Analysis of covariance matrix estimators for simple portfolio optimisation

Grant Holtes

November 19, 2023

Summary

This investigation evaluates the impact of different covariance matrix estimation methods in the context of portfolio optimisation. Focused on the role of covariance matrices, the analysis highlights limitations in the widely used sample covariance matrix. The examined alternative estimators are diagonal, constant correlation, simple shrinkage, constant correlation shrinkage, Ledoit-Wolf shrinkage, and oracle approximating shrinkage estimators. Notably, the Ledoit-Wolf shrinkage estimator outperforms others, demonstrating its effectiveness in optimizing portfolio returns across diverse asset universes. The key takeaway is the sample covariance matrix is generally prone to error, superior estimators exist which should be considered, and the difference between estimators can cause a material difference in returns when using the matracies in portfolio optimisation.

1 Motivation

1.1 Importance of covariance matrices

Covariance matrices are pivotal in finance, serving as a key input for portfolio management, risk assessment, and asset pricing. In portfolio construction, understanding the relationships between different assets is essential for achieving diversification and minimizing overall risk. Modern Portfolio Theory relies on covariance matrices to optimize portfolios based on the trade-off between expected return and volatility. Covariance matrices are also used in asset pricing models, aid in pricing derivatives accurately, and for risk management, facilitating the measurement of portfolio risk through metrics like VaR and CVaR. Given the wide range of models that require covariance matrices as an input and the sensitivity of the models to the input matrices, their estimation methods deserve some interrogation.

1.2 Issues with the sample covariance matrix

The obvious (or naive) choice to estimate covariance is the sample covariance matrix. This has a number of key issues that can make it unsuitable for use, including:

Error: While the estimator is unbiased, it is based on a finite sample and so does exhibit estimation error. This error can be significant, particularly when dealing with limited sample sizes, and it affects the accuracy of statistical inferences and predictions based on the estimated covariance matrix.

Sensitivity to noise: The sample covariance matrix is highly sensitive to outliers or extreme values in the data. Outliers can have a disproportionate impact on the estimated covariance, leading to distorted results

Singularity and Inversion: When the number of variables is equal to or greater than the number of observations, the sample covariance matrix becomes singular (non-invertible).

This paper will deal primarily with the first two issues of error and sensitivity.

2 Overview of available estimators

There are a range alternative approaches to covariance matrix estimation

Simplified sample estimators: These methods aim to produce a better matrix by applying simplifying assumptions. These include the diagonal covariance matrix, which assumes all variables are independent, and the constant correlation matrix, which assumes all variables have the same correlation.

Shrinkage estimators: Shrinkage estimators aim to produce more useful covariance matrices by combining 2 estimates: A unbiased but high error estimate, F, usually the sample covariance matrix, and a biased but low error estimate, S. These are combined in a weighted sum

$$\Sigma = (1 - \delta)S + \delta F$$

The shrinkage amount δ can be either a given input or be optimised by the estimator.

Time series estimators: As finacial data are often time series, these models aim to incorporate the information provided by the relationships between time periods into the estimated covariance matrix. The methods include the Exponentially Weighted Moving Averages (EWMA) model and GARCH (Generalized Autoregressive Conditional Heteroskedasticity) model.

Sparse covariance estimators: When dealing with high-dimensional data, methods like graphical lasso (L1 regularization) are used to estimate a sparse covariance matrix, assuming that most of the correlations are zero.

Maximum likelihood estimators: The Maximum Likelihood Covariance Estimator (MLCE) is an approach that seeks to find the covariance matrix that maximizes the log likelihood of the observed data. As with the sample covariance matrix, MLCE is an unbiased estimator, but is known to be sensitive to outliers in small sample sizes and estimation error when the number of variables exceeds the number of observations.

This investigation will focus on simplified sample estimators and shrinkage estimators.

3 Quantitative comparison

The impact of estimator choice is evaluated in the context of the use of the estimated matrix in optimising a portfolio of individual equities.

3.1 Experiment Design

The impact of the different covariance measures on the portfolio selection process is measured through a highly simplified simulation of an investing process with the following steps:

1. The investor accesses return data for their given asset universe, until the current as-at date.
2. The investor calculates the expected returns and and the covariance matrix using the method being tested.
3. The investor determines their portfolio weights using a simple optimiser, with the objective of maximising Sharpe ratio.
4. The as-at date is advanced by the simulation time step of one month, with the portfolio value incriminating by the realised return for the period.

