure 4: The m_B zone, shown in blue

From these results, a conclusion can be made that the optimal zone of m's that should be used to average over when using the grand average estimator of [1], given by (8), is within $[0.2n, 0.4n]$ or within $[0.3n, 0.4n]$. More on this in section 3.5. The results in the following 2 subsections will be computed using $\mathcal{M}_1$ representing $m \in [0.2, 0.4]$.

3.3 The classical approach and the new approach

In order to see the amplitude of the estimation error under the usual approach, a comparison is also made with the new approach of regularizing the eigenvalues from Abadir et al.[1]. Again here, the results of the Monte Carlo simulations for both the classical approach and the new one are shown in **Figure 5**.

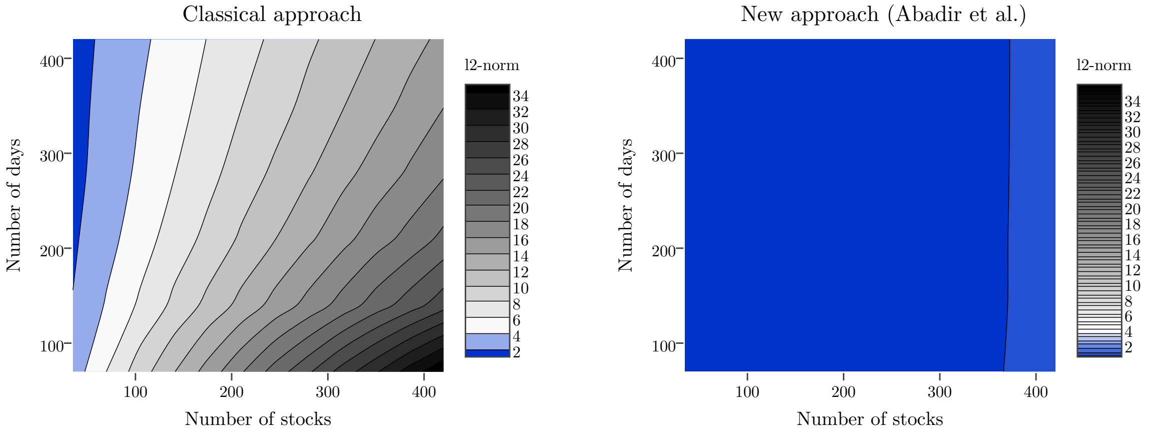

Figure 5: The classical approach versus the new approach

Similar results as in 3.1 are obtained: the range of errors under the usual approach, $[1.16, 35.82]$, is incredibly worst than the range under the new one, which is only $[0.08, 0.67]$. All of the errors obtained under the new approach are also always within the blue region here. Ideally, we would like to see if the performance of the new approach is also dependent on the number of days and/or the number of stocks used to compute $\hat{\boldsymbol{\Sigma}}$.

Again, since it is difficult to distinguish such a relation when plotted relatively to the classical approach, **Figure 6** shows the plot for the estimator under investigation on its own. Surprisingly, there is a different dependence: comparatively to the classical approach and the shrinkage approach, the new estimator is performing uniformly well over the whole space, without any dependence on the line $k < n$. There is a dependence over the k parameter, but the errors are very low compared to the other approaches. Clearly, the advantage of this estimator is that it performs the same, no mather if the ratio $\frac{k}{n}$ gets bigger or smaller. But to attest the superiority of this estimator over the shrinkage approach, further investigation is required, which is done in 3.4.

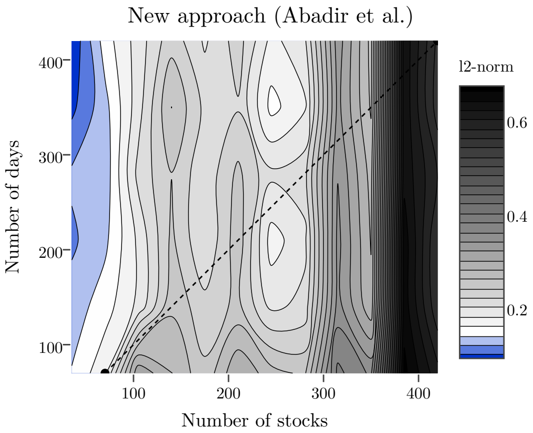

Figure 6: The Abadir et al. approach alone

3.4 The shrinkage approach and the new approach

The main purpose of this paper is to evaluate if the new approach proposed by Abadir et al.[1] is superior than the shrinkage approach. Following the same steps as in the previous subsections, a comparison is made between the shrinkage approach and the new one by showing the results of the Monte Carlo simulations in **Figure 7**.

