Portfolio Optimization

High-Order Portfolios

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

portfoliooptimizationbook.com

Latest update: March 14, 2025

Outline

Abstract

Markowitz's mean-variance portfolio theory, which balances expected return against risk measured by variance, faces challenges when applied to financial data due to its non-Gaussian distribution characterized by asymmetry and heavy tails. Incorporating higher-order moments such as skewness and kurtosis into portfolio design could address this issue but introduces significant difficulties. The computation, storage, and manipulation of these moments become exponentially more complex as they grow at a rate of N^4 with the number of assets. Additionally, the resulting portfolio formulations are nonconvex, complicating optimization further. Despite these challenges, advancements in computational power and techniques have only recently made high-order portfolios feasible for managing large asset numbers, reaching into the hundreds or thousands (Palomar 2025, chap. 9).

Outline

Introduction

- Markowitz's mean-variance portfolio optimization:
  - Balances expected return and risk (variance).
  - Optimization problem:

$$\begin{array}{ll}\underset{\boldsymbol{w}}{\text{maximize}} & \boldsymbol{w}^\mathsf{T}\boldsymbol{\mu} - \frac{\lambda}{2}\boldsymbol{w}^\mathsf{T}\boldsymbol{\Sigma}\boldsymbol{w}\\ \text{subject to} & \boldsymbol{w} \in \mathcal{W}\end{array}$$

  where
    - λ: risk-aversion hyper-parameter
    - $\mathcal{W}$: constraint set, e.g., $\mathcal{W} = \{\boldsymbol{w} \mid \mathbf{1}^\mathsf{T}\boldsymbol{w} = 1, \boldsymbol{w} \geq \mathbf{0}\}$.

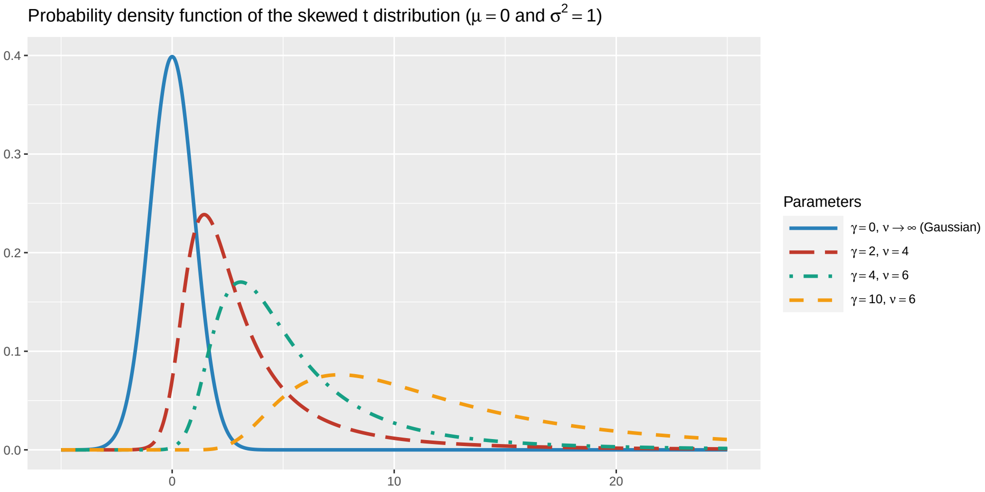
- **Beyond the Gaussian assumption:**
  - Empirical studies show financial data are not Gaussian.
  - Portfolio optimization should include higher-order moments:
    - skewness (third moment)
    - kurtosis (fourth moment)
  - Aim for higher skewness and lower kurtosis in portfolios.

- **Skewed t distribution:**
  - Next figure shows skewed t distribution with varying skewness (γ) and kurtosis (ν).
  - $\gamma = 0$: symmetric case.
  - $\nu \to \infty$: non-heavy-tailed case.

- **Incorporating higher moments in portfolio optimization:**
  - Investors may prefer to trade-off lower expected return/higher volatility for higher skewness/lower kurtosis.
  - Measures can be skew-adjusted; example of Sharpe ratio:

$$\text{skew-adjusted-SR} = \text{SR} \times \sqrt{1 + \frac{\text{skewness}}{3}\text{SR}}.$$

Illustration of skewness and kurtosis with the skewed t distribution:



Introduction

- **High-order portfolios overview:**
  - Third and fourth moments of a portfolio:
    - Third moment (skewness): $\boldsymbol{w}^\mathsf{T}\boldsymbol{\Phi}(\boldsymbol{w}\otimes\boldsymbol{w})$.
    - Fourth moment (kurtosis): $\boldsymbol{w}^\mathsf{T}\boldsymbol{\Psi}(\boldsymbol{w}\otimes\boldsymbol{w}\otimes\boldsymbol{w})$.
  - Co-skewness matrix: $\boldsymbol{\Phi}\in\mathbb{R}^{N\times N^2}$.
  - Co-kurtosis matrix: $\boldsymbol{\Psi}\in\mathbb{R}^{N\times N^3}$.
  - Challenges:
    - Computation, storage, and manipulation of high-order moments are difficult.
    - Nonconvex nature of the third moment complicates portfolio formulations.
- **Historical perspective:** (Palomar 2025, chap. 9)
  - Early attempts in the 1960s to incorporate high-order moments (Young and Trent 1969; Jean 1971).
  - Dimensionality and computational issues hindered progress.
  - Skepticism due to the impracticality of modeling high-order cross-moments (Brandt, Santa-Clara, and Valkanov 2009).
  - Recent advancements in estimation methods and computational techniques (Boudt, Lu, and Peeters 2014; Zhou and Palomar 2021; Wang et al. 2023).

Introduction

- **Estimation and computational challenges:**
  - High-order portfolio design can negatively impact out-of-sample performance without improved estimators.
  - Improved estimation methods introduce structure and shrinkage.
  - Parametric multivariate distributions reduce parameter estimation complexity.
- **Algorithmic developments:**
  - Nonconvex problems can be addressed with meta-heuristic optimization, but have high computational cost.
  - Local optimization methods offer practical solutions with acceptable costs.
  - Difference-of-convex (DC) programming and DC-SOS decomposition techniques.
  - Successive convex approximation (SCA) framework (Scutari et al. 2014) accelerates convergence for high-dimensional problems.
- **Advancements in high-order portfolio optimization:**
  - Significant reduction in computational cost through parametric models.
  - Development of faster numerical methods for large-scale portfolio optimization.
  - High-order portfolios now feasible for practical application.

Outline

High-order moments

- **Understanding high-order moments:**
  - High-order moments are crucial for non-Gaussian distributions.
  - First four moments of a random variable X:
    - Mean (first moment): $\bar{X} \triangleq \mathbb{E}[X]$
    - Variance (second moment): $\mathbb{E}[(X - \bar{X})^2]$
    - Skewness (third moment): $\mathbb{E}[(X - \bar{X})^3]$
    - Kurtosis (fourth moment): $\mathbb{E}[(X - \bar{X})^4]$.

- **Interpretation:**
  - Mean: indicates location.
  - Variance: measures spread.
  - Skewness: assesses asymmetry.
  - Kurtosis: characterizes the tail thickness.

High-order moments: Non-parametric case

- **Portfolio moments:**
  - Portfolio return with N assets: $\boldsymbol{w}^\mathsf{T}\boldsymbol{r}$.
  - First four moments of portfolio return:
    - Mean: $\phi_1(\boldsymbol{w}) = \boldsymbol{w}^\mathsf{T}\boldsymbol{\mu}$
    - Variance: $\phi_2(\boldsymbol{w}) = \boldsymbol{w}^\mathsf{T}\boldsymbol{\Sigma}\boldsymbol{w}$
    - Skewness: $\phi_3(\boldsymbol{w}) = \boldsymbol{w}^\mathsf{T}\boldsymbol{\Phi}(\boldsymbol{w}\otimes\boldsymbol{w})$
    - Kurtosis: $\phi_4(\boldsymbol{w}) = \boldsymbol{w}^\mathsf{T}\boldsymbol{\Psi}(\boldsymbol{w}\otimes\boldsymbol{w}\otimes\boldsymbol{w})$.
  - Parameters:
    - $\boldsymbol{\mu}$: mean vector
    - $\boldsymbol{\Sigma}$: covariance matrix
    - $\boldsymbol{\Phi}$: co-skewness matrix
    - $\boldsymbol{\Psi}$: co-kurtosis matrix.