This simulation is run for each of the in-scope covariance measures and for a range of asset universes, being the top-n members of the SP500. Using the current largest members of an index as a investing universe naturally biases the analysis, but this bias is assumed to affect all of the covariance measures equally and so can be ignored.

4 Methods to be analysed

4.1 Sample covariance matrix

The sample covariance matrix is also tested and will act as the baseline that the other methods can be compared against. This is given by Equation 1, where X is a matrix of row oriented observations and $\overline{X}$ is the sample mean vector across all observations.

$$S = \frac{1}{n-1}(X - \overline{X})(X - \overline{X})' \tag{1}$$

This is implemented by the python code below

```
def sample_covariance_matrix(X):
    n = X.shape[1] - 1 # degrees of freedom adj
    avg = np.average(X, axis=1) # Get average of each column
    X = X - avg[:,np.newaxis] # Center data on 0.
    c = np.dot(X, X.T) # Get sum of elementwise differences
    c *= np.true_divide(1, n)
    return c
```

4.2 Diagonal Covariance Matrix

The diagonal covariance matrix simply sets all the non-diagonal elements of the sample covariance matrix to zero. This has the effect of assuming that all variables are independent, with insignificant correlations. Variance (diagonal) elements are left unchanged.

```
def diagonal_covariance_matrix(X):
    F = sample_covariance_matrix(X)
    n = F.shape[0]
    return F * np.identity(n)
```

4.3 Constant Correlation Covariance Matrix

The constant correlation covariance matrix re-scales each non-diagonal element such that every pair of variables have the same correlation. The chosen correlation is the average correlation, calculated as the average of the non-diagonal elements of the sample correlation matrix. Variance (diagonal) elements are left unchanged.

```
def constant_correlation_estimator(X):
    F = sample_covariance_matrix(X)
    n = F.shape[0]
    # Calculate correlation
    corr_ = np.ones((n,n))
    for i in range(n):
        for j in range(n):
            corr_[i,j] = F[i,j] / (np.sqrt(F[i,i] * F[j,j]))
    # Get non-diagonal elements
    indicies = np.where(~np.eye(n,dtype=bool))
    avg_corr = np.mean(corr_[indicies])
    # Scale covariances
    S = np.ones((n,n))
    for i in range(n):
        for j in range(n):
            if i != j: # Scale off-diagonal componants
                k = avg_corr / corr_[i,j]
                S[i,j] = F[i,j] * k
            else:
                S[i,j] = F[i,j] # Leave diagonal unchanged
    return S
```

4.4 Simple Shrinkage Estimator

The simple shrinkage estimator interpolates linearly between the values of the values of the sample covariance matrix F and the diagonal covariance matrix S.

$$\Sigma = (1-\delta)S + \delta F \tag{2}$$

A shrinkage factor δ of 0.5 is used.

```
def simple_shrinkage_estimator(X, shrinkage = 0.5):
    F = sample_covariance_matrix(X)
```

```
    S = diagonal_covariance_matrix(X)
    return (1-shrinkage)*F + shrinkage*S
```

4.5 Constant correlation shrinkage estimator

This estimator is similar to the simple shrinkage estimator, but rather than shrinking towards the diagonal covariance matrix, it uses the constant correlation matrix.

4.6 Ledoit-Wolf Shrinkage Estimator

The Ledoit-Wolf estimator is an extension of the simple shrinkage estimator where an optimal shrinkage factor is estimated based on the input data. The estimator is designed to be computationally efficient and provides a robust estimate, especially when the number of variables is large relative to the number of observations.

```
from sklearn.covariance import LedoitWolf
def ledoit_wolf_shrinkage_estimator(X):
    cov = LedoitWolf().fit(X.T)
    return cov.covariance_
```

The scikit-learn implementation is used for simplicity.

4.7 Oracle Shrinkage Estimator

Similar to the Ledoit-Wolf estimator, the Oracle estimator extends the simple shrinkage estimator by estimating an optimal shrinkage factor. While the Ledoit-Wolf method typically uses a diagonal matrix as the target, assuming most variables are uncorrelated, the Oracle method allows for more flexibility by incorporating a target matrix derived from prior information or assumptions about the data structure. The Oracle Covariance Estimator also explicitly optimizes the shrinkage parameter to minimize the mean squared error, taking into account the target matrix. These extensions make it more computationally intensive that the Ledoit-Wolf method.

```
from sklearn.covariance import OAS
def ledoit_wolf_shrinkage_estimator(X):
    cov = OAS().fit(X.T)
    return cov.covariance_
```

The scikit-learn implementation is used for simplicity.

