Portfolio Optimization
Financial Data: IID Modeling

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

Outline

Abstract

The efficient-market hypothesis suggests that a security's price reflects its intrinsic value, incorporating all available information, and 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, ranging from simple sample estimators to more advanced robust non-Gaussian estimators that incorporate prior information through shrinkage, factor modeling, or prior views (Palomar 2025, chap. 3).

Outline

IID model

- **Introduction to financial data modeling:**
  - Financial data modeling is crucial for understanding and predicting market behaviors.
  - Involves *N* securities or assets from various classes (e.g., bonds, equities, commodities).

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

- **i.i.d. model for returns:**
  - Returns modeled as:

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

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

IID model

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

- **Random walk model:**
  - i.i.d. model corresponds to the random walk model on log-prices $\boldsymbol{y}_t$.
  - Log-prices defined as: $\boldsymbol{y}_t \triangleq \log \boldsymbol{p}_t$.
  - Leads to 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:**
  - 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: (Meucci 2005), (Tsay 2010), (Ruppert and Matteson 2015), (Lütkepohl 2007), (Tsay 2013).

IID model

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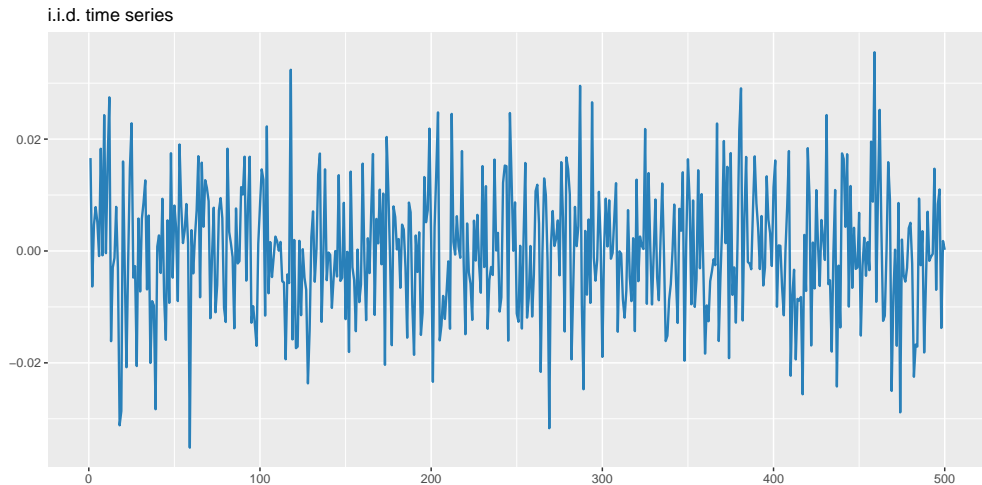
i.i.d. time series

Outline

Sample estimators

- **Estimating i.i.d. model parameters:**
  - Parameters $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ are estimated using historical data $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T$.
  - Utilizes T past observations for estimation.
- **Sample mean estimator:**

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

  - Represents the average of past observations.
- **Sample covariance matrix estimator:**

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

  - Measures variability around the sample mean.
- **Unbiasedness of estimators:**
  - Sample mean and covariance estimators are unbiased.
  - 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.
  - Results in an underestimate: $\mathbb{E}\big[\hat{\boldsymbol{\Sigma}}\big] = \big(1 - \frac{1}{T}\big)\boldsymbol{\Sigma}$.

- **Consistency of estimators:**
  - Both estimators are consistent, converging to true values as $T \to \infty$.
  - Supported by the law of large numbers.

- **Estimation error reduction:**
  - Estimation error decreases with increasing T.
  - Illustrated next through synthetic Gaussian data with $N = 100$.
  - Normalized error approaches zero as sample size grows.

Sample estimators

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

Estimation error of sample mean

Estimation error of sample covariance matrix
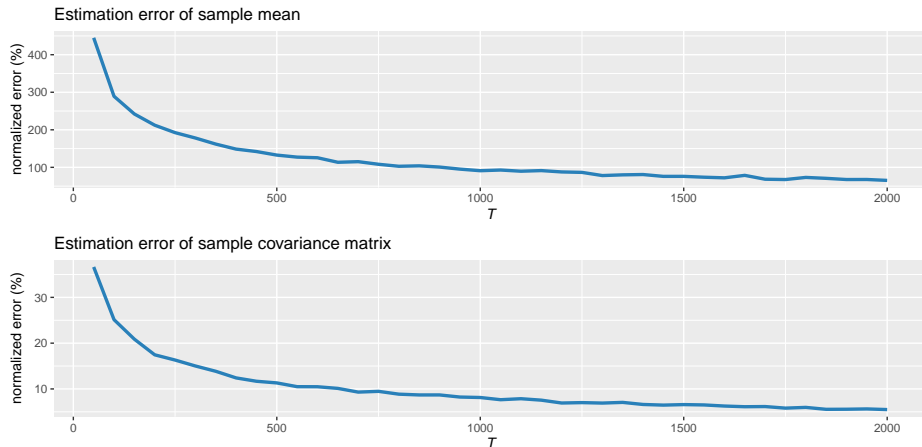
Sample estimators

