Prior Information: Shrinkage and Black-Litterman

Prof. Daniel P. Palomar

The Hong Kong University of Science and Technology (HKUST)

MAFS6010R - Portfolio Optimization with R
MSc in Financial Mathematics
Fall 2019-20, HKUST, Hong Kong

Outline

Outline

Returns

- Let us denote the log-returns of N assets at time t with the vector $\mathbf{r}_t \in \mathbb{R}^N$.
- 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 $\mathbf{r}_t$ conditional on $\mathcal{F}_{t-1}$.
- $\mathbf{r}_t$ is a multivariate stochastic process with **conditional** mean and covariance matrix denoted as[1]

$$\begin{aligned} \boldsymbol{\mu}_t &\triangleq \mathsf{E}\left[\mathbf{r}_t \mid \mathcal{F}_{t-1}\right] \\ \boldsymbol{\Sigma}_t &\triangleq \mathsf{Cov}\left[\mathbf{r}_t \mid \mathcal{F}_{t-1}\right] = \mathsf{E}\left[(\mathbf{r}_t - \boldsymbol{\mu}_t)(\mathbf{r}_t - \boldsymbol{\mu}_t)^T \mid \mathcal{F}_{t-1}\right]. \end{aligned}$$

[1] Y. 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 $\mathbf{r}_t$ follows an i.i.d. distribution (which is not very inacurate in general).
- That is, both the conditional mean and conditional covariance are constant

$$\boldsymbol{\mu}_t = \boldsymbol{\mu},$$
$$\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}.$$

- Very simple model, however, it is one of the most fundamental assumptions for many important works, e.g., the Nobel prize-winning Markowitz portfolio theory[2].

[2] H. 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 $T \to \infty$ 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 estimatior 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 $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

[3] A. 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:

$$\mathbf{r}_t = \boldsymbol{\alpha} + \mathbf{B}\mathbf{f}_t + \mathbf{w}_t,$$

  where
  - $\boldsymbol{\alpha}$ denotes a constant vector
  - $\mathbf{f}_t \in \mathbb{R}^K$ with $K \ll N$ is a vector of a few factors that are responsible for most of the randomness in the market,
  - $\mathbf{B} \in \mathbb{R}^{N \times K}$ denotes how the low dimensional factors affect the higher dimensional market;
  - $\mathbf{w}_t$ 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: $\boldsymbol{\Sigma} = \mathbf{B}\mathbf{B}^T + \boldsymbol{\Psi}$.

Outline

Small Sample Regime

- In the large sample regime, i.e., when the number of observations T is large, then the estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ 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.

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

Shrinkage

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

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

  where $\boldsymbol{\theta}^{\mathsf{target}}$ is a known target, which amounts to some prior information, and ρ is the shrinkage trade-off parameter.
- There are two main problems here:
  - choosing the target $\boldsymbol{\theta}^{\mathsf{target}}$: 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.

[5] Y. 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 philophies 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 $\hat{\boldsymbol{\theta}}$ 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

Shrinkage for the Mean

- Consider the sample mean estimator:

$$\hat{\boldsymbol{\mu}} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{r}_t$$

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

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

and the MSE is

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

- 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 1956[6] that in terms of MSE this approach is suboptimal.

[6] C. 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 1956[7] and was later improved by James and Stein in 1961[8].
- 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

$$\hat{\boldsymbol{\mu}}^{\mathsf{JS}} = (1-\rho)\,\hat{\boldsymbol{\mu}} + \rho\mathbf{t}$$

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

[7] C. 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.

