Portfolio Optimization

Robust Portfolios

Daniel P. Palomar (2024). *Portfolio Optimization: Theory and Application.* Cambridge University Press.

portfoliooptimizationbook.com

Latest update: November 09, 2024

Outline

Abstract

Markowitz's mean-variance portfolio optimizes the trade-off between expected return and risk, but it requires prior estimation of the mean vector and covariance matrix of assets, which often leads to significant errors in practice, making the approach less popular among practitioners. This issue, known as the Markowitz optimization enigma, arises because the naive approach ignores estimation errors, resulting in unstable and sensitive solutions. To address this, two main approaches are explored: robust optimization, which formulates problems to account for parameter errors, and resampling and bootstrapping methods, which use data resampling to create stable aggregated solutions. Both approaches are mature and maintain the convexity of the original portfolio formulation (Palomar 2024, chap. 14).

Outline

Introduction

- **Markowitz's mean-variance portfolio:**
  - Formulates portfolio design as a trade-off between expected return $\mathbf{w}^\mathsf{T}\boldsymbol{\mu}$ and risk measured by variance $\mathbf{w}^\mathsf{T}\boldsymbol{\Sigma}\mathbf{w}$:

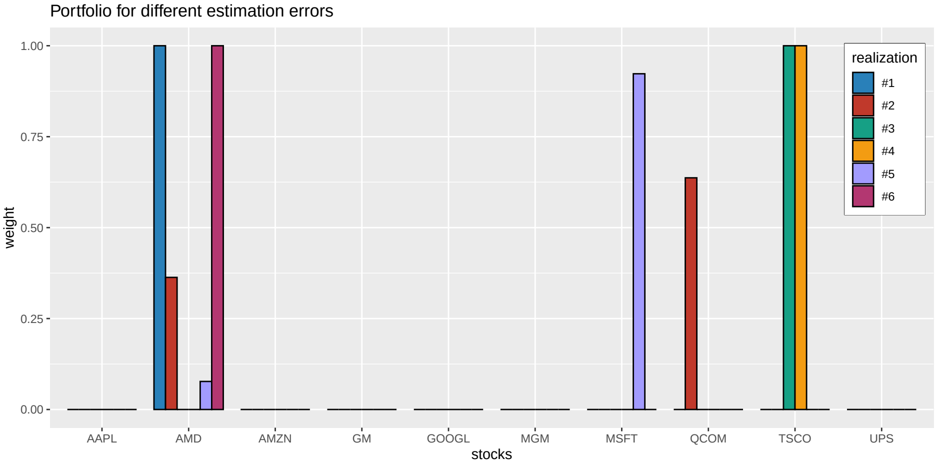
$$\begin{array}{ll} \underset{\mathbf{w}}{\text{maximize}} & \mathbf{w}^\mathsf{T}\boldsymbol{\mu} - \frac{\lambda}{2}\mathbf{w}^\mathsf{T}\boldsymbol{\Sigma}\mathbf{w} \\ \text{subject to} & \mathbf{w} \in \mathcal{W} \end{array}$$

  - Parameters $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ represent expected returns and covariance matrix.
  - λ is a hyper-parameter controlling investor's risk aversion.
  - $\mathcal{W}$ denotes constraint set, e.g., $\mathcal{W} = \{\mathbf{w} \mid \mathbf{1}^\mathsf{T}\mathbf{w} = 1, \mathbf{w} \geq \mathbf{0}\}$.
- **Parameter estimation:**
  - Parameters $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ are unknown and estimated using historical data $\mathbf{x}_1, \dots, \mathbf{x}_T$.
  - Estimators range from simple sample estimators to sophisticated shrinkage heavy-tailed maximum likelihood estimators.
  - Estimation error depends on the number of observations.
  - Limited historical data and lack of stationarity lead to noisy estimates $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$.
  - Noisy estimates particularly affect $\hat{\boldsymbol{\mu}}$ (Michaud 1989; Best and Grauer 1991; Chopra and Ziemba 1993).

- **Achilles' heel of portfolio optimization:**
  - Estimation noise in $\hat{\mu}$ and $\hat{\Sigma}$ leads to erratic portfolio designs.
  - Known as "Markowitz optimization enigma" (Michaud 1989):
    - Portfolio optimization problems are "estimation-error maximizers."
    - "Optimal" portfolios are financially meaningless (absence of significant investment value).

- **Sensitivity illustration:**
  - Next figure shows sensitivity of mean-variance portfolio for six different realizations of estimation error.
  - Behavior is erratic due to sensitivity to parameter errors.
  - Each realization differs significantly, which is impractical for portfolio allocation.

Sensitivity of the naive mean–variance portfolio:



Outline

Outline

Robust optimization

- **General mathematical optimization problem:** with optimization variable $\boldsymbol{x}$ and parameter $\boldsymbol{\theta}$:

$$\begin{array}{lll} \underset{\boldsymbol{x}}{\text{minimize}} & f_0(\boldsymbol{x};\boldsymbol{\theta}) & \\ \text{subject to} & f_i(\boldsymbol{x};\boldsymbol{\theta}) \leq 0, & i = 1,\dots,m \\ & h_i(\boldsymbol{x};\boldsymbol{\theta}) = 0, & i = 1,\dots,p, \end{array}$$

  - f_0: objective function
  - $f_i,\ i = 1,\dots,m$: inequality constraint functions
  - $h_i,\ i = 1,\dots,p$: equality constraint functions
  - $\boldsymbol{\theta}$: external parameter (e.g., $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$ in mean-variance portfolio)
  - Solution denoted by $\boldsymbol{x}^{\star}(\boldsymbol{\theta})$

Robust optimization

- **Parameter estimation:**
  - $\boldsymbol{\theta}$ is unknown and estimated as $\hat{\boldsymbol{\theta}}$
  - Solution using estimated parameter: $\mathbf{x}^{\star}(\hat{\boldsymbol{\theta}})$
  - Different from desired solution: $\mathbf{x}^{\star}(\hat{\boldsymbol{\theta}}) \neq \mathbf{x}^{\star}(\boldsymbol{\theta})$
  - Question of approximate equality: $\mathbf{x}^{\star}(\hat{\boldsymbol{\theta}}) \approx \mathbf{x}^{\star}(\boldsymbol{\theta})$
  - For mean-variance portfolio, solutions can be quite different as previously seen.

- **Making the problem robust to parameter errors:**
  - **Stochastic optimization:**
    - Relies on probabilistic modeling of the parameter (Prekopa 1995; Ruszczynski and Shapiro 2003; Birge and Louveaux 2011).
    - Includes: expectations (average behavior) and chance constraints (probabilistic constraints).
  - **Worst-case robust optimization:**
    - Relies on defining an uncertainty set for the parameter (Ben-Tal, El Ghaoui, and Nemirovski 2009; Ben-Tal and Nemirovski 2008; D. Bertsimas and Caramanis 2011; Lobo 2000).

Stochastic optimization

- **Stochastic robust optimization:**
  - Estimated parameter $\hat{\boldsymbol{\theta}}$ modeled as a random variable fluctuating around true value $\boldsymbol{\theta}$.
  - True value modeled as:
$$\boldsymbol{\theta} = \hat{\boldsymbol{\theta}} + \boldsymbol{\delta},$$
where $\boldsymbol{\delta}$ is the estimation error, a zero-mean random variable (e.g., Gaussian distribution).
  - Importance of choosing the correct covariance matrix (or shape) for the error term.
- **Average constraint:**
  - Instead of using the "naive constraint":
$$f\left(\boldsymbol{x}; \hat{\boldsymbol{\theta}}\right) \leq \alpha$$
  - Use the "average constraint":
$$\mathbb{E}_{\boldsymbol{\theta}}\left[f\left(\boldsymbol{x}; \boldsymbol{\theta}\right)\right] \leq \alpha,$$
where the expectation is with respect to the random variable $\boldsymbol{\theta}$.
  - Interpretation: Constraint satisfied on average, a relaxation of the true constraint.
  - Preserves convexity: If $f(\cdot; \boldsymbol{\theta})$ is convex for each $\boldsymbol{\theta}$, its expected value over $\boldsymbol{\theta}$ is also convex.

Stochastic optimization: Implementation

- **Brute-force sampling:** Sample S times the random variable $\boldsymbol{\theta}$ and use the constraint:

$$\frac{1}{S}\sum_{i=1}^{S} f(\mathbf{x};\boldsymbol{\theta}_i) \leq \alpha.$$

- **Adaptive sampling:** Sample the random variable $\boldsymbol{\theta}$ efficiently at each iteration while solving the problem.

- **Closed-form expression:** Compute the expected value in closed form when possible.