- **Challenges with sample estimators:**
  - Simple and cost-effective but require a large number of observations T for accuracy.
  - Sample mean $\hat{\mu}$ is particularly inefficient, leading to noisy estimates.
- **High estimation error with limited data:**
  - For $N = 100$ and $T = 500$, the normalized error of $\hat{\mu}$ can exceed 100%.
  - In words: the error magnitude is as large as the true value of μ!
- **Practical limitations for large T:**
  1. **Lack of available historical data:**
     - Ideal data span (e.g., 20 years for $N = 500$) often exceeds available records.
  2. **Lack of stationarity:**
     - Financial markets evolve, making long-term historical data less relevant.
- **Implications for portfolio optimization:**
  - Limited data leads to noisy estimates of $\hat{\mu}$ and $\hat{\Sigma}$.
  - Estimation noise undermines the reliability of portfolio designs.
 - 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.

Location estimators

Illustration of different location estimators:

Scatter plot of data points with different location estimators

location
- true
- sample mean
- median
- spatial median
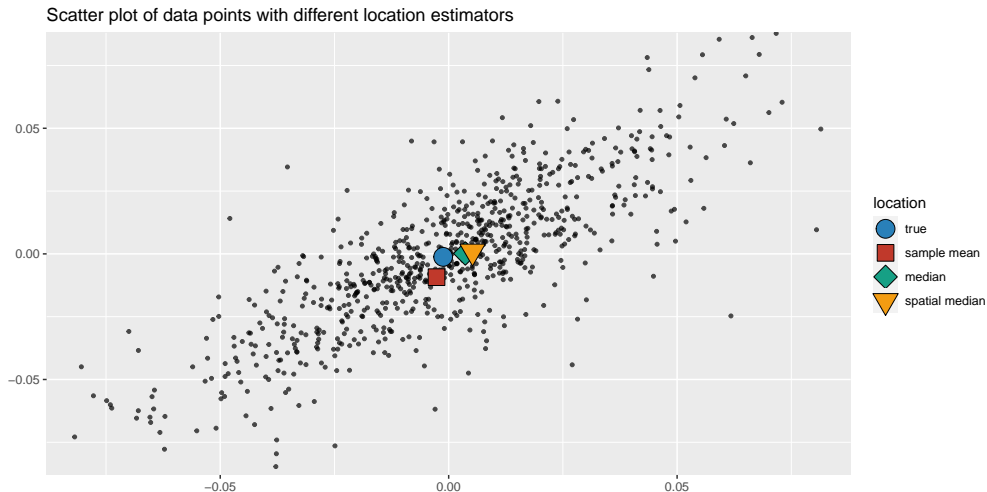
Least squares (LS) estimator

- **Least squares (LS) estimator:**
    - Originates from Gauss's work in 1795 on planetary motions.
    - Involves minimizing the squared difference between observed and predicted values.
    - Formulated as:

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

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

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

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

    - 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 sample mean:**
  - Lacks robustness against outliers and heavy-tailed distributions.
  - Vulnerability to contaminated points affects reliability of $\hat{\mu}$.

- **Robust estimation needs:**
  - Due to the limitations of LS and sample mean, alternative robust estimators have been explored.
  - Aim to improve reliability and accuracy in the presence of outliers and non-normal distributions.
  - Study of robust multivariate location estimators dates back to the 1960s.

Median estimator

- **Median as a robust estimator:**
  - Median separates the higher half from the lower half of a sample.
  - Considered the "middle" value, providing a typical representation of the data.
  - Unaffected by extreme values, unlike the mean.
  - Provides robustness against outliers.
  - 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 i.i.d. model:**
  - Can be formulated as the optimization problem:

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

  - 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.
  - Formulated as the optimization problem:

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

  - Uses the ℓ_2-norm (Euclidean distance) as the measure of error.
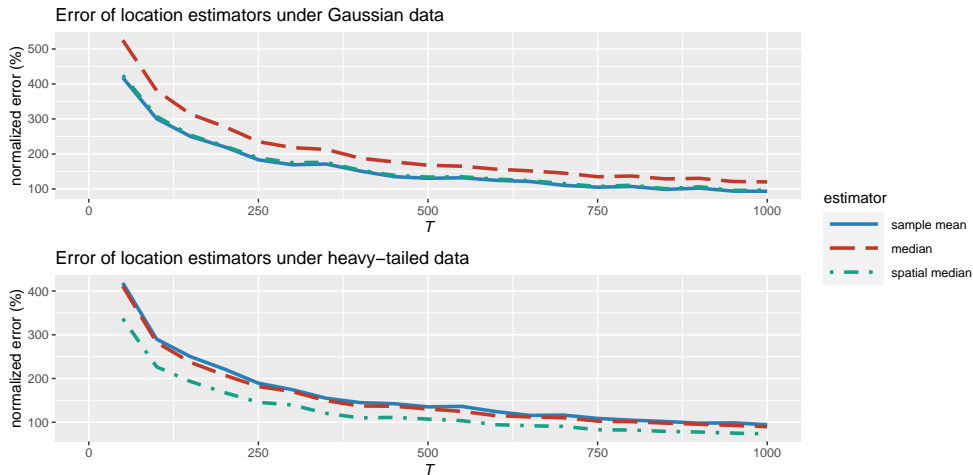
- **Characteristics of the spatial median:**
  - The estimator for each element is not independent of other elements.
  - For $N = 1$, it coincides with the univariate median.

