

Prior Information: Shrinkage and Black-Litterman

Prof. Daniel P. Palomar

The Hong Kong University of Science and Technology (HKUST)

MAFS5310 - Portfolio Optimization with R

MSc in Financial Mathematics

Fall 2020-21, HKUST, Hong Kong

Outline

1 The Need for Prior Information

2 Shrinkage

- Shrinkage for μ
- Shrinkage for Σ
- Random Matrix Theory (RMT)

3 Black-Litterman Model

Outline

1 The Need for Prior Information

2 Shrinkage

- Shrinkage for μ
- Shrinkage for Σ
- Random Matrix Theory (RMT)

3 Black-Litterman Model

Returns

- Let us denote the log-returns of N assets at time t with the vector $r_t \in \mathbb{R}^N$.
- The time index t can denote any arbitrary period such as days, weeks, months, 5-min intervals, etc.
- \mathcal{F}_{t-1} denotes the previous historical data.
- Financial modeling aims at modeling r_t conditional on \mathcal{F}_{t-1} .
- r_t is a multivariate stochastic process with **conditional** mean and covariance matrix denoted as¹

$$\mu_t \triangleq \mathbb{E}[r_t \mid \mathcal{F}_{t-1}]$$

$$\Sigma_t \triangleq \text{Cov}[r_t \mid \mathcal{F}_{t-1}] = \mathbb{E}[(r_t - \mu_t)(r_t - \mu_t)^T \mid \mathcal{F}_{t-1}].$$

¹Y. Feng and D. P. Palomar, *A Signal Processing Perspective on Financial Engineering*. Foundations and Trends[®] in Signal Processing, Now Publishers Inc., 2016.

I.I.D. Model

- For simplicity we will assume that r_t follows an i.i.d. distribution (which is not very inaccurate in general).
- That is, both the conditional mean and conditional covariance are constant

$$\begin{aligned}\mu_t &= \mu, \\ \Sigma_t &= \Sigma.\end{aligned}$$

- Very simple model, however, it is one of the most fundamental assumptions for many important works, e.g., the Nobel prize-winning Markowitz portfolio theory².

²H. Markowitz, "Portfolio selection", *J. Financ.*, vol. 7, no. 1, pp. 77–91, 1952.

Sample Estimators

- Consider the i.i.d. model:

$$\mathbf{r}_t = \boldsymbol{\mu} + \mathbf{w}_t,$$

where $\boldsymbol{\mu} \in \mathbb{R}^N$ is the mean and $\mathbf{w}_t \in \mathbb{R}^N$ is an i.i.d. process with zero mean and constant covariance matrix $\boldsymbol{\Sigma}$.

- The sample estimators (i.e., sample mean and sample covariance matrix) based on T observations are

$$\hat{\boldsymbol{\mu}} = \frac{1}{T} \sum_{t=1}^T \mathbf{r}_t$$
$$\hat{\boldsymbol{\Sigma}} = \frac{1}{T-1} \sum_{t=1}^T (\mathbf{r}_t - \hat{\boldsymbol{\mu}})(\mathbf{r}_t - \hat{\boldsymbol{\mu}})^T.$$

- Note that the factor $1/(T-1)$ is used instead of $1/T$ to get an unbiased estimator (asymptotically for $T \rightarrow \infty$ they coincide).

So What is the Problem?

- The sample estimates are only good for large T .
- The sample mean is particularly a very inefficient estimator, with very noisy estimates.³
- In practice, T is not large enough due to either:
 - unavailability of data
 - lack of stationarity of data which precludes the use of too much of it.
- As a consequence, the sample estimates are really bad due to estimator errors and a portfolio design (e.g., Markowitz mean-variance) based on those estimates can be fatal.
- Indeed, this is why Markowitz portfolio and similar are rarely used by practitioners.
- One solution is to merge those estimates with whatever prior information we may have on μ and Σ .

³A. Meucci, *Risk and Asset Allocation*. Springer, 2005.

Factor Models

- Factor models can be seen as a way to include some prior information either based on explicit factors or some low-rank structural constraints on the covariance matrix.
- Recall that factor models assumes the following structure for the returns:

$$r_t = \alpha + Bf_t + w_t,$$

where

- α denotes a constant vector
- $f_t \in \mathbb{R}^K$ with $K \ll N$ is a vector of a few factors that are responsible for most of the randomness in the market,
- $B \in \mathbb{R}^{N \times K}$ denotes how the low dimensional factors affect the higher dimensional market;
- w_t is a white noise residual vector that has only a marginal effect.
- The factors can be explicit or implicit.
- Widely used by practitioners (they buy factors at a high premium).
- Observe that the covariance matrix will be of the form of a low-rank matrix plus some residual diagonal matrix: $\Sigma = BB^T + \Psi$.

Outline

1 The Need for Prior Information

2 Shrinkage

- Shrinkage for μ
- Shrinkage for Σ
- Random Matrix Theory (RMT)

3 Black-Litterman Model

Small Sample Regime

- In the large sample regime, i.e., when the number of observations T is large, then the estimators of μ and Σ are already good enough.
- However, in the small sample regime, i.e., when the number of observations T is small (compared to the dimension of the observations N), then the estimators become noisy and unreliable.
- The error of an estimator can be separated into two terms: the bias and the variance of the estimator.
- In the small sample regime, the main source of error comes from the variance of the estimator (intuitively, because the estimator is based on a small number of random samples, it is also too random).⁴
- It is well-known in the estimation literature that lower estimation errors can be achieved by allowing some bias in exchange of a smaller variance.
- This can be implemented by shrinking the estimator to some known target values.

⁴A. Meucci, *Risk and Asset Allocation*. Springer, 2005.

Shrinkage

- Let θ denote the parameter to be estimated (in our case, either the mean vector or covariance matrix) and $\hat{\theta}$ some estimation (e.g., the sample mean or the sample covariance matrix).⁵
- A shrinkage estimator is typically defined as

$$\hat{\theta}^{\text{sh}} = (1 - \rho) \hat{\theta} + \rho \theta^{\text{target}}$$

where θ^{target} is the target (prior information) and ρ is the shrinkage trade-off parameter.

