Estimation of Covariance Matrix

Refael Lasry

February 2017

Contents

1 Introduction

1.1 Covariance Matrix

Let X be a $p \times 1$ random vector, where each entry $X_1, X_2, ..., X_p$ is a random variable.

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_j \\ \vdots \\ X_p \end{bmatrix}$$

The $p \times p$ covariance matrix Σ defined as follow,

$$\Sigma = Cov(X) := E[[X - E(X)][X - E(X)]^T] = \begin{bmatrix} (X_1 - E(X_1))^2 & ... & (X_1 - E(X_1))(X_p - E(X_p)) \\ (X_2 - E(X_2))(X_1 - E(X_1)) & ... & (X_2 - E(X_2))(X_p - E(X_p)) \\ \vdots & \ddots & \vdots \\ (X_p - E(X_p))(X_1 - E(X_1) & ... & (X_p - E(X_p))^2 \end{bmatrix}$$

it can be seen that,

$$\Sigma = Cov(X) = E[[X - E(X)][X - E(X)]^T] = E[[X - E[X]][X^T - E[X]^T]]$$

$$= E[XX^T - XE(X)^T - E(X)X^T + E(X)E(X)^T]$$

$$= E[XX^T] - E[XE(X)^T] - E[E(X)X^T] + E[E(X)E(X)^T]$$

$$= E(XX^T) - E(X)E(X)^T - E(X)E(X)^T + E[X]E[X]^T$$

$$= E[XX^T] - E[X]E[X]^T$$

The covariance matrix Σ sometimes called the population covariance matrix and the estimator $\hat{\Sigma}$ sometimes called sample covariance matrix. Using the phrase "population" comes to emphasis that Σ is containing the variance and the covariance of the "real world", while S is based on samples from the random vector X.

1.2 MLE Estimator

Given n samples from the random vector X,

$$X_1 = \begin{bmatrix} x_{11} \\ x_{12} \\ \vdots \\ x_{1p} \end{bmatrix}, X_2 = \begin{bmatrix} x_{21} \\ x_{22} \\ \vdots \\ x_{2p} \end{bmatrix}, \dots, X_n = \begin{bmatrix} x_{n1} \\ x_{n2} \\ \vdots \\ x_{np} \end{bmatrix}$$

each entry x_{ij} is representing the sample i of the parameter j.

Putting all the vectors together we get the sample matrix X.

$$X = \begin{bmatrix} | & | & & | \\ X_1 & X_2 & \cdots & X_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} x_{11} & x_{21} & \dots & x_{n1} \\ x_{12} & x_{22} & \dots & x_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1p} & x_{2p} & \dots & x_{np} \end{bmatrix}$$

The MLE unbiased estimator for the covariance matrix Σ, denote by S is,

$$S = \frac{1}{N-1} X (I - \frac{1}{N} \mathbf{1}\mathbf{1}') X'$$

where $\mathbf{1}$ is $N \times 1$ vector, $\mathbf{1} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$ and I is $p \times p$ identity matrix.

Number Of Parameters To Estimate including the diagonal is $\frac{p(p+1)}{2}$.

1.3 Centering The Sample Matrix

Removing the average happens all the time while computing the estimator for the covariance matrix. To make life easier, from now and on, the sample matrix X assumed to be centered. From each variable the average of the parameter is removed, x_{ij} will become $x_{ij} - \frac{1}{n}\sum_{i=1}^{n} x_{ij}$

$$X = \begin{bmatrix} x_{11} - \bar{x}_{\cdot 1} & x_{21} - \bar{x}_{\cdot 1} & \dots & x_{n1} - \bar{x}_{\cdot n} \\ x_{12} - \bar{x}_{\cdot 2} & x_{22} - \bar{x}_{\cdot 2} & \dots & x_{n2} - \bar{x}_{\cdot p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1p} - \bar{x}_{\cdot p} & x_{2p} - \bar{x}_{\cdot p} & \dots & x_{np} - \bar{x}_{\cdot p} \end{bmatrix}$$

The MLE estimator will get the form of, $S = E[XX^T] - 0 = E[XX^T]$.