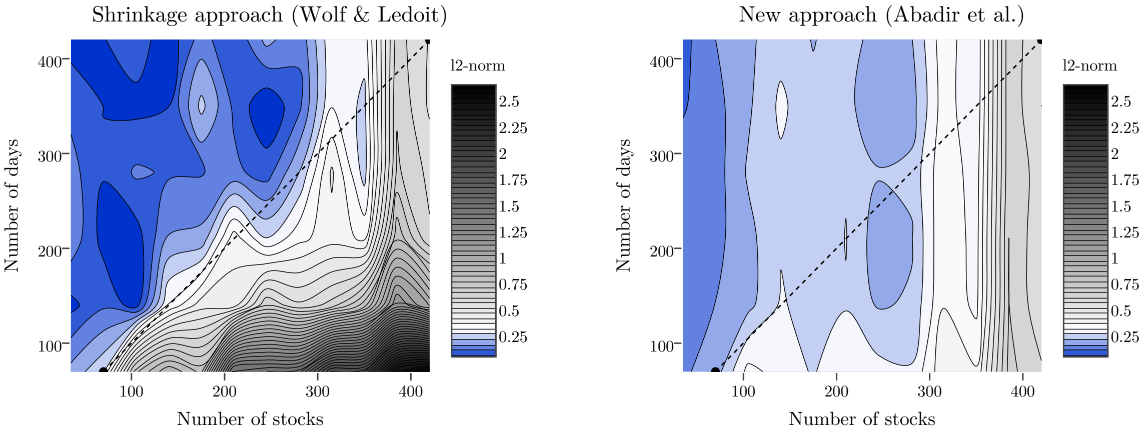

Figure 7: The shrinkage approach versus the new approach

Now the results are quite interesting. The range of errors under the shrinkage approach, $[0.01, 2.6]$, possess the lowest and the highest boundaries of both estimators, with the range for the new estimator being $[0.08, 0.67]$.

Clearly, the Abadir et al.[1] grand average estimator is more stable than the shrinkage approach, with it's range of errors lying in the lower half of the range of errors under the shrinkage one, and being non-dependent over the $k < n$ line. But in order to better assess the behavior of both estimators compared to each other, the relative errors are shown in **Figure 8**, that is :

$$\varepsilon_{relative} = \frac{\varepsilon_{\hat{\boldsymbol{\Sigma}}_{shrink}}}{\varepsilon_{\tilde{\boldsymbol{\Sigma}}_{M,S}}} - 1$$

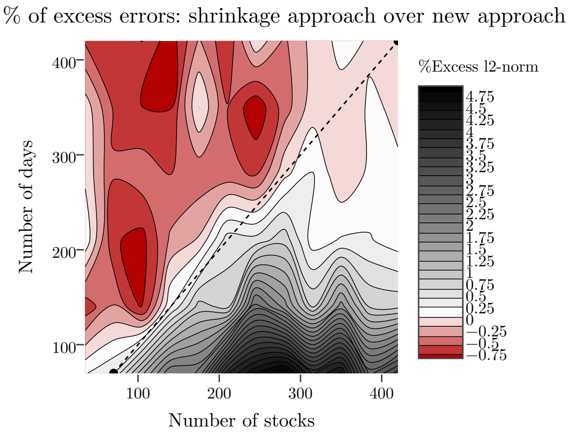

Figure 8: $\varepsilon_{relative}$: Abadir et al. approach is superior when $n < k$