[8] W. 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 shown[9] that a choice of ρ so that $E\left[\left\|\hat{\boldsymbol{\mu}}^{\mathsf{JS}} - \boldsymbol{\mu}\right\|^2\right] \leq E\left[\left\|\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\right\|^2\right]$ is

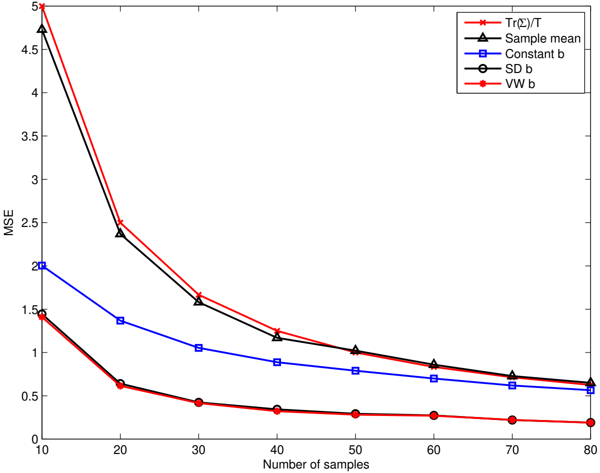
$$\rho = \frac{1}{T}\frac{N\bar{\lambda} - 2\lambda_{\mathsf{max}}}{\left\|\hat{\boldsymbol{\mu}} - \mathbf{t}\right\|^2}$$

  where $\bar{\lambda} = \frac{1}{N}\mathrm{Tr}(\boldsymbol{\Sigma})$ and λ_{max} are the average and maximum values, respectively, of the eigenvalues of $\boldsymbol{\Sigma}$.
- 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 $\mathbf{t} = \mathbf{0}$ or $\mathbf{t} = 0.1 \times \mathbf{1}$
  - grand mean: $\mathbf{t} = \frac{\mathbf{1}^T\hat{\boldsymbol{\mu}}}{N} \times \mathbf{1}$
  - volatility-weighted grand mean: $\mathbf{t} = \frac{\mathbf{1}^T\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}}{\mathbf{1}^T\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}} \times \mathbf{1}$

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

Example of James-Stein Estimator

Comparison of $\mathbf{t} = 0.2 \times \mathbf{1}$, the grand mean, and the volatility grand mean:[10]



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

Outline

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

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

  where $\hat{\boldsymbol{\Sigma}}$ is the sample covariance matrix, $\mathbf{T}$ 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 $\mathbf{T} = \mathbf{I}$
  - scaled identity: $\mathbf{T} = \frac{1}{N}\mathsf{Tr}(\hat{\boldsymbol{\Sigma}}) \times \mathbf{I}$
  - diagonal with variances: $\mathbf{T} = \mathsf{Diag}(\hat{\boldsymbol{\Sigma}})$
- 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 2003[11] and 2004[12] the use of RMT in financial econometrics.
- They considered shrinkage of the sample covariance matrix $\hat{\boldsymbol{\Sigma}}$ towards the identity matrix:

$$\hat{\boldsymbol{\Sigma}}^{\mathsf{sh}} = (1-\rho)\,\hat{\boldsymbol{\Sigma}} + \rho\mathbf{I}$$

- More precisely, they considered the following formulation:

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

  whose objective is uncomputable since it requires knowledge of the true $\boldsymbol{\Sigma}$!

[11] O. 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.

[12] O. 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

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

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

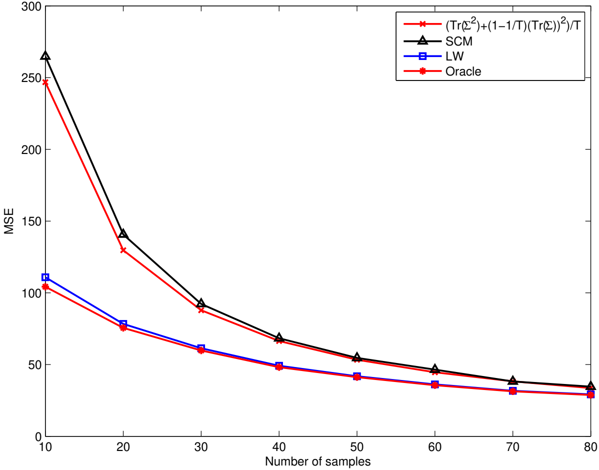
- Obviously the previous solution is useless as it requires knowledge of the true $\boldsymbol{\Sigma}$.
- One could be tempted to simply use the sample covariance matrix $\hat{\boldsymbol{\Sigma}}$ in lieu of $\boldsymbol{\Sigma}$. However, that would be a big mistake since it would lead to a non-consistent 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 $\boldsymbol{\Sigma}$.