- **Solving for the spatial median:**
  - The problem is a convex second-order cone problem (SOCP).
  - Can be solved using SOCP solvers or iterative algorithms.
  - Efficient iterative algorithms use the majorization-minimization (MM) method (Sun, Babu, and Palomar 2017).

Numerical experiments

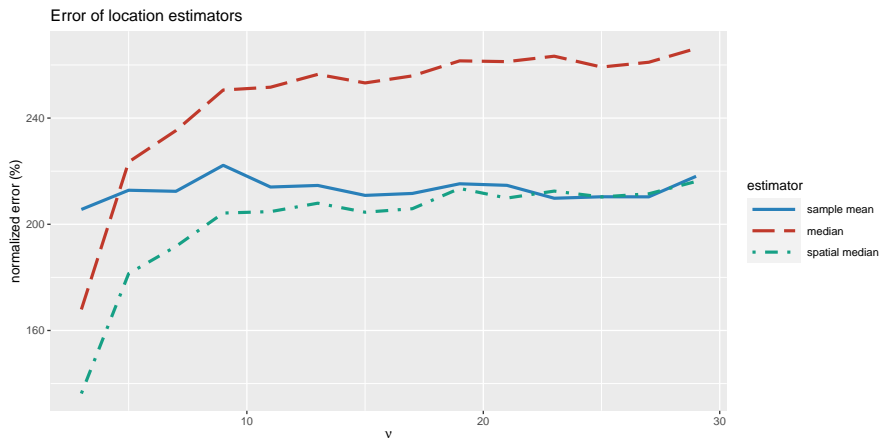Estimation error of location estimators versus number of observations (with $N = 100$):

Error of location estimators under Gaussian data

Error of location estimators under heavy−tailed data

estimator

— sample mean
– – median
–·– spatial median

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

Error of location estimators

Outline

Preliminaries on ML estimation

- **Maximum likelihood estimation (MLE):**
  - A fundamental technique in estimation theory.
  - Involves selecting the parameter vector $\boldsymbol{\theta}$ that maximizes the likelihood of observing a given set of data.

- **Concept of MLE:**
  - Based on the probability distribution function (pdf) f of 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:**
  - Formulated as:

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

  - Equivalently, maximizing the log-likelihood:

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

- **Theoretical properties of MLE:**
  - Asymptotically unbiased: The MLE's bias diminishes as $T \to \infty$.
  - Asymptotically efficient: 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.

Gaussian ML estimation

- **PDF for i.i.d. model with Gaussian residuals:**
  - Assuming residuals follow a multivariate normal distribution:

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

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

- **MLE formulation for Gaussian i.i.d. model:**

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

Gaussian ML estimation

- **Deriving MLE for μ and Σ:**
    - Set gradient with respect to μ and Σ to zero.
    - Results in the estimators:

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

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

- **Comparison with sample estimators:**
    - MLE coincides with sample mean and covariance estimators, except for the factor $1/T$ instead of $1/(T-1)$.
    - MLE of 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, optimal ML estimators will be different!
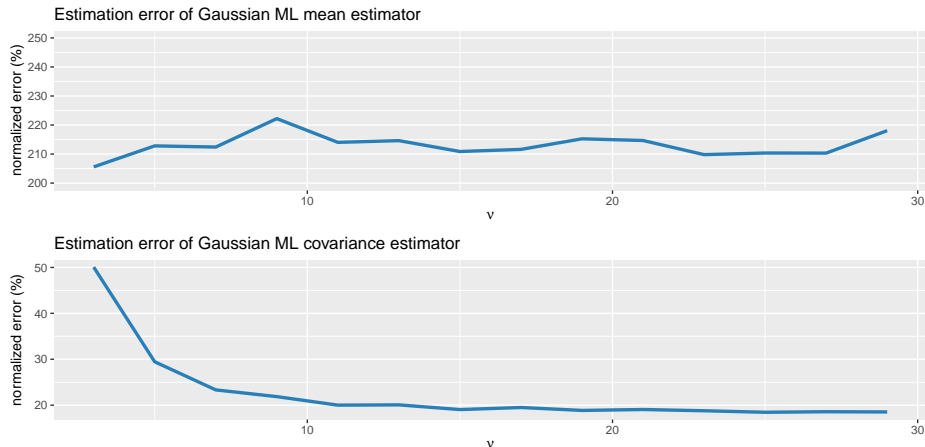
Estimation error of Gaussian ML estimators versus number of observations (with $N = 100$):
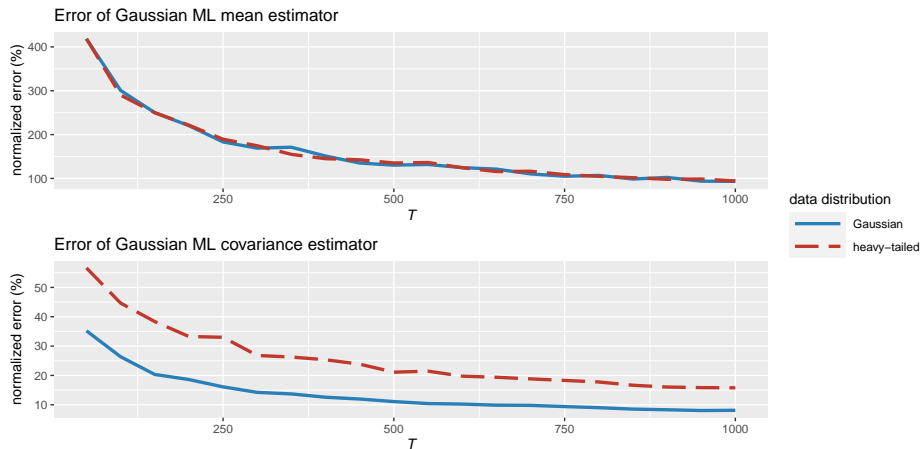
Estimation error of Gaussian ML estimators versus degrees of freedom in a t distribution (with $T = 200$ and $N = 100$):

