Portfolio Optimization

Financial Data: I.I.D. Modeling

Daniel P. Palomar (2025). *Portfolio Optimization: Theory and Application.*
Cambridge University Press.

portfoliooptimizationbook.com

Latest update: 2025-09-28

Outline

Executive Summary

- The efficient-market hypothesis suggests that a security's price reflects its intrinsic value, incorporating all available information.
- Then, prices can be modeled as a random walk with returns being independent and identically distributed (i.i.d.) random variables.
- These slides explore various methods to characterize the multivariate i.i.d. distribution of returns.
- Methods range from simple sample estimators to more advanced robust non-Gaussian estimators.
- Advanced estimators incorporate prior information through:
    - Shrinkage techniques
    - Factor modeling
    - Prior views
- Reference: (Palomar 2025, chap. 8)

Outline

I.I.D. Model

Introduction to Financial Data Modeling

- Financial data modeling is crucial for understanding and predicting market behaviors.
- It involves N securities or assets from various classes (e.g., bonds, equities, commodities).

Random Returns Representation

- Random returns of assets at time t are denoted by $\boldsymbol{x}_t$ or $\boldsymbol{r}_t$.
- The time index t can represent different periods (minutes to years).

I.I.D. Model for Returns

- Returns are modeled as:

$$\boldsymbol{x}_t = \boldsymbol{\mu} + \boldsymbol{\epsilon}_t.$$

- $\boldsymbol{\mu}$ represents the expected return, and $\boldsymbol{\epsilon}_t$ is the residual with zero mean.
- $\boldsymbol{\Sigma}$ denotes the covariance matrix of residuals.

I.I.D. Model

Efficient-Market Hypothesis

- The i.i.d. model is motivated by the efficient-market hypothesis (EMH).
- Eugene F. Fama, a proponent of EMH, won the Nobel Prize in 2013.

Random Walk Model

- The i.i.d. model corresponds to the random walk model on log-prices $\boldsymbol{y}_t$.
- Log-prices are defined as: $\boldsymbol{y}_t \triangleq \log \boldsymbol{p}_t$.
- This leads to the i.i.d. model when considering log-returns: $\boldsymbol{x}_t = \boldsymbol{y}_t - \boldsymbol{y}_{t-1}$.

Limitations of the I.I.D. Model

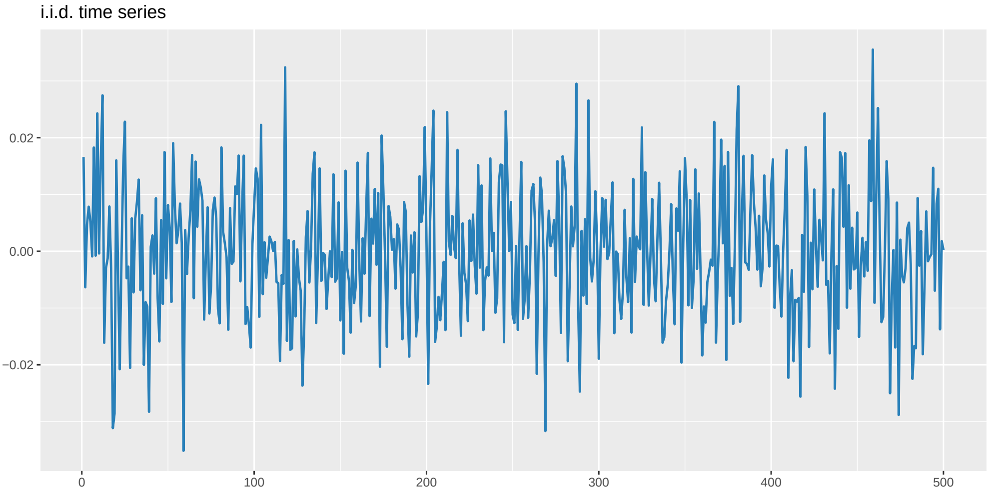
- It ignores temporal structure or dependency in financial data.

Sophisticated Time Series Models

- There is a myriad of different time series models developed over the past seven decades that attempt to capture the temporal structure.
- Recommended textbooks for financial data modeling are: (Meucci 2005), (Tsay 2010), (Ruppert and Matteson 2015), (Lütkepohl 2007), (Tsay 2013).

I.I.D. model

Example of a synthetic Gaussian i.i.d. time series:



Outline

Sample Estimators

Estimating I.I.D. Model Parameters

- The parameters $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ are estimated using historical data $\mathbf{x}_1, \ldots, \mathbf{x}_T$.
- It utilizes T past observations for estimation.

Sample Mean Estimator:

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

- It represents the average of past observations.

Sample Covariance Matrix Estimator:

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

- It measures variability around the sample mean.

Unbiasedness of Estimators

- The sample mean and covariance estimators are unbiased.
- The expected values of $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ equal the true values $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

Sample Estimators

Bias in Sample Covariance With $1/T$

- Using $1/T$ instead of $1/(T-1)$ in covariance estimation introduces bias.
- It results in an underestimate: $\mathbb{E}[\hat{\boldsymbol{\Sigma}}] = \left(1 - \frac{1}{T}\right)\boldsymbol{\Sigma}$.

Consistency of Estimators

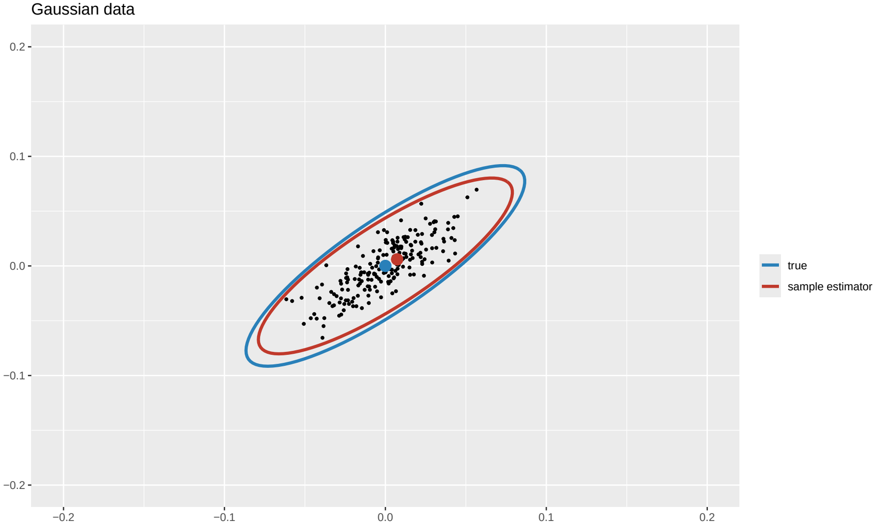
- Both estimators are consistent, converging to true values as $T \to \infty$.
- This is supported by the law of large numbers.

Estimation Error Reduction

- The estimation error decreases with increasing T.
- This is illustrated next through synthetic Gaussian data with $N = 100$.
- The normalized error approaches zero as the sample size grows.

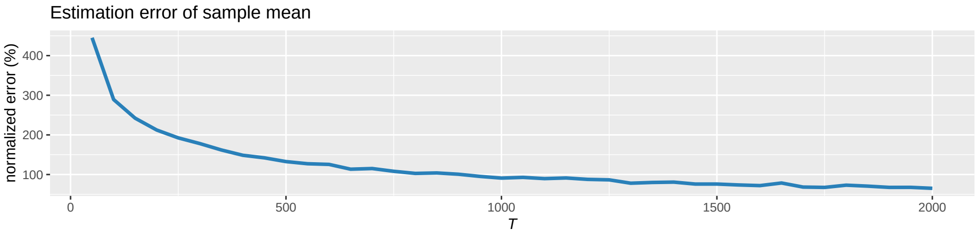
Sample Estimators

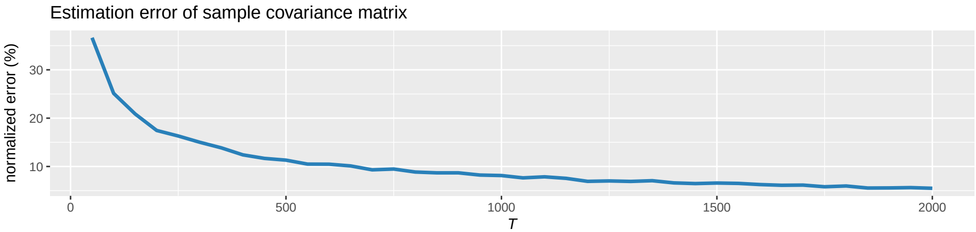
Illustration of sample mean and sample covariance matrix:



Sample Estimators

Estimation error of sample estimators versus number of observations (for Gaussian data with $N = 100$):





Sample Estimators

Challenges With Sample Estimators

- Simple and cost-effective but require a large number of observations T for accuracy.
- The sample mean $\hat{\boldsymbol{\mu}}$ is particularly inefficient, leading to noisy estimates.