1.4 Estimation while $n < p$

An important case of the estimation is where the number of samples is smaller then the number of the parameters ($n < p$). The MLE estimator has bad properties and a solution is needed. Some of the solutions are tapering, regularizations, linear and non-linear shrinkage .The following section will review the solution of linear shrinkage.

2 Shrinkage Estimation

First method to discuss is shrinkage, it will review through the article **ledoit˙wolf˜2003**. Shrinkage method is combining two estimators unbiased and bias with a proper weight. The bias estimator called shrinkage target. The estimator in the article is combined from the unbiased MLE and the biased estimator of the covariance matrix of the Single Index Model. The main part of the article is to find the optimal shrinkage parameter, it done by using asymptotic theory.

2.1 Shrinkage Target - Covariance of Single Index Model

Collection of assets is called a portfolio, each one of them has a return in some period of time. The single index model try to describe the components which affect on the return of assets in a portfolio. The idea of the model is that each return is effected from the total return of the portfolio. Basically it's a linear regression. For a specific period of time t, the return of an asset in time t is the target variable and the total return of the portfolio in time t is the explanatory variable.

2.1.1 Single-index Model

Continue the notation from the introduction, x_{ij} is observation i of the return of asset j. The number of the observations is n and the number of the parameters is p. The vector $X_i^T = [x_{i1}, x_{i2}, \dots, x_{ip}]$ is the observation i of p returns from the portfolio.

The model assumed that x_{ij} are generated by,

$$x_{ij} = \alpha_j + \beta_j x_{i\cdot} + e_{ij}$$

$$\forall j = 1, 2, ..., p; \forall i = 1, 2, ..., n$$

$$Var(e_{ij}) := \delta_{jj}$$

the variance of the residuals δ_{jj} assumed to be uncorrelated with x_{ij} and to one another.

Each return x_{ij} is regress on $x_{i\cdot} = \sum_{j=1}^{p} x_{ij}$ the market return of observation i. The return of the market is computed for each observation i,

$$x_{i\cdot} = [x_{1\cdot}, x_{2\cdot}, \dots, x_{n\cdot}] = \left[\sum_{j=1}^{p} x_{1j}, \sum_{j=1}^{p} x_{2j}, ..., \sum_{j=1}^{p} x_{nj}\right]$$

The model get $n \times p$ observed matrix,

$$X = \begin{bmatrix} x_{11} & x_{21} & \dots & x_{n1} \\ x_{12} & x_{22} & \dots & x_{n2} \\ \vdots & \vdots & \dots & \vdots \\ x_{1p} & x_{2p} & \dots & x_{np} \end{bmatrix}$$

and replace each entry by the linear equations,

$$X_{\hat{SIM}} = \begin{bmatrix} \alpha_1 + \beta_1 x_{1\cdot} + e_{11} & \alpha_1 + \beta_1 x_{2\cdot} + e_{21} & \dots & \alpha_1 + \beta_1 x_{n\cdot} + e_{n1} \\ \alpha_2 + \beta_2 x_{1\cdot} + e_{12} & \alpha_2 + \beta_2 x_{2\cdot} + e_{22} & \dots & \alpha_2 + \beta_2 x_{n\cdot} + e_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_p + \beta_p x_{1\cdot} + e_{1p} & \alpha_p + \beta_p x_{2\cdot} + e_{2p} & \dots & \alpha_p + \beta_p x_{n\cdot} + e_{np} \end{bmatrix}$$

