Portfolio Optimization

Prof. Daniel P. Palomar

ELEC5470/IEDA6100A - Convex Optimization
The Hong Kong University of Science and Technology (HKUST)
Fall 2020-21

Outline

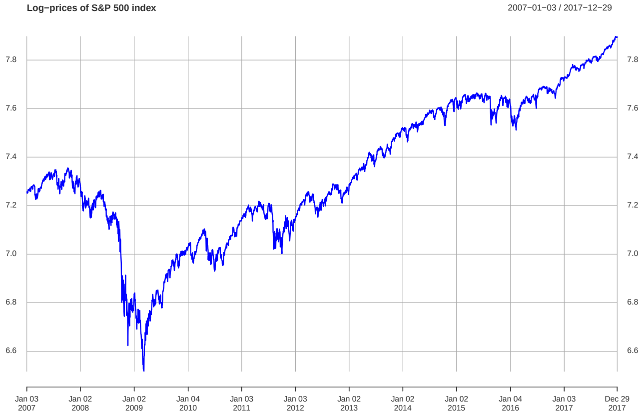
Outline

1. **Primer on Financial Data**
2. Modeling the Returns
3. Portfolio Basics
4. Heuristic Portfolios
5. Markowitz's Modern Portfolio Theory (MPT)
   - Mean-variance portfolio (MVP)
   - Global minimum variance portfolio (GMVP)
   - Maximum Sharpe ratio portfolio (MSRP)

Asset log-prices

- Let p_t be the price of an asset at (discrete) time index t.
- The fundamental model is based on modeling the log-prices $y_t \triangleq \log p_t$ as a random walk:

$$y_t = \mu + y_{t-1} + \epsilon_t$$



Asset returns

- For stocks, returns are used for the modeling since they are "stationary" (as opposed to the previous random walk).
- **Simple return** (a.k.a. linear or net return) is

$$R_t \triangleq \frac{p_t - p_{t-1}}{p_{t-1}} = \frac{p_t}{p_{t-1}} - 1.$$

- **Log-return** (a.k.a. continuously compounded return) is

$$r_t \triangleq y_t - y_{t-1} = \log \frac{p_t}{p_{t-1}} = \log\left(1 + R_t\right).$$

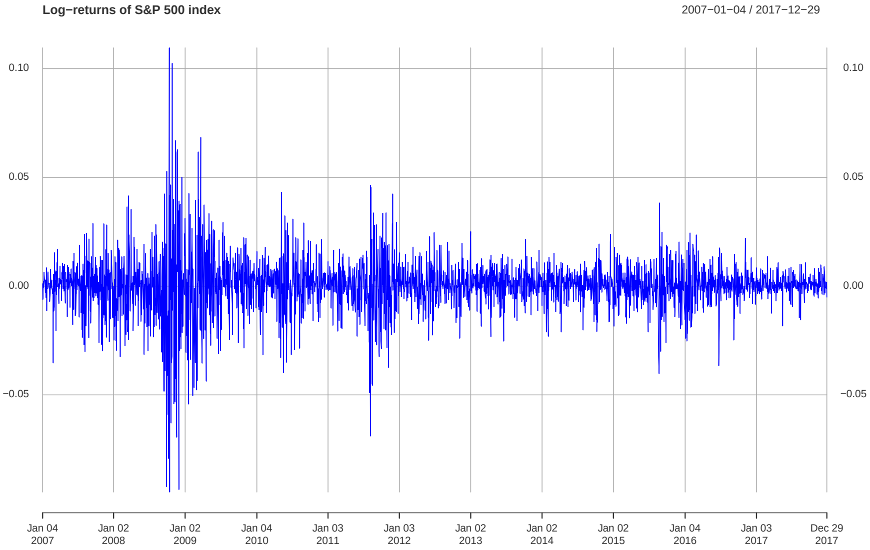
- Observe that the **log-return is "stationary"**:

$$r_t = y_t - y_{t-1} = \mu + \epsilon_t$$

- Note that $r_t \approx R_t$ when R_t is small (i.e., the changes in p_t are small) (Ruppert 2010)[1].