Ledoit-Wolf Estimator

- Ledoit and Wolf further derived the consistent estimator (termed LW estimator):
$$\hat{\boldsymbol{\Sigma}}^{\mathsf{sh}} = (1-\rho)\,\hat{\boldsymbol{\Sigma}} + \rho\mathbf{T}$$
with
$$\mathbf{T} = \frac{1}{N}\mathsf{Tr}(\hat{\boldsymbol{\Sigma}}) \times \mathbf{I}$$
$$\rho = \min\left(1, \frac{\frac{1}{T^2}\sum_{t=1}^{T}||\hat{\boldsymbol{\Sigma}} - \mathbf{r}_t\mathbf{r}_t^T||_F^2}{||\hat{\boldsymbol{\Sigma}} - \mathbf{T}||_F^2}\right).$$

Example of Ledoit-Wolf Estimator

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



[13] Y. 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:

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

- Aiming at minimizing the MSE in the estimation of $\boldsymbol{\Sigma}$, $E\left[\left\|\hat{\boldsymbol{\Sigma}}^{\mathsf{sh}} - \boldsymbol{\Sigma}\right\|_F^2\right]$, 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 $\left\|(\hat{\boldsymbol{\Sigma}}^{\mathsf{sh}})^{-1} - \boldsymbol{\Sigma}^{-1}\right\|_F^2$ as formulated by Zhang et al.[14]

[14] M. 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]

$$\begin{array}{ll} \underset{\rho\geq 0, \mathbf{W}\succeq \mathbf{0}}{\text{minimize}} & \frac{1}{N}\left\|\left(\hat{\boldsymbol{\Sigma}}^{\text{sh}}\right)^{-1}-\boldsymbol{\Sigma}^{-1}\right\|_F^2 \\ \text{subject to} & \hat{\boldsymbol{\Sigma}}^{\text{sh}}=\rho\mathbf{I}+\frac{1}{T}\mathbf{R}\mathbf{W}\mathbf{R}^T \\ & \mathbf{W} \quad \text{diagonal} \end{array}$$

  where $\mathbf{R} = \left[\begin{array}{ccc} \mathbf{r}_1 & \cdots & \mathbf{r}_T \end{array}\right]$ is the $N \times T$ data matrix and $\mathbf{W}$ is a $T \times T$ diagonal matrix that allows for a weighting of the different samples.

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

[15] M. 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 proved[16] that the optimal weights are $\mathbf{W} = \alpha\mathbf{I}$, so no need for different weights, and the following is an asymptotic consistent formulation (without $\boldsymbol{\Sigma}$):

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

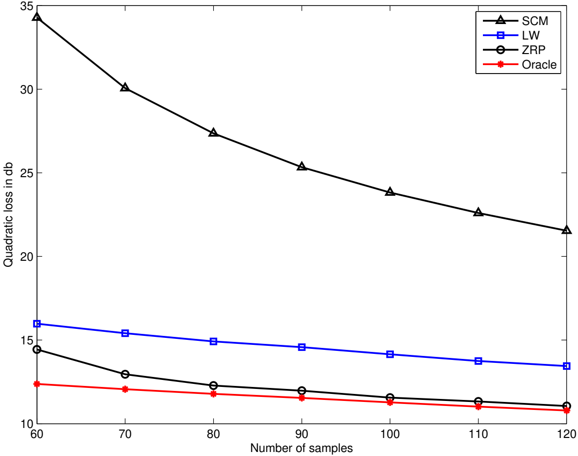
  where $c_N = N/T$.

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

[16] M. 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]



[17]Y. 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 accuraty 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:

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

- The portfolio that maximizes the Sharpe ratio is

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

- 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 $\hat{\boldsymbol{\Sigma}}^{\mathsf{sh}} = \rho_1 \mathbf{I} + \rho_2 \hat{\boldsymbol{\Sigma}}$.