- There are two main problems here:
 - choosing the target θ^{target} : this is problem dependent and may come from side information or some discretionary views on the market
 - choosing the shrinkage factor ρ : even though it looks like a simple problem, tons of ink have been devoted to it.
- Note that the above shrinkage model is actually a linear model and more sophisticated nonlinear models can be considered at the expense of mathematical complication and/or computational increase.

⁵Y. Feng and D. P. Palomar, *A Signal Processing Perspective on Financial Engineering*. Foundations and Trends® in Signal Processing, Now Publishers Inc., 2016.

Shrinkage Factor

- The choice of the shrinkage factor ρ is critical for the success of the shrinkage estimator.
- Of course the target is also important, but ironically even when the target is something totally uninformative, the results can still be surprisingly good.
- There are two main philosophies for the choice of ρ :
 - **Cross-validation**: this is a practical approach widely used in machine learning to choose many of the parameters that usually have to be tuned. The idea is simple: 1) compute the estimate $\hat{\theta}$ from the training data, 2) try different values of ρ and assess its performance using another set of data called cross-validation data to choose the best value, and 3) use the best ρ in yet a different set of new data called test data for the actual final performance.
 - **Random Matrix Theory (RMT)**: this is based on a heavy dose of mathematics going back to Wigner in 1955 who introduced the topic to model the nuclei of heavy atoms. This approach allows for a clean computation of ρ which is valid under a number of assumptions and in the limit of large T and N .

Outline

1 The Need for Prior Information

2 Shrinkage

- Shrinkage for μ
- Shrinkage for Σ
- Random Matrix Theory (RMT)

3 Black-Litterman Model

Shrinkage for the Mean

- Consider the sample mean estimator:

$$\hat{\mu} = \frac{1}{T} \sum_{t=1}^T r_t$$

- It is well-known from the central limit theorem that

$$\hat{\mu} \sim \mathcal{N}\left(\mu, \frac{1}{T}\Sigma\right)$$

and the MSE is

$$E\left[\|\hat{\mu} - \mu\|^2\right] = \frac{1}{T}\text{Tr}(\Sigma).$$

- The sample mean estimator is the least square solution as well as the maximum likelihood estimator under a Gaussian distribution.
- However, it was a shock when Stein proved in 1956⁶ that in terms of MSE this approach is suboptimal.

⁶C. Stein, "Inadmissibility of the usual estimator for the mean of a multivariate normal distribution", *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, no. 399, pp. 197–206, 1956.

James-Stein Estimator

- Stein developed the estimator in 1956⁷ and was later improved by James and Stein in 1961⁸.
- It can be shown that the James-Stein estimator dominates the least squares estimator, i.e., that it has a lower mean square error (at the expense of some bias).
- The James-Stein estimator is a member of a class of Bayesian estimators that dominate the maximum likelihood estimator.
- The James-Stein estimator is

$$\hat{\mu}^{\text{JS}} = (1 - \rho) \hat{\mu} + \rho t$$

where t is the shrinkage target and $0 \leq \rho \leq 1$ is the amount of shrinkage.

⁷C. Stein, "Inadmissibility of the usual estimator for the mean of a multivariate normal distribution", *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, no. 399, pp. 197–206, 1956.

⁸W. James and C. Stein, "Estimation with quadratic loss", in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, 1961, pp. 361–379.

James-Stein Estimator

- It can be shown⁹ that a choice of ρ so that $E \left[\|\hat{\mu}^{\text{JS}} - \mu\|^2 \right] \leq E \left[\|\hat{\mu} - \mu\|^2 \right]$ is

$$\rho = \frac{1}{T} \frac{N\bar{\lambda} - 2\lambda_{\max}}{\|\hat{\mu} - \mathbf{t}\|^2}$$

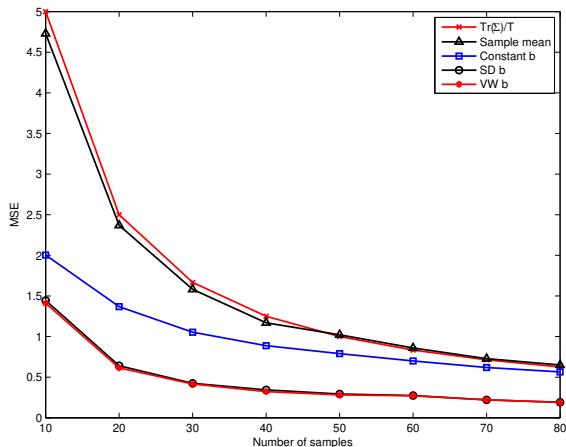
where $\bar{\lambda} = \frac{1}{N} \text{Tr}(\Sigma)$ and λ_{\max} are the average and maximum values, respectively, of the eigenvalues of Σ .

- Observe that ρ vanishes as T increases and the shrinkage estimator gets closer to the sample mean.
- Choices for the target include:
 - any arbitrary choice: for example $\mathbf{t} = 0$ or $\mathbf{t} = 0.1 \times \mathbf{1}$
 - grand mean: $\mathbf{t} = \frac{1^T \hat{\mu}}{N} \times \mathbf{1}$
 - volatility-weighted grand mean: $\mathbf{t} = \frac{1^T \hat{\Sigma}^{-1} \hat{\mu}}{1^T \hat{\Sigma}^{-1} \mathbf{1}} \times \mathbf{1}$

⁹A. Meucci, *Risk and Asset Allocation*. Springer, 2005.

Example of James-Stein Estimator

Comparison of $t = 0.2 \times 1$, the grand mean, and the volatility grand mean:¹⁰



¹⁰Y. Feng and D. P. Palomar, *A Signal Processing Perspective on Financial Engineering*. Foundations and Trends® in Signal Processing, Now Publishers Inc., 2016.

Outline

1 The Need for Prior Information

2 Shrinkage

- Shrinkage for μ
- Shrinkage for Σ
- Random Matrix Theory (RMT)

3 Black-Litterman Model

Shrinkage for the Covariance Matrix

- We will now assume that the mean is known and the goal is to estimate the covariance matrix or scatter matrix.
- The shrinkage estimator has the form

$$\hat{\Sigma}^{\text{sh}} = (1 - \rho) \hat{\Sigma} + \rho T$$

where $\hat{\Sigma}$ is the sample covariance matrix, T is the shrinkage target, and $0 \leq \rho \leq 1$ is the amount of shrinkage.