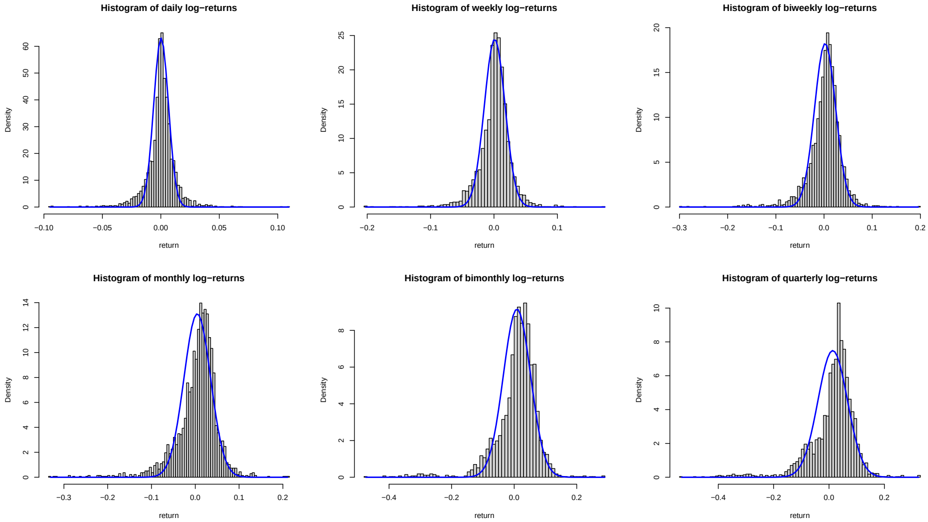
[1] D. Ruppert, *Statistics and Data Analysis for Financial Engineering*. Springer, 2010.

S&P 500 index - Log-returns



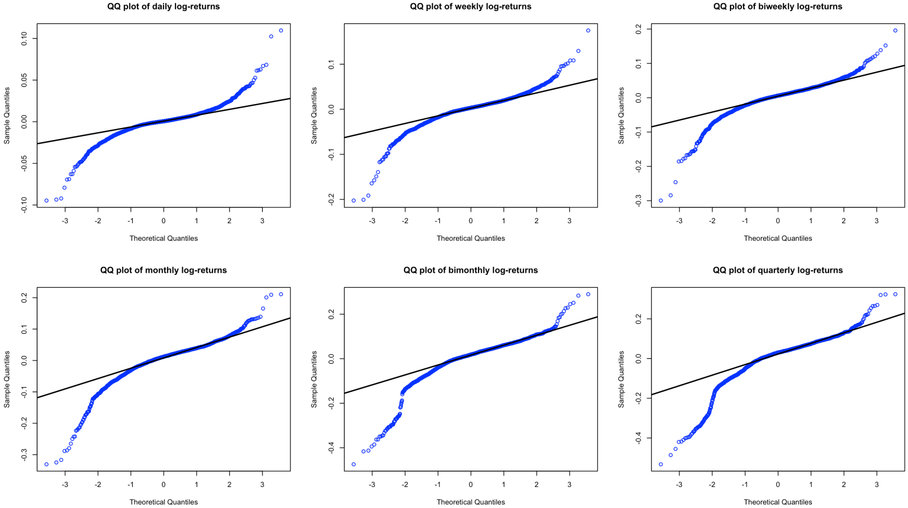
Non-Gaussianity and asymmetry

Histograms of S&P 500 log-returns:



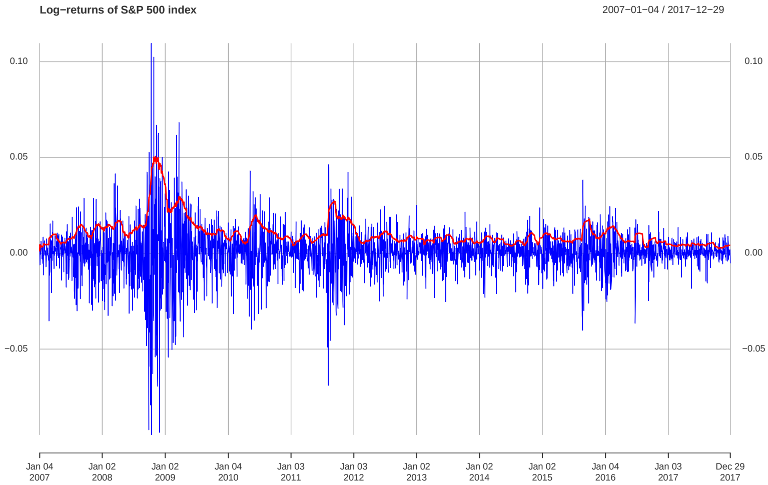
Heavy-tailness

QQ plots of S&P 500 log-returns:



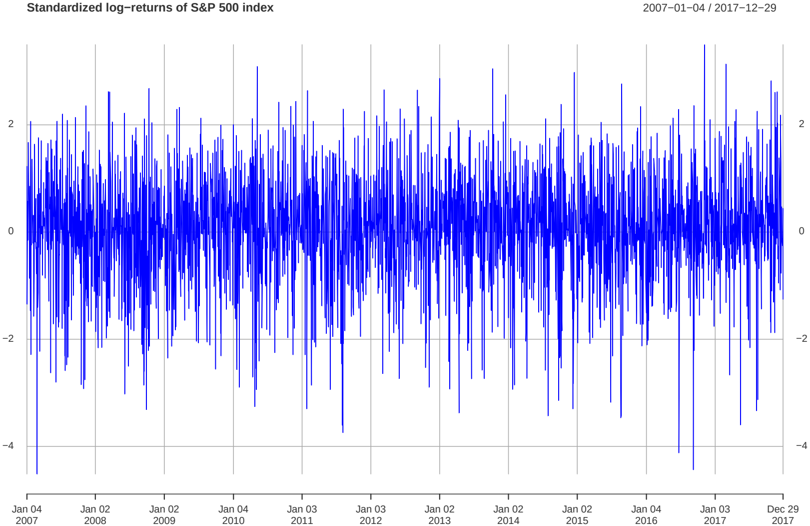
Volatility clustering

S&P 500 log-returns:



Volatility clustering removed

Standardized S&P 500 log-returns:



Frequency of data

- **Low frequency (weekly, monthly)**: Gaussian distributions seems to fit reality after correcting for volatility clustering (except for the asymmetry), but the nonstationarity is a big issue

- **Medium frequency (daily)**: definitely heavy tails even after correcting for volatility clustering, as well as asymmetry

- **High frequency (intraday, 30min, 5min, tick-data)**: below 5min the noise microstructure starts to reveal itself

Outline

Returns of the universe

- In practice, we don't just deal with one asset but with a whole universe of N assets.
- We denote the log-returns of the N assets at time t with the vector $\mathbf{r}_t \in \mathbb{R}^N$.
- The time index t can denote any arbitrary period such as days, weeks, months, 5-min intervals, etc.
- $\mathcal{F}_{t-1}$ denotes the previous historical data.
- Econometrics aims at modeling $\mathbf{r}_t$ conditional on $\mathcal{F}_{t-1}$.
- $\mathbf{r}_t$ is a multivariate stochastic process with **conditional mean and covariance matrix** denoted as (Feng and Palomar 2016)[2]

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

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

i.i.d. model

- For simplicity we will assume that $\mathbf{r}_t$ follows an i.i.d. distribution (which is not very innacurate in general).
- That is, both the **conditional mean and conditional covariance are constant**:

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

- Very simple model, however, it is one of the most fundamental assumptions for many important works, e.g., the Nobel prize-winning Markowitz's portfolio theory (Markowitz 1952)[3].

[3] H. Markowitz, "Portfolio selection," *J. Financ.*, vol. 7, no. 1, pp. 77–91, 1952.

Factor models

- Factor models are special cases of the i.i.d. model with the covariance matrix being decomposed into two parts: **low dimensional factors** and **marginal noise**.
- The factor model is

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

  where
  - $\boldsymbol{\alpha}$ denotes a constant vector
  - $\mathbf{f}_t \in \mathbb{R}^K$ with $K \ll N$ is a vector of a few factors that are responsible for most of the randomness in the market
  - $\mathbf{B} \in \mathbb{R}^{N \times K}$ denotes how the low dimensional factors affect the higher dimensional market assets
  - $\mathbf{w}_t$ is a white noise residual vector that has only a marginal effect.
- The factors can be **explicit or implicit**.
- **Widely used by practitioners** (they buy factors at a high premium).
- Connections with **Principal Component Analysis (PCA)** (Jolliffe 2002)[4].

[4] I. Jolliffe, *Principal Component Analysis*. Springer-Verlag, 2002.

Time-series models

- The previous models are i.i.d., but there are hundreds of other models attempting to capture the time correlation or time structure of the returns, as well as the volatility clustering or heteroskedasticity.
- To capture the **time correlation** we have mean models: VAR, VMA, VARMA, VARIMA, VECM, etc.
- To capture the **volatility clustering** we have covariance models: ARCH, GARCH, VEC, DVEC, BEKK, CCC, DCC, etc.
- Standard textbook references (Lütkepohl 2007; Tsay 2005, 2013):

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

  R. S. Tsay. Multivariate Time Series Analysis: With R and Financial Applications. John Wiley & Sons, 2013.
- Simple introductory reference (Feng and Palomar 2016):

  Y. Feng and D. P. Palomar. A Signal Processing Perspective on Financial Engineering. Foundations and Trends in Signal Processing, Now Publishers, 2016.

Sample estimates

- Consider the i.i.d. model:

$$\mathbf{r}_t = \boldsymbol{\mu} + \mathbf{w}_t,$$

where $\boldsymbol{\mu} \in \mathbb{R}^N$ is the mean and $\mathbf{w}_t \in \mathbb{R}^N$ is an i.i.d. process with zero mean and constant covariance matrix $\boldsymbol{\Sigma}$.