Variance of asset j in observation i

$$\phi_{jj} := Var(x_{ij}) = Var(\alpha_j + \beta_j x_{i\cdot} + e_{ij}) = Var(\beta_j x_{i\cdot} + e_{ij}) = Var(\beta_j x_{i\cdot}) + Var(e_{ij}) + 2Cov(\beta_j x_{i\cdot}, e_{ij})$$

$$= Var(\beta_j x_{i\cdot}) + Var(e_{ij}) = \beta_j^2 Var(x_{i\cdot}) + \delta_{jj} = \beta_j^2 \sigma_{00}^2 + \delta_{jj}$$

while σ_{00}^2 is the variance of $x_{i\cdot}$

Covariance of asset j and asset k

$$\phi_{jk} := Cov(x_{ij}, x_{ik}) = Cov(\alpha_j + \beta_j x_{i\cdot} + e_{ij}, \alpha_k + \beta_k x_{i\cdot} + e_{ik}) = Cov(\beta_j x_{i\cdot} + e_{ij}, \beta_k x_{i\cdot} + e_{ik})$$

$$= Cov(\beta_j x_{i\cdot}, \beta_k x_{i\cdot}) + Cov(\beta_j x_{i\cdot}, e_{ik}) + Cov(e_{ij}, \beta_k x_{i\cdot}) + Cov(e_{ij}, e_{ik}) = Cov(\beta_j x_{i\cdot}, \beta_k x_{i\cdot})$$

$$= \beta_j \beta_k Var(x_{i\cdot}) = \beta_j \beta_k \sigma_{00}^2$$

It can be seen that the covariance is not depend on the observation.

Matrix form Writing together the variance of x_{ij} and the covariance of x_{ij} with x_{ik} in matrix form it obtain the covariance matrix of the model,

$$\Phi = \sigma_{00}^2 \boldsymbol{\beta}\boldsymbol{\beta}' + \Delta$$

while, $\boldsymbol{\beta} = [\beta_1, \beta_2, ..., \beta_i, .., \beta_p]^T$, and Δ is a diagonal matrix containing residual variances δ_{jk}

$$\Phi = \begin{bmatrix} \phi_{11} & \phi_{12} & \dots & \phi_{1j} & \dots & \phi_{1p} \\ \phi_{21} & \phi_{22} & \dots & \phi_{2j} & \dots & \phi_{2p} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \phi_{i1} & \phi_{i2} & \dots & \phi_{ij} & \dots & \phi_{ip} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \phi_{p1} & \phi_{p2} & \dots & \phi_{pj} & \dots & \phi_{pp} \end{bmatrix} + \Delta = Var(x_{i\cdot}) \begin{bmatrix} \beta_1\beta_1 & \beta_1\beta_2 & \dots & \beta_1\beta_j & \dots & \beta_1\beta_p \\ \beta_2\beta_1 & \beta_2\beta_2 & \dots & \beta_2\beta_j & \dots & \beta_2\beta_p \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \beta_j\beta_1 & \beta_j\beta_2 & \dots & \beta_j\beta_j & \dots & \beta_j\beta_p \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \beta_p\beta_1 & \beta_p\beta_2 & \dots & \beta_p\beta_j & \dots & \beta_p\beta_p \end{bmatrix} + \Delta$$

$$= \frac{1}{Var(x_{i\cdot})} \begin{bmatrix} Cov(x_{i\cdot}, x_{i1})Cov(x_{i\cdot}, x_{i1}) & \dots & Cov(x_{i\cdot}, x_{i1})Cov(x_{i\cdot}, x_{ij}) & \dots & Cov(x_{i\cdot}, x_{i1})Cov(x_{i\cdot}, x_{ip}) \\ Cov(x_{i\cdot}, x_{i2})Cov(x_{i\cdot}, x_{i1}) & \dots & Cov(x_{i\cdot}, x_{i2})Cov(x_{i\cdot}, x_{ij}) & \dots & Cov(x_{i\cdot}, x_{i2})Cov(x_{i\cdot}, x_{ip}) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ Cov(x_{i\cdot}, x_{ij})Cov(x_{i\cdot}, x_{i1}) & \dots & Cov(x_{i\cdot}, x_{ij})Cov(x_{i\cdot}, x_{ij}) & \dots & Cov(x_{i\cdot}, x_{ij})Cov(x_{i\cdot}, x_{ip}) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ Cov(x_{i\cdot}, x_{ip})Cov(x_{i\cdot}, x_{i1}) & \dots & Cov(x_{i\cdot}, x_{ip})Cov(x_{i\cdot}, x_{ij}) & \dots & Cov(x_{i\cdot}, x_{ip})Cov(x_{i\cdot}, x_{ip}) \end{bmatrix} + \Delta$$