Maximizing the Sharpe Ratio

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

$$\hat{\boldsymbol{\Sigma}}^{\mathsf{sh}} = \rho_1 \mathbf{I} + \hat{\boldsymbol{\Sigma}}$$

- The optimal portfolio becomes then

$$\mathbf{w}^{\mathsf{SR}} \propto \left(\hat{\boldsymbol{\Sigma}}^{\mathsf{sh}}\right)^{-1} \hat{\boldsymbol{\mu}}.$$

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

$$\mathsf{SR} = \frac{\hat{\boldsymbol{\mu}}^T \left(\hat{\boldsymbol{\Sigma}}^{\mathsf{sh}}\right)^{-1} \boldsymbol{\mu}}{\sqrt{\hat{\boldsymbol{\mu}}^T \left(\hat{\boldsymbol{\Sigma}}^{\mathsf{sh}}\right)^{-1} \boldsymbol{\Sigma} \left(\hat{\boldsymbol{\Sigma}}^{\mathsf{sh}}\right)^{-1} \hat{\boldsymbol{\mu}}}}.$$

Maximizing the Sharpe Ratio

- We can finally formulate the problem as[18]

$$\begin{array}{ll} \underset{\rho_1 \geq 0}{\text{maximize}} & \dfrac{\hat{\boldsymbol{\mu}}^T (\hat{\boldsymbol{\Sigma}}^{\text{sh}})^{-1} \boldsymbol{\mu}}{\sqrt{\hat{\boldsymbol{\mu}}^T (\hat{\boldsymbol{\Sigma}}^{\text{sh}})^{-1} \boldsymbol{\Sigma} (\hat{\boldsymbol{\Sigma}}^{\text{sh}})^{-1} \hat{\boldsymbol{\mu}}}} \\ \text{subject to} & \hat{\boldsymbol{\Sigma}}^{\text{sh}} = \rho_1 \mathbf{I} + \hat{\boldsymbol{\Sigma}} \end{array}$$

- Again, this problem formulation is useless in practice because it requires knowledge of the true $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.
- But again this is where the magic of RMT comes into play...

[18] M. 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 estimator[19]

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

  where $\mathbf{W} = \mathbf{I} - \frac{1}{T} \mathbf{1}\mathbf{1}^T$.
- The interpretation is that one uses the estimations $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ in lieu of the true unknown quantities $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, 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.

[19] M. 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.

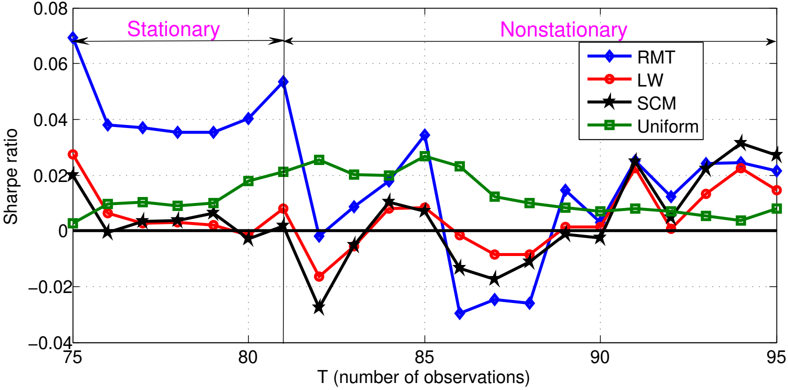
Example of Sharpe Ratio based Estimator

- 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 $T = 75, 76, \ldots, 95$ 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.