- As usual with shrinkage, we need to determine both the target and ρ .
- Choices for the target include:
 - any arbitrary choice: for example, the identity matrix $T = I$
 - scaled identity: $T = \frac{1}{N} \text{Tr}(\hat{\Sigma}) \times I$
 - diagonal with variances: $T = \text{Diag}(\hat{\Sigma})$
- To determine ρ one can use an empirical approach like cross-validation or a more mathematical-based approach like RMT.

Shrinkage Factor via RMT

- RMT can be used to determine ρ in a theoretical way, which becomes valid for large T and N .
- The first step is to choose some criterion to minimize and then one can try to use the RMT tools.
- We will consider the following criteria (but the literature on other criteria is very extensive):
 - MSE of covariance matrix
 - Quadratic loss of precision matrix
 - Sharpe ratio.

MSE of Covariance Matrix

- Ledoit and Wolf made popular in 2003¹¹ and 2004¹² the use of RMT in financial econometrics.
- They considered shrinkage of the sample covariance matrix $\hat{\Sigma}$ towards the identity matrix:

$$\hat{\Sigma}^{\text{sh}} = (1 - \rho) \hat{\Sigma} + \rho I$$

- More precisely, they considered the following formulation:

$$\begin{array}{ll} \underset{\rho_1, \rho_2}{\text{minimize}} & E \left[\left\| \hat{\Sigma}^{\text{sh}} - \Sigma \right\|_F^2 \right] \\ \text{subject to} & \hat{\Sigma}^{\text{sh}} = \rho_1 I + \rho_2 \hat{\Sigma} \end{array}$$

whose objective is uncomputable since it requires knowledge of the true Σ !

¹¹O. Ledoit and M. Wolf, “Improved estimation of the covariance matrix of stock returns with an application to portfolio selection”, *Journal of Empirical Finance*, vol. 10, no. 5, pp. 603–621, 2003.

¹²O. Ledoit and M. Wolf, “A well-conditioned estimator for large-dimensional covariance matrices”, *Journal of multivariate analysis*, vol. 88, no. 2, pp. 365–411, 2004.

MSE of Covariance Matrix

- If we ignore this little detail (lol), they obtained the optimal solution (termed oracle estimator) as

$$\hat{\Sigma}^{\text{sh}} = (1 - \rho) \hat{\Sigma} + \rho T$$

with $T = \frac{1}{N} \text{Tr}(\Sigma) \times I$ and $\rho = \frac{E[\|\hat{\Sigma} - \Sigma\|_F^2]}{E[\|\hat{\Sigma} - T\|_F^2]}$.

- Obviously the previous solution is useless as it requires knowledge of the true Σ .
- One could be tempted to simply use the sample covariance matrix $\hat{\Sigma}$ in lieu of Σ . However, that would be a big mistake since it would lead to a non-consistent estimator (in fact, in this particular case it would lead to $\rho = 0$!).
- This is where the magic of RMT comes into play: it turns out that asymptotically for large T and N , one can derive a consistent estimator that does not require knowledge of Σ .

Ledoit-Wolf Estimator

- Ledoit and Wolf further derived the consistent estimator (termed LW estimator):

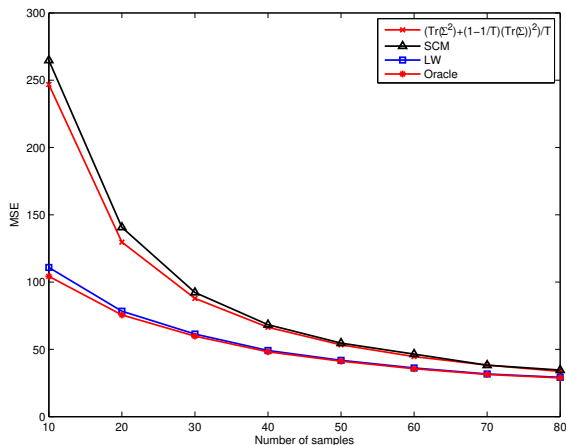
$$\hat{\Sigma}^{\text{sh}} = (1 - \rho) \hat{\Sigma} + \rho \mathbf{T}$$

with

$$\mathbf{T} = \frac{1}{N} \text{Tr}(\hat{\Sigma}) \times \mathbf{I}$$
$$\rho = \min \left(1, \frac{\frac{1}{T^2} \sum_{t=1}^T \|\hat{\Sigma} - \mathbf{r}_t \mathbf{r}_t^T\|_F^2}{\|\hat{\Sigma} - \mathbf{T}\|_F^2} \right).$$

Example of Ledoit-Wolf Estimator

Comparison of sample covariance matrix, oracle estimator, and LW estimator:¹³



¹³Y. Feng and D. P. Palomar, *A Signal Processing Perspective on Financial Engineering*. Foundations and Trends® in Signal Processing, Now Publishers Inc., 2016.

Quadratic Loss of Precision Matrix

- In many cases, it is the precision matrix (i.e., the inverse of the covariance matrix) that we really care about. For example, if our goal is to design a portfolio like the minimum variance portfolio:

$$\mathbf{w}^{\text{MV}} = \frac{\hat{\Sigma}^{-1} \mathbf{1}}{\mathbf{1}^T \hat{\Sigma}^{-1} \mathbf{1}}.$$

- Aiming at minimizing the MSE in the estimation of Σ , $E \left[\left\| \hat{\Sigma}^{\text{sh}} - \Sigma \right\|_F^2 \right]$, may not be the best strategy if one really cares about its inverse since the inversion operation can dramatically amplify the estimation error.
- It is more sensible to minimize the estimation error in the precision matrix directly $\left\| (\hat{\Sigma}^{\text{sh}})^{-1} - \Sigma^{-1} \right\|_F^2$ as formulated by Zhang et al.¹⁴

¹⁴M. Zhang, F. Rubio, and D. P. Palomar, “Improved calibration of high-dimensional precision matrices”, *IEEE Transactions on Signal Processing*, vol. 61, no. 6, pp. 1509–1519, 2013.