while Δ is,

$$\Delta = \begin{bmatrix} \delta_{11} & 0 & \dots & 0 & \dots & 0 \\ 0 & \delta_{22} & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \delta_{kk} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & \dots & \delta_{pp} \end{bmatrix}$$

2.1.2 Fitting the Model

The model is fitted by running a simple linear regression for each return j separately. In order to understand how the process is done, let's look on a specific return $j = 1$.

Each return has n samples $x_{1,1}, x_{2,1}, ..., x_{n,1}$ following linear equation of the model for each samples is,

$$x_{i1} = \alpha_1 + \beta_1 x_{i\cdot} + e_{1j} \quad \forall i = 1, 2, \dots, n$$

the target variable is x_{i1}, the explanatory variable is $x_{i\cdot}$ the slope of the regression is β_1, the constant of the regression is α_1.

Since there is n samples, there is n realizations of the linear equation. Similarly to simple linear regression the slope can be estimated by,

$$\beta_1 = \frac{Cov(x_{i\cdot}, x_{i1})}{Var(x_{i\cdot})}$$

the same fitting procedure that done for $j = 1$ is done for each return $j = 1, 2, \dots, p$, the vector that contain all the slopes of the regression is $\boldsymbol{\beta}$.

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} = \begin{bmatrix} \frac{Cov(x_{i\cdot}, x_{i1})}{Var(x_{i\cdot})} \\ \frac{Cov(x_{i\cdot}, x_{i2})}{Var(x_{i\cdot})} \\ \vdots \\ \frac{Cov(x_{i\cdot}, x_{ip})}{Var(x_{i\cdot})} \end{bmatrix}$$

The estimation of $\boldsymbol{\beta}$ is denote by $\boldsymbol{b} = (b_1, b_2, .., b_n)$. The estimation of $\Phi = \sigma_{00}^2 \boldsymbol{\beta}\boldsymbol{\beta'} + \Delta$ is denote by $\boldsymbol{F} = s_{00}^2 bb' + D$ where s_{00}^2 is the estimated variance of $x_{i\cdot}$ and D is a diagonal matrix with entries of the variance of the residuals.

2.1.3 Covariance matrix of the single index model

Focusing on specific entry, may help to understand the nature of Φ.

$$\phi_{21} = \frac{1}{Var(x_{i\cdot})} Cov(x_{i\cdot}, x_{i2}) Cov(x_{i\cdot}, x_{i1})$$

$x_{i\cdot} = [x_{1\cdot}, x_{2\cdot}, \ldots, x_{n\cdot}]$ - each element is sum of all the returns in different observation (time).

$x_{i2} = [x_{12}, x_{22}, \ldots, x_{n2}]$ - each element of the vector is return of asset 2 in different observation (time)

$x_{i1} = [x_{11}, x_{21}, \ldots, x_{n1}]$ - each element of the vector is return of asset 1 in different observation (time)

$Cov(x_{i\cdot}, x_{i2})$ - covariance of asset 2 and the total returns.

$Cov(x_{i\cdot}, x_{i1})$ - covariance of asset 1 and the total returns.

$Var(x_{i\cdot})$ - variance of the total returns.