High-order moments: Non-parametric case

- **Computational aspects:**
  - We will optimize some combination of these four moments of the portfolio.
  - This means that we will need to compute the gradient and Hessian of these moments.

- **Gradients and hessians:**
  - Gradients:
    - $\nabla\phi_1(\boldsymbol{w}) = \boldsymbol{\mu}$.
    - $\nabla\phi_2(\boldsymbol{w}) = 2\boldsymbol{\Sigma}\boldsymbol{w}$.
    - $\nabla\phi_3(\boldsymbol{w}) = 3\boldsymbol{\Phi}(\boldsymbol{w}\otimes\boldsymbol{w})$.
    - $\nabla\phi_4(\boldsymbol{w}) = 4\boldsymbol{\Psi}(\boldsymbol{w}\otimes\boldsymbol{w}\otimes\boldsymbol{w})$.
  - Hessians:
    - $\nabla^2\phi_1(\boldsymbol{w}) = \boldsymbol{0}$.
    - $\nabla^2\phi_2(\boldsymbol{w}) = 2\boldsymbol{\Sigma}$.
    - $\nabla^2\phi_3(\boldsymbol{w}) = 6\boldsymbol{\Phi}(\boldsymbol{I}\otimes\boldsymbol{w})$.
    - $\nabla^2\phi_4(\boldsymbol{w}) = 12\boldsymbol{\Psi}(\boldsymbol{I}\otimes\boldsymbol{w}\otimes\boldsymbol{w})$.

High-order moments: Non-parametric case

- **Complexity analysis:**
  - Complexity of parameters and their computation grows with N:
    - $\boldsymbol{\mu}$: $O(N)$
    - $\boldsymbol{\Sigma}$: $O(N^2)$
    - $\boldsymbol{\Phi}$: $O(N^3)$
    - $\boldsymbol{\Psi}$: $O(N^4)$.
  - Gradients and Hessians complexity:
    - $\nabla\phi_1(\boldsymbol{w})$ and $\nabla^2\phi_1(\boldsymbol{w})$: $O(1)$ and $O(1)$
    - $\nabla\phi_2(\boldsymbol{w})$ and $\nabla^2\phi_2(\boldsymbol{w})$: $O(N^2)$ and $O(1)$
    - $\nabla\phi_3(\boldsymbol{w})$ and $\nabla^2\phi_3(\boldsymbol{w})$: $O(N^3)$ and $O(N^3)$
    - $\nabla\phi_4(\boldsymbol{w})$ and $\nabla^2\phi_4(\boldsymbol{w})$: $O(N^4)$ and $O(N^4)$.
  - Memory requirement example: Storing $\boldsymbol{\Psi}$ for $N = 200$ needs $\sim$12 GB.

- **Implications for practical application:**
  - Markowitz's portfolio complexity: $O(N^2)$.
  - Incorporating third and fourth moments increases complexity to $O(N^4)$.
  - High complexity limits practical application to portfolios with a small number of assets.

High-order moments: Structured moments

- **Structured moments in portfolio analysis:**
  - Introducing structure to high-order moment matrices can reduce parameter estimation.
  - Factor modeling introduces structure at the expense of higher complexity in the estimation procedure due to intricate matrix structures.

- **Single market-factor model:**
  - Returns modeled as:

$$\boldsymbol{r}_t = \boldsymbol{\alpha} + \boldsymbol{\beta} r_t^{\mathsf{mkt}} + \boldsymbol{\epsilon}_t.$$

  - Moments expressed as:
    - Mean vector: $\boldsymbol{\mu} = \boldsymbol{\alpha} + \boldsymbol{\beta}\phi_1^{\mathsf{mkt}}$
    - Covariance matrix: $\boldsymbol{\Sigma} = \boldsymbol{\beta}\boldsymbol{\beta}^\mathsf{T}\phi_2^{\mathsf{mkt}} + \boldsymbol{\Sigma}_\epsilon$
    - Co-skewness matrix: $\boldsymbol{\Phi} = \boldsymbol{\beta}\left(\boldsymbol{\beta}^\mathsf{T} \otimes \boldsymbol{\beta}^\mathsf{T}\right)\phi_3^{\mathsf{mkt}} + \boldsymbol{\Phi}_\epsilon$
    - Co-kurtosis matrix: $\boldsymbol{\Psi} = \boldsymbol{\beta}\left(\boldsymbol{\beta}^\mathsf{T} \otimes \boldsymbol{\beta}^\mathsf{T} \otimes \boldsymbol{\beta}^\mathsf{T}\right)\phi_4^{\mathsf{mkt}} + \boldsymbol{\Psi}_\epsilon$
    - ϕ_i^{mkt}: ith moment of the market factor
    - $\boldsymbol{\Sigma}_\epsilon$, $\boldsymbol{\Phi}_\epsilon$, $\boldsymbol{\Psi}_\epsilon$: covariance, co-skewness, and co-kurtosis matrices of residuals.

High-order moments: Parametric case

- **Overview of multivariate distributions:**
  - Multivariate normal distribution characterized by mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$.
  - Financial data often exhibit non-Gaussian features like skewness and kurtosis.

- **Multivariate normal mixture distributions:**
  - Introduce randomness into covariance and mean for more general distributions.
  - Variance mixtures affect covariance but not mean.
  - Mean-variance mixtures affect both mean and covariance.

- **Normal variance mixture:**
  - Example: Multivariate t distribution with inverse gamma distribution for w.
  - Models heavy tails but not asymmetry.
  - Hierarchical structure:

$$\boldsymbol{x} \mid \tau \sim \mathcal{N}\left(\boldsymbol{\mu}, \frac{1}{\tau}\boldsymbol{\Sigma}\right),$$

$$\tau \sim \text{Gamma}\left(\frac{\nu}{2}, \frac{\nu}{2}\right).$$

High-order moments: Parametric case

- **Normal mean-variance mixture:**
  - Example: Multivariate generalized hyperbolic (GH) distribution.
  - Models both heavy tails and asymmetry.
  - Hierarchical structure for skewed t distribution:

$$\boldsymbol{x} \mid \tau \sim \mathcal{N}\left(\boldsymbol{\mu} + \frac{1}{\tau}\boldsymbol{\gamma}, \frac{1}{\tau}\boldsymbol{\Sigma}\right),$$

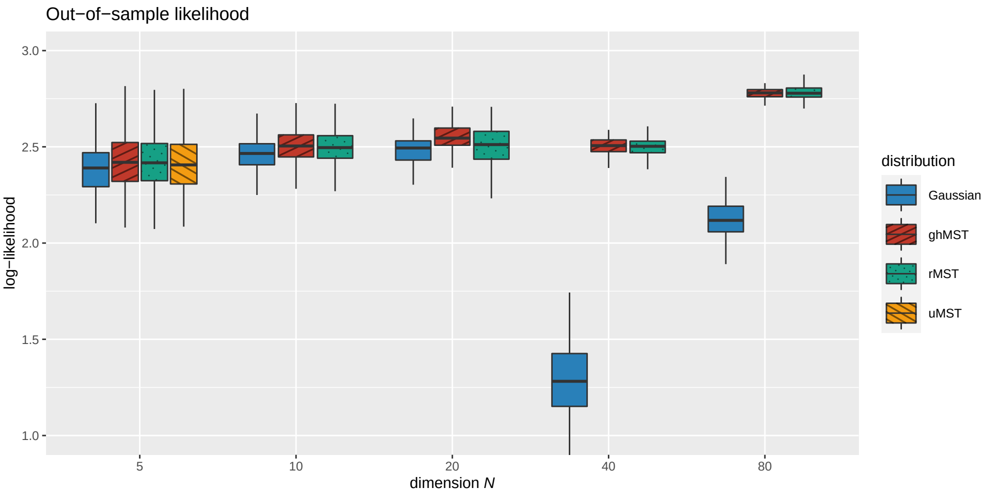
$$\tau \sim \mathsf{Gamma}\left(\frac{\nu}{2}, \frac{\nu}{2}\right).$$

- **Complexity in fitting distributions:**
  - More complex distributions like unrestricted multivariate skewed t (uMST) are hard to fit due to computational complexity.
  - Skewed t distribution offers a good balance between fitting financial data asymmetries and computational simplicity.