Outline

Heavy-tailed ML estimators

- **Gaussian ML estimators and heavy-tailed distributions:**
  - Optimal for Gaussian-distributed data, but financial data often exhibit heavy tails.
  - Need to assess the impact of heavy-tailed distributions on these estimators.

- **Properties of Gaussian ML estimators:**
  - Coincide with sample estimators for Gaussian data.
  - Unbiased and consistent, which are beneficial characteristics.

- **Evaluating estimator performance with heavy tails:**
  - Despite being unbiased and consistent, may not be the best choice for heavy-tailed data.
  - Important to consider the efficiency and robustness of estimators under such conditions.
  - 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.
  - 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 $N = 2$ (ratio $T/N = 5$), the sample covariance matrix is sensitive to these effects.
- **Consequences for Gaussian ML estimators:**
  - Perform well for Gaussian data but poorly with outliers or heavy-tailed data.
  - Nevertheless, sample covariance matrix is widely used by practitioners due to its simplicity!
- **Challenges in practice:**
  - Prevalence of heavy-tailed distributions in financial data complicates accurate 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:
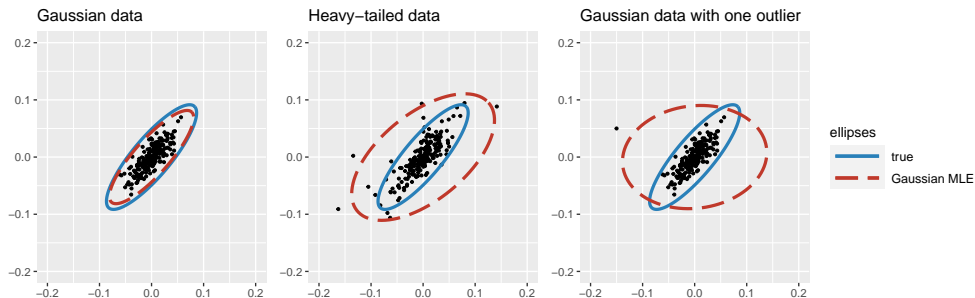
Heavy-tailed ML estimation

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

- **PDF for multivariate t distribution:**

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

  - Parameters: $\boldsymbol{\mu}$ (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 μ and Σ:**
  - Set gradient with respect to μ and Σ to zero.
  - Results in the fixed-point equations for μ and Σ:

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

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

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

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

- **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

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

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

2. $k \leftarrow k + 1$

Until: convergence

Numerical experiments

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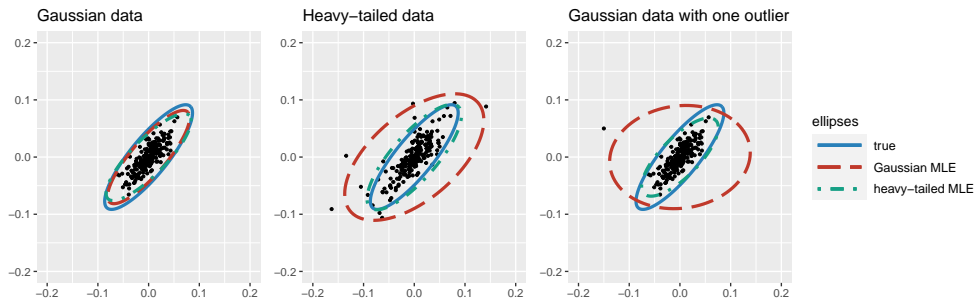
Robust estimators

- **Robust estimators overview:**
  - Estimators resilient to outliers and deviations from the assumed distribution.
  - 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, 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:**
  - Median offers greater robustness with a breakdown point of ~ 0.5.
  - Heavy-tailed ML estimators effectively handle deviations, enhancing robustness.