Now, it possible to see what ϕ_{21} is capturing. The numerator is the covariance of asset 1 with total returns multiply by the covariance of asset 2 with total returns. The denominator is the variance of total returns. ϕ_{21} is capturing the share of asset 1 and asset 2 variances from the total variance of the returns.

Number Of Parameters To Estimate For each asset, there is one parameter to estimate and one parameter of the index, total of $p + 1$ parameters to estimate. Comparing to the MLE with $\frac{(p+1)}{2}p$ parameters to estimate, the SIM has less parameters to estimate, therefore the variance expected to be lower from the variance of S.

2.1.4 Shrinkage Estimator

The suggested improved estimator is done by shrinkage method. In order to perform this method, two estimators is needed. The first is the MLE unbiased estimator S. The second estimator is the shrinkage target, it will be the covariance matrix of the single index model F. The weight between the estimators is, $0 < \alpha < 1$.

$$\alpha\boldsymbol{F} + (1-\alpha)\,\boldsymbol{S}$$

2.1.5 The Optimal Shrinkage Intensity

Choosing the optimal shrinkage parameter is done by minimizing the risk function.

The loss function is define with Forbenius norm,

$$L(\alpha) = \|\alpha\boldsymbol{F} + (1-\alpha)\,\boldsymbol{S} - \Sigma\|_f^2$$

The risk function,

$$R(\alpha) = E(L(\alpha)) = E\left[\|\alpha\boldsymbol{F} + (1-\alpha)\,\boldsymbol{S} - \Sigma\|_f^2\right]$$

$$= E\left[\sum_{i=1}^{p}\sum_{j=1}^{p}(\alpha f_{ij} + (1-\alpha)s_{ij} - \sigma_{ij})^2\right] = \sum_{i=1}^{p}\sum_{j=1}^{p}E[(\alpha f_{ij} + (1-\alpha)s_{ij} - \sigma_{ij})^2]$$

$$= \sum_{i=1}^{p}\sum_{j=1}^{p}Var(\alpha f_{ij} + (1-\alpha)s_{ij} - \sigma_{ij}) + (E\left[\alpha f_{ij} + (1-\alpha)s_{ij} - \sigma_{ij}\right])^2$$

$$= \sum_{i=1}^{p}\sum_{j=1}^{p}Var(\alpha f_{ij} + (1-\alpha)s_{ij}) + [\alpha E(f_{ij}) + E(s_{ij}) - \alpha E(s_{ij}) - E(\sigma_{ij})]^2$$

$$= \sum_{i=1}^{p}\sum_{j=1}^{p}Var(\alpha f_{ij}) + Var((1-\alpha)s_{ij})) + 2Cov(\alpha f_{ij}, (1-\alpha)s_{ij})) + [\alpha\phi_{ij} + \sigma_{ij} - \alpha\sigma_{ij} - \sigma_{ij}]^2$$

$$= \sum_{i=1}^{p}\sum_{j=1}^{p}\alpha^2 Var(f_{ij}) + (1-\alpha)^2 Var(s_{ij}) + 2\alpha(1-\alpha)Cov(f_{ij}, s_{ij}) + \alpha^2(\phi_{ij} - \sigma_{ij})^2$$

first derivation,

$$R'(\alpha) = 2\sum_{i=1}^{p}\sum_{j=1}^{p}\alpha Var(f_{ij}) - (1-\alpha)Var(s_{ij}) + (1-2\alpha)Cov(f_{ij}, s_{ij}) + \alpha(\phi_{ij} - \sigma_{ij})^2$$

$$= 2\sum_{i=1}^{p}\sum_{j=1}^{p}\alpha\left[Var(f_{ij}) + Var(s_{ij}) - 2Cov(f_{ij}, s_{ij}) + (\phi_{ij} - \sigma_{ij})^2\right] - Var(s_{ij}) + Cov(f_{ij}, s_{ij})$$