High Estimation Error With Limited Data

- For $N = 100$ and $T = 500$, the normalized error of $\hat{\boldsymbol{\mu}}$ can exceed 100%.
- In words: the error magnitude is as large as the true value of $\boldsymbol{\mu}$.

Practical Limitations for Large T

- Lack of available historical data: Ideal data span (e.g., 20 years for $N = 500$) often exceeds available records.
- Lack of stationarity: Financial markets evolve, making long-term historical data less relevant.

Implications for Portfolio Optimization

- Limited data leads to noisy estimates of $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$.
- Estimation noise undermines the reliability of portfolio designs.
- It challenges the practical adoption of Markowitz's portfolio theory.

Outline

Location Estimators

Interpreting μ in the I.I.D. Model

- μ represents the central location of the distribution of random points.

Estimating the Central Location

- Various methods exist beyond classical sample estimators.
- Sample estimators are sensitive to extreme values and missing data.

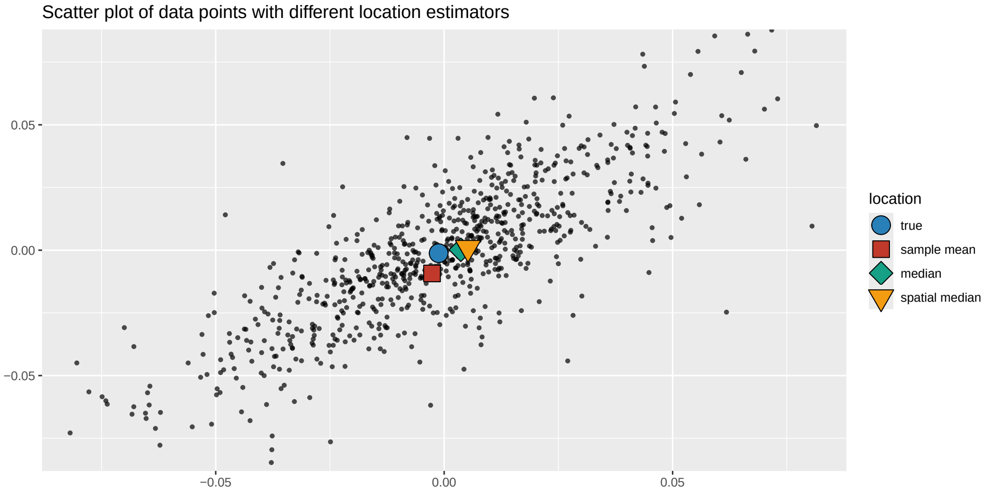
Robust Multivariate Location Estimators

- Developed due to the limitations of sample estimators and least squares.
- Aim to be less affected by outliers and missing values.
- Historical research dates back to the 1960s.

Some Methods to Estimate This Center

- Classical approach: least squares (LS).
- Median estimator.
- Spatial median estimator.

Illustration of different location estimators:



Least Squares (LS) Estimator

- It originates from Gauss's work in 1795 on planetary motions.
- It involves minimizing the squared difference between observed and predicted values.
- It is formulated as:

$$\underset{\boldsymbol{x}}{\text{minimize}} \quad \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2$$

- The closed-form solution is: $\boldsymbol{x}^\star = \left(\boldsymbol{A}^\mathsf{T}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^\mathsf{T}\boldsymbol{y}$.

Application to the I.I.D. Model

- Estimating $\boldsymbol{\mu}$ in the i.i.d. model can be seen as an LS problem:

$$\underset{\boldsymbol{\mu}}{\text{minimize}} \quad \sum_{t=1}^{T} \|\boldsymbol{x}_t - \boldsymbol{\mu}\|_2^2$$

- The solution coincides with the sample mean:

$$\hat{\boldsymbol{\mu}} = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{x}_t.$$

Least Squares (LS) Estimator

Challenges With the Sample Mean

- It lacks robustness against outliers and heavy-tailed distributions.
- The vulnerability to contaminated points affects the reliability of $\hat{\boldsymbol{\mu}}$.

Robust Estimation Needs

- Due to the limitations of LS and the sample mean, alternative robust estimators have been explored.
- They aim to improve reliability and accuracy in the presence of outliers and non-normal distributions.
- The study of robust multivariate location estimators dates back to the 1960s.

Median Estimator

Median as a Robust Estimator

- The median separates the higher half from the lower half of a sample.
- It is considered the "middle" value, providing a typical representation of the data.
- It is unaffected by extreme values, unlike the mean.
- It provides robustness against outliers.
- It represents a more "typical" value of the dataset.

Multivariate Median

- There are multiple ways to extend the median to a multivariate setting.
- Elementwise median is a straightforward extension.

Elementwise Median in the I.I.D. Model

- It can be formulated as the optimization problem:

$$\underset{\boldsymbol{\mu}}{\text{minimize}} \quad \sum_{t=1}^{T} \|\boldsymbol{x}_t - \boldsymbol{\mu}\|_1$$

- It uses the ℓ_1-norm to measure error, offering robustness against outliers.

Spatial Median Estimator

Spatial or Geometric Median

- An extension of the univariate median to multivariate data.
- It is formulated as the optimization problem:

$$\underset{\boldsymbol{\mu}}{\text{minimize}} \quad \sum_{t=1}^{T} \|\boldsymbol{x}_t - \boldsymbol{\mu}\|_2$$

- It uses the ℓ_2-norm (Euclidean distance) as the measure of error.

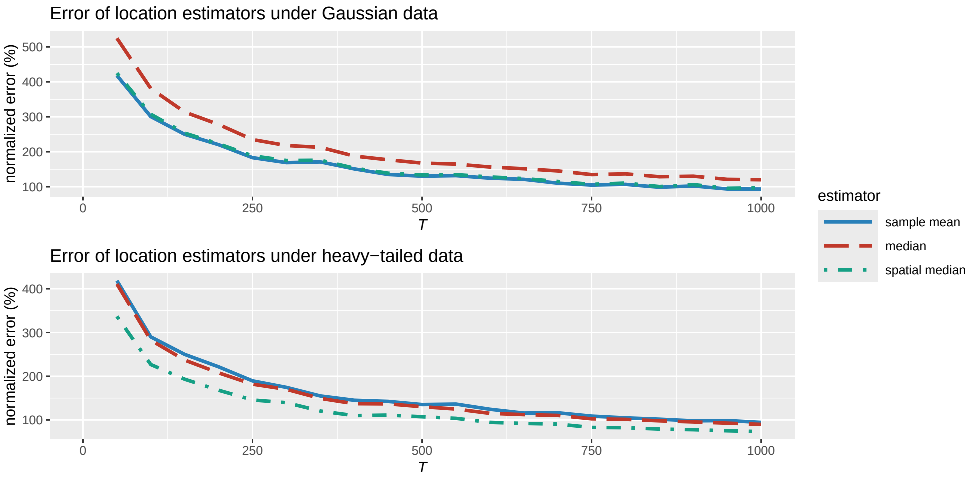
Characteristics of the Spatial Median

- The estimator for each element is not independent of other elements.
- For $N = 1$, it coincides with the univariate median.

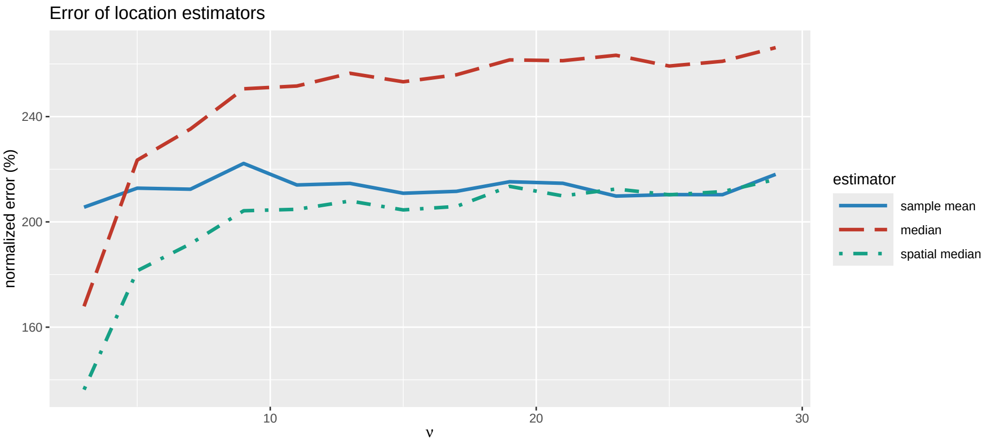
Solving for the Spatial Median

- The problem is a convex second-order cone problem (SOCP).
- It can be solved using SOCP solvers or iterative algorithms.
- Efficient iterative algorithms use the majorization-minimization (MM) method (Sun, Babu, and Palomar 2017).