- The mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ have to be estimated in practice based on T observations.
- The simplest estimators are the sample estimators:
  - **sample mean**: $\hat{\boldsymbol{\mu}} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{r}_t$
  - **sample covariance matrix**: $\hat{\boldsymbol{\Sigma}} = \frac{1}{T-1}\sum_{t=1}^{T}(\mathbf{r}_t - \hat{\boldsymbol{\mu}})(\mathbf{r}_t - \hat{\boldsymbol{\mu}})^T$.
    Note that the factor $1/(T-1)$ is used instead of $1/T$ to get an unbiased estimator (asymptotically for $T \to \infty$ they coincide).
- Many more sophisticated estimators exist, namely: shrinkage estimators, Black-Litterman estimators, etc.

Least-Square (LS) estimator

- Minimize the least-square error in the T observed i.i.d. samples:

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

- The optimal solution is the sample mean:

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

- The sample covariance of the residuals $\hat{\mathbf{w}}_t = \mathbf{r}_t - \hat{\boldsymbol{\mu}}$ is the sample covariance matrix:

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

Maximum Likelihood Estimator (MLE)

- Assume $\mathbf{r}_t$ are i.i.d. and follow a Gaussian distribution:

$$f(\mathbf{r}) = \frac{1}{\sqrt{(2\pi)^N |\boldsymbol{\Sigma}|}} e^{-\frac{1}{2}(\mathbf{r}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{r}-\boldsymbol{\mu})}.$$

  where
  - $\boldsymbol{\mu} \in \mathbb{R}^N$ is a mean vector that gives the location
  - $\boldsymbol{\Sigma} \in \mathbb{R}^{N \times N}$ is a positive definite covariance matrix that defines the shape.

MLE

- Given the T i.i.d. samples $\mathbf{r}_t$, $t = 1, \ldots, T$, the negative log-likelihood function is

$$\begin{aligned}
\ell(\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= -\log \prod_{t=1}^{T} f(\mathbf{r}_t) \\
&= \frac{T}{2} \log |\boldsymbol{\Sigma}| + \sum_{t=1}^{T} \frac{1}{2} (\mathbf{r}_t - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{r}_t - \boldsymbol{\mu}) + \text{const.}
\end{aligned}$$

- Setting the derivative of $\ell(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ w.r.t. $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}^{-1}$ to zeros and solving the equations yield:

$$\begin{aligned}
\hat{\boldsymbol{\mu}} &= \frac{1}{T} \sum_{t=1}^{T} \mathbf{r}_t \\
\hat{\boldsymbol{\Sigma}} &= \frac{1}{T} \sum_{t=1}^{T} (\mathbf{r}_t - \hat{\boldsymbol{\mu}}) (\mathbf{r}_t - \hat{\boldsymbol{\mu}})^T.
\end{aligned}$$

Outline

Portfolio return

- Suppose the capital budget is B dollars.
- The portfolio $\mathbf{w} \in \mathbb{R}^N$ denotes the normalized dollar weights of the N assets such that $\mathbf{1}^T\mathbf{w} = 1$ (so $B\mathbf{w}$ denotes dollars invested in the assets).
- For each asset i, the initial wealth is Bw_i and the end wealth is

$$Bw_i \left(p_{i,t}/p_{i,t-1}\right) = Bw_i \left(R_{it} + 1\right).$$

- Then the **portfolio return** is

$$R_t^p = \frac{\sum_{i=1}^N Bw_i \left(R_{it} + 1\right) - B}{B} = \sum_{i=1}^N w_i R_{it} \approx \sum_{i=1}^N w_i r_{it} = \mathbf{w}^T \mathbf{r}_t$$

- The portfolio expected return and variance are $\mathbf{w}^T\boldsymbol{\mu}$ and $\mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w}$, respectively.

Performance measures

- **Expected return**: $\mathbf{w}^T\boldsymbol{\mu}$
- **Volatility**: $\sqrt{\mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w}}$
- **Sharpe Ratio (SR)**: expected excess return per unit of risk

$$\mathrm{SR} = \frac{\mathbf{w}^T\boldsymbol{\mu} - r_f}{\sqrt{\mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w}}}$$

  where r_f is the risk-free rate (e.g., interest rate on a three-month U.S. Treasury bill).
- **Information Ratio (IR)**: SR with respect to a benchmark (e.g., the market index): $\mathrm{IR} = \frac{\mathrm{E}\left[\mathbf{w}^T\mathbf{r}_t - r_{b,t}\right]}{\sqrt{\mathrm{Var}\left[\mathbf{w}^T\mathbf{r}_t - r_{b,t}\right]}}$.
- **Drawdown**: decline from a historical peak of the cumulative profit $X(t)$:

$$D(T) = \max_{t\in[0,T]} X(t) - X(T)$$

- **VaR (Value at Risk)**: quantile of the loss.
- **ES (Expected Shortfall) or CVaR (Conditional Value at Risk)**: expected value of the loss above some quantile.