Quadratic Loss of Precision Matrix

- Consider then the following formulation:¹⁵

$$\begin{array}{ll}\text{minimize} & \frac{1}{N} \left\| (\hat{\Sigma}^{\text{sh}})^{-1} - \Sigma^{-1} \right\|_F^2 \\ \text{subject to} & \hat{\Sigma}^{\text{sh}} = \rho I + \frac{1}{T} R W R^T \\ & W \text{ diagonal}\end{array}$$

where $R = \begin{bmatrix} r_1 & \cdots & r_T \end{bmatrix}$ is the $N \times T$ data matrix and W is a $T \times T$ diagonal matrix that allows for a weighting of the different samples.

- Note that here the target matrix is $T = \frac{1}{T} R W R^T$, i.e., a weighted sample covariance matrix.
- This formulation is much harder because, even if Σ was known, there is no closed-form solution as before. We will use the magic of RMT...

¹⁵M. Zhang, F. Rubio, and D. P. Palomar, "Improved calibration of high-dimensional precision matrices", *IEEE Transactions on Signal Processing*, vol. 61, no. 6, pp. 1509–1519, 2013.

Quadratic Loss of Precision Matrix

- It was proved¹⁶ that the optimal weights are $W = \alpha I$, so no need for different weights, and the following is an asymptotic consistent formulation (without Σ):

$$\begin{aligned} & \underset{\rho, \alpha \geq 0, \delta}{\text{minimize}} && \frac{1}{N} \left\| (\hat{\Sigma}^{\text{sh}})^{-1} - \hat{\Sigma}^{-1} \right\|_F^2 \\ & && + \frac{2}{N} \text{Tr} \left(\rho^{-1} \left(\delta (\hat{\Sigma}^{\text{sh}})^{-1} - (1 - c_N) \hat{\Sigma}^{-1} \right) + \hat{\Sigma}^{-1} (\hat{\Sigma}^{\text{sh}})^{-1} \right) \\ & && - (2c_N - c_N^2) \frac{1}{N} \text{Tr}(\hat{\Sigma}^{-2}) \\ & && - (c_N - c_N^2) \left(\frac{1}{N} \text{Tr}(\hat{\Sigma}^{-1}) \right)^2 \\ & \text{subject to} && \hat{\Sigma}^{\text{sh}} = \rho I + \alpha \hat{\Sigma} \\ & && \delta = \alpha \left(1 - \frac{1}{T} \text{Tr}(\alpha \hat{\Sigma} (\hat{\Sigma}^{\text{sh}})^{-1}) \right) \end{aligned}$$

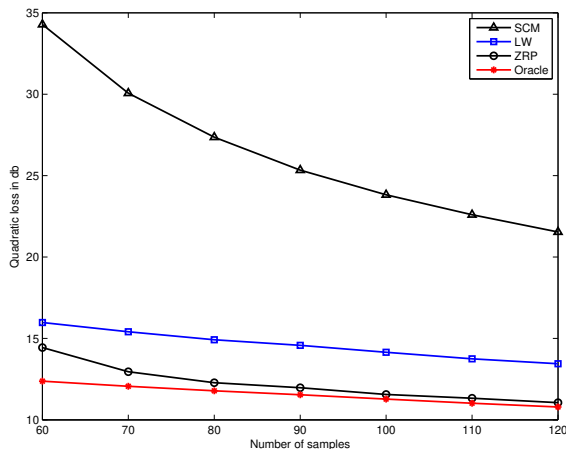
where $c_N = N/T$.

- The problem is highly nonconvex but it can be easily solved in practice via exhaustive search over ρ and α .

¹⁶M. Zhang, F. Rubio, and D. P. Palomar, "Improved calibration of high-dimensional precision matrices", *IEEE Transactions on Signal Processing*, vol. 61, no. 6, pp. 1509–1519, 2013.

Example of Precision Matrix Estimator

Comparison of sample covariance matrix, LW estimator, the previous estimator (ZRP), and the oracle:¹⁷



¹⁷Y. Feng and D. P. Palomar, *A Signal Processing Perspective on Financial Engineering*. Foundations and Trends® in Signal Processing, Now Publishers Inc., 2016.

Maximizing the Sharpe Ratio

- The previous formulations were based on selecting the shrinkage trade-off parameter ρ to improve the covariance or precision estimation accuracy based on some measure of error (e.g., the Frobenius norm).
- However, the ultimate goal of estimating the covariance matrix is to employ it for some portfolio design that is supposed to have a good out-of-sample performance.
- Since the most common way to measure the performance of a portfolio is the Sharpe ratio, we can precisely use it as our criterion of interest to choose ρ :

$$SR = \frac{\mathbf{w}^T \boldsymbol{\mu}}{\sqrt{\mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w}}}.$$

- The portfolio that maximizes the Sharpe ratio is

$$\mathbf{w}^{SR} \propto \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}.$$

- In practice, of course $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are unknown and one must use some estimates, for example, the sample mean $\hat{\boldsymbol{\mu}}$ and a shrinkage estimator for the covariance matrix $\hat{\boldsymbol{\Sigma}}^{sh} = \rho_1 \mathbf{I} + \rho_2 \hat{\boldsymbol{\Sigma}}$.

Maximizing the Sharpe Ratio

- Since the Sharpe ratio is invariant in w , we can arbitrarily set $\rho_2 = 1$ to eliminate one parameter to be chosen:

$$\hat{\Sigma}^{\text{sh}} = \rho_1 I + \hat{\Sigma}$$

- The optimal portfolio becomes then

$$w^{\text{SR}} \propto (\hat{\Sigma}^{\text{sh}})^{-1} \hat{\mu}.$$

- And the realized out-of-sample Sharpe ratio is

$$\text{SR} = \frac{\hat{\mu}^T (\hat{\Sigma}^{\text{sh}})^{-1} \mu}{\sqrt{\hat{\mu}^T (\hat{\Sigma}^{\text{sh}})^{-1} \Sigma (\hat{\Sigma}^{\text{sh}})^{-1} \hat{\mu}}}.$$

Maximizing the Sharpe Ratio

- We can finally formulate the problem as¹⁸

$$\begin{array}{ll}\text{maximize} & \frac{\hat{\mu}^T (\hat{\Sigma}^{\text{sh}})^{-1} \mu}{\sqrt{\hat{\mu}^T (\hat{\Sigma}^{\text{sh}})^{-1} \Sigma (\hat{\Sigma}^{\text{sh}})^{-1} \hat{\mu}}} \\ \text{subject to} & \hat{\Sigma}^{\text{sh}} = \rho_1 I + \hat{\Sigma}\end{array}$$

- Again, this problem formulation is useless in practice because it requires knowledge of the true μ and Σ .
- But again this is where the magic of RMT comes into play...