Estimation error of location estimators versus number of observations (with $N = 100$):



Estimation error of location estimators versus degrees of freedom in a t distribution (with $T = 200$ and $N = 100$):



Outline

Preliminaries on ML Estimation

Maximum Likelihood Estimation (MLE)

- A fundamental technique in estimation theory.
- It involves selecting the parameter vector $\boldsymbol{\theta}$ that maximizes the likelihood of observing a given set of data.

Concept of MLE

- It is based on the probability distribution function (pdf) f of the random variable $\boldsymbol{x}$.
- For T independent observations $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T$, the likelihood is $f(\boldsymbol{x}_1) \times \cdots \times f(\boldsymbol{x}_T)$.
- MLE chooses the parameter $\boldsymbol{\theta}$ that maximizes this product for a family of distributions $f_{\boldsymbol{\theta}}$.

Optimization Problem in MLE

- It is formulated as:

$$\underset{\boldsymbol{\theta}}{\text{maximize}} \quad f_{\boldsymbol{\theta}}(\boldsymbol{x}_1) \times \cdots \times f_{\boldsymbol{\theta}}(\boldsymbol{x}_T).$$

- Equivalently, maximizing the log-likelihood:

$$\underset{\boldsymbol{\theta}}{\text{maximize}} \quad \sum_{t=1}^{T} \log f_{\boldsymbol{\theta}}(\boldsymbol{x}_t).$$

Preliminaries on ML Estimation

Theoretical Properties of MLE

- Asymptotically unbiased: The MLE's bias diminishes as $T \to \infty$.
- Asymptotically efficient: It achieves the Cramer-Rao bound, representing the lowest variance for an unbiased estimator.

Practical Considerations

- The effectiveness of MLE's asymptotic properties depends on the size of T.
- Determining how large T needs to be for these properties to hold is crucial in practice.

PDF for I.I.D. Model with Gaussian Residuals

- Assuming residuals follow a multivariate normal distribution:

$$f(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^N|\boldsymbol{\Sigma}|}}\exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\mathsf{T}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right).$$

- The model parameters are $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

MLE Formulation for Gaussian I.I.D. Model

$$\underset{\boldsymbol{\mu},\boldsymbol{\Sigma}}{\text{minimize}} \quad \log\det(\boldsymbol{\Sigma}) + \frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{x}_t-\boldsymbol{\mu})^\mathsf{T}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_t-\boldsymbol{\mu}).$$

Deriving MLE for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$

- Set the gradient with respect to $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ to zero.
- This results in the estimators:

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

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

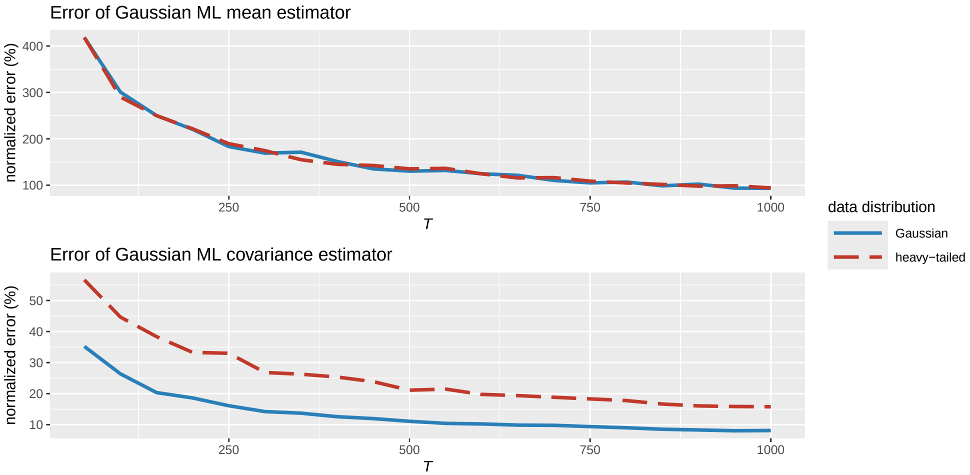
Comparison With Sample Estimators

- The MLE coincides with the sample mean and covariance estimators, except for the factor $1/T$ instead of $1/(T-1)$.
- The MLE of the covariance matrix is biased but asymptotically unbiased as $T \to \infty$.

Implications of MLE

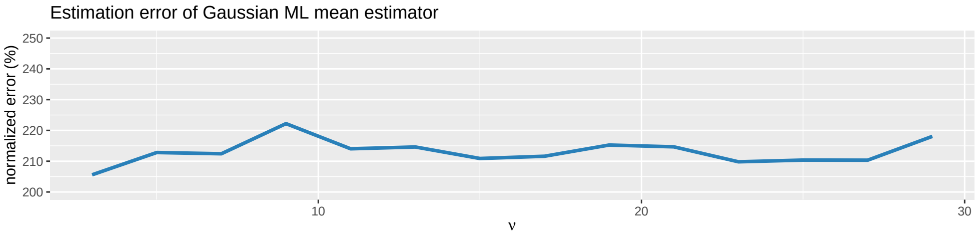
- Sample estimators are optimal for Gaussian-distributed data.
- For non-Gaussian distributions, the optimal ML estimators will be different.

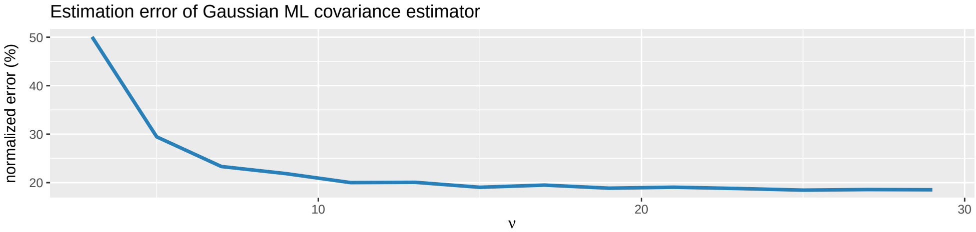
Estimation error of Gaussian ML estimators versus number of observations (with $N = 100$):



Numerical Experiments

Estimation error of Gaussian ML estimators versus degrees of freedom in a t distribution (with $T = 200$ and $N = 100$):





Outline

Heavy-Tailed ML Estimators

Gaussian ML Estimators and Heavy-Tailed Distributions

- Optimal for Gaussian-distributed data, but financial data often exhibit heavy tails.
- There is a need to assess the impact of heavy-tailed distributions on these estimators.

Properties of Gaussian ML Estimators

- They coincide with sample estimators for Gaussian data.
- They are unbiased and consistent, which are beneficial characteristics.

Evaluating Estimator Performance With Heavy Tails

- Despite being unbiased and consistent, they may not be the best choice for heavy-tailed data.
- It is important to consider the efficiency and robustness of estimators under such conditions.
- There is potential for improved estimators that better handle the peculiarities of financial data distributions.

The Failure of Gaussian ML Estimators

Impact of Heavy Tails on Estimation

- Heavy tails significantly affect covariance matrix estimation but not mean estimation.
- The error varies with tail heaviness: smaller ν means heavier tails and larger estimation error.

Visualizing Detrimental Effects

- The following scatter plots illustrate the impact of heavy tails and outliers on estimation.
- With $T = 10$ and $N = 2$, the sample covariance matrix is sensitive to these effects.

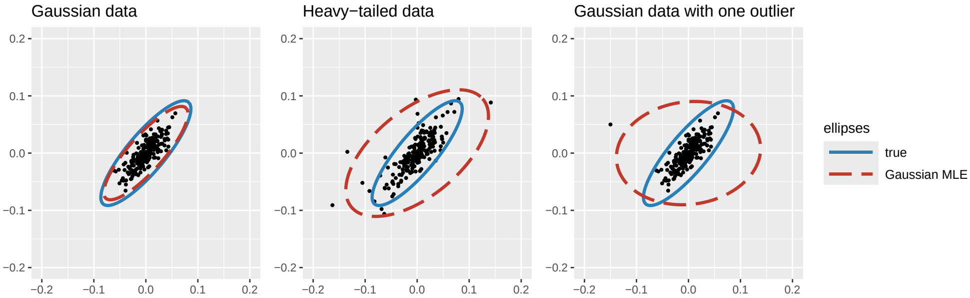
Consequences for Gaussian ML Estimators

- They perform well for Gaussian data but poorly with outliers or heavy-tailed data.
- Due its simplificty, the sample covariance matrix is widely used by practitioners.

Challenges in Practice

- The prevalence of heavy-tailed distributions in financial data complicates estimation.
- A single outlier can lead to significantly skewed estimations of the covariance matrix.
- There is a need for more robust estimation techniques that can handle outliers and non-Gaussian distributions effectively.

Effect of heavy tails and outliers in the Gaussian ML covariance matrix estimator:



Heavy-Tailed ML Estimation

MLE for Heavy-Tailed Distributions

- Gaussian MLE is not optimal for heavy-tailed data, which is common in finance.
- The Student t distribution is used to model heavy tails with parameter ν.