- **Stochastic programming:** Various numerical methods developed for stochastic optimization (Prekopa 1995; Ruszczynski and Shapiro 2003; Birge and Louveaux 2011).

Example: Stochastic average constraint in closed form

- Quadratic constraint $f(\mathbf{x};\boldsymbol{\theta}) = (\mathbf{c}^\mathsf{T}\mathbf{x})^2$, with parameter $\boldsymbol{\theta} = \mathbf{c}$.
- Modeled as $\mathbf{c} = \hat{\mathbf{c}} + \boldsymbol{\delta}$, where $\boldsymbol{\delta}$ is zero-mean with covariance matrix $\mathbf{Q}$.
- Expected value:

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\theta}}\left[f(\mathbf{x};\boldsymbol{\theta})\right] &= \mathbb{E}_{\boldsymbol{\delta}}\left[\left((\hat{\mathbf{c}} + \boldsymbol{\delta})^\mathsf{T}\mathbf{x}\right)^2\right] \\
&= \mathbb{E}_{\boldsymbol{\delta}}\left[(\hat{\mathbf{c}}^\mathsf{T}\mathbf{x})^2 + \mathbf{x}^\mathsf{T}\boldsymbol{\delta}\boldsymbol{\delta}^\mathsf{T}\mathbf{x}\right] \\
&= (\hat{\mathbf{c}}^\mathsf{T}\mathbf{x})^2 + \mathbf{x}^\mathsf{T}\mathbf{Q}\mathbf{x},
\end{aligned}$$

- Interpretation: the robust expected value has the form of the naive version plus a quadratic regularization term.

Stochastic optimization

- **Limitations of average constraints:**
  - No control over specific realizations of the estimation error.
  - Constraint may be violated in many instances.
  - Worst-case approach addresses this but can be too conservative.

- **Chance constraints:**
  - Compromise between being too relaxed or overly conservative.
  - Focus on satisfying the constraint with high probability (e.g., 95%).
  - Replace naive constraint with:

    $$\Pr\left[f(\boldsymbol{x};\boldsymbol{\theta}) \leq \alpha\right] \geq \epsilon,$$

    where ϵ is the confidence level (e.g., $\epsilon = 0.95$ for 95%).
  - Generally hard to deal with, often requiring approximations (Ben-Tal and Nemirovski 2008; Ben-Tal, El Ghaoui, and Nemirovski 2009).

Worst-case robust optimization

- **Worst-case robust optimization:**
  - Parameter $\boldsymbol{\theta}$ assumed to lie within an uncertainty region near the estimated value:
  $$\boldsymbol{\theta} \in \mathcal{U}_{\boldsymbol{\theta}},$$
  where $\mathcal{U}_{\boldsymbol{\theta}}$ is the uncertainty set centered at $\hat{\boldsymbol{\theta}}$.
  - Critical to choose the shape and size of the uncertainty set.

- **Typical choices for uncertainty set shape (size ϵ):**
  - **Spherical set:**
  $$\mathcal{U}_{\boldsymbol{\theta}} = \left\{\boldsymbol{\theta} \mid \|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\|_2 \leq \epsilon\right\};$$
  - **Box set:**
  $$\mathcal{U}_{\boldsymbol{\theta}} = \left\{\boldsymbol{\theta} \mid \|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\|_\infty \leq \epsilon\right\};$$
  - **Ellipsoidal set:**
  $$\mathcal{U}_{\boldsymbol{\theta}} = \left\{\boldsymbol{\theta} \mid \left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\right)^\mathsf{T} \boldsymbol{S}^{-1} \left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\right) \leq \epsilon^2\right\},$$
  where $\boldsymbol{S} \succ \mathbf{0}$ defines the shape of the ellipsoid.

Worst-case robust optimization

- **Worst-case constraint:**
  - Instead of using:
$$f\left(\boldsymbol{x};\hat{\boldsymbol{\theta}}\right) \leq \alpha,$$
  - Use the worst-case constraint:
$$\sup_{\boldsymbol{\theta}\in\mathcal{U}_{\hat{\boldsymbol{\theta}}}} f(\boldsymbol{x};\boldsymbol{\theta}) \leq \alpha.$$
  - Interpretation: Constraint satisfied for any point inside the uncertainty set, a conservative approach.
  - Preserves convexity: If $f(\cdot;\boldsymbol{\theta})$ is convex for each $\boldsymbol{\theta}$, its worst-case over $\boldsymbol{\theta}$ is also convex.

- **Variation: Distributionally robust optimization:**
  - Applies worst-case uncertainty philosophy to probability distributions.
  - Referred to as distributional uncertainty models or distributionally robust optimization (D. Bertsimas and Caramanis 2011).

Worst-case robust optimization: Implementation

- **Brute-force sampling:**
  - Sample S times the uncertainty set $\mathcal{U}_{\boldsymbol{\theta}}$ and use the constraint:

$$\max_{i=1,\dots,S} f(\mathbf{x};\boldsymbol{\theta}_i) \leq \alpha$$

  - Or equivalently, include S constraints:

$$f(\mathbf{x};\boldsymbol{\theta}_i) \leq \alpha, \qquad i = 1,\dots,S;$$

- **Adaptive sampling algorithms:** Sample the uncertainty set $\mathcal{U}_{\hat{\boldsymbol{\theta}}}$ efficiently at each iteration while solving the problem.
- **Closed-form expression:** Compute the supremum in closed form when possible.
- **Via Lagrange duality:** Rewrite the worst-case supremum as an infimum that can be combined with the outer portfolio optimization layer.
- **Saddle-point optimization:** Use numerical methods designed for minimax problems or related saddle-point problems (Bertsekas 1999; Tütüncü and Koenig 2004).

Example: Worst-case constraint in closed form

- Quadratic constraint $f(\boldsymbol{x};\boldsymbol{\theta}) = (\boldsymbol{c}^\mathsf{T}\boldsymbol{x})^2$, with parameter $\boldsymbol{\theta} = \boldsymbol{c}$.
- Belongs to a spherical uncertainty set:

$$\mathcal{U} = \left\{\boldsymbol{c} \mid \|\boldsymbol{c} - \hat{\boldsymbol{c}}\|_2 \leq \epsilon\right\}.$$

- Worst-case value:

$$\begin{aligned} \sup_{\boldsymbol{c}\in\mathcal{U}} |\boldsymbol{c}^\mathsf{T}\boldsymbol{x}| &= \sup_{\|\boldsymbol{\delta}\|\leq\epsilon} |(\hat{\boldsymbol{c}} + \boldsymbol{\delta})^\mathsf{T}\boldsymbol{x}| \\ &= |\hat{\boldsymbol{c}}^\mathsf{T}\boldsymbol{x}| + \sup_{\|\boldsymbol{\delta}\|\leq\epsilon} |\boldsymbol{\delta}^\mathsf{T}\boldsymbol{x}| \\ &= |\hat{\boldsymbol{c}}^\mathsf{T}\boldsymbol{x}| + \epsilon\|\boldsymbol{x}\|_2, \end{aligned}$$

  where we have used the triangle inequality and Cauchy-Schwarz's inequality (upper bound achieved by $\boldsymbol{\delta} = \epsilon\, \boldsymbol{x}/\|\boldsymbol{x}\|_2$).
- Interpretation: the worst-case expression has the form of the naive version plus a regularization term.

Outline

Worst-case mean vector $\boldsymbol{\mu}$

- **Global maximum return portfolio (GMRP):** For an estimated mean vector $\hat{\boldsymbol{\mu}}$:

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{maximize}} & \mathbf{w}^T \hat{\boldsymbol{\mu}} \\ \text{subject to} & \mathbf{1}^T \mathbf{w} = 1, \quad \mathbf{w} \geq \mathbf{0} \end{array}$$

  - Known to be highly sensitive to estimation errors.

- **Worst-case formulation:** Instead of the naive GMRP, use the worst-case formulation:

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{maximize}} & \inf_{\boldsymbol{\mu} \in \mathcal{U}_{\boldsymbol{\mu}}} \mathbf{w}^T \boldsymbol{\mu} \\ \text{subject to} & \mathbf{1}^T \mathbf{w} = 1, \quad \mathbf{w} \geq \mathbf{0} \end{array}$$

  - Typical choices for the uncertainty region $\mathcal{U}_{\boldsymbol{\mu}}$:
    - **Ellipsoidal set:**

$$\mathcal{U}_{\boldsymbol{\mu}} = \left\{ \boldsymbol{\mu} \mid \left(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}} \right)^T \mathbf{S}^{-1} \left(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}} \right) \leq \epsilon^2 \right\}$$

    - **Box set:**

$$\mathcal{U}_{\boldsymbol{\mu}} = \{ \boldsymbol{\mu} \mid \| \boldsymbol{\mu} - \hat{\boldsymbol{\mu}} \|_\infty \leq \epsilon \}$$

Worst-case mean vector $\boldsymbol{\mu}$

Worst-case mean vector under ellipsoidal uncertainty set

- **Ellipsoidal uncertainty set for $\boldsymbol{\mu}$:**

$$\mathcal{U}_{\boldsymbol{\mu}} = \left\{ \boldsymbol{\mu} = \hat{\boldsymbol{\mu}} + \kappa \boldsymbol{S}^{1/2}\boldsymbol{u} \mid \|\boldsymbol{u}\|_2 \leq 1 \right\},$$

  where:

  - $\boldsymbol{S}^{1/2}$ is the symmetric square-root matrix of the shape $\boldsymbol{S}$, e.g., $\boldsymbol{S} = (1/T)\boldsymbol{\Sigma}$;
  - κ determines the size of the ellipsoid.