setting $R'(\alpha)$ to zero, the α that bring $R(\alpha)$ to minimum is,

$$\alpha^* = \frac{\sum_{i=1}^{p}\sum_{j=1}^{p} Var(s_{ij}) - Cov(f_{ij}, s_{ij})}{\sum_{i=1}^{p}\sum_{j=1}^{p} Var(f_{ij} - s_{ij}) + (\phi_{ij} - \sigma_{ij})^2}$$

second derivation,

$$R''(\alpha) = 2\sum_{i=1}^{p}\sum_{j=1}^{p} Var(f_{ij}) + Var(s_{ij}) - 2Cov(f_{ij}, s_{ij}) + (\phi_{ij} - \sigma_{ij})^2$$

$R''(\alpha) > 0$, therefore it indeed a minimum point.

2.1.6 Shrinkage Parameter

The next step is using asymptotic theory and finding the $\alpha*$ without unobserved components.

$$n\alpha^* = \frac{\sum_{i=1}^{p}\sum_{j=1}^{p} Var(\sqrt{n}s_{ij}) - Cov(\sqrt{n}f_{ij}, \sqrt{n}s_{ij})}{\sum_{i=1}^{p}\sum_{j=1}^{p} Var(f_{ij} - s_{ij}) + (\phi_{ij} - \sigma_{ij})^2}$$

Theorem 1 Let the following denotation,

$$\pi = \sum_{i=1}^{p}\sum_{j=1}^{p}\lim_{n\to\infty}\sum_{n=1}^{n} Var(\sqrt{n}s_{ij})$$

while, $s_{ij} = \frac{1}{N}\sum_{n=1}^{N} x_{in}x_{jn}$

$$\rho = \sum_{i=1}^{p}\sum_{j=1}^{p}\lim_{N\to\infty}\sum_{n=1}^{N} Cov(\sqrt{N}f_{ij}, \sqrt{N}s_{ij})$$

$$\gamma = \sum_{i=1}^{p}\sum_{j=1}^{p}(\phi_{ij} - \sigma_{ij})^2$$

then the optimal shrinkage α^* satisfies,

$$\alpha^* = \frac{1}{N}\frac{\pi - \rho}{\gamma} + O\left(\frac{1}{N^2}\right)$$

Lemma 1 A consist estimator for π

$$\hat{\pi} = \sum_{i=1}^{p} \sum_{j=1}^{p} \hat{\pi}_{ij}$$

$$\hat{\pi}_{ij} = \frac{1}{N} \sum_{n=1}^{N} \left(x_{in} x_{jn} - s_{ij}\right)^2$$

Lemma 2 A consist estimator for ρ

$$\hat{\rho} = \sum_{i=1}^{p} \sum_{j=1}^{p} \hat{\rho}_{ij}$$

for $i = i$,

$$\hat{\rho}_{ii} = \hat{\pi}_{ii}$$

for $i \neq j$,

$$\hat{\rho}_{ij} = \frac{1}{N} \sum_{n=1}^{N} \hat{\rho}_{ijn}$$

$$\hat{\rho}_{ijn} = \frac{s_{j\cdot} s_{\cdot\cdot} x_{in} + s_{i\cdot} s_{\cdot\cdot} x_{jn} - s_{i\cdot} s_{j\cdot} \left(x_{\cdot n} - x_{\cdot\cdot}\right)}{s_{\cdot\cdot}^2} \left(x_{\cdot n} - x_{\cdot\cdot}\right) x_{in} x_{jn} - f_{ij} s_{ij}$$

while,

$$s_{\cdot\cdot}^2 = var(x_{i\cdot})$$

$$s_{k\cdot} = cov(x_{i\cdot}, x_{ik})$$

Lemma 3 A consist estimator for γ_{ij}

$$\hat{\gamma} = \sum_{i=1}^{p} \sum_{j=1}^{p} \left(f_{ij} - s_{ij}\right)^2$$

2.1.7 The Estimator

Combine theorem 1, lemmas 1 2 and 3 :

$$\frac{k}{N} \boldsymbol{F} + \left(1 - \frac{k}{N}\right) \boldsymbol{S}$$

where, $k = \frac{\hat{\pi} - \hat{\rho}}{\hat{\gamma}}$ is called the constant shrinkage parameter.