¹⁸M. Zhang, F. Rubio, D. P. Palomar, and X. Mestre, “Finite-sample linear filter optimization in wireless communications and financial systems”, *IEEE Transactions on Signal Processing*, vol. 61, no. 20, pp. 5014–5025, 2013.

Maximizing the Sharpe Ratio

- The following formulation is computable and leads to a consistent estimator¹⁹

$$\begin{aligned} & \underset{\rho_1 \geq 0}{\text{maximize}} && \frac{\hat{\mu}^T (\hat{\Sigma}^{\text{sh}})^{-1} \hat{\mu} - \delta}{\sqrt{b \hat{\mu}^T (\hat{\Sigma}^{\text{sh}})^{-1} \hat{\Sigma} (\hat{\Sigma}^{\text{sh}})^{-1} \hat{\mu}}} \\ & \text{subject to} && \hat{\Sigma}^{\text{sh}} = \rho_1 \mathbf{I} + \hat{\Sigma} \\ & && \delta = D / (1 - D) \\ & && D = \frac{1}{T} \text{Tr}(\hat{\Sigma} (\hat{\Sigma}^{\text{sh}})^{-1}) \\ & && b = \frac{T}{\text{Tr}(\mathbf{W}(\mathbf{I} + \delta \mathbf{W})^{-2})} \end{aligned}$$

where $\mathbf{W} = \mathbf{I} - \frac{1}{T} \mathbf{1} \mathbf{1}^T$.

- The interpretation is that one uses the estimations $\hat{\mu}$ and $\hat{\Sigma}$ in lieu of the true unknown quantities μ and Σ , but then some corrections terms are needed, i.e., δ in the numerator and b in the denominator.
- This problem is now computable but it is nonconvex. However, it is easy to solve it via an exhaustive search over the scalar ρ_1 .

¹⁹M. Zhang, F. Rubio, D. P. Palomar, and X. Mestre, "Finite-sample linear filter optimization in wireless communications and financial systems", *IEEE Transactions on Signal Processing*, vol. 61, no. 20, pp. 5014–5025, 2013.

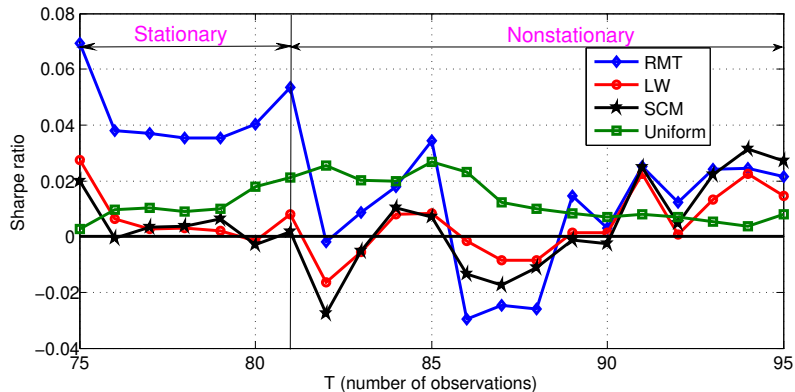
Example of Sharpe Ratio based Estimator

- Consider the daily returns of 45 stocks under the Hang Seng Index from 03-Jun-2009 to 31-Jul-2011.
- The portfolio is updated on a rolling window basis every 10 days and the past $T = 75, 76, \dots, 95$ days are used to design the portfolios at each update period.
- We compare the following portfolios:²⁰
 - based on the proposed method (RMT)
 - based on LW estimator
 - based on the sample covariance matrix
 - uniform portfolio.

²⁰Y. Feng and D. P. Palomar, *A Signal Processing Perspective on Financial Engineering*. Foundations and Trends[®] in Signal Processing, Now Publishers Inc., 2016.

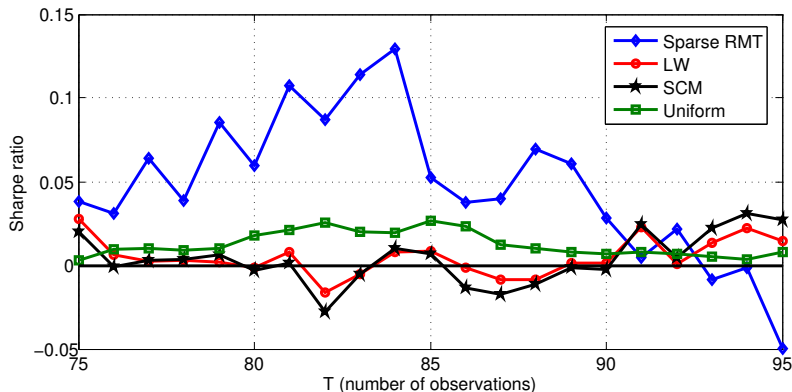
Example of Sharpe Ratio based Estimator

- The proposed method is the best, but note that, for $T > 81$, the performance starts to degrade. This is probably because the lack of stationarity.



Example of Sharpe Ratio based Estimator

- A sparse portfolio was considered (forcing to zero all the portfolio weights that had an absolute value less than 5% of the summed absolute values):



Beyond Linear Shrinkage

- Recall the the shrinkage covariance matrix estimation

$$\hat{\Sigma}^{\text{sh}} = (1 - \rho) \hat{\Sigma} + \rho I$$

- It can be interpreted as a linear shrinkage of the eigenvalues (while keeping the same eigenvectors) towards one:

$$\lambda_i(\hat{\Sigma}^{\text{sh}}) = (1 - \rho) \lambda_i(\hat{\Sigma}) + \rho 1$$

- One can wonder whether a more general shrinkage of the eigenvalues is possible.
- Precisely, recent promising results have been in the direction of nonlinear shrinkage of eigenvalues based on very sophisticated RMT:

J. Bun, J.-P. Bouchaud, and M. Potters, "Cleaning correlation matrices", *Risk Management*, 2006

Outline

1 The Need for Prior Information

2 Shrinkage

- Shrinkage for μ
- Shrinkage for Σ
- Random Matrix Theory (RMT)

3 Black-Litterman Model

What is RMT Anyway?