Robust estimators: M-estimators[*]

- **Introduction to M-estimators:**
    - Term dates back to the 1960s, introduced by (Huber 1964).
    - Generalization of maximum likelihood estimators.
    - Defined by fixed-point equations for robust estimation of location and scatter parameters.
- **Fixed-point equations for M-estimators:**
    - Location parameter $\boldsymbol{\mu}$ and scatter parameter $\boldsymbol{\Sigma}$ are estimated by:

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

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

(1)

    - Weight functions $u_1(\cdot)$ and $u_2(\cdot)$ must satisfy certain conditions.
- **Properties and robustness of M-estimators:** (Maronna 1976)
    - Weighted sample mean and covariance matrix.
    - Bounded influence function for robustness.
    - 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:

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

Robust estimators: Tyler's estimator

- **Tyler's estimator for scatter matrix:**
  - Introduced by Tyler in 1987 for heavy-tailed distributions (Tyler 1987).
  - Most robust version of an *M*-estimator.

- **Elliptical distribution assumption:**
  - Assumes zero-mean elliptical distribution for *x*.
  - If mean is not zero, it must be estimated and subtracted from observations.

- **Normalization and ML estimation:**
  - Observations normalized:

$$\boldsymbol{s}_t = \frac{\boldsymbol{x}_t}{\|\boldsymbol{x}_t\|_2}$$

  - ML based on normalized points, with pdf (angular distribution):

$$f(\boldsymbol{s}) \propto \frac{1}{\sqrt{|\boldsymbol{\Sigma}|}} \left(\boldsymbol{s}^\mathsf{T} \boldsymbol{\Sigma}^{-1} \boldsymbol{s}\right)^{-N/2}$$

Robust estimators: Tyler's estimator

- **MLE formulation:**
  - Given T observations, MLE is formulated as

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

  - Leads to fixed-point equation:

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

  with weights

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

- **Robustness and weights:**
  - Weights down-weight outliers, enhancing robustness.
  - 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.
  - Computational cost of robust estimators 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.
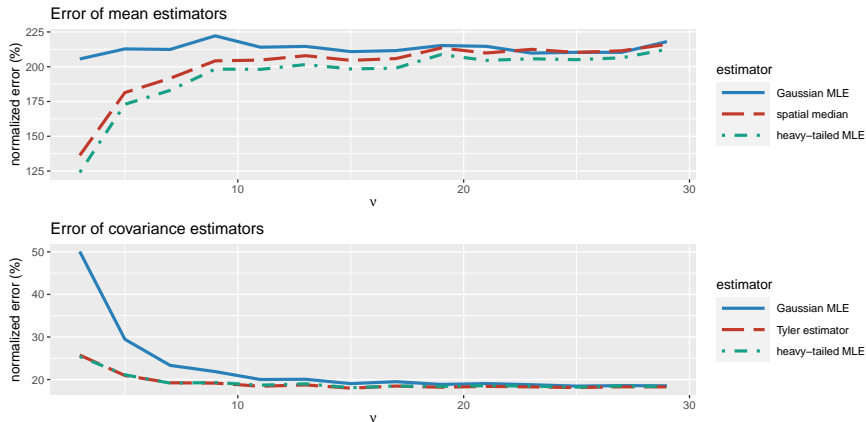
Estimation error of different ML estimators versus number of observations (for t-distributed heavy-tailed data with $\nu = 4$ and $N = 100$):
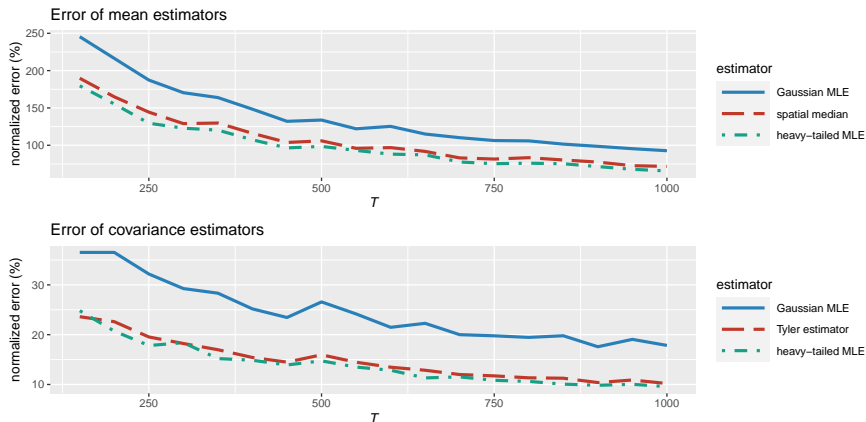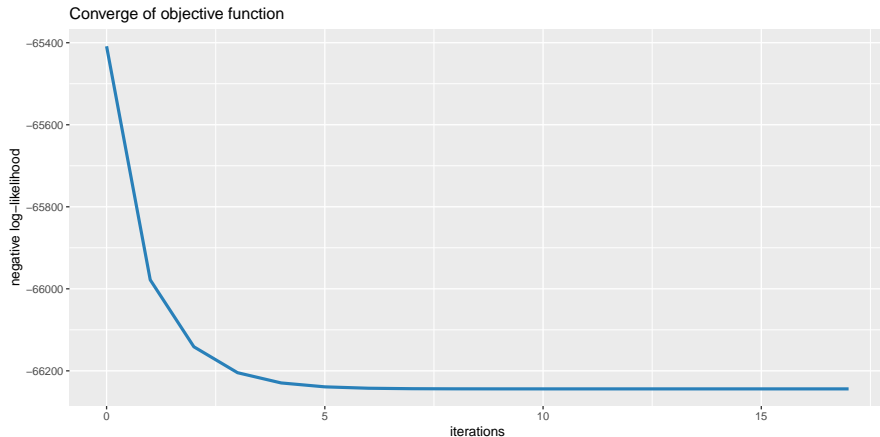
Estimation error of different ML estimators versus degrees of freedom in a t distribution (with $T = 200$ and $N = 100$):