- **Worst-case value of $\boldsymbol{w}^\mathsf{T}\boldsymbol{\mu}$:**

$$\begin{aligned}
\inf_{\boldsymbol{\mu} \in \mathcal{U}_{\boldsymbol{\mu}}} \boldsymbol{w}^\mathsf{T}\boldsymbol{\mu} &= \inf_{\|\boldsymbol{u}\| \leq 1} \boldsymbol{w}^\mathsf{T}(\hat{\boldsymbol{\mu}} + \kappa \boldsymbol{S}^{1/2}\boldsymbol{u}) \\
&= \boldsymbol{w}^\mathsf{T}\hat{\boldsymbol{\mu}} + \kappa \inf_{\|\boldsymbol{u}\| \leq 1} \boldsymbol{w}^\mathsf{T}\boldsymbol{S}^{1/2}\boldsymbol{u} \\
&= \boldsymbol{w}^\mathsf{T}\hat{\boldsymbol{\mu}} - \kappa \|\boldsymbol{S}^{1/2}\boldsymbol{w}\|_2.
\end{aligned}$$

Worst-case mean vector $\boldsymbol{\mu}$

Worst-case mean vector under box uncertainty set

- **Box uncertainty set for μ:**

$$\mathcal{U}_{\boldsymbol{\mu}} = \left\{\boldsymbol{\mu} \mid -\boldsymbol{\delta} \leq \boldsymbol{\mu} - \hat{\boldsymbol{\mu}} \leq \boldsymbol{\delta}\right\},$$

  where $\boldsymbol{\delta}$ is the half-width of the box in all dimensions.

- **Worst-case value of $\boldsymbol{w}^\mathsf{T}\mu$:**

$$\inf_{\mu \in \mathcal{U}_{\boldsymbol{\mu}}} \boldsymbol{w}^\mathsf{T}\mu = \boldsymbol{w}^\mathsf{T}\hat{\mu} - |\boldsymbol{w}|^\mathsf{T}\delta.$$

Worst-case mean vector $\boldsymbol{\mu}$

- **Robust GMRP with ellipsoidal uncertainty set:**
  - Robust version of the GMRP under ellipsoidal uncertainty set:

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{maximize}} & \mathbf{w}^T\hat{\boldsymbol{\mu}} - \kappa \left\|\mathbf{S}^{1/2}\mathbf{w}\right\|_2 \\ \text{subject to} & \mathbf{1}^T\mathbf{w} = 1, \quad \mathbf{w} \geq \mathbf{0}, \end{array}$$

  - Still a convex problem.
  - Complexity increased to a second-order cone program (from a simple linear program in the naive formulation).
- **Robust GMRP with box uncertainty set:**
  - Robust version of the GMRP under box uncertainty set:

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{maximize}} & \mathbf{w}^T\hat{\boldsymbol{\mu}} - |\mathbf{w}|^T\boldsymbol{\delta} \\ \text{subject to} & \mathbf{1}^T\mathbf{w} = 1, \quad \mathbf{w} \geq \mathbf{0}, \end{array}$$

  - Still a convex problem.
  - Can be rewritten as a linear program after removing the absolute value.
  - Under constraints $\mathbf{1}^T\mathbf{w} = 1$ and $\mathbf{w} \geq \mathbf{0}$, the problem reduces to a naive GMRP where $\hat{\boldsymbol{\mu}}$ is replaced by $\hat{\boldsymbol{\mu}} - \boldsymbol{\delta}$.

Quintile portfolio as a robust portfolio

- **Other uncertainty sets:**
  - Various uncertainty sets can be considered, such as the ℓ_1-norm ball.

- **Example: Quintile portfolio as a robust portfolio**
  - **Quintile portfolio:**
    - A heuristic portfolio widely used by practitioners.
    - Selects the top fifth of the assets (or a different fraction) and equally allocates capital among them.
    - Common-sense heuristic portfolio.
  - **Formal derivation as a robust portfolio:**
    - The quintile portfolio can be shown to be the optimal solution to the worst-case GMRP with an ℓ_1-norm ball uncertainty set around the estimated mean vector (Zhou and Palomar 2020):

$$\mathcal{U}_{\boldsymbol{\mu}} = \left\{ \hat{\boldsymbol{\mu}} + \boldsymbol{e} \mid \|\boldsymbol{e}\|_1 \leq \epsilon \right\}.$$

Worst-case covariance matrix $\boldsymbol{\Sigma}$

- **Global minimum variance portfolio (GMVP):** For an estimated covariance matrix $\hat{\boldsymbol{\Sigma}}$:

$$\begin{array}{ll} \underset{\boldsymbol{w}}{\text{minimize}} & \boldsymbol{w}^{\mathsf{T}}\hat{\boldsymbol{\Sigma}}\boldsymbol{w} \\ \text{subject to} & \mathbf{1}^{\mathsf{T}}\boldsymbol{w} = 1, \quad \boldsymbol{w} \geq \mathbf{0} \end{array}$$

  - Sensitive to estimation errors (though less so than errors in $\boldsymbol{\mu}$).
- **Worst-case formulation:** Instead of the naive GMVP, use the worst-case formulation:

$$\begin{array}{ll} \underset{\boldsymbol{w}}{\text{minimize}} & \sup_{\boldsymbol{\Sigma}\in\mathcal{U}_{\boldsymbol{\Sigma}}} \boldsymbol{w}^{\mathsf{T}}\boldsymbol{\Sigma}\boldsymbol{w} \\ \text{subject to} & \mathbf{1}^{\mathsf{T}}\boldsymbol{w} = 1, \quad \boldsymbol{w} \geq \mathbf{0} \end{array}$$

  - Typical choices for the uncertainty region $\mathcal{U}_{\boldsymbol{\Sigma}}$:
    - **Spherical set:**

$$\mathcal{U}_{\boldsymbol{\Sigma}} = \left\{\boldsymbol{\Sigma} \mid \|\boldsymbol{\Sigma} - \hat{\boldsymbol{\Sigma}}\|_F \leq \epsilon\right\}$$

    - **Ellipsoidal set:**

$$\mathcal{U}_{\boldsymbol{\Sigma}} = \left\{\boldsymbol{\Sigma} \mid \text{vec}(\boldsymbol{\Sigma} - \hat{\boldsymbol{\Sigma}})^{\mathsf{T}}\boldsymbol{S}^{-1}\text{vec}(\boldsymbol{\Sigma} - \hat{\boldsymbol{\Sigma}}) \leq \epsilon^2\right\}$$

    - **Box set:**

$$\mathcal{U}_{\boldsymbol{\Sigma}} = \left\{\boldsymbol{\Sigma} \mid \|\boldsymbol{\Sigma} - \hat{\boldsymbol{\Sigma}}\|_\infty \leq \epsilon\right\}$$

Worst-case covariance matrix Σ

Worst-case covariance matrix under a data spherical uncertainty set

- **Spherical uncertainty set for data matrix $\boldsymbol{X}$:**

$$\mathcal{U}_{\boldsymbol{X}} = \left\{ \boldsymbol{X} \mid \|\boldsymbol{X} - \hat{\boldsymbol{X}}\|_{\mathsf{F}} \leq \epsilon \right\},$$

  where ϵ determines the size of the sphere.