- Linear shrinkage of the covariance matrix $\hat{\Sigma}^{\text{sh}} = (1 - \rho) \hat{\Sigma} + \rho I$ can be seen in terms of eigenvalues:

$$\lambda_i(\hat{\Sigma}^{\text{sh}}) = (1 - \rho) \lambda_i(\hat{\Sigma}) + \rho$$

- And it is precisely about distribution of eigenvalues that RMT has a lot to say.
- The topic is too mathematically involved to survey here, but it is interesting to see the starting point of the whole theory.
- A good reference of RMT applied to the cleaning of covariance and correlation matrices with the financial application in mind is:

J. Bun, J.-P. Bouchaud, and M. Potters, *Cleaning Large Correlation Matrices: Tools from Random Matrix Theory*. Oxford Univ. Press, 2016

Wishart Matrix

- A Wishart matrix is a random symmetric matrix M of the form (i.e., a sample covariance matrix):

$$M = \frac{1}{T} X^T X$$

where X is an $T \times N$ random matrix of i.i.d. Gaussian elements $X_{ij} \sim \mathcal{N}(0, 1)$.

- The population matrix of the data is $\Sigma \triangleq E[M] = I$, i.e., it has all eigenvalues identical to 1.
- Matrix M is clearly random so, in principle, there is not much we can say about it.
- However, for a fixed dimension N and in the limit of large T (i.e., $T \gg N$), we can say that $M \rightarrow \Sigma = I$
- But when N is not small compared to T , then this convergence result does not hold anymore. In fact, for $T, N \rightarrow \infty$ the matrix M is still random and does not converge to anything.

Wishart Matrix

- RMT precisely considers the case when $T, N \rightarrow \infty$ but their ratio $q = N/T$ is not vanishingly small. This is often called the large dimension limit.
- In the case $q = 0$, such as the case of fixed N , we have already seen that the sample eigenvalues converge to the population eigenvalues.
- But what happens when $q > 0$?
- The first result is due to the seminal work of Marcenko and Pastur in 1967.²¹
- It turns out that the sample eigenvalues become noisy estimators of the “true” (population) eigenvalues no matter how large T is!
- Note that one specific element of the covariance matrix can be estimated with vanishing error for large T , but because we have more and more entries as N also grows, the eigenvalues always have some nonvanishing error.
- This is also called “the curse of dimensionality”.

²¹V. A. Marcenko and L. A. Pastur, “Distribution of eigenvalues for some sets of random matrices”, *Mat. Sb.*, vol. 72, no. 4, pp. 507–536, 1967.

Marcenko-Pastur Law for Wishart Matrices

- Marcenko and Pastur showed in 1967²² that in the limit when $T, N \rightarrow \infty$ while N/T converges to a fixed value $q \in (0, 1)$, the empirical distribution of eigenvalues of $M = \frac{1}{T}X^T X$ converges almost surely to

$$\rho_{\text{MP}}(\nu) = \frac{1}{2\pi} \frac{\sqrt{(\nu_+ - \nu)(\nu - \nu_-)}}{q\nu}, \quad \nu \in [\nu_-, \nu_+]$$

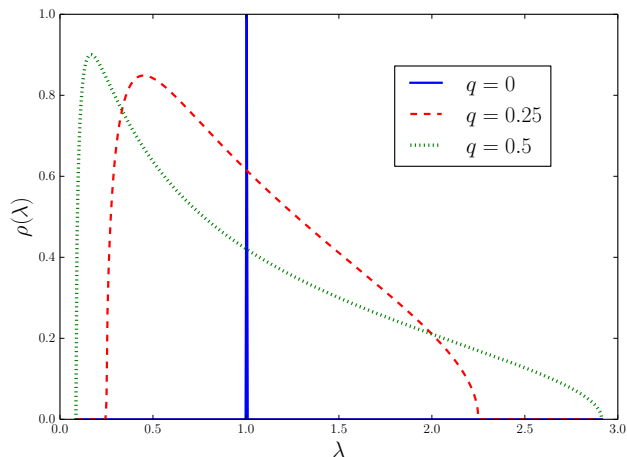
where $\nu_{\pm} = (1 \pm \sqrt{q})^2$.

- Whereas for $q \geq 1$, it is clear that M is a singular matrix with $N - T$ zero eigenvalues, which contribute $(1 - q^{-1}) \delta(\nu)$ to the density above:

$$\rho_{\text{MP}}(\nu) = \max(1 - q^{-1}, 0) \delta(\nu) + \frac{1}{2\pi} \frac{\sqrt{(\nu_+ - \nu)(\nu - \nu_-)}}{q\nu} 1_{[\nu_-, \nu_+]}$$

²²V. A. Marcenko and L. A. Pastur, "Distribution of eigenvalues for some sets of random matrices", *Mat. Sb.*, vol. 72, no. 4, pp. 507–536, 1967.

Marcenko-Pastur Law for Wishart Matrices

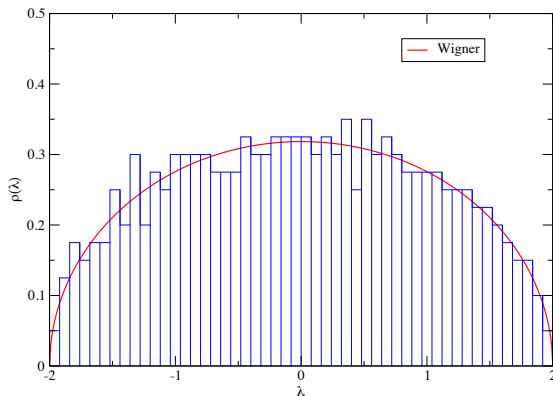


- Observe that the distortion becomes more and more substantial as q becomes large.

Wigner's Semicircle Law for Gaussian Matrices

- Wigner's semi-circle law from 1951 states that the empirical distribution of the eigenvalues of X converges almost surely to

$$\rho_W(\nu) = \frac{1}{2\pi} \sqrt{4 - \nu^2}, \quad |\nu| < 2$$



- The Marcenko-Pastur law has clearly relevance in finance because a key quantity in portfolio design is the covariance matrix of the log-returns, which could be modeled as Gaussian.
- In fact, even for non-Gaussian distributions with heavier tails like in finance, the Marcenko-Pastur law still seems to hold if one uses robust estimators against heavy tails.
- However, from factor modeling, we know that returns have a strong market component and perhaps other few factors plus the idiosyncratic component:
 - the idiosyncratic component, called the “bulk”, has a distribution that follows the Marcenko-Pastur law
 - the market (and other strong factors) are sometimes referred to as outliers and are totally separated from the bulk.