Practical constraints

- **Capital budget constraint**:
$$\mathbf{1}^T\mathbf{w} = 1.$$
- **Long-only constraint**:
$$\mathbf{w} \geq 0.$$
- **Dollar-neutral or self-financing constraint**:
$$\mathbf{1}^T\mathbf{w} = 0.$$
- **Holding constraint**:
$$\mathbf{l} \leq \mathbf{w} \leq \mathbf{u}$$
where $\mathbf{l} \in \mathbb{R}^N$ and $\mathbf{u} \in \mathbb{R}^N$ are lower and upper bounds of the asset positions, respectively.

Practical constraints

- **Leverage constraint**:
$$\|\mathbf{w}\|_1 \leq L.$$
- **Cardinality constraint**:
$$\|\mathbf{w}\|_0 \leq K.$$
- **Turnover constraint**:
$$\|\mathbf{w} - \mathbf{w}_0\|_1 \leq u$$
where $\mathbf{w}_0$ is the currently held portfolio.
- **Market-neutral constraint**:
$$\boldsymbol{\beta}^T \mathbf{w} = 0.$$

Outline

Heuristic portfolios

- Heuristic portfolios are not formally derived from a sound mathematical foundation. Instead, they are intuitive and based on common sense.

- We will explore the following simple and heuristic portfolios:
  - **Buy & Hold (B&H)**
  - Buy & Rebalance
  - **equally weighted portfolio (EWP)** or $1/N$ **portfolio**
  - **quintile portfolio**
  - **global maximum return portfolio (GMRP)**.

Buy & Hold (B&H)

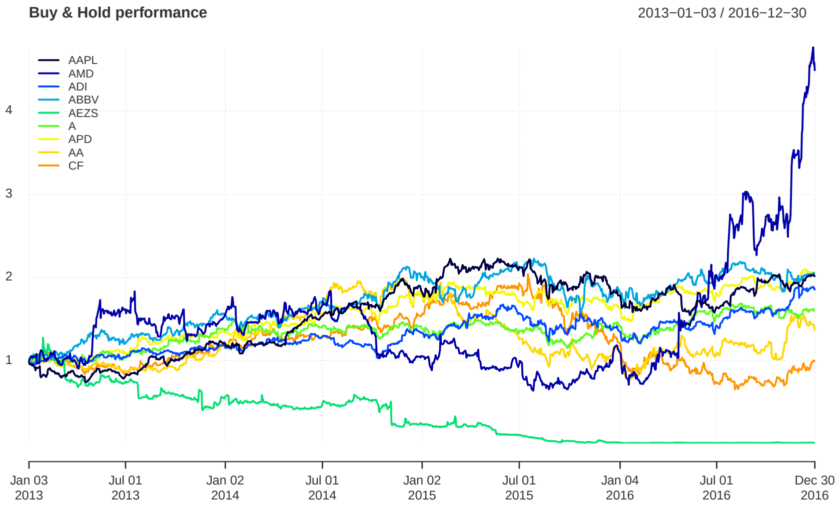
- The simplest investment strategy consists of selecting just one asset, allocating the whole budget B to it:
  - **Buy & Hold (B&H)**: chooses one asset and sticks to it forever.
  - **Buy & Rebalance**: chooses one asset but it reevaluates that choice regularly.
- The belief behind such investment is that the asset will increase gradually in value over the investment period.
- There is no diversification in this strategy.
- One can use different methods (like fundamental analysis or technical analysis) to make the choice.
- Mathematically, it can be expressed as

$$\mathbf{w} = \mathbf{e}_i$$

where $\mathbf{e}_i$ denotes the canonical vector with a 1 on the ith position and 0 elsewhere.

Buy & Hold (B&H)

Cumulative PnL of 9 possible B&H for 9 assets:



Equally weighted portfolio (EWP) or $1/N$ portfolio

- One of the most important goals of quantitative portfolio management is to realize the goal of diversification across different assets in a portfolio.
- A simple way to achieve diversification is by allocating the capital equally across all the assets.
- This strategy is called **equally weighted portfolio (EWP)**, $1/N$ **portfolio**, uniform portfolio, or maximum deconcentration portfolio:

$$\mathbf{w} = \frac{1}{N}\mathbf{1}.$$

- It has been called "Talmudic rule" (Duchin and Levy 2009) since the Babylonian Talmud recommended this strategy approximately 1,500 years ago: "A man should always place his money, one third in land, a third in merchandise, and keep a third in hand."
- It has gained much interest due to superior historical performance and the emergence of several equally weighted ETFs (DeMiguel et al. 2009). For example, Standard & Poor's has developed many S&P 500 equal weighted indices.