PDF for Multivariate t Distribution

- The probability density function is:

$$f(\boldsymbol{x}) = \frac{\Gamma((\nu+N)/2)}{\Gamma(\nu/2)\sqrt{(\nu\pi)^N|\boldsymbol{\Sigma}|}}\left(1+\frac{1}{\nu}(\boldsymbol{x}-\boldsymbol{\mu})^\mathsf{T}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right)^{-(\nu+N)/2}$$

- The parameters are $\boldsymbol{\mu}$ (location), $\boldsymbol{\Sigma}$ (scatter matrix), and ν (degrees of freedom).

MLE Formulation With t Distribution

- For fixed $\nu = 4$, the MLE problem simplifies to:

$$\underset{\boldsymbol{\mu},\boldsymbol{\Sigma}}{\text{minimize}} \quad \log\det(\boldsymbol{\Sigma}) + \frac{\nu+N}{T}\sum_{t=1}^{T}\log\left(1+\frac{1}{\nu}(\boldsymbol{x}_t-\boldsymbol{\mu})^\mathsf{T}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_t-\boldsymbol{\mu})\right)$$

Heavy-Tailed ML Estimation

Deriving MLE for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$

- Set the gradient with respect to $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ to zero.
- This results in the fixed-point equations for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$:

$$\boldsymbol{\mu} = \frac{\frac{1}{T}\sum_{t=1}^{T} w_t(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \times \boldsymbol{x}_t}{\frac{1}{T}\sum_{t=1}^{T} w_t(\boldsymbol{\mu}, \boldsymbol{\Sigma})}$$

$$\boldsymbol{\Sigma} = \frac{1}{T}\sum_{t=1}^{T} w_t(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \times (\boldsymbol{x}_t - \boldsymbol{\mu})(\boldsymbol{x}_t - \boldsymbol{\mu})^\mathsf{T}$$

where the weights $w_t(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ are defined as

$$w_t(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{\nu + N}{\nu + (\boldsymbol{x}_t - \boldsymbol{\mu})^\mathsf{T}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_t - \boldsymbol{\mu})}.$$

Advantages of Heavy-Tailed MLE

- Provides a more robust estimation for datasets with heavy tails.
- Weights observations differently, reducing the influence of outliers.

Heavy-Tailed ML Estimation

MM-based method to solve the heavy-tailed ML fixed-point equations

Initialization:

- Choose initial point $(\boldsymbol{\mu}^0, \boldsymbol{\Sigma}^0)$.
- Set iteration counter $k \leftarrow 0$.

Repeat (kth iteration):

1. Iterate the weighted sample mean and sample covariance matrix as

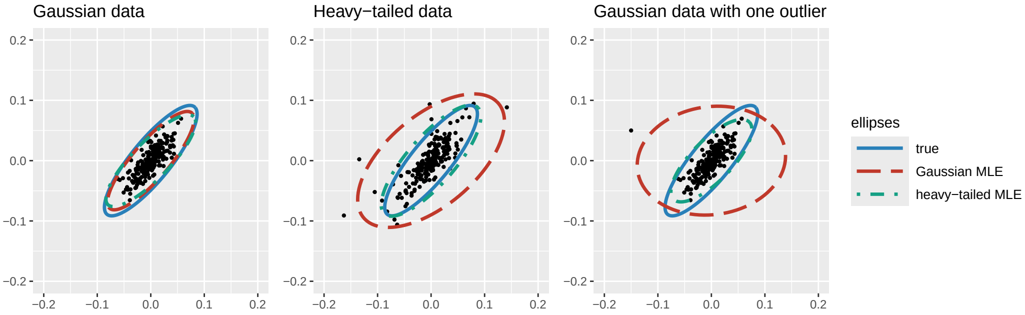
$$\boldsymbol{\mu}^{k+1} = \frac{\frac{1}{T}\sum_{t=1}^{T} w_t(\boldsymbol{\mu}^k, \boldsymbol{\Sigma}^k) \times \mathbf{x}_t}{\frac{1}{T}\sum_{t=1}^{T} w_t(\boldsymbol{\mu}^k, \boldsymbol{\Sigma}^k)}$$

$$\boldsymbol{\Sigma}^{k+1} = \frac{1}{T}\sum_{t=1}^{T} w_t(\boldsymbol{\mu}^{k+1}, \boldsymbol{\Sigma}^k) \times (\mathbf{x}_t - \boldsymbol{\mu}^{k+1})(\mathbf{x}_t - \boldsymbol{\mu}^{k+1})^\mathsf{T}$$

2. $k \leftarrow k + 1$

Until: convergence

Effect of heavy tails and outliers in heavy-tailed ML covariance matrix estimator:



Robust Estimators

Robust Estimators Overview

- Estimators resilient to outliers and deviations from the assumed distribution.
- They are reliable under non-ideal conditions.
- References: (Huber 1964), (Maronna 1976), (Maronna, Martin, and Yohai 2006), (Huber 2011), (Wiesel and Zhang 2014), (Zoubir et al. 2018, chap. 4), (Palomar 2025, chap. 3).

Measuring Robustness

- Influence function: Assesses the impact of deviations on the estimator.
- Breakdown point: Minimum fraction of contaminated data that compromises the estimator, with higher values indicating better robustness.

Sensitivity of Gaussian ML Estimators

- Gaussian-based estimators lack robustness and are highly sensitive to distribution tails.
- A single outlier can hinder the sample mean or covariance (breakdown point of $1/T$).

Robustness of Median and Heavy-Tailed ML Estimators

- The median offers greater robustness with a breakdown point of approximately 0.5.
- Heavy-tailed ML estimators effectively handle deviations, enhancing robustness.

Robust Estimators: M-Estimators*

Introduction to M-Estimators

- The term dates back to the 1960s, introduced by Huber (1964).
- They are a generalization of maximum likelihood estimators.
- They are defined by fixed-point equations for robust estimation of location and scatter.

Fixed-Point Equations for M-Estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$:

$$\frac{1}{T}\sum_{t=1}^{T} u_1\left(\sqrt{(\mathbf{x}_t-\boldsymbol{\mu})^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}_t-\boldsymbol{\mu})}\right)(\mathbf{x}_t-\boldsymbol{\mu}) = \mathbf{0}$$

$$\frac{1}{T}\sum_{t=1}^{T} u_2\left((\mathbf{x}_t-\boldsymbol{\mu})^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}_t-\boldsymbol{\mu})\right)(\mathbf{x}_t-\boldsymbol{\mu})(\mathbf{x}_t-\boldsymbol{\mu})^{\mathsf{T}} = \boldsymbol{\Sigma}$$

- The weight functions $u_1(\cdot)$ and $u_2(\cdot)$ must satisfy certain conditions.

Properties and Robustness of M-Estimators

- They are weighted sample mean and covariance matrix.
- They have a bounded influence function for robustness.
- The breakdown point is relatively low, despite robustness.

Robust Estimators: M-Estimators*

Other Robust Estimators (with higher breakdown points)

- Minimum volume ellipsoid (MVE).
- Minimum covariance determinant (MCD).

Gaussian ML Estimators as M-Estimators

- Trivial M-estimators with weight functions:

$$u_1(s) = u_2(s) = 1.$$

Relation to Heavy-Tailed ML Estimators

- M-estimators relate to heavy-tailed ML estimators with the choice:

$$u_1(s) = u_2(s^2) = \frac{\nu + N}{\nu + s^2}.$$

Tyler's Estimator for Scatter Matrix

- It was introduced by Tyler (1987) for heavy-tailed distributions.
- It is the most robust version of an M-estimator.

Elliptical Distribution Assumption

- It assumes a zero-mean elliptical distribution for $\boldsymbol{x}$.
- If the mean is not zero, it must be estimated and subtracted from the observations.

Normalization and ML Estimation

- Observations are normalized as
$$\boldsymbol{s}_t = \frac{\boldsymbol{x}_t}{\|\boldsymbol{x}_t\|_2}$$
.
- ML estimation is based on the normalized points, with the pdf (angular distribution) given by
$$f(\boldsymbol{s}) \propto \frac{1}{\sqrt{|\boldsymbol{\Sigma}|}} \left(\boldsymbol{s}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{s}\right)^{-N/2}$$
.

Robust Estimators: Tyler's Estimator

MLE Formulation

- Given T observations, the MLE is formulated as

$$\underset{\boldsymbol{\Sigma}}{\text{minimize}} \quad \log\det(\boldsymbol{\Sigma}) + \frac{N}{T}\sum_{t=1}^{T}\log\left(\mathbf{x}_t^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{x}_t\right)$$

- This leads to the fixed-point equation

$$\boldsymbol{\Sigma} = \frac{1}{T}\sum_{t=1}^{T} w_t(\boldsymbol{\Sigma}) \times \mathbf{x}_t\mathbf{x}_t^{\mathsf{T}},$$

  with weights given by

$$w_t(\boldsymbol{\Sigma}) = \frac{N}{\mathbf{x}_t^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{x}_t}.$$

Robustness and Weights

- The weights enhance robustness by down-weighting outliers.
- A solution exists if $T > N$.