Outline

1 The Need for Prior Information

2 Shrinkage

- Shrinkage for μ
- Shrinkage for Σ
- Random Matrix Theory (RMT)

3 Black-Litterman Model

Black-Litterman Model

- The Black-Litterman model²³ allows to incorporate investor's views about the expected return μ .
- **Market Equilibrium:** One source of information for μ is the market, e.g., the sample estimate $\hat{\mu} = \frac{1}{T} \sum_{t=1}^T r_t$. We can then explicitly write the estimate $\pi = \hat{\mu}$ in terms of the actual μ and the estimation error:

$$\pi = \mu + w, \quad w \sim \mathcal{N}(0, \tau \Sigma)$$

where the error has been statistically modeled with a covariance matrix equal to a scaled Σ (which is assumed known for simplicity).

- **Investor's View:** Suppose we have K views summarized from some investors written in the following form:

$$v = P\mu + e, \quad e \sim \mathcal{N}(0, \Omega)$$

where $P \in \mathbb{R}^{K \times N}$ and $v \in \mathbb{R}^K$ characterize the absolute or relative K views and $\Omega \in \mathbb{R}^{K \times K}$ measures the uncertainty in the views.

²³F. Black and R. Litterman, "Asset allocation: Combining investor views with market equilibrium", *The Journal of Fixed Income*, vol. 2, no. 1, pp. 7–18, 1991.

Example of Investor's Views

- Suppose there are $N = 5$ stocks and two independent views on them:²⁴
 - Stock 1 will have a return of 1.5% with standard deviation of 1%
 - Stock 3 will outperform Stock 2 by 4% with a standard deviation of 1%
- Mathematically, we can express these two views as

$$\begin{bmatrix} 1.5\% \\ 4\% \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \end{bmatrix} \mu + e$$

where $e \sim \mathcal{N}(0, \Omega)$ and $\Omega = \begin{bmatrix} 1\%^2 & 0 \\ 0 & 1\%^2 \end{bmatrix}$.

- The parameter τ also has to be specified: some researchers set $\tau \in [0.01, 0.05]$, others $\tau = 1$, while some suggest $\tau = 1/T$ (i.e., the more observations the less uncertainty on the market equilibrium).²⁵

²⁴F. J. Fabozzi, S. M. Focardi, and P. N. Kolm, *Quantitative Equity Investing: Techniques and Strategies*. Wiley, 2010.

²⁵T. M. Idzorek, "A step-by-step guide to the Black-Litterman model", *Forecasting Expected Returns in the Financial Markets*, p. 17, 2002.

Example of Investor's Views

- In some occasions, the investor may only have qualitative views (as opposed to quantitative ones), i.e., only P is available.
- Then, one can choose:²⁶

$$v_i = (P\pi)_i + \eta_i \sqrt{(P\Sigma P^T)_{ii}}, \quad i = 1, \dots, N$$

where $\eta_i \in \{-\beta, -\alpha, +\alpha, +\beta\}$ defines “very bearish”, “bearish”, “bullish”, and “very bullish” views, respectively. Typical choices are $\alpha = 1$ and $\beta = 2$.

- As for the uncertainty:

$$\Omega = \frac{1}{c} P \Sigma P^T$$

where the scatter structure of uncertainty is inherited from the market volatilities and correlations and $c \in (0, \infty)$ represents the overall level of confidence in the views.

²⁶A. Meucci, *Risk and Asset Allocation*. Springer, 2005.

Alternative to Marquet Equilibrium: CAPM

- An alternative market equilibrium can be obtained from CAPM:

$$E[r_{i,t}] - r_f = \beta_i (E[r_{M,t}] - r_f)$$

where r_f is the return of the risk-free asset and $r_{M,t}$ is the market return which can be expressed as $r_{M,t} = w_M^T r_t$ with w_M being the market portfolio.

- Then

$$\pi = \hat{\mu}_{\text{mkt}} - r_f = \beta (E[r_{M,t}] - r_f)$$

with

$$\beta = \text{Cov}(r_t, r_{M,t}) / \text{Var}(r_{M,t})$$

- Thus

$$\pi = \delta \text{Cov}(r_t, r_{M,t}) = \delta \Sigma w_M$$

with $\delta = (E[r_{M,t}] - r_f) / \text{Var}(r_{M,t})$.

Black-Litterman Model - Weighted LS Approach

- Let us combine the two equations

$$\boldsymbol{\pi} = \boldsymbol{\mu} + \mathbf{w}, \quad \mathbf{w} \sim \mathcal{N}(\mathbf{0}, \tau \boldsymbol{\Sigma})$$

and

$$\mathbf{v} = \mathbf{P}\boldsymbol{\mu} + \mathbf{e}, \quad \mathbf{e} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega})$$

in a more compact form as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\mu} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{V})$$

$$\text{with } \mathbf{y} = \begin{bmatrix} \boldsymbol{\pi} \\ \mathbf{v} \end{bmatrix}, \mathbf{X} = \begin{bmatrix} \mathbf{I} \\ \mathbf{P} \end{bmatrix}, \text{ and } \mathbf{V} = \begin{bmatrix} \tau \boldsymbol{\Sigma} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega} \end{bmatrix}.$$

- We can now estimate $\boldsymbol{\mu}$ from the observations $\mathbf{y} = \mathbf{X}\boldsymbol{\mu} + \boldsymbol{\epsilon}$ (a Bayesian interpretation is also possible).
- This is just a weighted least squares (LS) problem:²⁷

$$\underset{\boldsymbol{\mu}}{\text{minimize}} \quad (\mathbf{y} - \mathbf{X}\boldsymbol{\mu})^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\mu})$$

²⁷Y. Feng and D. P. Palomar, *A Signal Processing Perspective on Financial Engineering*. Foundations and Trends[®] in Signal Processing, Now Publishers Inc., 2016.