Quintile Portfolio

- The quintile portfolio is widely used by practitioners.
- Two types: **long-only quintile portfolio** and **long-short quintile portfolio**.
- Basic idea: 1) rank the N stocks according to some criterion, 2) divide them into five parts, and 3) long the top part (and possibly short the bottom part).
- One can rank the stocks in a multitude of ways (typically based on expensive factors that investment funds buy at a premium price).
- If we restric to price data, three common possible **rankings** are according to:
  1. $\boldsymbol{\mu}$
  2. $\dfrac{\boldsymbol{\mu}}{\mathrm{diag}(\boldsymbol{\Sigma})}$
  3. $\dfrac{\boldsymbol{\mu}}{\sqrt{\mathrm{diag}(\boldsymbol{\Sigma})}}$

Global maximum return portfolio (GMRP)

- Another simple way to make an investment from the N assets is to only invest on the one with the highest return.
- Mathematically, the global maximum return portfolio (GMRP) is formulated as

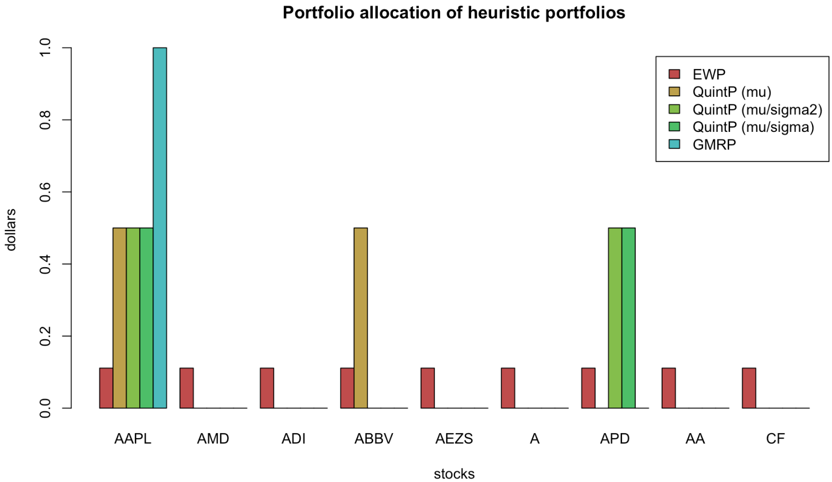
$$\begin{array}{ll} \underset{\mathbf{w}}{\text{maximize}} & \mathbf{w}^T \boldsymbol{\mu} \\ \text{subject to} & \mathbf{1}^T \mathbf{w} = 1, \quad \mathbf{w} \geq \mathbf{0}. \end{array}$$

- This problem is convex and can be optimally solved, but of course the solution is trivial: to allocate all the budget to the asset with maximum return.
- However, this seemingly good portfolio lacks diversification and performs poorly because past performance is not a guarantee of future performance.
- In addition, the estimation of $\boldsymbol{\mu}$ is extremely noisy in practice (Chopra and Ziemba 1993)[5].

[5] V. Chopra and W. Ziemba, "The effect of errors in means, variances and covariances on optimal portfolio choice," *Journal of Portfolio Management*, 1993.

Example

Dollar allocation for EWP, QuintP, and GMRP:



Outline

Risk control

- In finance, the expected return $\mathbf{w}^T\boldsymbol{\mu}$ is very relevant as it quantifies the average benefit.
- However, in practice, the **average performance is not enough** to characterize an investment and one needs to control the **probability of going bankrupt**.
- Risk measures control how risky an investment strategy is.
- The most basic **measure of risk** is given by the **variance** (Markowitz 1952)[6]: a higher variance means that there are large peaks in the distribution which may cause a big loss.
- There are **more sophisticated risk measures** such as **downside risk**, **VaR**, **ES**, etc.

[6] H. Markowitz, "Portfolio selection," *J. Financ.*, vol. 7, no. 1, pp. 77–91, 1952.

Outline

Mean-variance tradeoff

- The **mean return** $\mathbf{w}^T\boldsymbol{\mu}$ and the **variance** (risk) $\mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w}$ (equivalently, the standard deviation or **volatility** $\sqrt{\mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w}}$) constitute two important performance measures.