Numerical Experiments

Comparison of Estimators for Mean and Covariance Matrix

- Gaussian MLE
- Tyler's estimator for covariance matrix (paired with spatial median for location)
- Heavy-tailed MLE

Observations

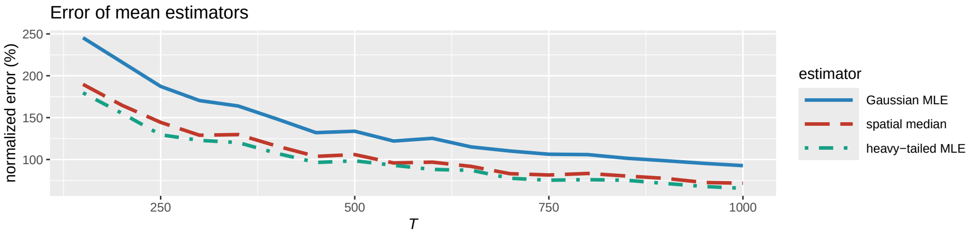
- Gaussian MLE and heavy-tailed MLE perform similarly for Gaussian tails.
- As tails become heavier (smaller ν), heavy-tailed MLE significantly outperforms Gaussian MLE.
- Tyler's estimator (with spatial median for the mean) is also not bad.

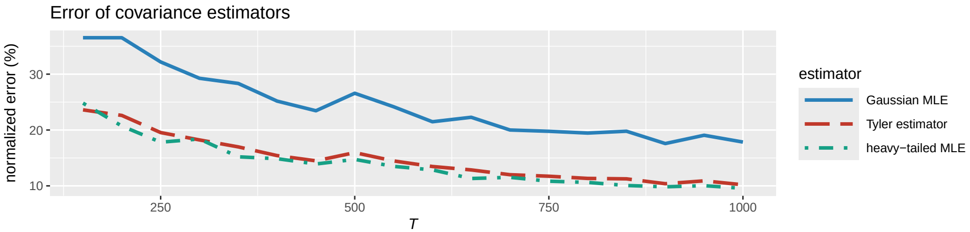
Conclusion

- Historical data-based mean vector $\boldsymbol{\mu}$ estimation errors can be substantial.
- Financial data's heavy-tailed nature necessitates robust heavy-tailed ML estimators.
- The computational cost of robust estimators is comparable to traditional sample estimators, with convergence in $3 \sim 5$ iterations.
- Practitioners often use factors from data providers for $\boldsymbol{\mu}$ estimation or opt for portfolio designs that bypass $\boldsymbol{\mu}$ estimation, such as GMVP or RPP.

Numerical Experiments

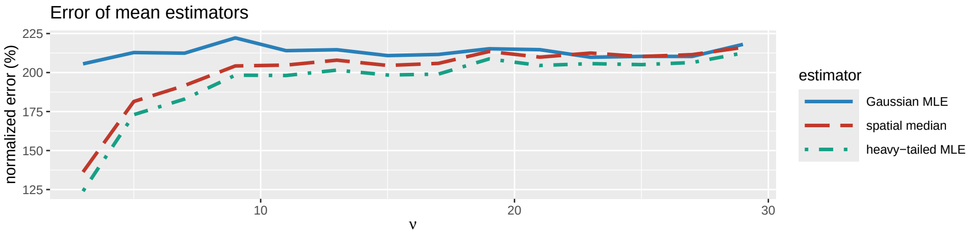
Estimation error of different ML estimators versus number of observations (for t-distributed heavy-tailed data with $\nu = 4$ and $N = 100$):

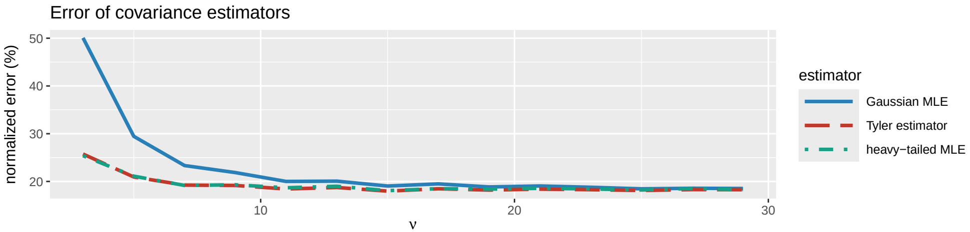




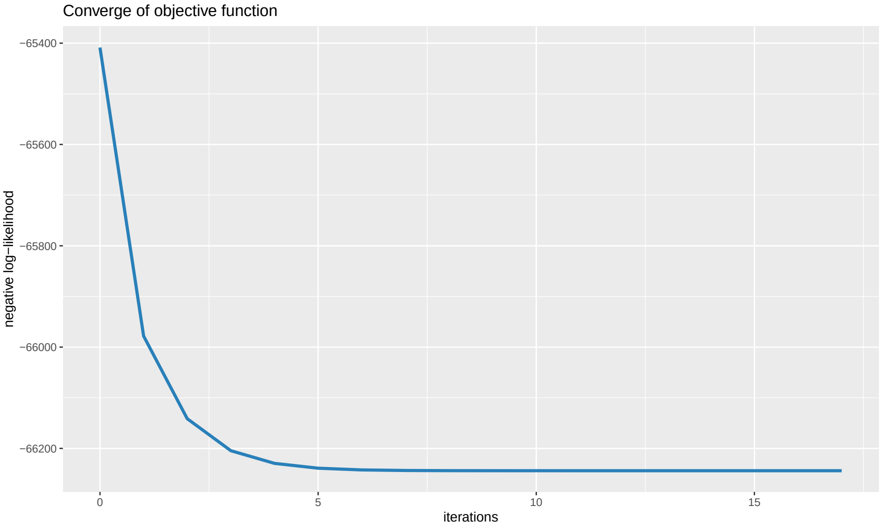
Numerical Experiments

Estimation error of different ML estimators versus degrees of freedom in a t distribution (with $T = 200$ and $N = 100$):





Convergence of robust heavy-tailed ML estimator:



Outline

Prior Information

Challenges With Historical Data

- A limited number of observations (T) can lead to high estimation errors.
- Practical settings often lack sufficient data for accurate parameter estimation.

Improving Estimators With Prior Information

- Researchers and practitioners have developed methods to integrate prior knowledge to enhance estimators.

Three Popular Methods to Incorporate Prior Information

- **Shrinkage** integrates prior knowledge through parameter targets and aims to improve estimation by pulling estimates towards a target.
- **Factor modeling** utilizes structural information about the data, which helps in reducing dimensionality and improving parameter estimation.
- The **Black-Litterman** approach merges historical data with subjective views, balancing empirical data with investor-specific insights.

Outline

Shrinkage

Introduction to Shrinkage

- This is a technique to reduce estimation error by introducing bias.
- It originated with Stein in 1955 and has been popular in finance for covariance matrix shrinkage since the early 2000s.

Bias-Variance Trade-off

- The mean squared error (MSE) of an estimator is the sum of its variance and squared bias.
- Small sample sizes lead to high variance, while larger samples may see bias dominate the error.

Stein's Seminal Contribution (Stein 1955)

- Demonstrated the benefit of introducing bias for overall error reduction.
- Shrinkage involves moving an estimator towards a target value to minimize error.

Shrinkage Estimator Formula

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

where ρ is the shrinkage factor, $\hat{\boldsymbol{\theta}}$ denotes the original estimator, and $\boldsymbol{\theta}^{\mathrm{tgt}}$ is the target.

Shrinkage

Implementation Considerations

- The **choice of target** (θ^{tgt}) represents prior information or market views.
- The **choice of shrinkage factor** (ρ) is critical for balancing the weight of the target in the estimator.

Choosing the Shrinkage Factor

- An **empirical choice** is based on cross-validation.
- An **analytical choice** utilizes advanced mathematical techniques (e.g., random matrix theory).

Application in Finance

- The parameter $\boldsymbol{\theta}$ could be the mean vector $\boldsymbol{\mu}$ or the covariance matrix $\boldsymbol{\Sigma}$.
- Estimators like the sample mean or covariance matrix can be adjusted using shrinkage for better accuracy.

Shrinking the Mean Vector

Sample Mean and Its Properties

- The sample mean $\hat{\boldsymbol{\mu}}$ is an unbiased estimator of the mean vector $\boldsymbol{\mu}$.
- Its distribution is characterized by $\hat{\boldsymbol{\mu}} \sim \mathcal{N}\left(\boldsymbol{\mu}, \frac{1}{T}\boldsymbol{\Sigma}\right)$.
- The mean squared error (MSE) is $\mathbb{E}\left[\left\|\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\right\|^2\right] = \frac{1}{T}\mathrm{Tr}(\boldsymbol{\Sigma})$.