Black-Litterman Model - Weighted LS Approach

- The solution is simply

$$\hat{\mu}_{BL} = \left(X^T V^{-1} X \right)^{-1} X^T V^{-1} y$$

- We can substitute the expressions for y , X , and V , leading to

$$\hat{\mu}_{BL} = \left((\tau \Sigma)^{-1} + P^T \Omega^{-1} P \right)^{-1} \left((\tau \Sigma)^{-1} \pi + P^T \Omega^{-1} v \right)$$

- Consider two extremes:

- $\tau = 0$: we give total accuracy to the market equilibrium view and indeed

$$\hat{\mu}_{BL} = \pi \triangleq \hat{\mu}_{mkt}$$

- $\tau \rightarrow \infty$: we give no accuracy at all to the market equilibrium view and therefore the investor's views dominate

$$\hat{\mu}_{BL} = \left(P^T \Omega^{-1} P \right)^{-1} P^T \Omega^{-1} v \triangleq \hat{\mu}_{views}$$

Black-Litterman Model - Weighted LS Approach

- We can now rewrite the solution as

$$\begin{aligned}\hat{\mu}_{BL} &= \left((\tau \Sigma)^{-1} + P^T \Omega^{-1} P \right)^{-1} \left((\tau \Sigma)^{-1} \pi + P^T \Omega^{-1} v \right) \\ &= \left((\tau \Sigma)^{-1} + P^T \Omega^{-1} P \right)^{-1} \left((\tau \Sigma)^{-1} \pi + P^T \Omega^{-1} P \hat{\mu}_{\text{views}} \right) \\ &= W_{\text{mkt}} \hat{\mu}_{\text{mkt}} + W_{\text{views}} \hat{\mu}_{\text{views}}\end{aligned}$$

where $W_{\text{mkt}} = \left((\tau \Sigma)^{-1} + P^T \Omega^{-1} P \right)^{-1} (\tau \Sigma)^{-1}$ and

$$W_{\text{views}} = \left((\tau \Sigma)^{-1} + P^T \Omega^{-1} P \right)^{-1} P^T \Omega^{-1} P.$$

- Note that $W_{\text{mkt}} + W_{\text{views}} = I$, so the Black-Litterman solution $\hat{\mu}_{BL}$ is a combination of the two extreme solutions $\hat{\mu}_{\text{mkt}}$ and $\hat{\mu}_{\text{views}}$.
- The Black-Litterman model is similar to the previous James-Stein shrinkage estimator where the target comes now from the investor's views $\hat{\mu}_{\text{views}}$ and the shrinkage scalar parameter is now a matrix.

Black-Litterman Model - Bayesian Approach 1

- This is actually the original formulation by Black and Litterman²⁸.
- We model the returns as

$$r \sim \mathcal{N}(\mu, \Sigma)$$

where the covariance Σ can be estimated from past returns but μ cannot be known with certainty.

- BL then models μ as a random variable normally distributed

$$\mu \sim \mathcal{N}(\pi, \tau \Sigma)$$

where π represents the best guess for μ and $\tau \Sigma$ the uncertainty on this guess. Note that then $r \sim \mathcal{N}(\pi, (1 + \tau) \Sigma)$.

- The views are modeled as

$$P\mu \sim \mathcal{N}(v, \Omega)$$

²⁸F. Black and R. Litterman, "Asset allocation: Combining investor views with market equilibrium", *The Journal of Fixed Income*, vol. 2, no. 1, pp. 7–18, 1991.

Black-Litterman Model - Bayesian Approach 1

- Then the posterior distribution for μ is obtained from Bayes formula:

$$\mu \mid v, \Omega \sim \mathcal{N}(\mu_{\text{BL}}, \Sigma_{\text{BL}}^{\mu})$$

where

$$\mu_{\text{BL}} = \left((\tau \Sigma)^{-1} + P^T \Omega^{-1} P \right)^{-1} \left((\tau \Sigma)^{-1} \pi + P^T \Omega^{-1} v \right)$$

and

$$\Sigma_{\text{BL}}^{\mu} = \left((\tau \Sigma)^{-1} + P^T \Omega^{-1} P \right)^{-1}.$$

- But we really want the posterior for the returns

$$r \mid v, \Omega \sim \mathcal{N}(\mu_{\text{BL}}, \Sigma_{\text{BL}})$$

where $\Sigma_{\text{BL}} = \Sigma_{\text{BL}}^{\mu} + \Sigma$.

- Using the matrix inversion lemma, we can further rewrite

$$\mu_{\text{BL}} = \pi + \tau \Sigma P^T (\tau P \Sigma P^T + \Omega)^{-1} (v - P \pi)$$

$$\Sigma_{\text{BL}} = (1 + \tau) \Sigma - \tau^2 \Sigma P^T (\tau P \Sigma P^T + \Omega)^{-1} P \Sigma.$$

Black-Litterman Model - Bayesian Approach 2

- In this case, μ is not modeled as a random variable but simply as $\mu = \pi$.²⁹
- The views are modeled on the random returns rather than on μ : $v = Pr + e$.
- The conditional distribution is modeled as

$$v \mid r \sim \mathcal{N}(Pr, \Omega)$$

- Applying Bayes we get

$$r \mid v, \Omega \sim \mathcal{N}(\mu_{BL}^m, \Sigma_{BL}^m)$$

where

$$\mu_{BL}^m = \pi + \Sigma P^T (P \Sigma P^T + \Omega)^{-1} (v - P \pi)$$

$$\Sigma_{BL}^m = \Sigma - \Sigma P^T (P \Sigma P^T + \Omega)^{-1} P \Sigma.$$

²⁹A. Meucci, *Risk and Asset Allocation*. Springer, 2005.

Beyond Black-Litterman

The following references by Meucci are recommended for more sophisticated ways to incorporate views in the portfolio design:

- Meucci, Attilio. Beyond Black-Litterman: Views on Non-Normal Markets. November 2005, Available at SSRN: <http://ssrn.com/abstract=848407>
- Meucci, Attilio. Beyond Black-Litterman in Practice: A Five-Step Recipe to Input Views on non-Normal Markets. May 2006, Available at SSRN: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=872577
- Meucci, Attilio. The Black-Litterman Approach: Original Model and Extensions. April 2008, Available at SSRN: <http://ssrn.com/abstract=1117574>

Thanks

For more information visit:

<https://www.danielppalomar.com>