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

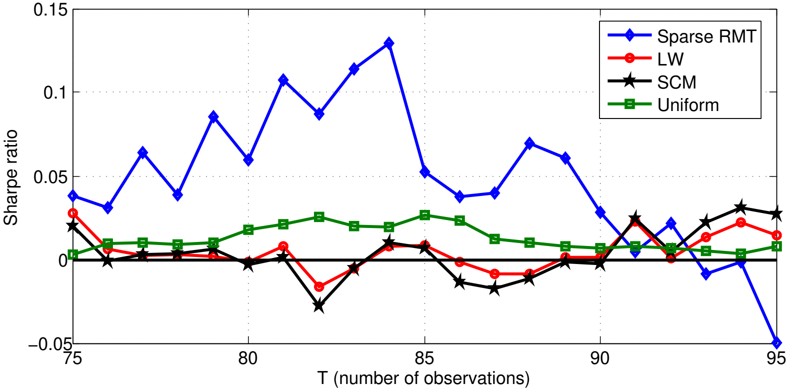
Example of Sharpe Ratio based Estimator

- 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

$$\hat{\boldsymbol{\Sigma}}^{\mathsf{sh}} = (1-\rho)\,\hat{\boldsymbol{\Sigma}} + \rho\mathbf{I}$$

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

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

- 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

What is RMT Anyway?

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

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

- 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 matries 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 $\mathbf{M}$ of the form (i.e., a sample covariance matrix):

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

  where $\mathbf{X}$ 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 $\mathbf{\Sigma} \triangleq E[\mathbf{M}] = \mathbf{I}$, i.e., it has all eigenvalues identical to 1.
- Matrix $\mathbf{M}$ 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 $\mathbf{M} \to \mathbf{\Sigma} = \mathbf{I}$
- But when N is not small compared to T, then this convergence result does not hold anymore. In fact, for $T, N \to \infty$ the matrix $\mathbf{M}$ is still random and does not converge to anything.

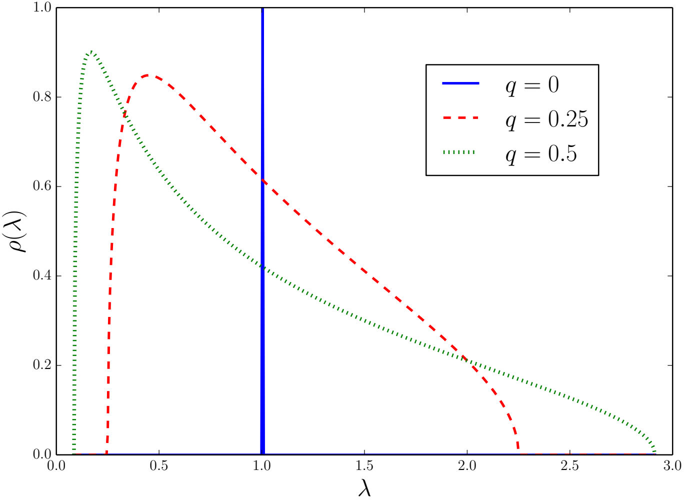
Wishart Matrix

- 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".

[21] V. 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

- In fact, the distortion becomes more and more substantial as q becomes large. See the limiting eigenvalue distribution:



Marcenko-Pastur Law for Wishart Matrices

- To be more precise, Marcenko and Pastur showed in 1967[22] that in the limit when $T, N \to \infty$ while N/T converges to a fixed value $q \in (0,1)$, the empirical distribution of eigenvalues of $\mathbf{M} = \frac{1}{T}\mathbf{X}^T\mathbf{X}$ converges almost surely to

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

  where $\nu_\pm = \left(1 \pm \sqrt{q}\right)^2$.
- Whereas for $q \geq 1$, it is clear that $\mathbf{M}$ is a singular matrix with $N - T$ zero eigenvalues, which contribute $\left(1 - q^{-1}\right)\delta(\nu)$ to the density above:

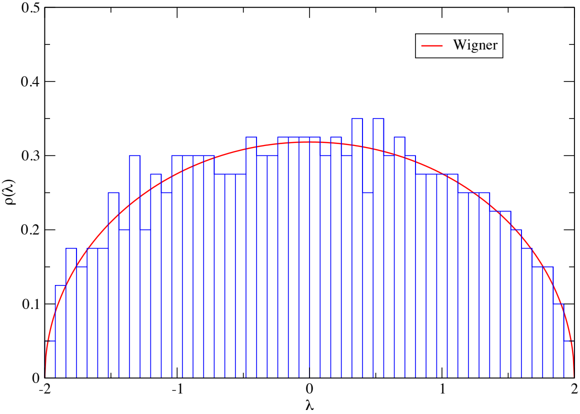
$$\rho_{\mathsf{MP}}(\nu) = \max\left(1 - q^{-1}, 0\right)\delta(\nu) + \frac{1}{2\pi}\frac{\sqrt{(\nu_+ - \nu)(\nu - \nu_-)}}{q\nu}1[\nu_-, \nu_+].$$

[22] V. 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.

Wigner's Semicircle Law for Gaussian Matrices

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

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



RMT in Finance

- 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 of 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

Black-Litterman Model

- The Black-Litterman model[23] allows to incorporate investor's views about the expected return $\boldsymbol{\mu}$.
- **Market Equilibrium**: One source of information for $\boldsymbol{\mu}$ is the market, e.g., the sample estimate $\hat{\boldsymbol{\mu}} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{r}_t$. We can then explicitly write the estimate $\boldsymbol{\pi} = \hat{\boldsymbol{\mu}}$ in terms of the actual $\boldsymbol{\mu}$ and the estimation error:

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

  where the error has been statistically modeled with a covariance matrix equal to a scaled $\boldsymbol{\Sigma}$ (which is assumed known for simplicity).
- **Investor's View**: Suppose we have K views summarized from some investors written in the following form:

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

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

[23] F. 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

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

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

- 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]

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

[25] T. 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 $\mathbf{P}$ is available.
- Then, one can choose:[26]

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

  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:

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

  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.

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

Alternative to Marquet Equilibrium: CAPM

- An alternative market equilibrium can be obtained from CAPM.
- Recall CAPM:

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

where r_f is the return of the risk-free asset and $r_{M,t}$ is the market return which can be expressed as $r_{M,t} = \mathbf{w}_M^T \mathbf{r}_t$

- Then

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

with

$$\boldsymbol{\beta} = \mathsf{Cov}\left(\mathbf{r}_t, r_{M,t}\right) / \mathsf{Var}\left(r_{M,t}\right)$$

- Thus

$$\boldsymbol{\pi} = \delta \mathsf{Cov}\left(\mathbf{r}_t, r_{M,t}\right) = \delta \boldsymbol{\Sigma} \mathbf{w}_M$$

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

Black-Litterman Model - Weighted LS Approach

- Let us combine the two equations

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

and

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

in a more compact form as

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

with $\mathbf{y} = \begin{bmatrix} \boldsymbol{\pi} \\ \mathbf{v} \end{bmatrix}$, $\mathbf{X} = \begin{bmatrix} \mathbf{I} \\ \mathbf{P} \end{bmatrix}$, 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]

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

[27] Y. 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

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

- We can substitute the expressions for $\mathbf{y}$, $\mathbf{X}$, and $\mathbf{V}$, leading to

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

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

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

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

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

Black-Litterman Model - Weighted LS Approach

- We can now rewrite the solution as

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

  where $\mathbf{W}_{\mathsf{mkt}} = \left((\tau\boldsymbol{\Sigma})^{-1} + \mathbf{P}^T\boldsymbol{\Omega}^{-1}\mathbf{P}\right)^{-1}(\tau\boldsymbol{\Sigma})^{-1}$ and

  $\mathbf{W}_{\mathsf{views}} = \left((\tau\boldsymbol{\Sigma})^{-1} + \mathbf{P}^T\boldsymbol{\Omega}^{-1}\mathbf{P}\right)^{-1}\mathbf{P}^T\boldsymbol{\Omega}^{-1}\mathbf{P}$.

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

Black-Litterman Model - Bayesian Approach 1

- This is actually the original formulation by Black and Litterman[28].
- We model the returns as

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

where the covariance $\boldsymbol{\Sigma}$ can be estimated from past returns but $\boldsymbol{\mu}$ cannot be known with certainty.