- **Worst-case value of $\sqrt{\boldsymbol{w}^\mathsf{T}\Sigma\boldsymbol{w}}$:** (assuming $\hat{\Sigma} = \frac{1}{T}\hat{\boldsymbol{X}}^\mathsf{T}\hat{\boldsymbol{X}}$)

$$\begin{aligned}
\sup_{\boldsymbol{X}\in\mathcal{U}_{\boldsymbol{X}}} \sqrt{\boldsymbol{w}^\mathsf{T}\left(\frac{1}{T}\boldsymbol{X}^\mathsf{T}\boldsymbol{X}\right)\boldsymbol{w}} &= \sup_{\|\boldsymbol{\Delta}\|_\mathsf{F}\leq\epsilon} \frac{1}{\sqrt{T}}\left\|(\hat{\boldsymbol{X}}+\boldsymbol{\Delta})\boldsymbol{w}\right\|_2 \\
&= \frac{1}{\sqrt{T}}\left\|\hat{\boldsymbol{X}}\boldsymbol{w}\right\|_2 + \sup_{\|\boldsymbol{\Delta}\|_\mathsf{F}\leq\epsilon} \frac{1}{\sqrt{T}}\|\boldsymbol{\Delta}\boldsymbol{w}\|_2 \\
&= \frac{1}{\sqrt{T}}\left\|\hat{\boldsymbol{X}}\boldsymbol{w}\right\|_2 + \frac{1}{\sqrt{T}}\epsilon\|\boldsymbol{w}\|_2.
\end{aligned}$$

Worst-case covariance matrix Σ

Worst-case covariance matrix under an ellipsoidal uncertainty set

- **Ellipsoidal uncertainty set for Σ:**

$$\mathcal{U}_{\boldsymbol{\Sigma}} = \left\{\boldsymbol{\Sigma} \succeq \mathbf{0} \mid (\mathrm{vec}(\boldsymbol{\Sigma}) - \mathrm{vec}(\hat{\boldsymbol{\Sigma}}))^{\mathsf{T}} \boldsymbol{S}^{-1} (\mathrm{vec}(\boldsymbol{\Sigma}) - \mathrm{vec}(\hat{\boldsymbol{\Sigma}})) \leq \epsilon^2 \right\},$$

  where:
  - $\mathrm{vec}(\cdot)$ denotes the vec operator that stacks the matrix argument into a vector.
  - matrix $\boldsymbol{S}$ gives the shape of the ellipsoid.
  - ϵ determines the size.
- **Worst-case value of $\boldsymbol{w}^{\mathsf{T}}\boldsymbol{\Sigma}\boldsymbol{w}$:** Given by the (convex) SDP (Palomar 2024, chap. 14):

$$\begin{array}{ll} \underset{\boldsymbol{Z}}{\text{minimize}} & \mathrm{Tr}\left(\hat{\boldsymbol{\Sigma}}\left(\boldsymbol{w}\boldsymbol{w}^{\mathsf{T}} + \boldsymbol{Z}\right)\right) + \epsilon \left\|\boldsymbol{S}^{1/2}\left(\mathrm{vec}(\boldsymbol{w}\boldsymbol{w}^{\mathsf{T}}) + \mathrm{vec}(\boldsymbol{Z})\right)\right\|_2 \\ \text{subject to} & \boldsymbol{Z} \succeq \mathbf{0}. \end{array}$$

Worst-case covariance matrix $\boldsymbol{\Sigma}$

Worst-case covariance matrix under a box uncertainty set

- **Box uncertainty set for $\boldsymbol{\Sigma}$:**

$$\mathcal{U}_{\boldsymbol{\Sigma}} = \left\{ \boldsymbol{\Sigma} \succeq \mathbf{0} \mid \underline{\boldsymbol{\Sigma}} \leq \boldsymbol{\Sigma} \leq \overline{\boldsymbol{\Sigma}} \right\},$$

  where $\underline{\boldsymbol{\Sigma}}$ and $\overline{\boldsymbol{\Sigma}}$ denote the elementwise lower and upper bounds, respectively.
- **Worst-case value of $\boldsymbol{w}^{\mathsf{T}}\boldsymbol{\Sigma}\boldsymbol{w}$:** Given by the (convex) SDP (Lobo 2000):

$$\begin{array}{ll}
\underset{\overline{\boldsymbol{\Lambda}}, \underline{\boldsymbol{\Lambda}}}{\text{minimize}} & \text{Tr}(\overline{\boldsymbol{\Lambda}}\, \overline{\boldsymbol{\Sigma}}) - \text{Tr}(\underline{\boldsymbol{\Lambda}}\, \underline{\boldsymbol{\Sigma}}) \\
\text{subject to} & \begin{bmatrix} \overline{\boldsymbol{\Lambda}} - \underline{\boldsymbol{\Lambda}} & \boldsymbol{w} \\ \boldsymbol{w}^{\mathsf{T}} & 1 \end{bmatrix} \succeq \mathbf{0} \\
& \overline{\boldsymbol{\Lambda}} \geq \mathbf{0}, \quad \underline{\boldsymbol{\Lambda}} \geq \mathbf{0}.
\end{array}$$

Worst-case covariance matrix Σ

- **Robust GMVP with spherical uncertainty set:**
  - Robust version of the GMVP under a spherical uncertainty set for the data matrix:

$$\begin{array}{ll}\underset{\boldsymbol{w}}{\text{minimize}} & \left(\left\|\hat{\boldsymbol{X}}\boldsymbol{w}\right\|_2 + \epsilon\|\boldsymbol{w}\|_2\right)^2\\ \text{subject to} & \boldsymbol{1}^T\boldsymbol{w} = 1, \quad \boldsymbol{w} \geq \boldsymbol{0}\end{array}$$

  - Still a convex problem.
  - Complexity increased to a second-order cone program (from a simple quadratic program).
- **Tikhonov regularization:**
  - Heuristic similar to the robust GMVP:

$$\begin{array}{ll}\underset{\boldsymbol{w}}{\text{minimize}} & \left\|\hat{\boldsymbol{X}}\boldsymbol{w}\right\|_2^2 + \epsilon\|\boldsymbol{w}\|_2^2\\ \text{subject to} & \boldsymbol{1}^T\boldsymbol{w} = 1, \quad \boldsymbol{w} \geq \boldsymbol{0}\end{array}$$

  - Objective function can be rewritten as:

$$\boldsymbol{w}^T \frac{1}{T}\left(\hat{\boldsymbol{X}}^T\hat{\boldsymbol{X}} + \epsilon\boldsymbol{I}\right)\boldsymbol{w}$$

  - Leads to a regularized sample covariance matrix.

Worst-case covariance matrix Σ

- **Robust GMVP with ellipsoidal uncertainty set:**
  - Robust version of the GMVP under an ellipsoidal uncertainty set for the covariance matrix:

$$
\begin{array}{ll}
\underset{\mathbf{w},\mathbf{W},\mathbf{Z}}{\text{minimize}} & \text{Tr}\left(\hat{\Sigma}\left(\mathbf{W}+\mathbf{Z}\right)\right)+\epsilon\left\|\mathbf{S}^{1/2}\left(\text{vec}(\mathbf{W})+\text{vec}(\mathbf{Z})\right)\right\|_2 \\
\text{subject to} & \begin{bmatrix}\mathbf{W} & \mathbf{w}\\ \mathbf{w}^T & 1\end{bmatrix}\succeq\mathbf{0} \\
& \mathbf{1}^T\mathbf{w}=1, \quad \mathbf{w}\geq\mathbf{0} \\
& \mathbf{Z}\succeq\mathbf{0}
\end{array}
$$

  - Still a convex problem.
  - Complexity increased to a semidefinite program (from a simple quadratic program in the naive formulation).
  - Note: the first matrix inequality is equivalent to $\mathbf{W}\succeq\mathbf{w}\mathbf{w}^T$ and, at an optimal point, it can be shown to be satisfied with equality $\mathbf{W}=\mathbf{w}\mathbf{w}^T$.

Worst-case mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$

- **Combined worst-case mean-variance portfolio:**
  - Combines uncertainty in the mean vector $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$ under the mean-variance portfolio formulation.
  - For illustration, consider the mean-variance worst-case portfolio formulation under box uncertainty sets for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

- **Formulation:**

$$\begin{array}{ll} \underset{\boldsymbol{w},\overline{\boldsymbol{\Lambda}},\underline{\boldsymbol{\Lambda}}}{\text{maximize}} & \boldsymbol{w}^{\mathsf{T}}\hat{\boldsymbol{\mu}} - |\boldsymbol{w}|^{\mathsf{T}}\boldsymbol{\delta} - \frac{\lambda}{2}\left(\text{Tr}(\overline{\boldsymbol{\Lambda}}\,\overline{\boldsymbol{\Sigma}}) - \text{Tr}(\underline{\boldsymbol{\Lambda}}\,\underline{\boldsymbol{\Sigma}})\right) \\ \text{subject to} & \mathbf{1}^{\mathsf{T}}\boldsymbol{w} = 1, \quad \boldsymbol{w} \geq \mathbf{0}, \\ & \begin{bmatrix} \overline{\boldsymbol{\Lambda}} - \underline{\boldsymbol{\Lambda}} & \boldsymbol{w} \\ \boldsymbol{w}^{\mathsf{T}} & 1 \end{bmatrix} \succeq \mathbf{0} \\ & \overline{\boldsymbol{\Lambda}} \geq \mathbf{0}, \quad \underline{\boldsymbol{\Lambda}} \geq \mathbf{0} \end{array}$$

  - This is a convex semidefinite problem.