Numerical experiments

Convergence of robust heavy-tailed ML estimator:

Converge of objective function

Outline

Prior information

- **Challenges with historical data:**
  - 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.
    - Aims to improve estimation by pulling estimates towards a target.
  - **Factor modeling:**
    - Utilizes structural information about the data.
    - Helps in reducing dimensionality and improving parameter estimation.
  - **Black-Litterman approach:**
    - Merges historical data with subjective views.
    - Balances empirical data with investor-specific insights.

Outline

Shrinkage

- **Introduction to shrinkage:**
  - Technique to reduce estimation error by introducing bias.
  - Originated with Stein in 1955, popular in finance for covariance matrix shrinkage since early 2000s.

- **Bias-variance trade-off:**
  - Mean squared error (MSE) of an estimator is the sum of its variance and squared bias.
  - Small sample sizes lead to high variance; larger samples may see bias dominate 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:**

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

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

Shrinkage

- **Implementation considerations:**
  - **Choice of target (θ^{tgt}):** Represents prior information or market views.
  - **Choice of shrinkage factor (ρ):** Critical for balancing the weight of the target in the estimator.

- **Choosing the shrinkage factor:**
  - **Empirical choice:** Based on cross-validation.
  - **Analytical choice:** Utilizes advanced mathematical techniques (e.g., random matrix theory).

- **Application in finance:**
  - Parameter θ could be mean vector μ or covariance matrix Σ.
  - Estimators like sample mean or covariance matrix can be adjusted using shrinkage for better accuracy.

Shrinking the mean vector

- **Sample mean and its properties:**
  - Sample mean $\hat{\boldsymbol{\mu}}$ is an unbiased estimator of the mean vector $\boldsymbol{\mu}$.
  - Distribution characterized by $\hat{\boldsymbol{\mu}} \sim \mathcal{N}\left(\boldsymbol{\mu}, \frac{1}{T}\boldsymbol{\Sigma}\right)$.
  - Mean squared error (MSE): $\mathbb{E}\left[\|\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\|^2\right] = \frac{1}{T}\mathsf{Tr}(\boldsymbol{\Sigma})$.

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

- **James-Stein estimator:**
  - Formula:
  $$\hat{\boldsymbol{\mu}}^{\mathsf{JS}} = (1 - \rho)\,\hat{\boldsymbol{\mu}} + \rho\,\boldsymbol{\mu}^{\mathsf{tgt}}$$
  - Improves MSE over the sample mean for any target $\boldsymbol{\mu}^{\mathsf{tgt}}$ with a properly chosen ρ:
  $$\rho = \frac{(N+2)}{(N+2) + T \times (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}^{\mathsf{tgt}})^{\mathsf{T}}\,\boldsymbol{\Sigma}^{-1}\,(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}^{\mathsf{tgt}})}$$

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

Shrinking the covariance matrix

- **Shrinkage in covariance matrix estimation:**
  - Introduces bias to reduce estimation error.
  - Shrinkage estimator formula:

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

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

- **Historical context:**
  - Concept used in the 1980s in wireless communications as "diagonal loading".
  - Gained popularity in finance in the early 2000s through work by 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:** Off-diagonal elements equal to the average cross-correlation.

Shrinking the covariance matrix

- **Determining the shrinkage factor ρ:**
  - Can be empirical (via cross-validation) or analytical (using random matrix theory, RMT).
  - Ledoit and Wolf popularized RMT approach to minimize error measure $\mathbb{E}\left[\left\|\hat{\boldsymbol{\Sigma}}^{\text{sh}} - \boldsymbol{\Sigma}\right\|_{\text{F}}^2\right]$.
  - Asymptotic RMT formula for ρ (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 inverse covariance matrix: $\mathbb{E}\left[\left\|\left(\hat{\boldsymbol{\Sigma}}^{\text{sh}}\right)^{-1} - \boldsymbol{\Sigma}^{-1}\right\|_{\text{F}}^2\right]$.
  - Maximizing Sharpe ratio: Choose ρ to maximize achieved Sharpe ratio.

- **Nonlinear shrinkage:**
  - Extends the idea to eigenvalues of the covariance matrix.
  - Requires increased mathematical sophistication.

- **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 based on MSE of estimators.
  - MSE may not be the optimal error measure in portfolio optimization contexts.
  - Importance of using more appropriate error measures highlighted.