- **Goodness of fit:**
  - Empirical comparison shows skewed t distribution a good choice for financial data.
  - Maintains simplicity while fitting data well, unlike more complex distributions.

High-order moments: Parametric case

Likelihood of different fitted multivariate distributions for S&P 500 daily stock returns:



High-order moments: Parametric case

- **Simplified moments under multivariate skewed t distribution:**
  - The parametric model simplifies moment computations significantly.
  - First four moments (Wang et al. 2023):
    - $\phi_1(\boldsymbol{w}) = \boldsymbol{w}^\mathsf{T}\boldsymbol{\mu} + a_1\boldsymbol{w}^\mathsf{T}\gamma$
    - $\phi_2(\boldsymbol{w}) = a_{21}\boldsymbol{w}^\mathsf{T}\boldsymbol{\Sigma}\boldsymbol{w} + a_{22}(\boldsymbol{w}^\mathsf{T}\gamma)^2$
    - $\phi_3(\boldsymbol{w}) = a_{31}(\boldsymbol{w}^\mathsf{T}\gamma)^3 + a_{32}(\boldsymbol{w}^\mathsf{T}\gamma)\boldsymbol{w}^\mathsf{T}\boldsymbol{\Sigma}\boldsymbol{w}$
    - $\phi_4(\boldsymbol{w}) = a_{41}(\boldsymbol{w}^\mathsf{T}\gamma)^4 + a_{42}(\boldsymbol{w}^\mathsf{T}\gamma)^2\boldsymbol{w}^\mathsf{T}\boldsymbol{\Sigma}\boldsymbol{w} + a_{43}(\boldsymbol{w}^\mathsf{T}\boldsymbol{\Sigma}\boldsymbol{w})^2$
  - Coefficients a_1, a_{21}, a_{22}, etc., are functions of the degrees of freedom ν.

- **Clarification on parameters:**
  - $\boldsymbol{\mu}$: Location vector, not the mean.
  - $\boldsymbol{\Sigma}$: Scatter matrix, not the covariance matrix.

High-order moments: Parametric case

- **Gradients and Hessians:**
  - Gradients:
    - $\nabla\phi_1(\mathbf{w}) = \boldsymbol{\mu} + a_1\boldsymbol{\gamma}$
    - $\nabla\phi_2(\mathbf{w}) = 2a_{21}\boldsymbol{\Sigma}\mathbf{w} + 2a_{22}(\mathbf{w}^\mathsf{T}\boldsymbol{\gamma})\boldsymbol{\gamma}$
    - $\nabla\phi_3(\mathbf{w}) = 3a_{31}(\mathbf{w}^\mathsf{T}\boldsymbol{\gamma})^2\boldsymbol{\gamma} + a_{32}\left((\mathbf{w}^\mathsf{T}\boldsymbol{\Sigma}\mathbf{w})\boldsymbol{\gamma} + 2(\mathbf{w}^\mathsf{T}\boldsymbol{\gamma})\boldsymbol{\Sigma}\mathbf{w}\right)$
    - $\nabla\phi_4(\mathbf{w}) = 4a_{41}(\mathbf{w}^\mathsf{T}\boldsymbol{\gamma})^3\boldsymbol{\gamma} + 2a_{42}\left((\mathbf{w}^\mathsf{T}\boldsymbol{\gamma})^2\boldsymbol{\Sigma}\mathbf{w} + (\mathbf{w}^\mathsf{T}\boldsymbol{\Sigma}\mathbf{w})(\mathbf{w}^\mathsf{T}\boldsymbol{\gamma})\boldsymbol{\gamma}\right) + 4a_{43}(\mathbf{w}^\mathsf{T}\boldsymbol{\Sigma}\mathbf{w})\boldsymbol{\Sigma}\mathbf{w}$
  - Hessians:
    - $\nabla^2\phi_1(\mathbf{w}) = \mathbf{0}$
    - $\nabla^2\phi_2(\mathbf{w}) = 2a_{21}\boldsymbol{\Sigma} + 2a_{22}\boldsymbol{\gamma}\boldsymbol{\gamma}^\mathsf{T}$
    - $\nabla^2\phi_3(\mathbf{w}) = 6a_{31}(\mathbf{w}^\mathsf{T}\boldsymbol{\gamma})\boldsymbol{\gamma}\boldsymbol{\gamma}^\mathsf{T} + 2a_{32}\left(\boldsymbol{\gamma}\mathbf{w}^\mathsf{T}\boldsymbol{\Sigma} + \boldsymbol{\Sigma}\mathbf{w}\boldsymbol{\gamma}^\mathsf{T} + (\mathbf{w}^\mathsf{T}\boldsymbol{\gamma})\boldsymbol{\Sigma}\right)$
    - $\nabla^2\phi_4(\mathbf{w}) =$
      $12a_{41}(\mathbf{w}^\mathsf{T}\boldsymbol{\gamma})^2\boldsymbol{\gamma}\boldsymbol{\gamma}^\mathsf{T} + 2a_{42}\left(2(\mathbf{w}^\mathsf{T}\boldsymbol{\gamma})\boldsymbol{\Sigma}\mathbf{w}\boldsymbol{\gamma}^\mathsf{T} + (\mathbf{w}^\mathsf{T}\boldsymbol{\gamma})^2\boldsymbol{\Sigma} + 2(\mathbf{w}^\mathsf{T}\boldsymbol{\gamma})\boldsymbol{\gamma}\mathbf{w}^\mathsf{T}\boldsymbol{\Sigma} + (\mathbf{w}^\mathsf{T}\boldsymbol{\Sigma}\mathbf{w})\boldsymbol{\gamma}\boldsymbol{\gamma}^\mathsf{T}\right) +$
      $4a_{43}\left(2\boldsymbol{\Sigma}\mathbf{w}\mathbf{w}^\mathsf{T}\boldsymbol{\Sigma} + (\mathbf{w}^\mathsf{T}\boldsymbol{\Sigma}\mathbf{w})\boldsymbol{\Sigma}\right)$

- **Take-home message under parametric modeling:**
  - No need to compute, store, and manipulate huge co-skewness and co-kurtosis matrices.
  - Can cheaply compute gradients and Hessians based on the parameters: $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, ν, and $\boldsymbol{\gamma}$.

High-order moments: L-moments

- **L-moments overview:**
  - L-moments characterize the distribution of a random variable and describe its properties such as location, dispersion, asymmetry, and tail thickness.
  - They are linear functions of order statistics, making them easier to estimate than traditional moments.
- **Definition of L-moments:**
  - Let X be a random variable and $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ be the *order statistics* of a random sample of size n drawn from the distribution of X.
  - L-moments for a random variable X are defined as:

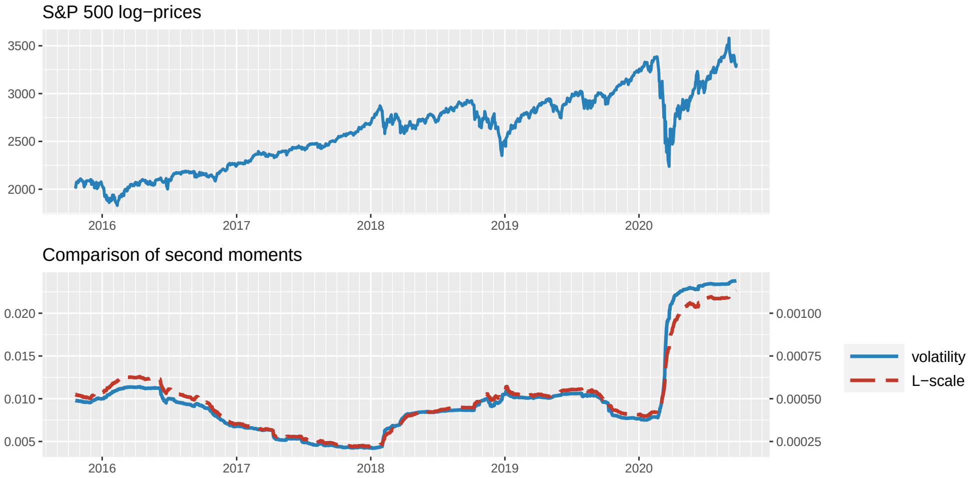
$$\lambda_r = \frac{1}{r} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} \mathbb{E}[X_{r-k:r}], \quad r = 1, 2, \ldots$$