- BL then models $\boldsymbol{\mu}$ as a random variable normally distributed

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

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

- The views are modeled as

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

[28] F. 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 $\boldsymbol{\mu}$ is obtained from Bayes formula:

$$\boldsymbol{\mu} \mid \mathbf{v}, \boldsymbol{\Omega} \sim \mathcal{N}\left(\boldsymbol{\mu}_{\mathsf{BL}}, \boldsymbol{\Sigma}_{\mathsf{BL}}^{\boldsymbol{\mu}}\right)$$

  where

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

  and

$$\boldsymbol{\Sigma}_{\mathsf{BL}}^{\boldsymbol{\mu}} = \left((\tau\boldsymbol{\Sigma})^{-1} + \mathbf{P}^T\boldsymbol{\Omega}^{-1}\mathbf{P}\right)^{-1}.$$

- But we really want the posterior for the returns

$$\mathbf{r} \mid \mathbf{v}, \boldsymbol{\Omega} \sim \mathcal{N}\left(\boldsymbol{\mu}_{\mathsf{BL}}, \boldsymbol{\Sigma}_{\mathsf{BL}}\right)$$

  where $\boldsymbol{\Sigma}_{\mathsf{BL}} = \boldsymbol{\Sigma}_{\mathsf{BL}}^{\boldsymbol{\mu}} + \boldsymbol{\Sigma}$.
- Using the matrix inversion lemma, we can further rewrite

$$\boldsymbol{\mu}_{\mathsf{BL}} = \boldsymbol{\pi} + \tau\boldsymbol{\Sigma}\mathbf{P}^T(\tau\mathbf{P}\boldsymbol{\Sigma}\mathbf{P}^T + \boldsymbol{\Omega})^{-1}(\mathbf{v} - \mathbf{P}\boldsymbol{\pi})$$
$$\boldsymbol{\Sigma}_{\mathsf{BL}} = (1+\tau)\boldsymbol{\Sigma} - \tau^2\boldsymbol{\Sigma}\mathbf{P}^T(\tau\mathbf{P}\boldsymbol{\Sigma}\mathbf{P}^T + \boldsymbol{\Omega})^{-1}\mathbf{P}\boldsymbol{\Sigma}.$$

Black-Litterman Model - Bayesian Approach 2

- In this case, $\boldsymbol{\mu}$ is not modeled as a random variable but simply as $\boldsymbol{\mu} = \boldsymbol{\pi}$.[29]
- The views are modeled on the random returns rather than on $\boldsymbol{\mu}$: $\mathbf{v} = \mathbf{P}\mathbf{r} + \mathbf{e}$.
- The conditional distribution is modeled as

$$\mathbf{v} \mid \mathbf{r} \sim \mathcal{N}\left(\mathbf{P}\mathbf{r}, \boldsymbol{\Omega}\right)$$

- Applying Bayes we get

$$\mathbf{r} \mid \mathbf{v}, \boldsymbol{\Omega} \sim \mathcal{N}\left(\boldsymbol{\mu}_{\mathsf{BL}}^{m}, \boldsymbol{\Sigma}_{\mathsf{BL}}^{m}\right)$$

where

$$\begin{aligned} \boldsymbol{\mu}_{\mathsf{BL}}^{m} &= \boldsymbol{\pi} + \boldsymbol{\Sigma}\mathbf{P}^{T}(\mathbf{P}\boldsymbol{\Sigma}\mathbf{P}^{T} + \boldsymbol{\Omega})^{-1}\left(\mathbf{v} - \mathbf{P}\boldsymbol{\pi}\right) \\ \boldsymbol{\Sigma}_{\mathsf{BL}}^{m} &= \boldsymbol{\Sigma} - \boldsymbol{\Sigma}\mathbf{P}^{T}(\mathbf{P}\boldsymbol{\Sigma}\mathbf{P}^{T} + \boldsymbol{\Omega})^{-1}\mathbf{P}\boldsymbol{\Sigma}. \end{aligned}$$

[29] A. 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