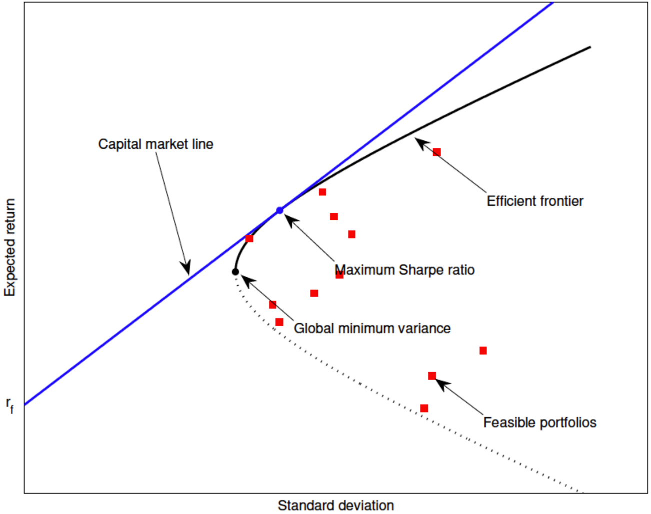
- Usually, the higher the mean return the higher the variance and vice-versa.

- Thus, we are faced with two objectives to be optimized: it is a **multi-objective optimization** problem.

- They define a fundamental **mean-variance tradeoff** curve (Pareto curve).

- The choice of a specific point in this tradeoff curve depends on how agressive or risk-averse the investor is.

Mean-variance tradeoff



Markowitz's mean-variance portfolio (1952)

- The idea of **Markowitz's mean-variance portfolio (MVP)** (Markowitz 1952)[7] is to find a trade-off between the expected return $\mathbf{w}^T\boldsymbol{\mu}$ and the risk of the portfolio measured by the variance $\mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w}$:

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{maximize}} & \mathbf{w}^T\boldsymbol{\mu} - \lambda\mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w} \\ \text{subject to} & \mathbf{1}^T\mathbf{w} = 1 \end{array}$$

  where $\mathbf{w}^T\mathbf{1} = 1$ is the capital budget constraint and λ is a parameter that controls how risk-averse the investor is.

- This is also referred to as Modern Portfolio Theory (MPT).

- This is a convex quadratic problem (QP) with only one linear constraint which admits a closed-form solution:

$$\mathbf{w}_{\text{MVP}} = \frac{1}{2\lambda}\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu} + \nu\mathbf{1}\right),$$

  where ν is the optimal dual variable $\nu = \frac{2\lambda - \mathbf{1}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}}$.

[7] H. Markowitz, "Portfolio selection," *J. Financ.*, vol. 7, no. 1, pp. 77–91, 1952.

Markowitz's mean-variance portfolio (1952)

- There are two alternative obvious reformulations for Markowitz's portfolio.
- **Maximization of mean return**:

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{maximize}} & \mathbf{w}^T\boldsymbol{\mu} \\ \text{subject to} & \mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w} \leq \alpha \\ & \mathbf{1}^T\mathbf{w} = 1. \end{array}$$

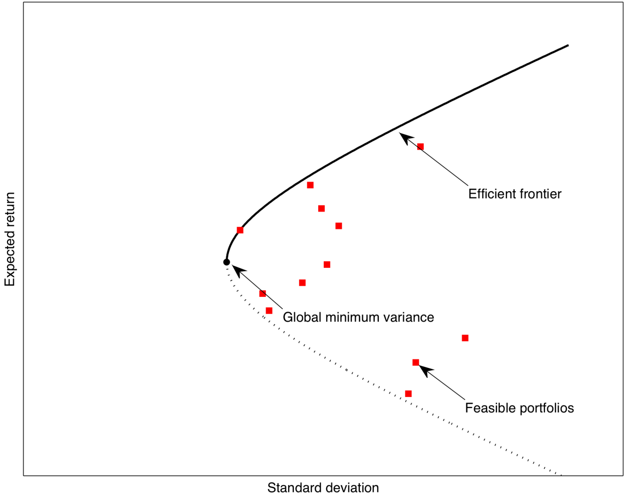
- **Minimization of risk**:

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{minimize}} & \mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w} \\ \text{subject to} & \mathbf{w}^T\boldsymbol{\mu} \geq \beta \\ & \mathbf{1}^T\mathbf{w} = 1. \end{array}$$

- The three formulations give different points on the Pareto optimal curve.
- They all require choosing one parameter (α, β, or λ).
- By sweeping over this parameter, one can recover the whole Pareto optimal curve.

Efficient frontier

The previous three problems result in the same mean-variance trade-off curve (Pareto curve):



Markowitz's portfolio with practical constraints

- A general Markowitz's portfolio with practical constraints could be:

$$
\begin{array}{lll}
\underset{\mathbf{w}}{\text{maximize}} & \mathbf{w}^T\boldsymbol{\mu} - \lambda\mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w} & \\
\text{subject to} & \mathbf{w}^T\mathbf{1} = 1 & \text{budget} \\
& \mathbf{w} \geq \mathbf{0} & \text{no shorting} \\
& \|\mathbf{w}\|_1 \leq \gamma & \text{leverage} \\
& \|\mathbf{w} - \mathbf{w}_0\|_1 \leq \tau & \text{turnover} \\
& \|\mathbf{w}\|_\infty \leq u & \text{max position} \\
& \|\mathbf{w}\|_0 \leq K & \text{sparsity}
\end{array}
$$

  where:
  - $\gamma \geq 1$ controls the amount of shorting and leveraging
  - $\tau > 0$ controls the turnover (to control the transaction costs in the rebalancing)
  - u limits the position in each stock
  - K controls the cardinality of the portfolio (to select a small set of stocks from the universe).
- Without the sparsity constraint, the problem can be rewritten as a QP.