  - First four L-moments:
    - $\lambda_1 = \mathbb{E}[X]$
    - $\lambda_2 = \frac{1}{2}\mathbb{E}[X_{2:2} - X_{1:2}]$
    - $\lambda_3 = \frac{1}{3}\mathbb{E}[X_{3:3} - 2X_{2:3} + X_{1:3}]$
    - $\lambda_4 = \frac{1}{4}\mathbb{E}[X_{4:4} - 3X_{3:4} + 3X_{2:4} - X_{1:4}]$

High-order moments: L-moments

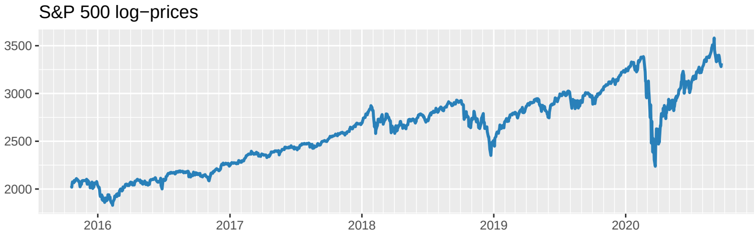
- **Descriptive information provided by L-moments:**
  - L-location (λ_1) is identical to the mean.
  - L-scale (λ_2) measures expected difference between any two realizations (like variance).
  - L-skewness (λ_3) provides a measure of asymmetry less sensitive to extreme tails.
  - L-kurtosis (λ_4) measures tail thickness, less sensitive to extreme tails.
- **Estimation of L-moments:**
  - Direct estimation from observations is computationally demanding.
  - Simplified estimators in terms of sample values in ascending order $x_{(i)}$:
    - $\hat{\lambda}_1 = \frac{1}{n}\sum_{i=1}^{n} x_{(i)}$
    - $\hat{\lambda}_2 = \frac{1}{2}\frac{1}{C_2^n}\sum_{i=1}^{n}\left(C_1^{i-1} - C_1^{n-i}\right)x_{(i)}$
    - $\hat{\lambda}_3 = \frac{1}{3}\frac{1}{C_3^n}\sum_{i=1}^{n}\left(C_2^{i-1} - 2C_1^{i-1}C_1^{n-i} + C_2^{n-i}\right)x_{(i)}$
    - $\hat{\lambda}_4 = \frac{1}{4}\frac{1}{C_4^n}\sum_{i=1}^{n}\left(C_3^{i-1} - 3C_2^{i-1}C_1^{n-i} + 3C_1^{i-1}C_2^{n-i} - C_3^{n-i}\right)x_{(i)}$
- **Comparison with traditional moments:**
  - L-moments convey similar information to traditional moments but are more stable.
  - This stability makes L-moments particularly useful for analyzing financial data, where they exhibit fewer jumps and provide a clearer picture.

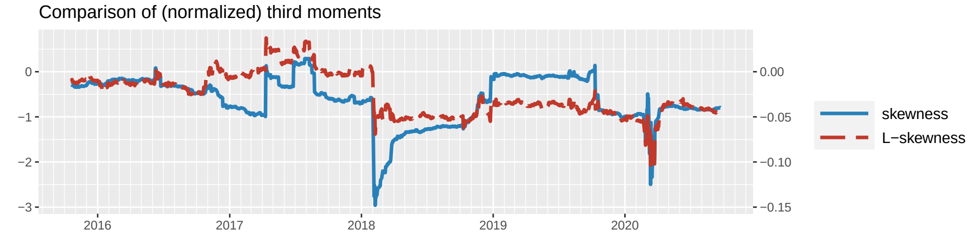
Moments and L-moments of the S&P 500 index in a rolling-window fashion:



High-order moments: L-moments

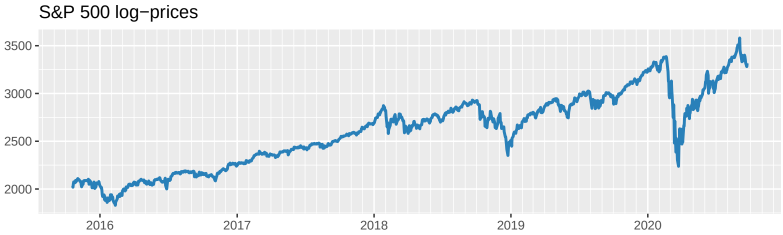
Moments and L-moments of the S&P 500 index in a rolling-window fashion:

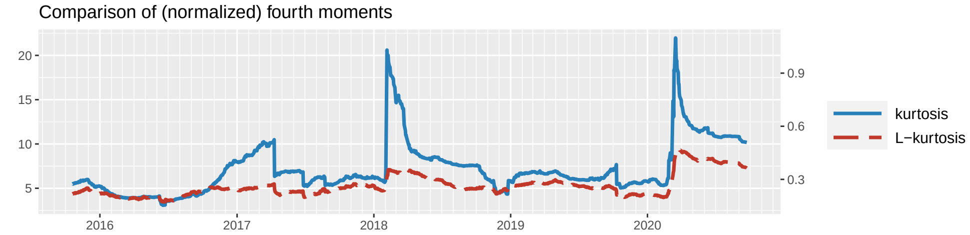




High-order moments: L-moments

Moments and L-moments of the S&P 500 index in a rolling-window fashion:





Outline

Portfolio formulations

- **Formulations involving High-Order Moments:**
  - Higher-order moments introduce nonconvexity into portfolio formulations.
  - Different expressions for moments $\phi_1(\boldsymbol{w})$, $\phi_2(\boldsymbol{w})$, $\phi_3(\boldsymbol{w})$, and $\phi_4(\boldsymbol{w})$ offer various formulation options.
  - MVSK portfolio: mean-variance-skewness-kurtosis.

- **Moment expression options:**
  - **Non-parametric:** Direct calculations from returns; computationally demanding and may need large data.
  - **Factor model structured:** Utilizes market or multiple factors to structure moments, reducing parameter count.
  - **Parametric:** Uses multivariate skewed t distribution, easing moment computations and balancing distributional capture with feasibility.
  - **L-moments:** Linear in order statistics, offering easier estimation and robustness to extremes, while conveying distribution insights.

MVSK portfolios

- **MVSK portfolio formulation:** Optimize a weighted combination of the first four moments:

$$\begin{array}{ll}\underset{\mathbf{w}}{\text{minimize}} & -\lambda_1\phi_1(\mathbf{w}) + \lambda_2\phi_2(\mathbf{w}) - \lambda_3\phi_3(\mathbf{w}) + \lambda_4\phi_4(\mathbf{w})\\ \text{subject to} & \mathbf{w} \in \mathcal{W}\end{array}$$

  where λ_i are hyper-parameters reflecting risk aversion.

- **Investor preferences:**
  - Seek higher mean and skewness: $\phi_1(\mathbf{w})$, $\phi_3(\mathbf{w})$.
  - Prefer lower variance and kurtosis: $\phi_2(\mathbf{w})$, $\phi_4(\mathbf{w})$.

MVSK portfolios

- **Alternative formulations:**
  - Constraints on moments:

$$
\begin{array}{ll}
\underset{\mathbf{w}}{\text{find}} & \mathbf{w} \\
\text{subject to} & \phi_1(\mathbf{w}) \geq \alpha_1 \\
& \phi_2(\mathbf{w}) \leq \alpha_2 \\
& \phi_3(\mathbf{w}) \geq \alpha_3 \\
& \phi_4(\mathbf{w}) \leq \alpha_4,
\end{array}
$$

  - α_i: Hyper-parameters denoting investor's targets.