4.8 Out of scope methods

The below methods were not analysed, but would be interesting to include in any further analysis.

Factor Models: Factor models are a class of models used in covariance matrix estimation that assume the observed variables can be decomposed into a linear combination of common factors and idiosyncratic, or specific, factors. The factor model is described as a linear regression of the assets X on their means μ and the common factors F. The residuals U are the idiosyncratic factors

$$X = \mu + BF + U \tag{3}$$

The covariance matrix of the observed variables X can be expressed as:

$$\Sigma = B\Lambda B' + \Psi \tag{4}$$

where

- Σ is the covariance matrix of X.
- Λ is a diagonal matrix containing the variances of the common factors.
- Ψ is a diagonal matrix containing the variances of the idiosyncratic factors.

If the factors are assumed to completely explain the relationships between the assets Ψ and U can be assumed to be zero.

5 Results

The impact of the covariance measures are measured through their impact of portfolio return. The time series of portfolio returns for the N asset universe show that the impact on return from the covariance measures are small relative to the overall market movements, as in Figure 1.
Analysing aggregate returns for all covariance measures and asset universe in Table 1 shows that despite returns being highly correlated between the covariance measures, there are significant differences between aggregate returns for a given asset universe.

Method	3	5	10	20	30	40
sample covariance matrix	201.4	201.4	197.8	102.0	156.5	205.3
diagonal covariance matrix	184.7	144.4	119.3	101.5	133.7	140.3
constant correlation estimator	191.3	196.4	180.8	93.0	159.5	200.5
simple shrinkage estimator	192.2	187.3	161.1	110.8	184.8	213.5
constant correlation shrinkage estimator	196.2	198.9	190.0	97.0	163.6	210.6
ledoit-wolf shrinkage estimator	202.0	202.0	195.7	107.2	161.9	208.5
oracle approximating shrinkage estimator	201.7	201.7	197.1	105.6	158.8	207.6

Table 1: Aggregate returns by covariance measure, for each asset universe size, percent

Comparing aggregate returns to the sample covariance baseline provides the following insights:

- The Ledoit-Wolf shrinkage estimator provides superior performance over almost all asset universe sizes.
- The simplified diagonal and constant correlation estimators do not provide increased performance.

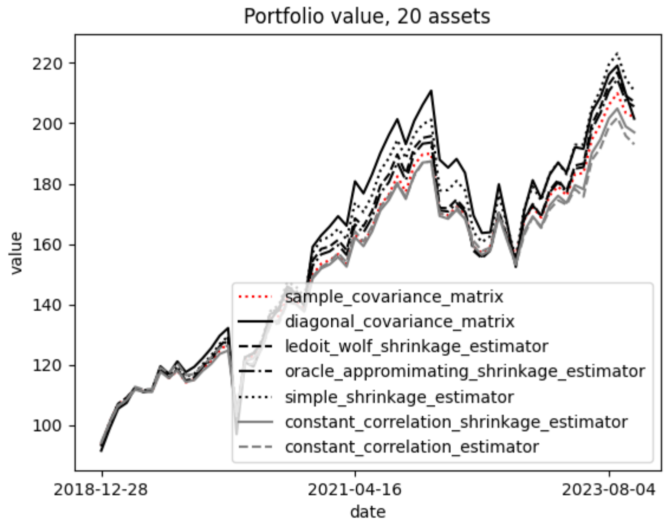

Figure 1: Portfolio value over 4 years for the 20 asset universe, by covariance measure

- When combined with the sample covariance matrix through shrinkage, the diagonal and constant correlation estimators provide better performance when the number of assets is high.
- All shrinkage estimators provide better performance on average when the number of assets is larger. This could be due to the error in the sample covariance matrix increasing with the number of assets, leading to larger gains through the error reduction that shrinkage provides

The relative performance of each covariance estimator compared to the sample covariance matrix can be seen in the time series of the difference in portfolio values between each measure and the sample covariance matrix in Figure 3 and Figure 2. The time series show how each covariance estimator performed well or poorly with fairly high consistency over the simulation.

The consistency of over or under-performance of each covariance estimator can be explored further by looking at the proportion of simulation time steps or investing decisions where each measure outperformed the sample covariance matrix, as in Table 2. This analysis shows that the Ledoit-Wolf and Oracle estimators outperformed the baseline in the majority of the decisions, over all asset universe sizes. The simple shrinkage estimator and the constant correlation estimator were less consistent and as in the aggregate returns analysis performed better on larger asset universe sizes.