The last figure summarizes pretty well how the Abadir et al.[1] grand average estimator is compared to the shrinkage one: the new estimator is considerably superior to the shrinkage approach when $n < k$, while the shrinkage approach is slightly superior to the new estimator when $k < n$. Slightly superior is the good choice of words, since both estimators are already generating pretty much negligible errors when $k < n$. To illustrate this point, the absolute excess errors between them are shown in **Figure 9**, that is:

$$\varepsilon_{absolute} = \varepsilon_{\hat{\boldsymbol{\Sigma}}_{shrink}} - \varepsilon_{\tilde{\boldsymbol{\Sigma}}_{M,S}}$$

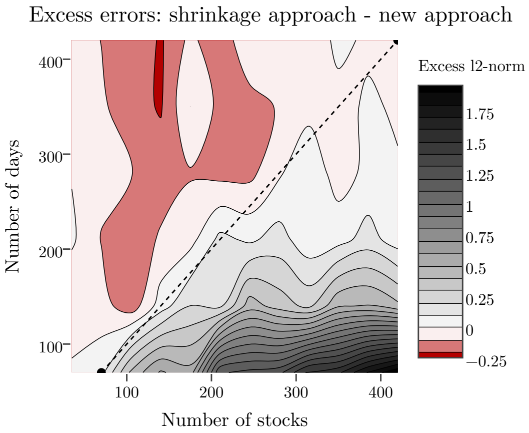

Figure 9: $\varepsilon_{absolute}$: Abadir et al. inferiority when $k < n$ is negligible

Given the superiority when $n < k$ and the fact that its weakness when $k < n$ is negligible, the new estimator by Abadir et al.[1] seems to be generally the superior one.

3.5 More on the optimal m's region

So far, the results in the sections 3.3 and 3.4 have been computed using the grand average estimator (8) averaged over $\mathcal{M}_1$, which represents $m \in [0.2, 0.4]$. Here, as it has been mentioned in 3.2, further investigation is done about whether $\mathcal{M}_2$, representing $m \in [0.3, 0.4]$, or $\mathcal{M}_1$ should be preferable. Following the same methodology as in the previous subsection, $\varepsilon_{relative}$ and $\varepsilon_{absolute}$ are again showned here for the two different m's regions in **Figure 10** and **Figure 11**.

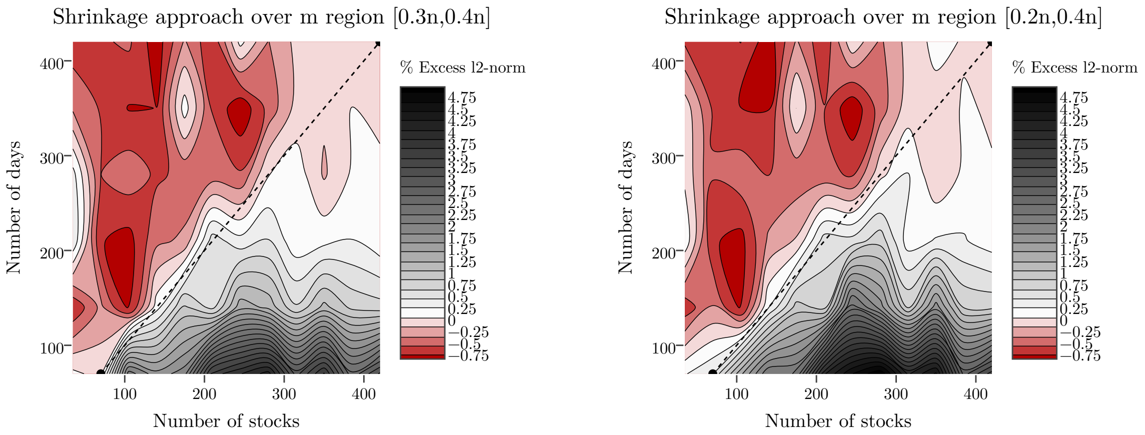

Figure 10: $\varepsilon_{relative}$:The m's regions $[0.3n, 0.4n]$ and $[0.2n, 0.4n]$ are equivalent