- **Numerical algorithms:**
  - General-purpose solvers are always an option.
  - Specialized algorithms for solving MVSK have been recently developed (Zhou and Palomar 2021; Wang et al. 2023).

MVSK portfolios

- **Expected utility in portfolio design:**
  - Focuses on maximizing the expected value of a utility function $\mathbb{E}\left[U(\boldsymbol{w}^\mathsf{T}\boldsymbol{r})\right]$.
  - Moves beyond the mean-variance objective.
- **Historical context:**
  - High-order portfolios considered since 1969.
  - Geometric mean approximation of returns:

$$\mathbb{E}\left[\log\left(1+\boldsymbol{w}^\mathsf{T}\boldsymbol{r}\right)\right] \approx \log\left(1+\phi_1(\boldsymbol{w})\right) - \frac{\phi_2(\boldsymbol{w})}{2\phi_1^2(\boldsymbol{w})} + \frac{\phi_3(\boldsymbol{w})}{3\phi_1^3(\boldsymbol{w})} - \frac{\phi_4(\boldsymbol{w})}{4\phi_1^4(\boldsymbol{w})},$$

  - Mean-variance portfolio arises from using the first two terms of the log-approximation.
- **Advancements in high-order approximations:**
  - High-order expansions for arbitrary expected utilities can be now considered.
  - Recent work includes high-order approximations with structured estimators for moments.

Making portfolios efficient

- **Shortage function in multi-objective optimization:**
  - Measures the distance between a portfolio's moments and the efficient frontier.
  - Utilized to optimize portfolios towards Pareto-optimal points.
- **Optimization using the shortage function:**
  - Given a reference portfolio $\boldsymbol{w}^0$ and direction vector $\boldsymbol{g}$.
  - Objective: Maximize δ, representing movement towards the efficient frontier.
  - Constraints ensure improvement in desired moments along $\boldsymbol{g}$.
  - Formulation:

$$
\begin{array}{ll}
\underset{\boldsymbol{w},\delta\geq 0}{\text{maximize}} & \delta \\
\text{subject to} & \phi_1(\boldsymbol{w}) \geq \phi_1(\boldsymbol{w}^0) + \delta g_1 \\
& \phi_2(\boldsymbol{w}) \leq \phi_2(\boldsymbol{w}^0) - \delta g_2 \\
& \phi_3(\boldsymbol{w}) \geq \phi_3(\boldsymbol{w}^0) + \delta g_3 \\
& \phi_4(\boldsymbol{w}) \leq \phi_4(\boldsymbol{w}^0) - \delta g_4.
\end{array}
$$

- **Feasibility and efficiency:**
  - Formulation is always feasible.
  - If $\boldsymbol{w}^0$ is on the efficient frontier, the solution is $\boldsymbol{w} = \boldsymbol{w}^0$ and $\delta = 0$.

Portfolio tilting

- **Extending portfolio improvement with optimality measure:**
  - **Reference portfolio $\boldsymbol{w}^0$ optimization:**
    - Obtained by minimizing a cost function $\xi(\boldsymbol{w})$.
    - Examples of $\xi(\cdot)$ include Herfindahl index, risk contributions equalization, diversification ratio, and tracking error.

- **MVSK portfolio tilting formulation:**
  - Objective: Maximize δ to make $\boldsymbol{w}^0$ more optimal.
  - Constraints include maintaining or improving moment values and allowing a controlled loss of optimality (κ) for closer efficient frontier alignment.
  - Formulation:

$$\begin{array}{ll}\underset{\boldsymbol{w},\delta\geq 0}{\text{maximize}} & \delta\\ \text{subject to} & \xi(\boldsymbol{w}) \leq \xi(\boldsymbol{w}^0) + \kappa\\ & \phi_1(\boldsymbol{w}) \geq \phi_1(\boldsymbol{w}^0) + g_1(\delta)\\ & \phi_2(\boldsymbol{w}) \leq \phi_2(\boldsymbol{w}^0) - g_2(\delta)\\ & \phi_3(\boldsymbol{w}) \geq \phi_3(\boldsymbol{w}^0) + g_3(\delta)\\ & \phi_4(\boldsymbol{w}) \leq \phi_4(\boldsymbol{w}^0) - g_4(\delta)\end{array}$$

Portfolio tilting

- **Cost function examples for portfolio tilting:**
  - **Herfindahl index:** $\xi(\boldsymbol{w}) = \sum_{i=1}^{N} w_i^2$; encourages diversity.
  - **Risk parity:** $\xi(\boldsymbol{w}) = \sum_{i=1}^{N} \left(\frac{w_i(\boldsymbol{\Sigma w})_i}{\boldsymbol{w}^\mathsf{T}\boldsymbol{\Sigma w}} - \frac{1}{N} \right)^2$; equalizes risk contributions.
  - **Diversification ratio:** $\xi(\boldsymbol{w}) = -\frac{\boldsymbol{w}^\mathsf{T}\boldsymbol{\sigma}}{\sqrt{\boldsymbol{w}^\mathsf{T}\boldsymbol{\Sigma w}}}$; maximizes return-to-risk ratio.
  - **Tracking error:** $\xi(\boldsymbol{w}) = \sqrt{(\boldsymbol{w} - \boldsymbol{w}^{\text{benchmark}})^\mathsf{T}\boldsymbol{\Sigma}(\boldsymbol{w} - \boldsymbol{w}^{\text{benchmark}})}$; minimizes benchmark deviation.

- **Optimization Parameters:**
  - $g_i(\delta)$: Functions increasing with δ.
  - κ: Allows optimality loss for efficient frontier approach, set as $0.01 \times \xi(\boldsymbol{w}^0)$.

- **Numerical algorithms:**
  - General-purpose solvers are always an option.
  - Efficient algorithms developed for solving the MVSK tilting portfolio formulation (Zhou and Palomar 2021).

Polynomial goal programming MVSK portfolio

- **Polynomial goal programming for moment trade-off:**
  - Objective: Minimize distance to reference moments with a polynomial.
  - Formulation:

$$\begin{array}{ll}\underset{\boldsymbol{w},\boldsymbol{d}\geq\boldsymbol{0}}{\text{minimize}} & \left|\frac{d_1}{\phi_1^0}\right|^{\lambda_1}+\left|\frac{d_2}{\phi_2^0}\right|^{\lambda_2}+\left|\frac{d_3}{\phi_3^0}\right|^{\lambda_3}+\left|\frac{d_4}{\phi_4^0}\right|^{\lambda_4}\\ \text{subject to} & \phi_1(\boldsymbol{w})+d_1\geq\phi_1^0\\ & \phi_2(\boldsymbol{w})-d_2\leq\phi_2^0\\ & \phi_3(\boldsymbol{w})+d_3\geq\phi_3^0\\ & \phi_4(\boldsymbol{w})-d_4\leq\phi_4^0\end{array}$$

  - $\boldsymbol{d}$: Deviation from aspired moment levels ϕ_i^0.
  - Aspired levels represent extreme values and are not jointly achievable.

- **Minkovski distance case:**
  - Special case of polynomial goal programming with exponents $\lambda_i = 1/p$.
  - Objective becomes a Minkovski distance minimization.