2.2 Simulations

Purpose of the shrinkage method is to gain better result from MLE the usual estimator. Testing the estimators require realizations of the random vector X, population covariance matrix of X and evaluation system.

2.2.1 Evaluation Of The Estimator

Given a true covariance matrix and an estimator $\Sigma, \hat{\Sigma}$ the following can be computed,

for specific entry,

$$\hat{\hat{\Sigma}}_{ij} = \frac{1}{L}\sum_{l=1}^{L}\hat{\Sigma}_{ij}$$

where L is the number of repetitions.

$$Bias^2(\hat{\Sigma}_{ij}) = \frac{1}{L}\sum_{l=1}^{L}\left(\hat{\Sigma}_{ij}^{l} - \Sigma_{ij}\right)^2$$

$$Var(\hat{\Sigma}_{ij}) = \sum_{l=1}^{L}\left(\hat{\Sigma}_{ij}^{l} - \hat{\hat{\Sigma}}_{ij}\right)^2$$

$$EMSE(\hat{\Sigma}_{ij}) = Bias^2(\hat{\Sigma}_{ij}) + Var(\hat{\Sigma}_{ij})$$

for the entire matrix,

$$Bias^2(\hat{\Sigma}) = \sum_{i=1}^{p}\sum_{j=1}^{p} Bias^2(\hat{\Sigma}_{ij})$$

$$Var(\hat{\Sigma}) = \sum_{i=1}^{p}\sum_{j=1}^{p} Var(\hat{\Sigma}_{ij})$$

$$EMSE(\hat{\Sigma}) = Bias^2(\hat{\Sigma}) + Var(\hat{\Sigma})$$

2.2.2 Data

Market Model Generating data according to simple market model,

$$r_{i,t} = \beta_i \cdot r_{M,t} + \epsilon_{i,j}$$

$$\beta_1 + \beta_2 + ... + \beta_p = 1$$

where $r_{i,t}$ is the return of asset i on time t. $r_{M,t}$ is the total return of the market in time t. $\epsilon_{i,j}$ is normal random variable.

Multi-Normal distribution Generating data with known covariance matrix which will take the role of the true covariance matrix in the simulations.

2.2.3 Results

Simulations details:

	S&P 500	**Market Model**
number of simulations	10	100
number of variables - p	491	100
number of observations - N	250	100
sample size	0.8	0.6
$\frac{p}{n}$	2.455	1.66
π - error on the sample covariance matrix	2.6316	63.4444
ρ - covariance between the estimation error of MLE and SIM	2.5782	62.2252
γ - misspecification of the single-index model	0.0775	1.0280
shrinkage constant k	0.6126	1.162
shrinkage parameter α	0.0030	0.01937

Evaluation of the estimators:

	S&P 500		**Market Model**	
	MLE	shrinkage	MLE	shrinkage
$Bias^2(\hat{\Sigma})$	0.0533	0.0477	0.6759	0.6583
$Var(\hat{\Sigma})$	0.0029	$3.4607 \cdot 10^{-7}$	0.4023	0.0001
$EMSE(\hat{\Sigma})$	0.05637	0.0477	1.0700	0.6584

Variance the shirnkage target vs variance of MLE:

	S&P 500	**Market Model**
variance of MLE	0.0028	0.4151
variance of shrinkage target	$3.2811 \cdot 10^{-7}$	$6.9581 \cdot 10^{-5}$

Known covariance with multi-normal distribution .

The result were not good enough. For $\frac{p}{n} > 1$ the shrinkage parameter was above 1.

2.3 Discussion

The shrinkage constant converge to zero as the number of observations goes up.

The method has good performance only when the eigenvectors are closed.