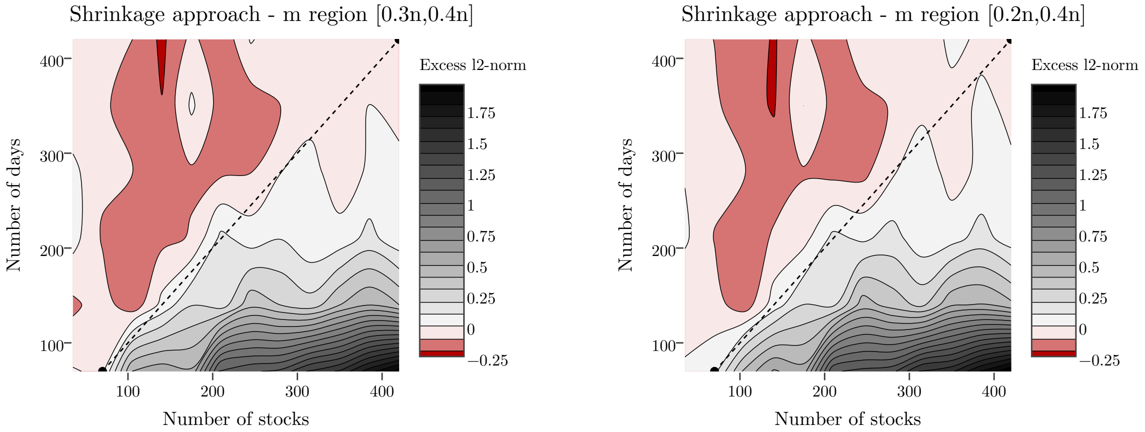

Figure 11: $\varepsilon_{absolute}$:The m's regions $[0.3n, 0.4n]$ and $[0.2n, 0.4n]$

The conclusion that one should extract from the above is that the two regions, $\mathcal{M}_1$ and $\mathcal{M}_2$, seem to be equivalent, and therefore both regions, the wider $[0.2n, 0.4n]$ or the narrower $[0.3n, 0.4n]$, can be used.

4 Conclusion

The main conclusion from this paper should be that the new estimator proposed by Abadir et al.(2014) [1] is not only superior to the classical approach to estimate $\boldsymbol{\Sigma}$, but also generally superior to the shrinkage estimator, especially when $n < k$. That being said, one must keep in mind that it is, however, negligibly inferior to the shrinkage approach when $k < n$. To obtain the best perfomance, one could be interested in combining the use of both approaches: the new approach whenever the problem has $n <= k$ and the shrinkage approach whenever the problem has $k < n$.

Moreover, an important finding that holds in the current Monte Carlo setting of this paper is that the region $\mathcal{M}$ that should be used in (8) can be narrowed to $\mathcal{M}_2$, that is $m \in [0.3, 0.4]$. It is an important finding, since the paper from Abadir et al. [1] proposes a region of $m \in [0.2, 0.8]$ instead, meaning that more errors could be introduced in the averaging process, as non-optimal regions are included in the bootstrapping process when computing the grand average estimator (8). Also, since the found region is more optimal, the computational resources required to get the same level of accuracy can be reduced.

Certainly, more profound investigation should be done to assert the current findings, but everyone facing the needs to compute covariance and correlation matrices should possibly be interested by the new estimator provided by Abadir et al. [1]. Specifically, further investigation about using the new estimator in a portfolio selection setting to compare it with the shrinkage approach could be of great use.