Outline

Algorithms

- **MVSK portfolio formulation and objective function:**
  - Focus on MVSK portfolio with nonconvex objective due to higher-order moments.
  - Objective function:

$$f(\mathbf{w}) = -\lambda_1\phi_1(\mathbf{w}) + \lambda_2\phi_2(\mathbf{w}) - \lambda_3\phi_3(\mathbf{w}) + \lambda_4\phi_4(\mathbf{w})$$

- **Decomposing MVSK objective into convex and nonconvex terms:**

$$\begin{aligned} f_{\text{cvx}}(\mathbf{w}) &= -\lambda_1\phi_1(\mathbf{w}) + \lambda_2\phi_2(\mathbf{w}) \\ f_{\text{ncvx}}(\mathbf{w}) &= -\lambda_3\phi_3(\mathbf{w}) + \lambda_4\phi_4(\mathbf{w}). \end{aligned}$$

- **Efficient numerical methods beyond general nonlinear solvers:**
  - Developing ad-hoc methods for greater efficiency based on the *successive convex approximation* (SCA) framework (Scutari et al. 2014) and the *majorization-minimization* (MM) framework (Sun, Babu, and Palomar 2017).
  - R package `highOrderPortfolios` (Zhou and Palomar 2021; Wang et al. 2023):

SCA: Preliminaries

- **Successive convex approximation (SCA) method:** (Scutari et al. 2014)
  - SCA approximates a difficult optimization problem with a series of simpler convex problems.
  - Iteratively produces a sequence $\mathbf{x}^0, \mathbf{x}^1, \mathbf{x}^2, \dots$ converging towards a solution $\mathbf{x}^\star$.
- **SCA iterative process:**
  - At each iteration k, uses a surrogate function $\tilde{f}\left(\mathbf{x}; \mathbf{x}^k\right)$ to approximate objective $f(\mathbf{x})$.
  - The update rule includes a smoothing step to ensure convergence:

$$\begin{aligned} \hat{\mathbf{x}}^{k+1} &= \underset{\mathbf{x}\in\mathcal{X}}{\operatorname{argmin}} \ \tilde{f}\left(\mathbf{x}; \mathbf{x}^k\right) \\ \mathbf{x}^{k+1} &= \mathbf{x}^k + \gamma^k \left(\hat{\mathbf{x}}^{k+1} - \mathbf{x}^k\right) \end{aligned} \qquad k = 0, 1, 2, \dots$$

  where $\{\gamma^k\}$ is a sequence with $\gamma^k \in (0, 1]$.
- **Convergence conditions:**
  - Surrogate function $\tilde{f}\left(\mathbf{x}; \mathbf{x}^k\right)$ must be strongly convex on $\mathcal{X}$.
  - Surrogate function must be differentiable with gradient equal to that of $f(\mathbf{x})$ at $\mathbf{x}^k$.

- **SCA framework for nonconvex optimization:**
  - Leave convex term $f_{\text{cvx}}(\boldsymbol{w})$ unchanged.
  - Approximate nonconvex term $f_{\text{ncvx}}(\boldsymbol{w})$ quadratically around $\boldsymbol{w} = \boldsymbol{w}^k$:

$$\begin{aligned}\tilde{f}_{\text{ncvx}}\left(\boldsymbol{w};\boldsymbol{w}^k\right) =& f_{\text{ncvx}}\left(\boldsymbol{w}^k\right) + \nabla f_{\text{ncvx}}\left(\boldsymbol{w}^k\right)^T\left(\boldsymbol{w}-\boldsymbol{w}^k\right)\\ &+\frac{1}{2}\left(\boldsymbol{w}-\boldsymbol{w}^k\right)^T\left[\nabla^2 f_{\text{ncvx}}\left(\boldsymbol{w}^k\right)\right]_{\text{PSD}}\left(\boldsymbol{w}-\boldsymbol{w}^k\right),\end{aligned}$$

  - Projection onto positive semidefinite matrices: $[\boldsymbol{\Xi}]_{\text{PSD}} = \boldsymbol{U}\text{Diag}\left(\boldsymbol{\lambda}^+\right)\boldsymbol{U}^T$.

- **Quadratic convex approximation of $f(\boldsymbol{w})$:**

$$\tilde{f}\left(\boldsymbol{w};\boldsymbol{w}^k\right) = \frac{1}{2}\boldsymbol{w}^T\boldsymbol{Q}^k\boldsymbol{w} + \boldsymbol{w}^T\boldsymbol{q}^k + \text{constant},$$

where:

$$\begin{aligned}\boldsymbol{Q}^k &= \lambda_2\nabla^2\phi_2(\boldsymbol{w}) + \left[\nabla^2 f_{\text{ncvx}}\left(\boldsymbol{w}^k\right)\right]_{\text{PSD}}\\ \boldsymbol{q}^k &= -\lambda_1\nabla\phi_1(\boldsymbol{w}) + \nabla f_{\text{ncvx}}\left(\boldsymbol{w}^k\right) - \left[\nabla^2 f_{\text{ncvx}}\left(\boldsymbol{w}^k\right)\right]_{\text{PSD}}\boldsymbol{w}^k.\end{aligned}$$

SCA: Algorithm

SCA-Q-MVSK method for MVSK portfolio optimization (Zhou and Palomar 2021)

Initialization:

- Choose initial point $\mathbf{w}^0 \in \mathcal{W}$ and sequence $\{\gamma^k\}$.
- Set iteration counter $k \leftarrow 0$.

Repeat (kth iteration):

1. Calculate $\nabla f_{\text{ncvx}}\left(\mathbf{w}^k\right)$ and $\left[\nabla^2 f_{\text{ncvx}}\left(\mathbf{w}^k\right)\right]_{\text{PSD}}$.
2. Solve the QP approximation problem and keep solution as $\hat{\mathbf{w}}^{k+1}$.
3. $\mathbf{w}^{k+1} \leftarrow \mathbf{w}^k + \gamma^k(\hat{\mathbf{w}}^{k+1} - \mathbf{w}^k)$
4. $k \leftarrow k + 1$

Until: convergence

MM: Preliminaries

- **Majorization-Minimization (MM) method overview:** (Hunter and Lange 2004; Sun, Babu, and Palomar 2017) (Palomar 2025, Appendix B)
  - MM simplifies complex optimization problems through iterative surrogate minimization.
  - Iteratively produces a sequence $\mathbf{x}^0, \mathbf{x}^1, \mathbf{x}^2, \ldots$ converging towards a solution $\mathbf{x}^\star$.
- **MM iterative process:**
  - At each iteration k, MM uses a surrogate function $u\left(\mathbf{x}; \mathbf{x}^k\right)$ to approximate $f(\mathbf{x})$.
  - The update rule is:

$$\mathbf{x}^{k+1} = \underset{\mathbf{x} \in \mathcal{X}}{\arg\min}\; u\left(\mathbf{x}; \mathbf{x}^k\right) \qquad k = 0, 1, 2, \ldots$$

- **Convergence conditions for MM:** Surrogate function $u\left(\mathbf{x}; \mathbf{x}^k\right)$ must satisfy:
  - *Upper bound property*: $u\left(\mathbf{x}; \mathbf{x}^k\right) \geq f\left(\mathbf{x}\right)$.
  - *Touching property*: $u\left(\mathbf{x}^k; \mathbf{x}^k\right) = f\left(\mathbf{x}^k\right)$.
  - *Tangent property*: Differentiable with $\nabla u\left(\mathbf{x}; \mathbf{x}^k\right) = \nabla f\left(\mathbf{x}\right)$.
- **Role of the surrogate function:**
  - Acts as a majorizer, providing an upper bound to the original function.
  - The method's name stems from the process of constructing and minimizing this majorizer.

MM: Algorithm

- **MM Framework for nonconvex optimization:**
  - Convex term $f_{\text{cvx}}(\boldsymbol{w})$ remains as is.
  - Construct a majorizer for nonconvex term $f_{\text{ncvx}}(\boldsymbol{w})$ around $\boldsymbol{w} = \boldsymbol{w}^k$:

$$\tilde{f}_{\text{ncvx}}\left(\boldsymbol{w}; \boldsymbol{w}^k\right) = f_{\text{ncvx}}\left(\boldsymbol{w}^k\right) + \nabla f_{\text{ncvx}}\left(\boldsymbol{w}^k\right)^{\mathsf{T}}\left(\boldsymbol{w} - \boldsymbol{w}^k\right) + \frac{\tau_{\text{MM}}}{2}\|\boldsymbol{w} - \boldsymbol{w}^k\|_2^2$$

where τ_{MM} is a positive constant ensuring $\tilde{f}_{\text{ncvx}}\left(\boldsymbol{w}; \boldsymbol{w}^k\right)$ is an upper-bound of $f_{\text{ncvx}}(\boldsymbol{w})$.