Stein's Insight on Bias and MSE

- Stein's 1955 paper (Stein 1955) showed that allowing bias can reduce the overall MSE.
- Shrinkage introduces bias towards a target to achieve a lower MSE.

James-Stein Estimator

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

- It improves the MSE over the sample mean for any target $\boldsymbol{\mu}^{\mathrm{tgt}}$ with a properly chosen ρ:

$$\rho = \frac{(N+2)}{(N+2) + T \times \left(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}^{\mathrm{tgt}}\right)^{\mathsf{T}} \boldsymbol{\Sigma}^{-1} \left(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}^{\mathrm{tgt}}\right)}$$

Shrinking the Mean Vector

Adaptability of ρ

- $\rho \to 0$ as T increases, favoring the original sample mean.
- $\rho \to 0$ if the target significantly differs from the sample mean, acting as a safety mechanism.

Choosing the Target $\boldsymbol{\mu}^{\text{tgt}}$

- The choice of target is flexible, but MSE improvement depends on the target's informativeness.
- Common choices include:
  - Zero: $\boldsymbol{\mu}^{\text{tgt}} = \mathbf{0}$.
  - Grand mean: $\boldsymbol{\mu}^{\text{tgt}} = \frac{\mathbf{1}^{\mathsf{T}}\hat{\boldsymbol{\mu}}}{N} \times \mathbf{1}$.
  - Volatility-weighted grand mean: $\boldsymbol{\mu}^{\text{tgt}} = \frac{\mathbf{1}^{\mathsf{T}}\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}}{\mathbf{1}^{\mathsf{T}}\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}} \times \mathbf{1}$.

Shrinking the Covariance Matrix

Shrinkage in Covariance Matrix Estimation

- Introduces bias to reduce estimation error.
- The shrinkage estimator formula is:

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

  where $\boldsymbol{\Sigma}^{\mathsf{tgt}}$ is the target and ρ is the shrinkage factor.

Historical Context

- The concept was used in the 1980s in wireless communications as "diagonal loading".
- It gained popularity in finance in the early 2000s through the work of Ledoit and Wolf (Ledoit and Wolf 2003, 2004).

Common Targets for Covariance Matrix

- Scaled identity matrix: $\boldsymbol{\Sigma}^{\mathsf{tgt}} = \frac{1}{N}\mathsf{Tr}(\hat{\boldsymbol{\Sigma}}) \times \boldsymbol{I}$.
- Diagonal matrix: $\boldsymbol{\Sigma}^{\mathsf{tgt}} = \mathsf{Diag}(\hat{\boldsymbol{\Sigma}})$.
- Equal-correlation matrix, where the off-diagonal elements are equal to the average cross-correlation.

Shrinking the Covariance Matrix

Determining the Shrinkage Factor ρ

- Can be empirical (via cross-validation) or analytical: random matrix theory (RMT).
- Ledoit and Wolf popularized the RMT approach to minimize $\mathbb{E}\left[\left\|\hat{\boldsymbol{\Sigma}}^{\text{sh}} - \boldsymbol{\Sigma}\right\|_{\text{F}}^2\right]$.
- The asymptotic RMT formula for ρ is (Ledoit and Wolf 2003, 2004):

$$\rho = \min\left(1, \frac{\frac{1}{T}\sum_{t=1}^{T}\left\|\hat{\boldsymbol{\Sigma}} - \boldsymbol{x}_t\boldsymbol{x}_t^{\mathsf{T}}\right\|_{\text{F}}^2}{\left\|\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^{\text{tgt}}\right\|_{\text{F}}^2}\right).$$

Extension to Heavy-Tailed Distributions

- Shrinkage factor ρ can be adapted for heavy-tailed distributions, enhancing robustness.

Alternative Error Measures

- Error in terms of the inverse covariance matrix: $\mathbb{E}\left[\left\|\left(\hat{\boldsymbol{\Sigma}}^{\text{sh}}\right)^{-1} - \boldsymbol{\Sigma}^{-1}\right\|_{\text{F}}^2\right]$.
- Choose ρ to maximize directly the achieved Sharpe ratio.

Nonlinear Shrinkage

- Extends the idea to the eigenvalues of the covariance matrix.
- Requires increased mathematical sophistication.

Numerical Experiments

Shrinkage Estimators and Observation Size

- Estimation error analysis for shrinkage estimators with synthetic Gaussian data.
- Clear improvement in mean vector estimation; modest improvement in covariance matrix estimation.
- Shrinkage benefits decrease as the number of observations (T) increases.

Shrinkage to Zero

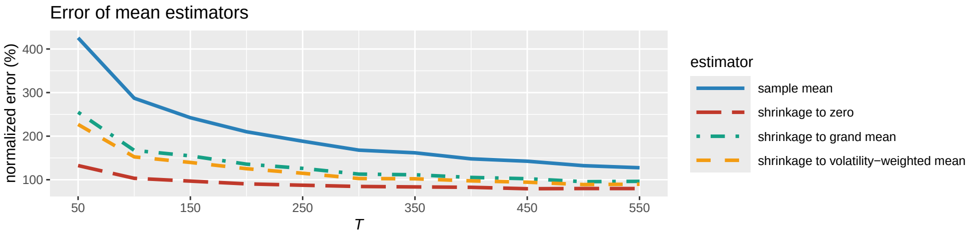
- Shrinkage to zero return vector yields the best results.
- Aligns with the efficient-market hypothesis, suggesting current prices reflect all available information.

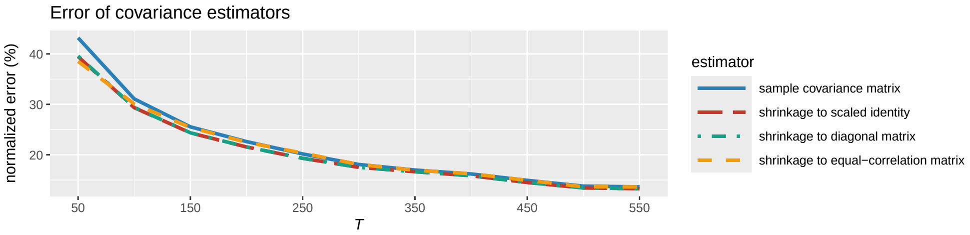
Consideration of MSE in Portfolio Optimization

- Numerical results are based on the MSE of estimators.
- MSE may not be the optimal error measure in portfolio optimization contexts.
- The importance of using more appropriate error measures is highlighted.

Numerical Experiments

Estimation error of different shrinkage estimators versus number of observations (for Gaussian data with $N = 100$):





Outline

Factor Models

Factor Modeling in Finance

- Incorporates prior information into asset return models.
- Found in numerous finance textbooks.

Single-Factor Model

- The simplest form of factor modeling.
- The equation is:

$$\boldsymbol{x}_t = \boldsymbol{\alpha} + \boldsymbol{\beta} f_t^{\text{mkt}} + \boldsymbol{\epsilon}_t$$

- $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ represent asset-specific alpha and beta.
- f_t^{mkt} is the market factor and $\boldsymbol{\epsilon}_t$ is the zero-mean residual.

Connection to CAPM

- The single-factor model is related to the Capital Asset Pricing Model (CAPM).
- CAPM assumes zero alpha and relates expected excess returns to market beta.

Factor Models

Multi-Factor Modeling

- A generalization of the single-factor model.
- The equation is:

$$\boldsymbol{x}_t = \boldsymbol{\alpha} + \boldsymbol{B}\boldsymbol{f}_t + \boldsymbol{\epsilon}_t$$

- $\boldsymbol{f}_t$ contains K factors and $\boldsymbol{B}$ has factor loadings.
- Dynamic factor models include time-dependency in the factors.

Idiosyncratic Component

- The residual term $\boldsymbol{\epsilon}_t$ is assumed to have a diagonal covariance matrix $\boldsymbol{\Psi}$.
- It captures asset-specific noise not explained by the factors.

Factor Models

Mean and Covariance Matrix:

$$\boldsymbol{\mu} = \boldsymbol{\alpha} + \boldsymbol{B}\boldsymbol{\mu}_f$$

$$\boldsymbol{\Sigma} = \boldsymbol{B}\boldsymbol{\Sigma}_f\boldsymbol{B}^\mathsf{T} + \boldsymbol{\Psi}$$

- Decomposes the covariance into low-rank and full-rank diagonal components.
- For a single-factor model, the covariance matrix is: $\boldsymbol{\Sigma} = \sigma_f^2 \boldsymbol{\beta}\boldsymbol{\beta}^\mathsf{T} + \boldsymbol{\Psi}$.

Parameter Reduction

- Factor models reduce the number of parameters that need to be estimated.
- For example, for $N = 500$ assets and $K = 3$ factors, the number of parameters is reduced from 125,750 to 1,503.