Outline

Numerical experiments

- **Effectiveness of robust portfolio formulations:**
  - Depends on the shape and size of the uncertainty region.
  - Parameters must be properly chosen and tuned to the type of data and nature of the financial market.
  - Risk of overfitting the model to the training data; extra care is needed.
- **Goal of robust design:**
  - Aim to make the solution more stable and less sensitive to error realization.
  - Not necessarily to improve performance.
  - Focus on gaining robustness rather than achieving the best performance in a given backtest compared to a naive design.
- **Evaluation of robust portfolios under errors in mean vector $\boldsymbol{\mu}$:**
  - We next consider robustness against errors in mean vector $\boldsymbol{\mu}$, which is paramount.
  - Robustness for the covariance matrix $\boldsymbol{\Sigma}$ is less critical, but should be also assesed.

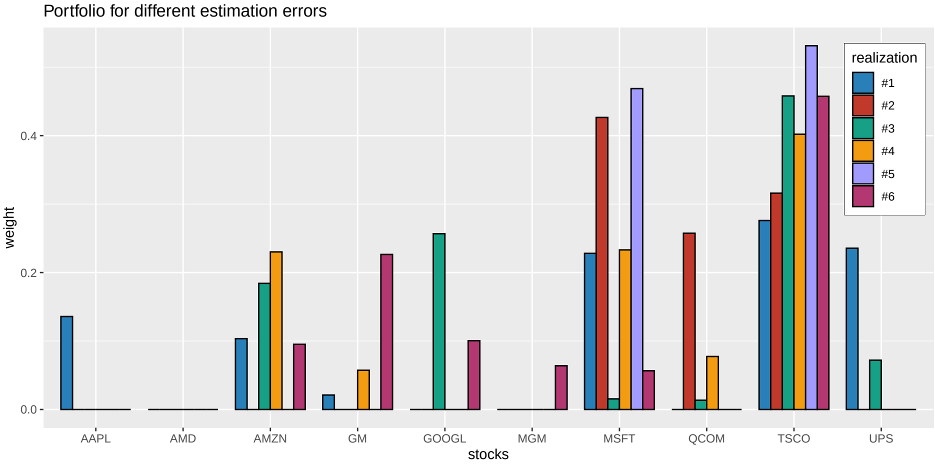
Numerical experiments

- **Sensitivity analysis:**
  - Sensitivity of a robust portfolio under an ellipsoidal uncertainty set for the mean vector $\boldsymbol{\mu}$ over six different realizations of the estimation error.
  - Compared to naive portfolio design, it is more stable and less sensitive.
  - Similar results can be observed with other variations in the robust formulation.

- **Performance assessment:**
  - Assess performance of mean-variance robust designs (with box and ellipsoidal uncertainty sets for the mean vector $\boldsymbol{\mu}$) compared to the naive design.
  - Backtests conducted for 50 randomly chosen stocks from the S&P 500 during 2017-2020.

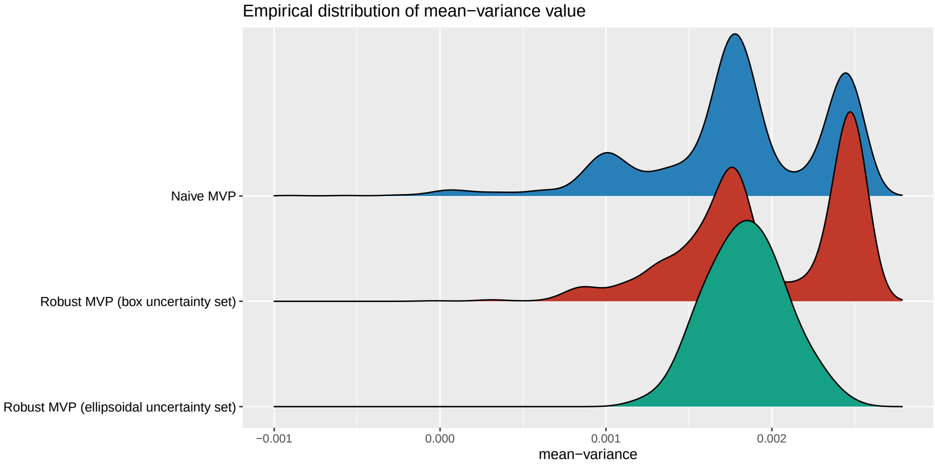
- **Empirical distribution analysis:**
  - Empirical distribution of the achieved mean-variance objective and the Sharpe ratio, calculated over 1,000 Monte Carlo noisy observations.
  - Robust designs avoid the worst-case realizations (left tail in the distributions) at the expense of not achieving the best-case realizations (right tail).
  - Robust designs are more stable and robust as expected.

Sensitivity of the robust mean–variance portfolio under an ellipsoidal uncertainty set for μ:



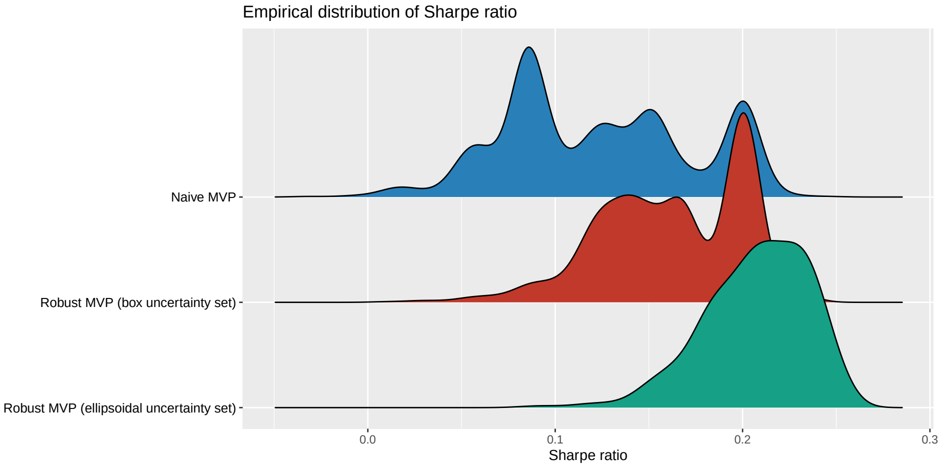
Numerical experiments

Empirical performance distribution of naive versus robust mean–variance portfolios:

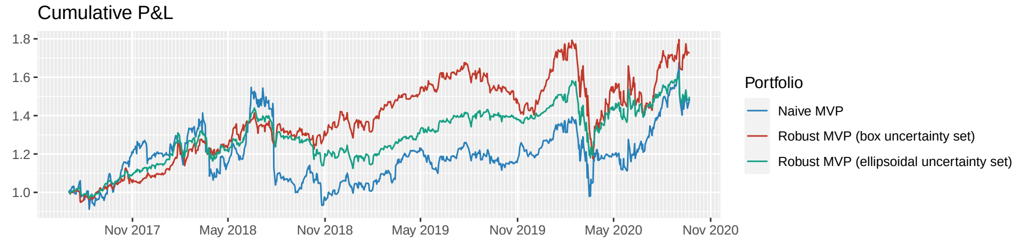


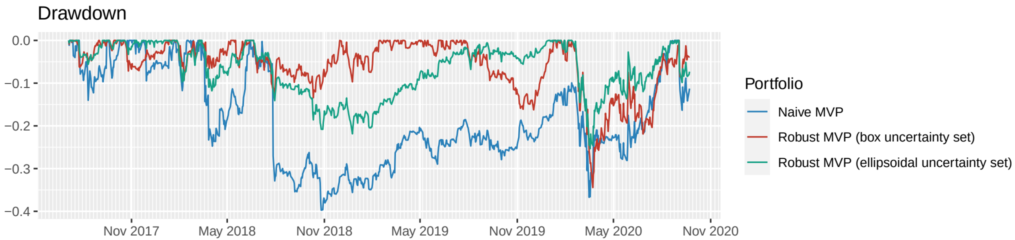
Numerical experiments

Empirical performance distribution of naive versus robust mean–variance portfolios:

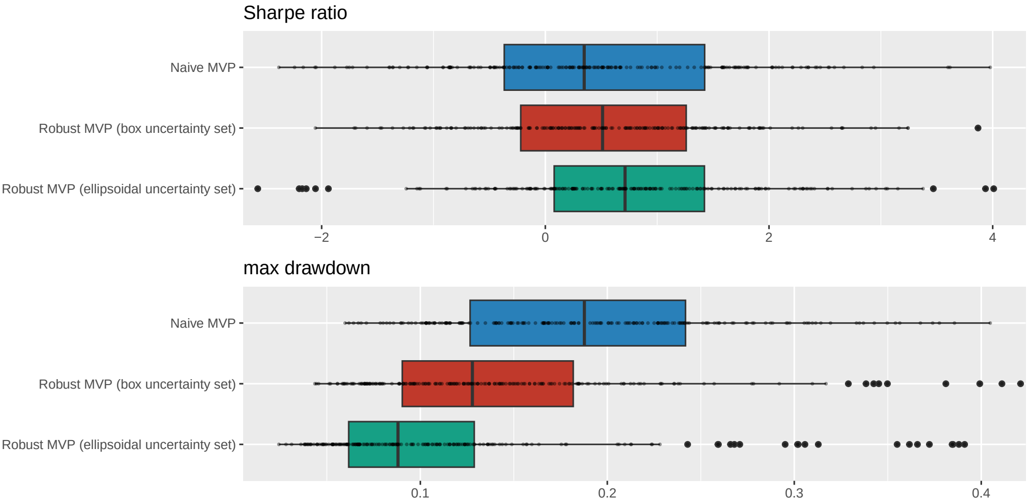


Backtest of naive versus robust mean–variance portfolios:





Multiple backtests of naive versus resampled mean–variance portfolios:



Outline

Outline

What are resampling methods?

- **Importance of estimating parameter accuracy:**
  - Estimating a parameter $\boldsymbol{\theta}$ is of little use without knowing the accuracy of the estimate.
  - Confidence intervals are key in statistical inference, allowing localization of the true parameter with a certain confidence level (e.g., 95%).
  - Traditionally, confidence intervals were derived using theoretical mathematics.
  - Resampling methods use computer-based numerical techniques to assess statistical accuracy without formulas (Efron and Tibshirani 1993).
- **Resampling methods:**
  - Creation of new samples based on a single observed sample block.
  - Suppose we have n observations, $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$, of a random variable $\boldsymbol{x}$ from which we estimate some parameters $\boldsymbol{\theta}$ as:

$$\hat{\boldsymbol{\theta}} = f(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n),$$

    where $f(\cdot)$ denotes the estimator.
  - The estimation $\hat{\boldsymbol{\theta}}$ is a random variable because it is based on n random variables.
  - Resampling methods help characterize the distribution of the estimation without needing more realizations of the random variable $\boldsymbol{x}$.

Resampling methods

- **Cross-Validation:**
  - Widely used in portfolio backtesting and machine learning.
  - Divides n observations into two groups: a training set for fitting the estimator $f(\cdot)$ and a validation set for assessing its performance.
  - Repeated multiple times to provide multiple realizations of the performance value.
  - **k-fold cross-validation:** Divides the set into k subsets, each held out in turn as the validation set while using the others for training.
  - **Leave-one-out cross-validation:** Extreme case where the original dataset of n observations is divided into $k = n$ subsets, with each subset holding out a single observation for validation.

Resampling methods

- **The Bootstrap:**
  - Proposed in 1979 by Efron, based on sound statistical theory (Efron 1979).
  - Mimics the original sampling process by sampling n times with replacement from the original n observations.
  - Repeated B times to obtain bootstrap samples:

$$\left(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\right) \rightarrow \left(\boldsymbol{x}_1^{*(b)}, \ldots, \boldsymbol{x}_n^{*(b)}\right), \quad b = 1, \ldots, B,$$

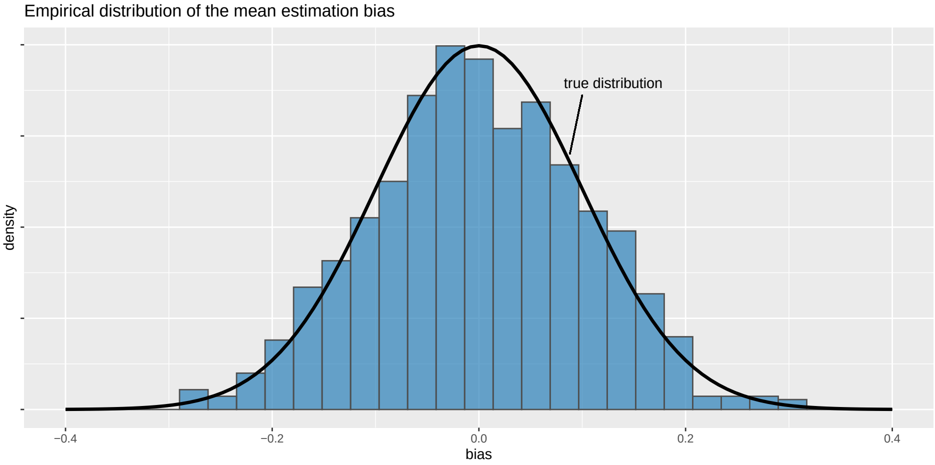
  - Leads to different realizations of the estimation (bootstrap replicates):

$$\hat{\boldsymbol{\theta}}^{*(b)} = f\left(\boldsymbol{x}_1^{*(b)}, \ldots, \boldsymbol{x}_n^{*(b)}\right), \quad b = 1, \ldots, B,$$

  - Measures of accuracy (bias, variance, confidence intervals, etc.) can be empirically obtained.
  - Theoretical result: Statistical behavior of the random resampled estimates $\hat{\boldsymbol{\theta}}^{*(b)}$ compared to $\hat{\boldsymbol{\theta}}$ faithfully represents the statistics of the random estimates $\hat{\boldsymbol{\theta}}$ compared to the true parameter $\boldsymbol{\theta}$.
  - Estimations of accuracy are asymptotically consistent as $B \rightarrow \infty$ (under some technical conditions) (Efron and Tibshirani 1993).

Resampling methods

Empirical distribution of the sample mean bias via the bootstrap:



Resampling methods

Several variations and extensions of the basic bootstrap have been developed over the years:

- **Parametric bootstrap:**
  - Assumes a specific distribution for the data (e.g., Gaussian) (Efron and Tibshirani 1993).
  - Steps:
    1. Assume a parametric distribution for the data.
    2. Estimate the distribution parameters from the observed data.
    3. Generate data using the estimated parametric distribution.
- **Block bootstrap:**
  - Addresses structural dependency in data (Lahiri 1999).
  - Involves resampling blocks of data rather than individual observations.
- **Bag of little bootstraps:**
  - Designed for large datasets with many observations (Kleiner et al. 2014).
  - Steps:
    1. Divide the dataset into smaller subsets.
    2. Apply the bootstrap method to each subset.
    3. Aggregate results to assess estimator quality.

Resampling methods

- **The Jackknife:**
  - Proposed in the mid-1950s by M. Quenouille.
  - Derived for estimating biases and standard errors of sample estimators.
  - Given n observations $\mathbf{x}_1, \ldots, \mathbf{x}_n$, the ith jackknife sample is obtained by removing the ith data point:
  $$\mathbf{x}_1, \ldots, \mathbf{x}_{i-1}, \mathbf{x}_{i+1}, \ldots, \mathbf{x}_n.$$
  - Produces $B = n$ bootstrap samples each with $n - 1$ observations.
  - An approximation to the bootstrap; makes a linear approximation to the bootstrap.
  - Accuracy depends on how "smooth" the estimator is; for highly nonlinear functions, the jackknife can be inefficient.

- **Bagging (bootstrap aggregating):**
  - Method for generating multiple versions of an estimator or predictor via the bootstrap and then using these to get an aggregated version (Breiman 1996; Hastie, Tibshirani, and Friedman 2009).
  - Improves accuracy of the basic estimator or predictor, which typically suffers from sensitivity to the realization of the random data.
  - Mathematically, bagging is a simple average of the bootstrap replicates:

$$\hat{\boldsymbol{\theta}}^{\text{bag}} = \frac{1}{B}\sum_{b=1}^{B}\hat{\boldsymbol{\theta}}^{*(b)}.$$

Outline

Portfolio resampling

- **Optimization problem in portfolio design:**
  - Formulated based on T observations of assets' returns $\boldsymbol{x}_1, \dots, \boldsymbol{x}_T$.
  - Solution is an optimal portfolio $\boldsymbol{w}$.
  - Sensitive to noise in observed data and estimated parameters (mean vector $\hat{\boldsymbol{\mu}}$ and covariance matrix $\hat{\Sigma}$).
- **Resampling techniques:**
  - Utilize statistical methods from the past half-century.
  - Techniques include bootstrap and bagging.
  - Improve portfolio design by reducing sensitivity to data noise.
- **Historical context:**
  - Resampling proposed in the 1990s to assess portfolio accuracy.
  - Naive approach: Use available data to design mean-variance portfolios and obtain the efficient frontier.
  - Issue: High sensitivity to data realization makes the computed efficient frontier unreliable.
- **Resampled efficient frontier:** (Jorion 1992; Michaud and Michaud 1998)
  - Resampling allows computation of a more reliable efficient frontier.
  - Identifies statistically equivalent portfolios.