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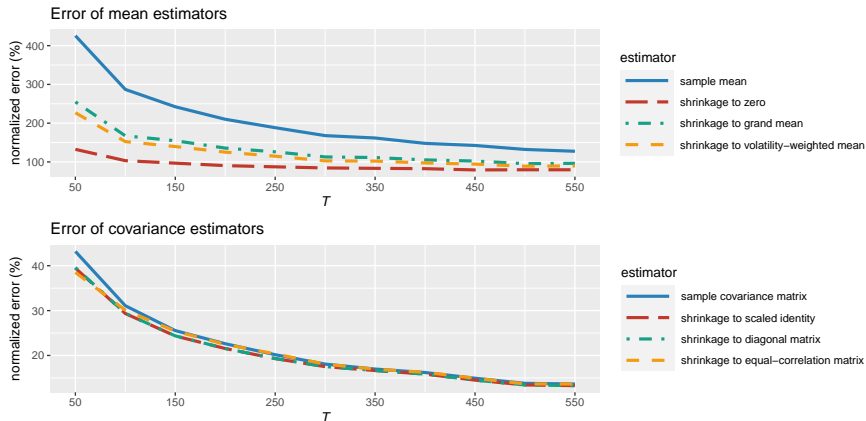
Outline

Factor models

- **Factor modeling in finance:**
  - Incorporates prior information into asset return models.
  - Found in numerous finance textbooks.

- **Single-factor model:**
  - Simplest form of factor modeling.
  - Equation:

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

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

- **Connection to CAPM:**
  - Single-factor model 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:**
  - Generalization of single-factor model.
  - Equation:

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

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

- **Idiosyncratic component:**
  - Residual term $\boldsymbol{\epsilon}_t$ assumed to have a diagonal covariance matrix $\boldsymbol{\Psi}$.
  - Captures asset-specific noise not explained by factors.

Factor models

- **Mean and covariance matrix estimation:**
  - Mean vector:
  $$\boldsymbol{\mu} = \boldsymbol{\alpha} + \boldsymbol{B}\boldsymbol{\mu}_f$$
  - Covariance matrix:
  $$\boldsymbol{\Sigma} = \boldsymbol{B}\boldsymbol{\Sigma}_f\boldsymbol{B}^\mathsf{T} + \boldsymbol{\Psi}$$
  - Decomposes covariance into low-rank and full-rank diagonal components.
  - Single-factor model: $\boldsymbol{\Sigma} = \sigma_f^2\beta\beta^\mathsf{T} + \boldsymbol{\Psi}$

- **Parameter reduction:**
  - Factor models reduce the number of parameters needed for estimation.
  - Example: For $N = 500$ assets and $K = 3$ factors, parameters reduce from 125,750 to 1,503.

- **Types of factor models:**
  1. **Macroeconomic factor models:** Observable economic factors, unknown loadings.
  2. **Fundamental factor models:** Loadings from asset characteristics, unknown factors.
  3. **Statistical factor models:** Both factors and loadings are unknown.

Macroeconomic factor models

- **Macroeconomic factor models overview:**
  - Allows for the integration of economic indicators into the analysis of asset returns.
  - Utilize 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:

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

  - 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 f_t.
    - Loadings B estimated similarly to macroeconomic models.
    - Original model had $K = 3$ factors: firm size, book-to-market values, excess market return (Fama and French 1992).
    - Extended to $K = 5$ factors including profitability and investment patterns (Fama and French 2015).
  - **Barra risk factor analysis approach:**
    - Loadings B constructed from asset characteristics.
    - Factors f_t estimated via regression, opposite of macroeconomic models.
    - Developed by Barra Inc. in 1975.

Statistical factor models

- **Statistical factor models overview:**
  - Both factors $\boldsymbol{f}_t$ and loading matrix $\boldsymbol{B}$ are unknown.
  - Introduce structure to the covariance matrix $\boldsymbol{\Sigma}$ as a low-rank plus diagonal matrix.

- **Covariance matrix structure:**
  - $\boldsymbol{\Sigma}$ decomposed into $\boldsymbol{B}\boldsymbol{\Sigma}_f\boldsymbol{B}^{\mathsf{T}}$ (low-rank) and $\boldsymbol{\Psi}$ (diagonal).
  - Factors assumed to be zero-mean and normalized for simplification.

- **Heuristic formulation:**
  - Approximate sample covariance matrix $\hat{\boldsymbol{\Sigma}}$ with desired structure:

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

Statistical factor models

- **ML estimation under the Gaussian assumption:**
  - Formulate ML estimation imposing covariance structure:

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

  - Nonconvex nature makes the problem 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:**
  - 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:**
  - Maximizes variance along direction $\boldsymbol{u}$: $\mathrm{Var}(\boldsymbol{u}^\mathsf{T}\boldsymbol{x}) = \boldsymbol{u}^\mathsf{T}\boldsymbol{\Sigma}\boldsymbol{u}$.
  - Solution found through eigenvalue decomposition: $\boldsymbol{\Sigma} \approx \boldsymbol{U}^{(K)}\boldsymbol{D}^{(K)}\boldsymbol{U}^{(K)\mathsf{T}}$.
  - $\boldsymbol{U}^{(K)}$ contains the first K eigenvectors; $\boldsymbol{D}^{(K)}$ has the largest K eigenvalues.