Types:

- **Macroeconomic factor models** use observable economic factors and have unknown loadings.
- **Fundamental factor models** use loadings from asset characteristics and have unknown factors.
- **Statistical factor models** have both unknown factors and loadings.

Macroeconomic Factor Models Overview

- Allows for the integration of economic indicators into the analysis of asset returns.
- Utilizes observable economic time series as factors (e.g., market index, GDP growth rate, interest rates, inflation rates).
- Factors are often proprietary, derived from complex analyses of various data sources.
- Investment funds may pay high premiums for access to these factors, whereas small investors might use publicly available data.

Parameter Estimation

- With known factors, model parameters ($\boldsymbol{\alpha}$ and $\boldsymbol{B}$) can be estimated through least squares regression:

$$\underset{\boldsymbol{\alpha}, \boldsymbol{B}}{\text{minimize}} \sum_{t=1}^{T} \left\| \boldsymbol{x}_t - (\boldsymbol{\alpha} + \boldsymbol{B}\boldsymbol{f}_t) \right\|_2^2,$$

- The mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ are derived from the estimated parameters as per the factor model equation.

Fundamental Factor Models

Fundamental Factor Models Overview

- Use observable asset characteristics, known as fundamentals, to define factors.
- Common fundamentals include industry classification, market capitalization, and style classification (value, growth).

Industry Approaches

- Fama-French approach:
  - Form portfolios based on asset characteristics to derive factors $\boldsymbol{f}_t$.
  - Loadings $\boldsymbol{B}$ are estimated similarly to macroeconomic models.
  - The original model had $K = 3$ factors: firm size, book-to-market values, and excess market return (Fama and French 1992).
  - It was extended to $K = 5$ factors, including profitability and investment patterns (Fama and French 2015).
- Barra risk factor analysis approach:
  - Loadings $\boldsymbol{B}$ are constructed from asset characteristics.
  - Factors $\boldsymbol{f}_t$ are estimated via regression, opposite of macroeconomic models.
  - This approach was developed by Barra Inc. in 1975.

Statistical Factor Models

Statistical Factor Models Overview

- Both the factors $\boldsymbol{f}_t$ and the loading matrix $\boldsymbol{B}$ are unknown.
- They introduce structure to the covariance matrix $\boldsymbol{\Sigma}$ as a low-rank plus diagonal matrix.

Covariance Matrix Structure

- $\boldsymbol{\Sigma}$ is decomposed into $\boldsymbol{B}\boldsymbol{\Sigma}_f\boldsymbol{B}^\mathsf{T}$ (low-rank) and $\boldsymbol{\Psi}$ (diagonal).
- The factors are assumed to be zero-mean and normalized for simplification.

Heuristic Formulation

Approximate the sample covariance matrix $\hat{\boldsymbol{\Sigma}}$ with the desired structure:

$$\underset{\boldsymbol{B},\boldsymbol{\psi}}{\text{minimize}} \quad \|\hat{\boldsymbol{\Sigma}} - \left(\boldsymbol{B}\boldsymbol{B}^\mathsf{T} + \text{Diag}(\boldsymbol{\psi})\right)\|_\mathsf{F}^2.$$

Statistical Factor Models

ML Estimation Under the Gaussian Assumption

- The ML estimation is formulated by imposing the covariance structure:

$$\begin{array}{ll}\underset{\boldsymbol{\alpha},\boldsymbol{\Sigma},\boldsymbol{B},\boldsymbol{\psi}}{\text{minimize}} & \log\det(\boldsymbol{\Sigma}) + \dfrac{1}{T}\displaystyle\sum_{t=1}^{T}(\boldsymbol{x}_t - \boldsymbol{\alpha})^\mathsf{T}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_t - \boldsymbol{\alpha}) \\ \text{subject to} & \boldsymbol{\Sigma} = \boldsymbol{B}\boldsymbol{B}^\mathsf{T} + \mathrm{Diag}(\boldsymbol{\psi}).\end{array}$$

- The nonconvex nature of the problem makes it challenging.
- Iterative algorithms have been developed to address the nonconvex optimization challenges.

Extensions

- Heavy-tailed distributions can be accommodated.
- The structure of financial data, such as nonnegative asset correlation, can be integrated into model formulations.

Principal Component Analysis (PCA)*

PCA in a Nutshell

- It is a statistical technique for handling high-dimensional datasets by focusing on the most significant variance components.
- It facilitates more efficient data analysis and interpretation.

PCA for Dimension Reduction

- PCA identifies the directions of maximum variance in high-dimensional data.
- It simplifies data analysis by reducing dimensions to a lower-dimensional subspace.

PCA Methodology

- It maximizes the variance along a direction $\boldsymbol{u}$: $\mathrm{Var}(\boldsymbol{u}^\top \boldsymbol{x}) = \boldsymbol{u}^\top \boldsymbol{\Sigma} \boldsymbol{u}$.
- The solution is found through eigenvalue decomposition: $\boldsymbol{\Sigma} \approx \boldsymbol{U}^{(K)} \boldsymbol{D}^{(K)} \boldsymbol{U}^{(K)\top}$.
- $\boldsymbol{U}^{(K)}$ contains the first K eigenvectors, and $\boldsymbol{D}^{(K)}$ has the largest K eigenvalues.

Principal Component Analysis (PCA)*

PCA in Statistical Factor Models

- It approximates the solution to the statistical factor model by performing PCA on the sample covariance matrix.
- As a heuristic, it keeps K principal components and uses a scaled identity matrix for the diagonal component $\boldsymbol{\Psi}$.
- The approximate solution to the factor model is:

$$\boldsymbol{B} = \boldsymbol{U}^{(K)}\mathrm{Diag}\left(\sqrt{\lambda_1 - \kappa}, \ldots, \sqrt{\lambda_K - \kappa}\right)$$

$$\boldsymbol{\Psi} = \kappa \boldsymbol{I}$$

where κ is the average of the $N - K$ smallest eigenvalues.

PCA Estimator for the Covariance Matrix

- The PCA estimator for the covariance matrix is:

$$\hat{\boldsymbol{\Sigma}} = \boldsymbol{U}\mathrm{Diag}\left(\lambda_1, \ldots, \lambda_K, \kappa, \ldots, \kappa\right)\boldsymbol{U}^\mathsf{T}.$$

- This achieves noise averaging of the smallest eigenvalues, similar to the concept of shrinkage.

Numerical Experiments

Evaluation of Covariance Matrix Estimation Under a Factor Model

- Estimation accuracy depends on how well the data follows the factor model structure.
- Incorrect assumptions about the factor model structure can lead to poorer results than using the sample covariance matrix.

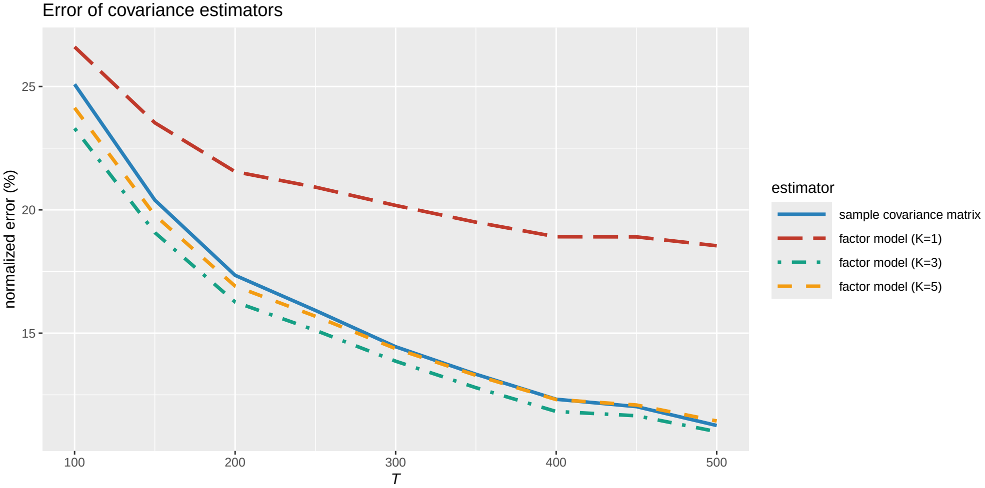
Importance of Model Choice

- The decision to use a factor model for covariance matrix estimation must be made cautiously.
- Factor modeling strategies and their implications on trading are elaborated in (Fabozzi, Focardi, and Kolm 2010).

Observations From Synthetic Data Analysis

- An estimation error comparison between factor model-based and sample covariance matrices is shown for synthetic Gaussian data.
- The data complies with a factor model structure having $K = 3$ principal components.
- Accurate estimation with the correct K improves results over the sample covariance matrix.
- Incorrect K values, such as $K = 1$, can significantly worsen estimation accuracy.

Estimation error of covariance matrix under factor modeling versus number of observations (with $N = 100$):



Outline

Black-Litterman Model Basics

- It combines historical data with an investor's prior information.
- It is a standard in finance, detailed in textbooks like (Fabozzi, Focardi, and Kolm 2010) and (Meucci 2005).