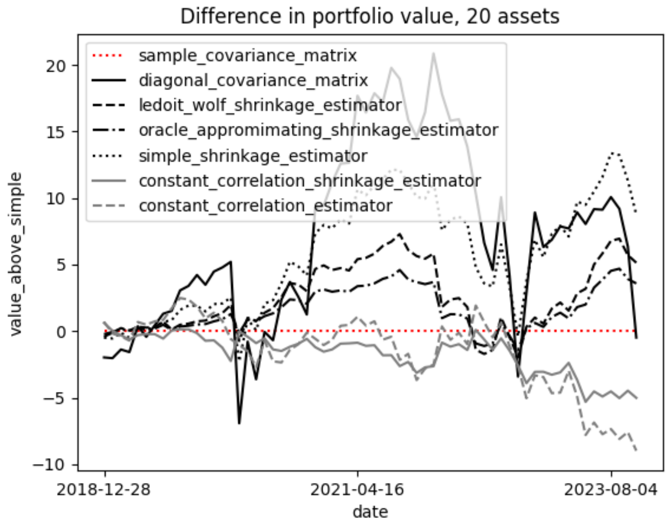

Figure 2: Increase in portfolio value compared to the sample covariance matrix baseline, 20 assets

Method	3	5	10	20	30	40
diagonal covariance matrix	48.4	35.9	35.9	51.6	48.4	45.3
constant correlation estimator	51.6	45.3	46.9	46.9	50.0	53.1
simple shrinkage estimator	48.4	37.5	35.9	50.0	62.5	51.6
constant correlation shrinkage estimator	50.0	50.0	46.9	48.4	50.0	53.1
ledoit wolf shrinkage estimator	57.8	57.8	50.0	64.1	62.5	56.2
oracle approximating shrinkage estimator	57.8	57.8	51.6	62.5	62.5	57.8

Table 2: Proportion of periods where returns were better than the sample covariance matrix, percent. Values over 50 show out-performance

6 Conclusions

6.1 Findings

There are a number of simple conclusions that can be drawn from these limited results, in the context of the usage of the covariance matrix in portfolio optimisation:

Sample covariance is usually inferior: The sample covariance matrix generally provided worse performance than the shrinkage estimators.

Shrinkage targets alone are not effective: The shrinkage targets themselves, the diagonal and constant correlation matrices, did not provide reliable benefits over the sample covariance matrix

Ledoit-Wolf and oracle estimators are effective: Ledoit-Wolf and oracle estimators provided fairly consistent out performance compared to the sample covariance matrix.

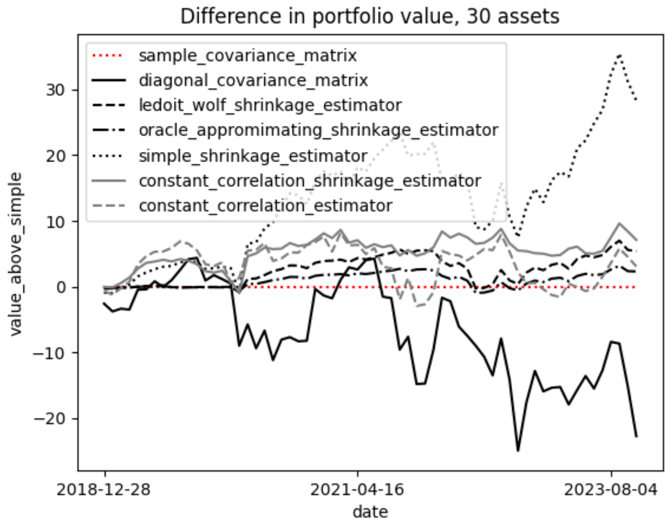

Figure 3: Increase in portfolio value compared to the sample covariance matrix baseline, 30 assets

6.2 Areas for further research

Factor Models Comparing the factor model covariance matrix to the estimators explored here.

Multi-asset cases Asset allocation over many asset classes provides a similar portfolio optimisation use case, but with an increase importance on estimating negative correlations, for example between equities and bonds. This will likely lead to measures such as the constant correlation estimator to significantly under-perform.

Other use cases Covariance matrices are used in many other applications outside of portfolio optimisation, which provide other avenues to investigate the impact of using different estimators.