- **Quadratic convex approximation of $f(\boldsymbol{w})$:**
  - Formulated as:

$$\tilde{f}\left(\boldsymbol{w}; \boldsymbol{w}^k\right) = -\lambda_1\phi_1(\boldsymbol{w}) + \lambda_2\phi_2(\boldsymbol{w}) + \nabla f_{\text{ncvx}}\left(\boldsymbol{w}^k\right)^{\mathsf{T}}\boldsymbol{w} + \frac{\tau_{\text{MM}}}{2}\|\boldsymbol{w} - \boldsymbol{w}^k\|_2^2 + \text{constant},$$

  - Gradient of f_{ncvx} obtained from gradients of $\phi_3(\boldsymbol{w})$ and $\phi_4(\boldsymbol{w})$.

MM: Algorithm

- **MM-based algorithm implementation:**
  - Solve convex problems successively:

$$\underset{\boldsymbol{w}}{\text{minimize}} - \lambda_1\phi_1(\boldsymbol{w}) + \lambda_2\phi_2(\boldsymbol{w}) + \nabla f_{\text{ncvx}}\left(\boldsymbol{w}^k\right)^\mathsf{T}\boldsymbol{w} + \frac{\tau_{\text{MM}}}{2}\|\boldsymbol{w} - \boldsymbol{w}^k\|_2^2,$$

  - Denote solution as $\text{MM}(\boldsymbol{w}^k)$, with $\boldsymbol{w}^{k+1} = \text{MM}(\boldsymbol{w}^k)$.

- **Acceleration technique - SQUAREM:**
  - Instead of direct update, use two steps and combine:

$$\begin{aligned}
\text{difference first update:} \quad & \boldsymbol{r}^k = \text{MM}(\boldsymbol{w}^k) - \boldsymbol{w}^k,\\
\text{difference of differences:} \quad & \boldsymbol{v}^k = R(\text{MM}(\boldsymbol{w}^k)) - R(\boldsymbol{w}^k),\\
\text{stepsize:} \quad & \alpha^k = -\max\left(1, \|\boldsymbol{r}^k\|_2/\|\boldsymbol{v}^k\|_2\right),\\
\text{actual step taken:} \quad & \boldsymbol{y}^k = \boldsymbol{w}^k - \alpha^k\boldsymbol{r}^k,\\
\text{final update on actual step:} \quad & \boldsymbol{w}^{k+1} = \text{MM}(\boldsymbol{y}^k).
\end{aligned}$$

  - Stepsize α^k can be refined for robustness and faster convergence.

MM: Algorithm

Acc-MM-L-MVSK method for MVSK portfolio optimization (Wang et al. 2023)

Initialization:

- Choose initial point $\boldsymbol{w}^0 \in \mathcal{W}$ and proper constant τ_{MM} for the majorized problem.
- Set iteration counter $k \leftarrow 0$.

Repeat (kth iteration):

1. Calculate $\nabla f_{\mathsf{ncvx}}\left(\boldsymbol{w}^k\right)$.
2. Compute the quantities $\boldsymbol{r}^k$, $\boldsymbol{v}^k$, α^k, $\boldsymbol{y}^k$, and current solution $\boldsymbol{w}^{k+1}$, which requires solving the majorized problem three times.
3. $k \leftarrow k+1$

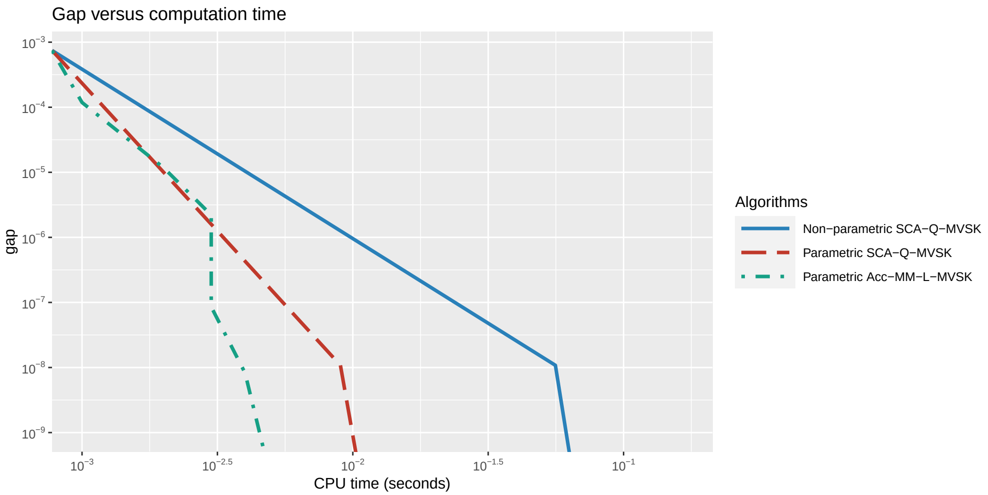
Until: convergence

Numerical experiments: Convergence and computation time

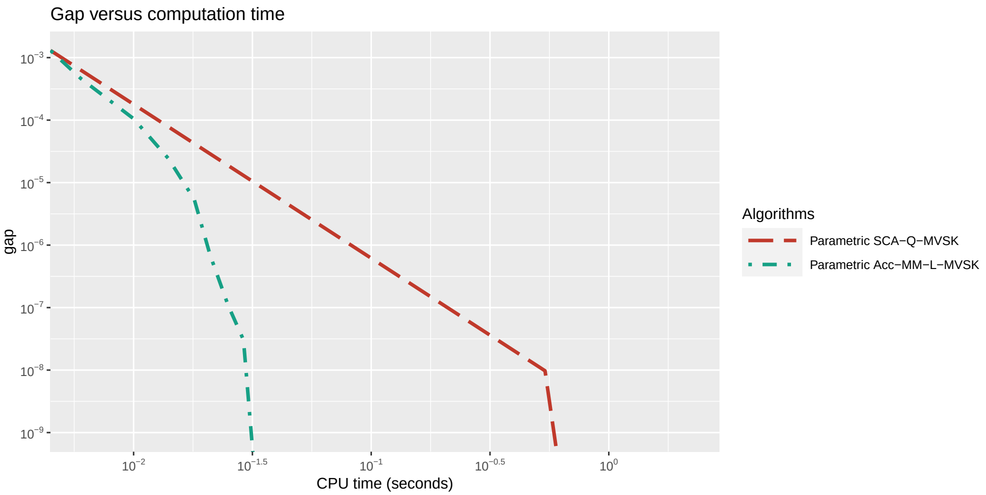
- **Empirical study summary:**
  - Assess computational efficiency of SCA-Q-MVSK versus Acc-MM-L-MVSK methods.
  - Both non-parametric and parametric calculations for gradients and Hessians were used.

- **Findings:**
  - Parametric calculations are faster and necessary for large N.
  - At $N = 100$, parametric is quicker (0.5s) than non-parametric (5s).
  - At $N = 400$, only parametric (1 min) is viable.
  - Enhanced Acc-MM-L-MVSK method omits τ_{MM} calculation, offering the quickest convergence.

Convergence of different MVSK portfolio optimization algorithms for $N = 100$:

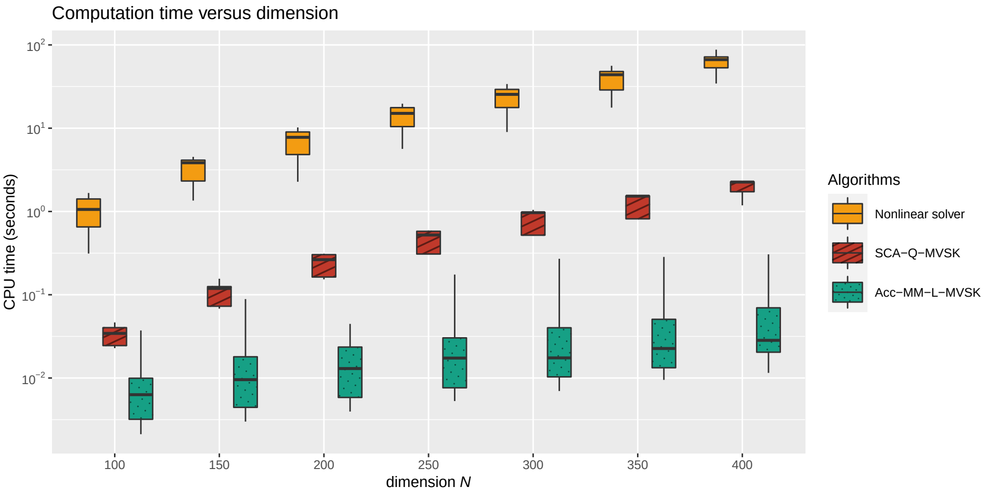


Convergence of different MVSK portfolio optimization algorithms for $N = 400$:



Numerical experiments: Convergence and computation time

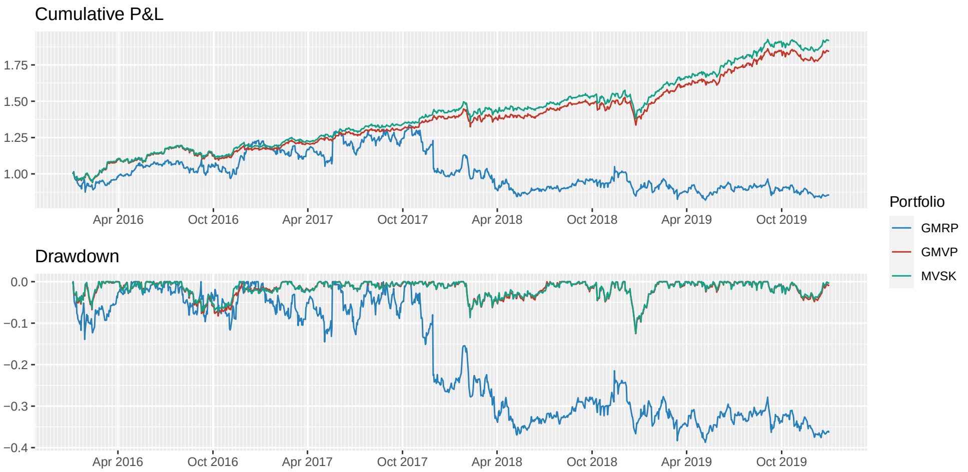
Computation time of different (parametric) MVSK portfolio optimization algorithms:



Numerical experiments: Portfolio backtest

- **Portfolios:**
  - Global Maximum Return (GMRP)
  - Global Minimum Variance (GMVP)
  - Modified MVSK (like GMVP with $\lambda_1 = 0$)
- **Evaluation:**
  - Period: 2016-2019
  - Universe: 20 S&P 500 stocks
- **Metrics:**
  - Cumulative P&L
  - Drawdown
- **Results:**
  - Cumulative P&L and drawdown shown in figure and table.
  - MVSK shows slight improvement over GMVP.
- **Insights:**
  - Incorporating skewness and kurtosis enhances performance.
  - Highlights the value of higher-order moments in optimization.

Backtest of high-order portfolios: cumulative P&L and drawdown:



Numerical experiments: Portfolio backtest

Backtest of high-order portfolios: performance measures:

Portfolio	Sharpe ratio	annual return	annual volatility	Sortino ratio	max drawdown	CVaR (0.95)
GMRP	-0.01	0%	27%	-0.02	39%	4%
GMVP	1.47	16%	11%	2.12	13%	2%
MVSK	1.56	17%	11%	2.26	12%	2%

Numerical experiments: Multiple portfolio backtests

- **Multiple randomized backtests overview:**
  - Dataset: $N = 20$ stocks, 2015-2020.
  - Method: 100 resamples with $N = 8$ stocks and random 2-year periods.
  - Backtest: Walk-forward with 1-year lookback, monthly reoptimization.

- **Results presentation:**
  - Performance measures for each portfolio across all backtests in next table.
  - Boxplots of Sharpe ratio and maximum drawdown.

- **Key observations:**
  - Modest performance improvement of MVSK portfolio over GMVP.
  - Highlights the potential benefits of incorporating higher-order moments in portfolio optimization strategies.

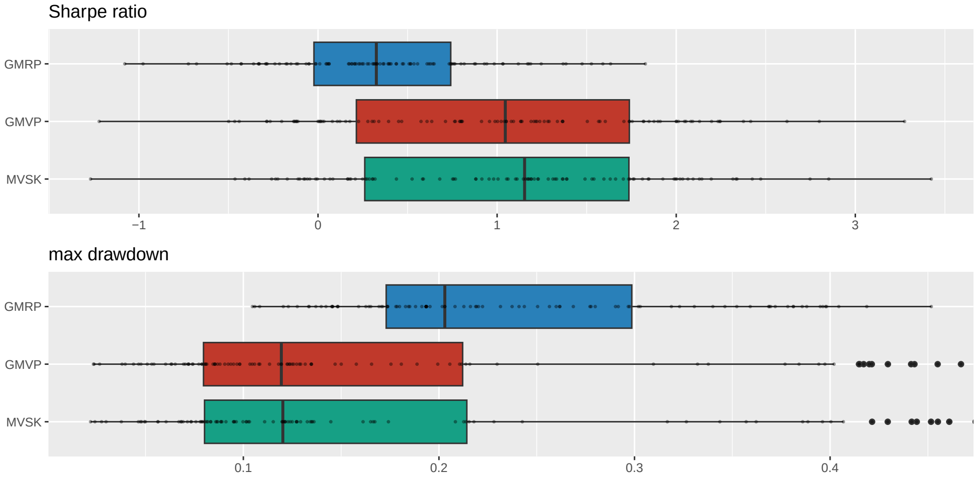
Numerical experiments: Multiple portfolio backtests

Multiple randomized backtest of high-order portfolios: performance measures:

Portfolio	Sharpe ratio	annual return	annual volatility	Sortino ratio	max drawdown	CVaR (0.95)
GMRP	0.33	9%	26%	0.45	20%	4%
GMVP	1.05	13%	13%	1.48	12%	2%
MVSK	1.15	14%	13%	1.58	12%	2%

Numerical experiments: Multiple portfolio backtests

Multiple randomized backtest of high-order portfolios: Sharpe ratio and maximum drawdown:



Outline

Summary

- Markowitz's portfolio, based on mean and variance, may not fully capture financial data's non-Gaussian traits, suggesting a need for higher moments.
- High-order portfolios include skewness and kurtosis to address the asymmetry and heavy tails in financial distributions.
- Initially conceptualized in the 1960s, high-order portfolios faced challenges due to the exponential increase in parameters (N^4) and nonconvex formulations, making early estimation and optimization difficult.
- Various high-order portfolio strategies exist, including MVSK portfolios and polynomial-goal formulations.
- Modern algorithms now facilitate efficient high-order portfolio management.
- Decades of research have made high-order portfolio design feasible for managing extensive asset collections, leaving the adoption of higher moments to traders' discretion.

References I

Boudt, K., W. Lu, and B. Peeters. 2014. “Higher Order Comoments of Multifactor Models and Asset Allocation.” *Finance Research Letters* 13: 225–33.

Brandt, M. W., P. Santa-Clara, and R. Valkanov. 2009. “Parametric Portfolio Policies: Exploiting Characteristics in the Cross Section of Equity Returns.” *Review of Financial Studies* 22 (9): 3411–47.

Hunter, D. R., and K. Lange. 2004. “A Tutorial on MM Algorithms.” *The American Statistician* 58: 30–37.

Jean, W. H. 1971. “The Extension of Portfolio Analysis to Three or More Parameters.” *Journal of Financial and Quantitative Analysis* 6 (1): 505–15.

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

Scutari, G., F. Facchinei, P. Song, D. P. Palomar, and J.-S. Pang. 2014. “Decomposition by Partial Linearization: Parallel Optimization of Multi-Agent Systems.” *IEEE Transactions on Signal Processing* 62 (3): 641–56.

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.

References II

Wang, X., R. Zhou, J. Ying, and D. P. Palomar. 2023. "Efficient and Scalable Parametric High-Order Portfolios Design via the Skew-t Distribution." *IEEE Transactions on Signal Processing* 71 (October): 3726–40.

Young, W. E., and R. H. Trent. 1969. "Geometric Mean Approximations of Individual Security and Portfolio Performance." *Journal of Financial and Quantitative Analysis* 4 (2): 179–99.

Zhou, R., and D. P. Palomar. 2021. "Solving High-Order Portfolios via Successive Convex Approximation Algorithms." *IEEE Transactions on Signal Processing* 69 (February): 892–904.