Portfolio bagging

- **Aggregating portfolios via bagging:** Technique considered in 1998 for portfolio aggregation using a bagging procedure (Michaud and Michaud 1998, 2007, 2008; Scherer 2002).
- **Steps in the bagging procedure:**
  1. **Resample the original data:**
     - Resample the original data $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T)$ B times using the bootstrap method.
     - Estimate B different versions of the mean vector and covariance matrix: $\hat{\boldsymbol{\mu}}^{*(b)}$ and $\hat{\boldsymbol{\Sigma}}^{*(b)}$.
  2. **Solve the optimal portfolio:**
     - Solve the optimal portfolio $\boldsymbol{w}^{*(b)}$ for each bootstrap sample.
  3. **Average the portfolios via bagging:**

$$\boldsymbol{w}^{\text{bag}} = \frac{1}{B}\sum_{b=1}^{B} \boldsymbol{w}^{*(b)}.$$

- **Observations:**
  - The bagging procedure for portfolio aggregation is straightforward.
  - The main bottleneck is the increase in computational cost by a factor of the number of bootstraps B compared to the naive approach.

Portfolio subset resampling

- **Objective:**
  - Reduce the computational cost of the portfolio bagging procedure.
- **Technique:**
  - Sample the asset dimension rather than the observation (temporal) dimension.
  - Similar to the technique used to develop random forests (Ho 1998).
- **Procedure:**
  1. **Randomly select a subset of assets:**
     - Instead of using all N assets, randomly select a subset.
     - Design a portfolio of reduced dimensionality, translating into reduced computational cost.
  2. **Rule-of-thumb for subset size:**
     - Select subsets of $\lceil N^{0.7} \rceil$ or $\lceil N^{0.8} \rceil$ assets.
     - Example: For $N = 50$, the subset sizes would be 16 or 23, respectively.
  3. **Repeat and aggregate:**
     - Repeat the procedure multiple times.
     - Aggregate all computed portfolios.
     - Note: Since the portfolios are of reduced dimensionality, zeros are implicitly used in the elements corresponding to the other dimensions prior to averaging.

Portfolio subset resampling

- **Benefits:**
  - **Reduced computational cost:**
    - Significant reduction in computational cost due to smaller subset size.
  - **Improved parameter estimation:**
    - Better estimation of parameters because the ratio of observations-to-dimensionality is increased.
    - Example: For $T = 252$ daily observations and $N = 50$ assets, the nominal ratio would be $T/N \approx 5$, but the subset resampling ratio is $T/N^{0.7} \approx 16$ or $T/N^{0.8} \approx 11$.
    - Higher ratio leads to more reliable parameter estimates.

- **Numerical experiments:**
  - Confirm that subset resampling is an effective technique in practice.

Portfolio subset bagging

- **Combination of techniques:**
  - Combine random subset resampling along the asset domain with the bootstrap along the temporal domain (Shen et al. 2019).
- **Procedure:**
  - Each bootstrap sample contains only a subset of the N assets.
  - This combination leverages the benefits of both resampling techniques to create more robust and computationally efficient portfolios.
- **Summary:**
  - **Computational savings and improved robustness:**
    - Portfolio subset resampling and portfolio subset bagging achieve significant computational savings.
    - These techniques improve the robustness and reliability of portfolios.
  - **Mitigating sensitivity to noise:**
    - Help mitigate sensitivity to noise and variability in the data.
    - Lead to more stable investment strategies.

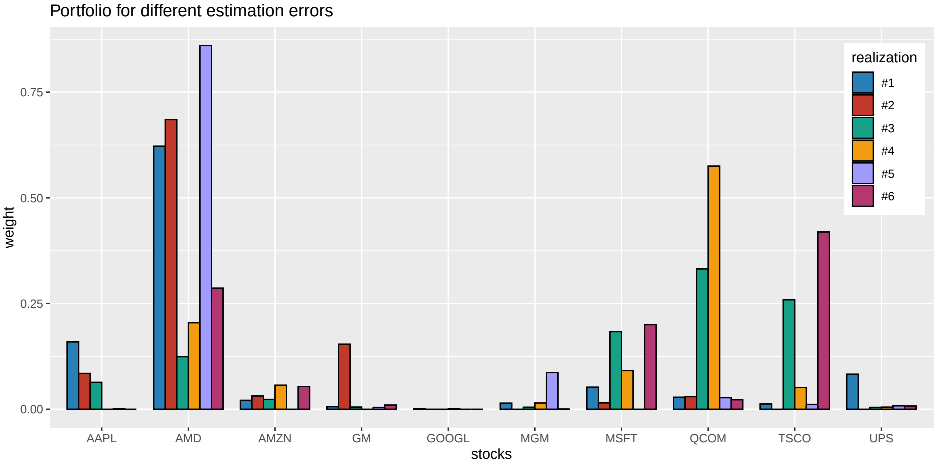
Outline

Numerical experiments

- **Sensitivity analysis:**
  - **Naive mean-variance portfolio (MVP):** Extremely sensitive to estimation errors as previously seen.
  - **Robust portfolio optimization:** Less sensitive to estimation errors as shown next as previously seen.
  - **Bagged portfolios:** More stable and less sensitive with $B = 200$ bootstrap samples.
- **Performance assessment:**
  - Assess performance of resampled portfolios (bagging, subset resampling, subset bagging) compared with naive design.
  - Backtests conducted for 50 randomly chosen stocks from the S&P 500 during 2017-2020.
- **Empirical distribution analysis:**
  - Empirical distribution of the achieved mean-variance objective and the Sharpe ratio, calculated over 1,000 Monte Carlo noisy observations.
  - Resampled portfolios are more stable, avoid extreme bad realizations (although the naive design can be superior in some cases).
  - Resampled portfolios seem to be superior in Sharpe ratio and drawdown.

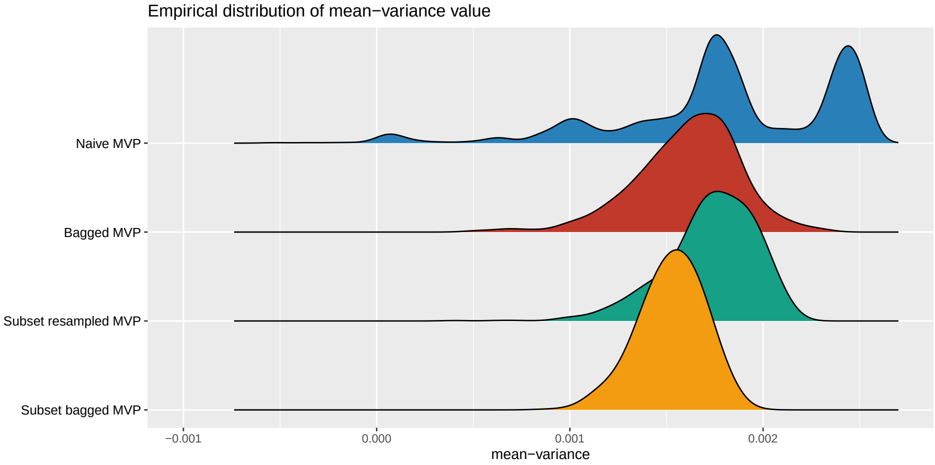
Numerical experiments

Sensitivity of the bagged mean–variance portfolio:

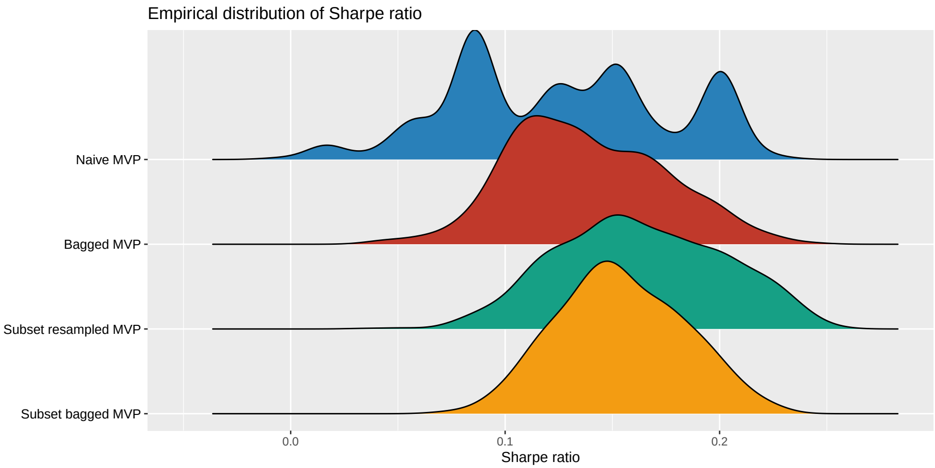


Numerical experiments

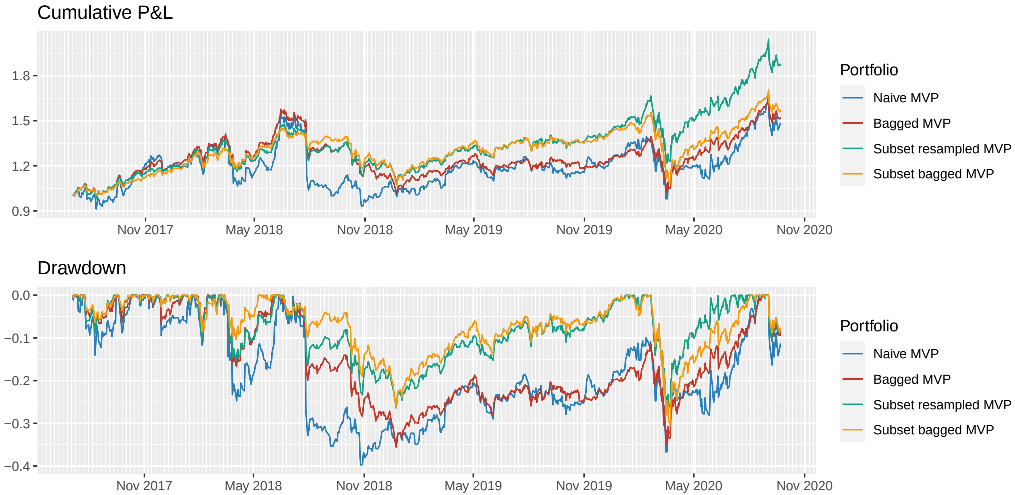
Empirical performance distribution of naive versus resampled mean–variance portfolios:



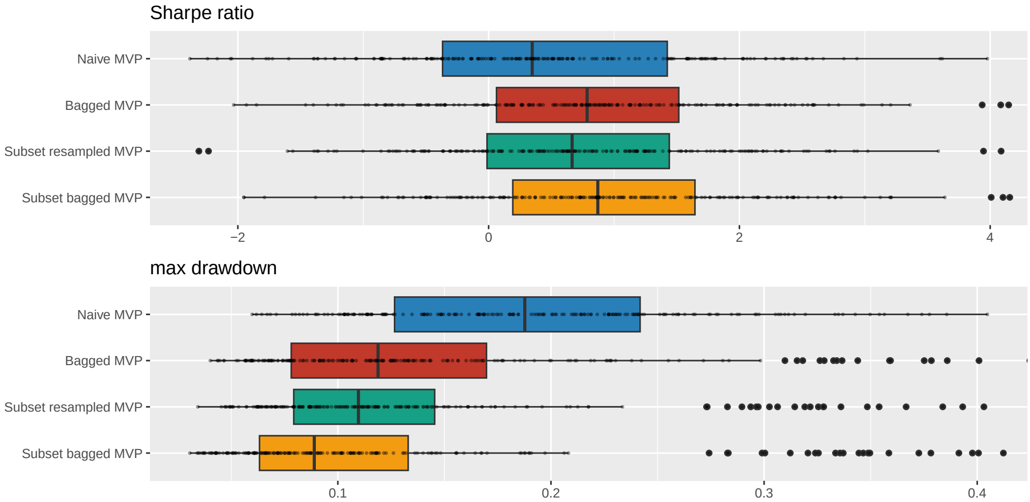
Empirical performance distribution of naive versus resampled mean–variance portfolios:



Backtest of naive versus resampled mean–variance portfolios:



Multiple backtests of naive versus resampled mean–variance portfolios:



Outline

Summary

- Ideally, a portfolio optimization solution should meet the desired objective within constraints.

- In practice, it often fails due to reliance on estimated parameters like the mean vector and covariance matrix, which contain errors from noisy and limited data.

- Ignoring these estimation errors can lead to disastrous results, earning the term "estimation-error maximizers" for such portfolio problems.

- Effective approaches to mitigate naive solutions include:
  - **Robust portfolios**: Incorporate parameter errors using robust optimization, a well-developed method in portfolio optimization.
  - **Resampled portfolios**: Use bootstrapping and resampling to aggregate multiple naive solutions into a more stable and reliable portfolio.

References I

Ben-Tal, A., L. El Ghaoui, and A. Nemirovski. 2009. *Robust Optimization*. Princeton University Press.

Ben-Tal, A., and A. Nemirovski. 2008. "Selected Topics in Robust Convex Optimization." *Mathematical Programming* 112 (1).

Bertsekas, D. P. 1999. *Nonlinear Programming*. Athena Scientific.

Best, M. J., and R. R. Grauer. 1991. "On the Sensitivity of Mean-Variance-Efficient Portfolios to Changes in Asset Means: Some Analytical and Computational Results." *Review of Financial Studies* 4 (2): 315–42.

Birge, J. R., and F. V. Louveaux. 2011. *Introduction to Stochastic Programming*. Springer.

Breiman, L. 1996. "Bagging Predictors." *Machine Learning* 24: 123–40.

Chopra, V., and W. Ziemba. 1993. "The Effect of Errors in Means, Variances and Covariances on Optimal Portfolio Choice." *Journal of Portfolio Management* 19 (2): 6–11.

D. Bertsimas, D. B. Brown, and C. Caramanis. 2011. "Theory and Applications of Robust Optimization." *SIAM Review* 53 (3).

Efron, B. 1979. "Bootstrap Methods: Another Look at the Jackknife." *The Annals of Statistics* 7 (1): 1–26.

References II

Efron, B., and R. J. Tibshirani. 1993. *An Introduction to the Bootstrap*. Springer.

Hastie, T., R. Tibshirani, and J. Friedman. 2009. *The Elements of Statistical Learning*. 2nd ed. Springer.

Ho, T. K. 1998. "The Random Subspace Method for Constructing Decision Forests." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20 (8): 832–44.

Jorion, P. 1992. "Portfolio Optimization in Practice." *Financial Analysts Journal* 48 (1): 68–74.

Kleiner, A., A. Talwalkar, P. Sarkar, and M. I. Jordan. 2014. "A Scalable Bootstrap for Massive Data." *Journal of the Royal Statistical Society. Series B: Statistical Methodology* 76 (4): 795–816.

Lahiri, S. N. 1999. "Theoretical Comparisons of Block Bootstrap Methods." *The Annals of Statistics* 27 (1): 386–404.

Lobo, M. S. 2000. "Robust and Convex Optimization with Applications in Finance." PhD thesis, Stanford University.

Michaud, R. O. 1989. "The Markowitz Optimization Enigma: Is 'Optimized' Optimal?" *Financial Analysts Journal* 45 (1): 31–42.

References III

Michaud, R. O., and R. O. Michaud. 1998. *Efficient Asset Management: A Practical Guide to Stock Portfolio Optimization and Asset Allocation*. Harvard Business School Press.

———. 2007. "Estimation Error and Portfolio Optimization: A Resampling Solution." *Journal of Investment and Management* 6 (1): 8–28.

———. 2008. *Efficient Asset Management: A Practical Guide to Stock Portfolio Optimization and Asset Allocation*. 2nd ed. Oxford University Press.

Palomar, D. P. 2024. *Portfolio Optimization: Theory and Application*. Cambridge University Press.

Prekopa, A. 1995. *Stochastic Programming*. Kluwer Academic Publishers.

Ruszczynski, A. P., and A. Shapiro. 2003. *Stochastic Programming*. Elsevier.

Scherer, B. 2002. "Portfolio Resampling: Review and Critique." *Financial Analysts Journal* 58 (6): 98–109.

Shen, W., B. Wang, J. Pu, and J. Wang. 2019. "The Kelly Growth Optimal Portfolio with Ensemble Learning." In *Proceedings of the AAAI Conference on Artificial Intelligence*, 33:1134–41. 1.

Tütüncü, R. H., and M. Koenig. 2004. "Robust Asset Allocation." *Annals of Operations Research* 132 (1): 157–87.

References IV

Zhou, R., and D. P. Palomar. 2020. "Understanding the Quintile Portfolio." *IEEE Transactions on Signal Processing* 68 (July): 4030–40.