Components of the Black-Litterman Model

- **Market equilibrium** is an estimate of $\boldsymbol{\mu}$ from the market, denoted as $\boldsymbol{\pi} = \hat{\boldsymbol{\mu}}$:

$$\boldsymbol{\pi} = \boldsymbol{\mu} + \boldsymbol{\epsilon}$$

  with $\boldsymbol{\epsilon}$ being zero-mean and having a covariance of $\tau\boldsymbol{\Sigma}$, where τ is often set to $1/T$.
- **Investor's views** consist of K views on asset returns, expressed as:

$$\boldsymbol{v} = \boldsymbol{P}\boldsymbol{\mu} + \boldsymbol{e}$$

  where $\boldsymbol{v}$ and $\boldsymbol{P}$ represent the views and their relation to asset returns, and the error term $\boldsymbol{e}$ is zero-mean with covariance $\boldsymbol{\Omega}$.

Views

Quantitative and Qualitative Views

- Quantitative views specify expected returns and their uncertainties.
- Qualitative views provide directional expectations (e.g., bullish, bearish) without specific return figures.

Example of Quantitative Investor's Views Two independent views on $N = 5$ stocks:

- View 1: stock 1 expected to return 1.5% with a standard deviation of 1%.
- View 2: stock 3 expected to outperform stock 2 by 4% with a 1% standard deviation.

Mathematical Representation

- The views are expressed as:

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

- The covariance of the error $\boldsymbol{e}$ is:

$$\boldsymbol{\Omega} = \begin{bmatrix} 1\%^2 & 0 \\ 0 & 1\%^2 \end{bmatrix}.$$

Combining Market Equilibrium and Investor's Views

- Various mathematical formulations (least squares, maximum likelihood, Bayesian) yield similar solutions for integrating market equilibrium with an investor's views.

Weighted Least Squares Formulation

- A compact representation is:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\mu} + \boldsymbol{n},$$

where $\boldsymbol{y} = \begin{bmatrix} \boldsymbol{\pi} \\ \boldsymbol{v} \end{bmatrix}$, $\boldsymbol{X} = \begin{bmatrix} \boldsymbol{I} \\ \boldsymbol{P} \end{bmatrix}$, and the noise covariance is $\boldsymbol{V} = \begin{bmatrix} \tau\boldsymbol{\Sigma} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Omega} \end{bmatrix}$.

- The problem is formulated as:

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

- The solution is:

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

Merging the Market Equilibrium with the Views*

Original Bayesian Formulation of the Black-Litterman Model

- Returns $\boldsymbol{x}$ are assumed to follow a normal distribution: $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.
- The mean $\boldsymbol{\mu}$ is also modeled as random with a Gaussian distribution, $\boldsymbol{\mu} \sim \mathcal{N}(\boldsymbol{\pi}, \tau\boldsymbol{\Sigma})$, where $\boldsymbol{\pi}$ is the best guess for $\boldsymbol{\mu}$ and $\tau\boldsymbol{\Sigma}$ represents the uncertainty.
- Views are modeled with a Gaussian distribution: $\boldsymbol{P}\boldsymbol{\mu} \sim \mathcal{N}(\boldsymbol{v}, \boldsymbol{\Omega})$.
- The posterior distribution of $\boldsymbol{\mu}$ given the views is $\boldsymbol{\mu} \mid \boldsymbol{v}, \boldsymbol{\Omega} \sim \mathcal{N}\left(\boldsymbol{\mu}_{\text{BL}}, \boldsymbol{\Sigma}_{\text{BL}}\right)$.
- The posterior mean $\boldsymbol{\mu}_{\text{BL}}$ matches the weighted least squares solution.
- The posterior covariance $\boldsymbol{\Sigma}_{\text{BL}}$ includes the original covariance $\boldsymbol{\Sigma}$ and the covariance of the posterior mean.
- The mean estimator is:

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

- The covariance matrix estimator is:

$$\boldsymbol{\Sigma}_{\text{BL}} = (1+\tau)\boldsymbol{\Sigma} - \tau^2\boldsymbol{\Sigma}\boldsymbol{P}^{\mathsf{T}}\left(\tau\boldsymbol{P}\boldsymbol{\Sigma}\boldsymbol{P}^{\mathsf{T}} + \boldsymbol{\Omega}\right)^{-1}\boldsymbol{P}\boldsymbol{\Sigma}.$$

Merging the Market Equilibrium with the Views*

Alternative Bayesian Formulation

- A variation introduced by (Meucci 2005) models views on random returns as $\boldsymbol{v} = \boldsymbol{P}\boldsymbol{x} + \boldsymbol{e}$, which differs from the original formulation where views are on $\boldsymbol{\mu}$.
- This approach leads to a posterior distribution of returns with results akin to the original Black-Litterman model.

Impact of Parameter τ on the Black-Litterman Estimator

- For $\tau = 0$, the market equilibrium is considered completely accurate, leading to $\boldsymbol{\mu}_{\text{BL}} = \boldsymbol{\pi}$.
- As $\tau \to \infty$, the market equilibrium is disregarded, and the investor's views solely influence the outcome, resulting in $\boldsymbol{\mu}_{\text{BL}} = \left(\boldsymbol{P}^{\mathsf{T}}\boldsymbol{\Omega}^{-1}\boldsymbol{P}\right)^{-1}\left(\boldsymbol{P}^{\mathsf{T}}\boldsymbol{\Omega}^{-1}\boldsymbol{v}\right)$.
- For $0 < \tau < \infty$, $\boldsymbol{\mu}_{\text{BL}}$ represents a blend of the market equilibrium and investor's views, embodying the principle of shrinkage.

Outline

Summary

Countless models for financial data exist in the literature. The i.i.d. model, while rough, is functional and widely used. Key points of the i.i.d. model for financial data include:

- **Sample estimators perform poorly** since Gaussian assumptions don't hold in practice.
- **Robust estimators are necessary** to handle outliers, like spatial median and Tyler estimator.
- **Heavy-tailed estimators suit financial data well** as they are naturally robust. Simple iterative algorithms can compute them.
- **Estimating mean vector from historical data is extremely noisy**. Practitioners use premium data provider factors for regression instead.
- **Prior information should be used when available** via shrinkage, factor modeling, or Black-Litterman information fusion.

References I

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

Fama, E. F., and K. R. French. 1992. "The Cross-Section of Expected Stock Returns." *The Journal of Finance* 47 (2): 427–65.

———. 2015. "A Five-Factor Asset Pricing Model." *The Journal of Finance* 116 (1): 1–22.

Huber, P. J. 1964. "Robust Estimation of a Location Parameter." *The Annals of Statistics* 53 (1): 73–101.

———. 2011. *Robust Statistics*. Springer.

Ledoit, O., and M. Wolf. 2003. "Improved Estimation of the Covariance Matrix of Stock Returns with an Application to Portfolio Selection." *Journal of Empirical Finance* 10 (5): 603–21.

References II

———. 2004. "A Well-Conditioned Estimator for Large-Dimensional Covariance Matrices." *Journal of Multivariate Analysis* 88 (2): 365–411.

Lütkepohl, H. 2007. *New Introduction to Multiple Time Series Analysis*. Springer.

Maronna, R. A. 1976. "Robust *M*-Estimators of Multivariate Location and Scatter." *The Annals of Statistics* 4 (1): 51–67.

Maronna, R. A., D. R. Martin, and V. J. Yohai. 2006. *Robust Statistics: Theory and Methods*. Wiley Series in Probability and Statistics. Wiley.

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

Palomar, D. P. 2025. *Portfolio Optimization: Theory and Application*. Cambridge University Press.

Ruppert, D., and S. D. Matteson. 2015. *Statistics and Data Analysis for Financial Engineering: With R Examples*. 2nd ed. Springer.

References III

Stein, C. 1955. "Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution." *Proceedings of the 3rd Berkeley Symposium on Probability and Statistics* 1: 197–206.

Sun, Y., P. Babu, and D. P. Palomar. 2017. "Majorization-Minimization Algorithms in Signal Processing, Communications, and Machine Learning." *IEEE Transactions on Signal Processing* 65 (3): 794–816.

Tsay, R. S. 2010. *Analysis of Financial Time Series*. 3rd ed. John Wiley & Sons.

———. 2013. *Multivariate Time Series Analysis: With R and Financial Applications*. John Wiley & Sons.

Tyler, D. E. 1987. "A Distribution-Free *M*-Estimator of Multivariate Scatter." *The Annals of Statistics* 15 (1): 234–51.

References IV

Wiesel, A., and T. Zhang. 2014. *Structured Robust Covariance Estimation*. Foundations and Trends in Signal Processing, Now Publishers.

Zoubir, A. M., V. Koivunen, E. Ollila, and M. Muma. 2018. *Robust Statistics for Signal Processing*. Cambridge University Press.