Principal component analysis (PCA)*

- **PCA in statistical factor models:**
  - Approximates the solution to the statistical factor model by performing PCA on the sample covariance matrix.
  - Heuristic: it keeps K principal components and uses a scaled identity matrix for the diagonal component $\mathbf{\Psi}$.
  - Approximate solution to factor model:

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

  $$\mathbf{\Psi} = \kappa\mathbf{I},$$

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

- **PCA estimator for covariance matrix:**
  - PCA estimator:

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

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

Numerical experiments

- **Evaluation of covariance matrix estimation under factor model:**
  - Estimation accuracy depends on how well 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:**
  - Estimation error comparison between factor model-based and sample covariance matrices shown for synthetic Gaussian data.
  - Data complies with a factor model structure having $K = 3$ principal components.
  - Accurate estimation with 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 $N = 100$):
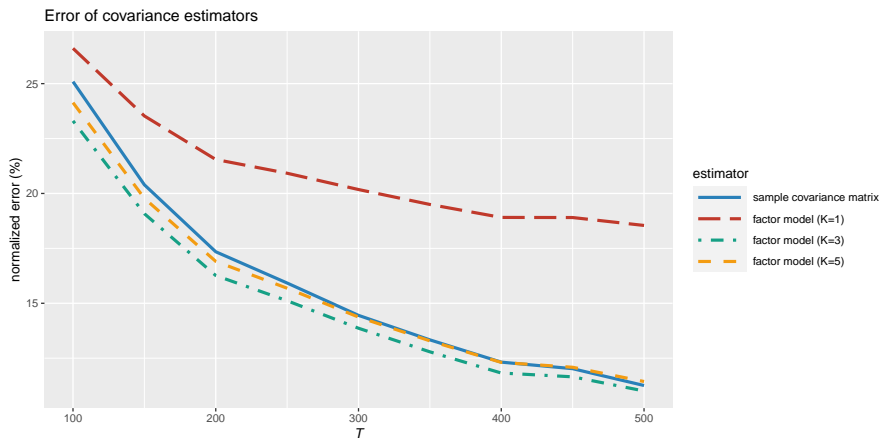
Error of covariance estimators

estimator
— sample covariance matrix
- - factor model (K=1)
-·- factor model (K=3)
- - factor model (K=5)

Outline

Black-Litterman model

- **Black-Litterman model basics:**
  - Combines historical data with investor's prior information.
  - Standard in finance, detailed in textbooks like (Fabozzi, Focardi, and Kolm 2010) and (Meucci 2005).

- **Components of Black-Litterman model:**
  - **Market equilibrium:** Estimate of μ from the market, denoted as $\pi = \hat{\mu}$:

$$\pi = \mu + \epsilon$$

  with ϵ zero-mean and covariance $\tau\Sigma$ (τ often set as $\tau = 1/T$).
  - **Investor's views:** K views on asset returns, expressed as

$$v = P\mu + e$$

    - v and P represent the views and their relation to asset returns.
    - Error term e is zero-mean with covariance Ω.

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 a set of $N = 5$ 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 standard deviation of 1%.

- **Mathematical representation:**
  - Views expressed as:
  $$\begin{bmatrix} 1.5\% \\ 4\% \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \end{bmatrix} \boldsymbol{\mu} + \boldsymbol{e}$$
  - Covariance of error $\boldsymbol{e}$:
  $$\boldsymbol{\Omega} = \begin{bmatrix} 1\%^2 & 0 \\ 0 & 1\%^2 \end{bmatrix}.$$

Merging the market equilibrium with the views

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

- **Weighted least squares formulation:**
  - Compact representation:

  $$y = X\mu + n,$$

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

  - Problem formulated as:

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

  - Solution:

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

Merging the market equilibrium with the views[*]

- **Original Bayesian formulation of Black-Litterman model:**
  - Returns x assumed to follow a normal distribution: $x \sim \mathcal{N}(\mu, \Sigma)$.
  - Mean μ also modeled as random with Gaussian distribution: $\mu \sim \mathcal{N}(\pi, \tau\Sigma)$.
    - π is the best guess for μ; $\tau\Sigma$ represents uncertainty.
  - Views modeled with Gaussian distribution: $P\mu \sim \mathcal{N}(v, \Omega)$.
  - Posterior distribution of μ given views: $\mu \mid v, \Omega \sim \mathcal{N}(\mu_{\mathrm{BL}}, \Sigma_{\mathrm{BL}})$.
  - Posterior mean μ_{BL} matches the weighted least squares solution.
  - Posterior covariance Σ_{BL} includes the original covariance Σ and the covariance of the posterior mean.
  - Mean estimator:

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

  - Covariance matrix estimator:

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

- **Alternative Bayesian formulation:**
  - Variation introduced by (Meucci 2005) models views on random returns: $\boldsymbol{v} = \boldsymbol{Px} + \boldsymbol{e}$, differing 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 Black-Litterman estimator:**
  - $\tau = 0$: Market equilibrium is considered completely accurate, leading to $\boldsymbol{\mu}_{\text{BL}} = \boldsymbol{\pi}$.
  - $\tau \to \infty$: Market equilibrium is disregarded, and 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 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.

———. 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.

References II

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.

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.

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.