5 Appendix

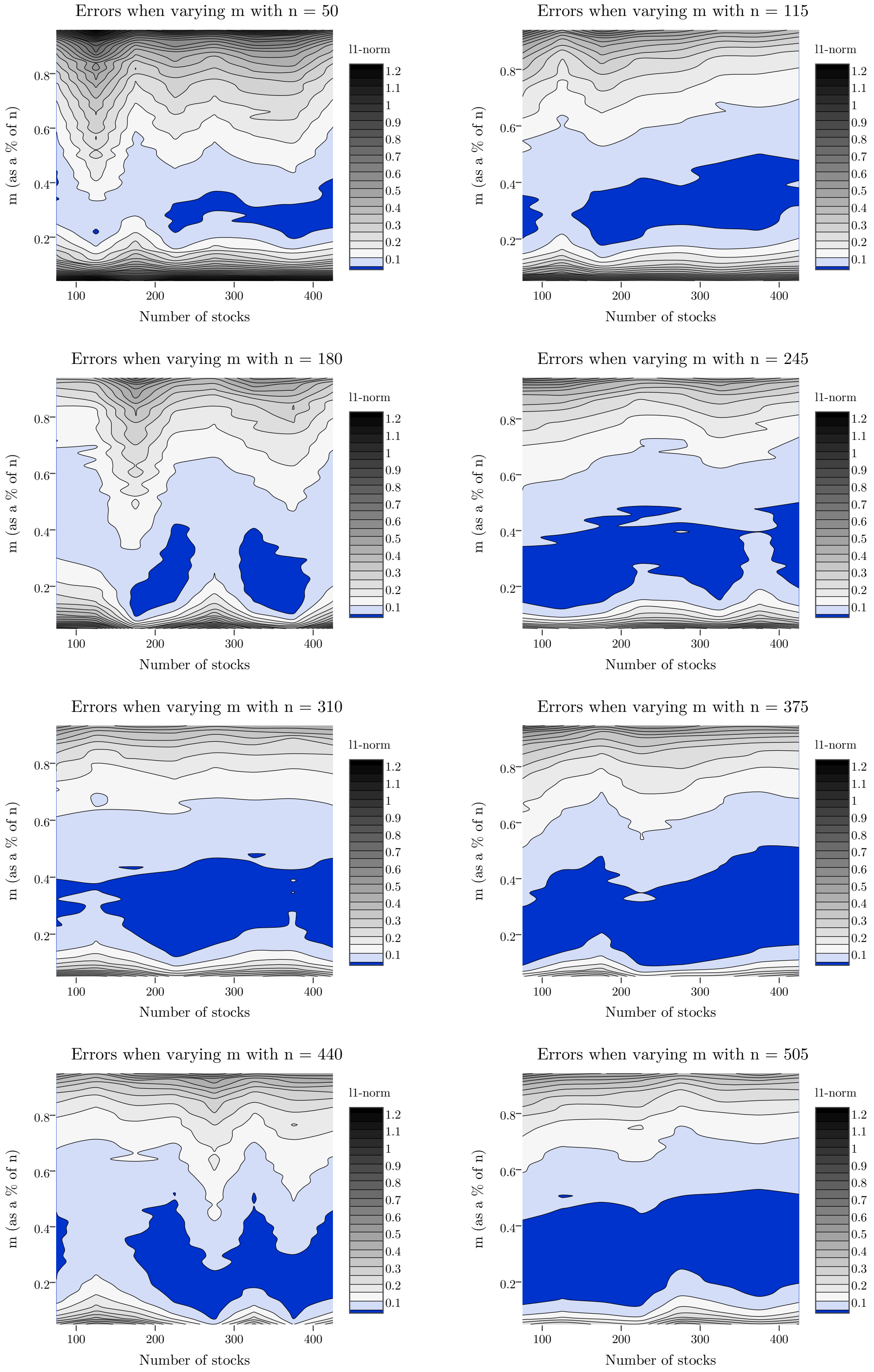

Figure 12: All the contour plots of every n from 3.2

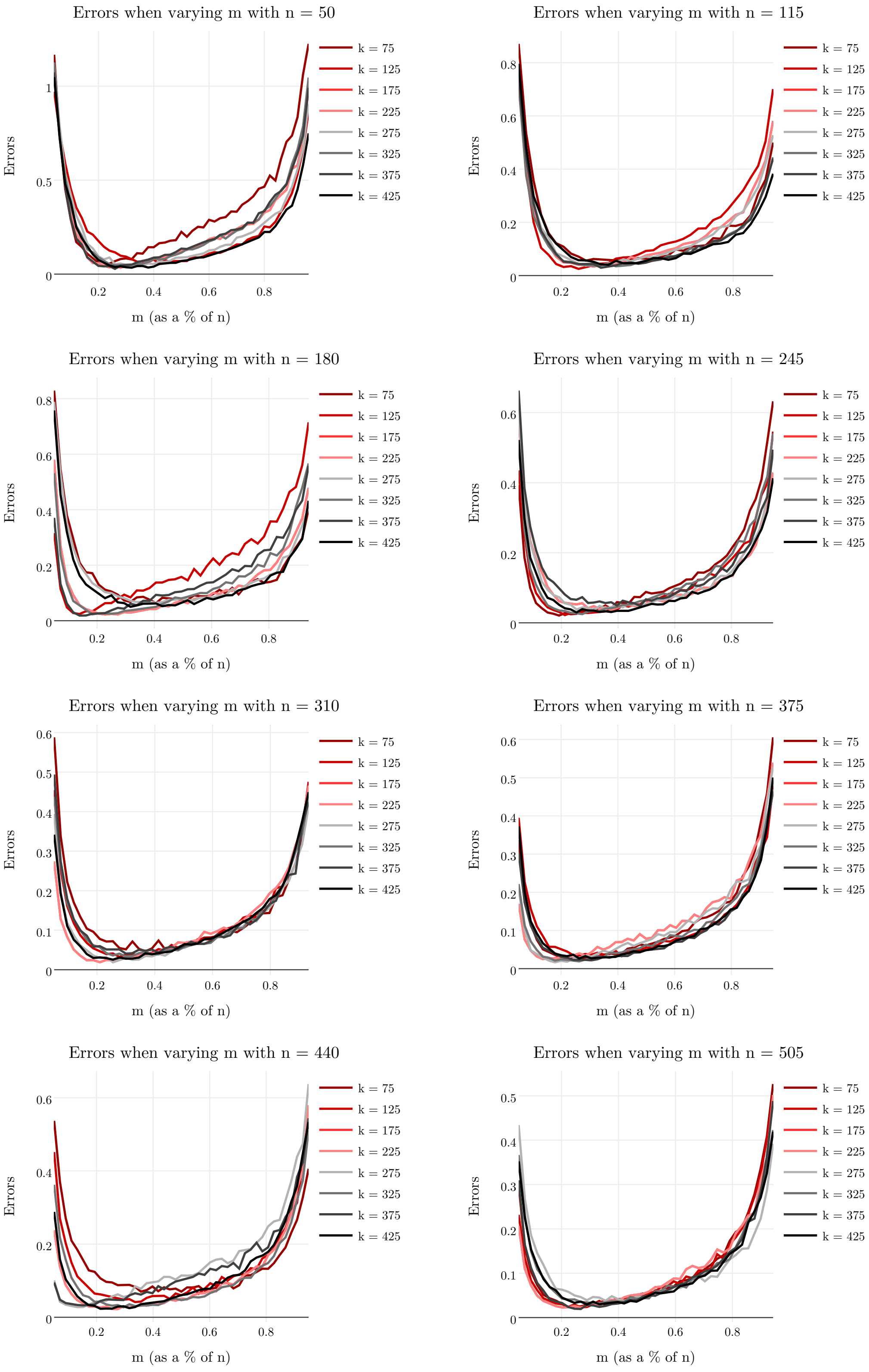

Figure 13: All the k's composing each of the contour plots presented above, from 3.2

References

[1] Distaso W. Abadir K. M. and F. Zikes. “Design-free estimation of variance matrices.” In: *Journal of Econometrics* 181.2 (2014), pp. 165–180.

[2] Ledoit O. and Wolf M. “A Well-Conditioned Estimator for Large Dimensional Covariance Matrices.” In: *Journal of Multivariate Analysis* 88 (2003), pp. 365–411.

[3] Ledoit O. and Wolf M. “Honey, I Shrunk the Sample Covariance Matrix.” In: *Journal of Portfolio Management* 31 (2004), pp. 1–22.