Markowitz's MVP as a regression

- Interestingly, the MVP formulation can be interpreted as a regression.
- The key observation is that the variance of the portfolio can be seen as an ℓ_2-norm error term:

$$\begin{aligned} \mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w} &= \mathbf{w}^T\mathsf{E}\left[(\mathbf{r}_t-\boldsymbol{\mu})(\mathbf{r}_t-\boldsymbol{\mu})^T\right]\mathbf{w} \\ &= \mathsf{E}\left[(\mathbf{w}^T(\mathbf{r}_t-\boldsymbol{\mu}))^2\right] \\ &= \mathsf{E}\left[(\mathbf{w}^T\mathbf{r}_t-\rho)^2\right] \end{aligned}$$

  where $\rho = \mathbf{w}^T\boldsymbol{\mu}$.
- The sample approximation of the expected value is

$$\mathsf{E}\left[(\mathbf{w}^T\mathbf{r}_t-\rho)^2\right] \approx \frac{1}{T}\sum_{t=1}^{T}(\mathbf{w}^T\mathbf{r}_t-\rho)^2 = \frac{1}{T}\|\mathbf{R}\mathbf{w}-\rho\mathbf{1}\|^2$$

  where $\mathbf{R} \triangleq [\mathbf{r}_1,\ldots,\mathbf{r}_T]^T$.

Markowitz's MVP as a regression

- Let's start by rewritting the MVP as a minimization:

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{minimize}} & \mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w} - \frac{1}{\lambda}\mathbf{w}^T\boldsymbol{\mu} \\ \text{subject to} & \mathbf{1}^T\mathbf{w} = 1 \end{array}$$

- Thus, we can finally write the MVP as the regression

$$\begin{array}{ll} \underset{\mathbf{w},\rho}{\text{minimize}} & \frac{1}{T}\|\mathbf{R}\mathbf{w} - \rho\mathbf{1}\|^2 - \frac{1}{\lambda}\rho \\ \text{subject to} & \rho = \mathbf{w}^T\boldsymbol{\mu} \\ & \mathbf{1}^T\mathbf{w} = 1 \end{array}$$

- If instead of the mean-variance formulation, we fix the expected return to some value and minimize the variance, then ρ simply becomes a fixed value rather than a variable.
- Intuitively, this reformulation is trying to achieve the expected return ρ with minimum variance in the ℓ_2-norm sense.

Outline

Global minimum variance portfolio (GMVP)

- Recall the risk minimization formulation:

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{minimize}} & \mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w} \\ \text{subject to} & \mathbf{w}^T\boldsymbol{\mu} \geq \beta \\ & \mathbf{1}^T\mathbf{w} = 1. \end{array}$$

- The global minimum variance portfolio (GMVP) ignores the expected return and focuses on the risk only:

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{minimize}} & \mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w} \\ \text{subject to} & \mathbf{1}^T\mathbf{w} = 1. \end{array}$$

- It is a simple convex QP with solution

$$\mathbf{w}_{\text{GMVP}} = \frac{1}{\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}}\boldsymbol{\Sigma}^{-1}\mathbf{1}.$$

- It is widely used in academic papers for simplicity of evaluation and comparison of different estimators of the covariance matrix $\boldsymbol{\Sigma}$ (while ignoring the estimation of $\boldsymbol{\mu}$).

GMVP with leverage constraints

- The GMVP is typically considered with no-short constraints $\mathbf{w} \geq \mathbf{0}$:

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

- However, if short-selling is allowed, one needs to limit the amount of leverage to avoid impractical solutions with very large positive and negative weights that cancel out.
- A sensible GMVP formulation with leverage is

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{minimize}} & \mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w} \\ \text{subject to} & \mathbf{1}^T\mathbf{w} = 1, \;\; \|\mathbf{w}\|_1 \leq \gamma \end{array}$$

  where $\gamma \geq 1$ is a parameter that **controls the amount of leverage**:
  - $\gamma = 1$ means no shorting (so equivalent to $\mathbf{w} \geq \mathbf{0}$)
  - $\gamma > 1$ allows some shorting as well as leverage in the longs, e.g., $\gamma = 1.5$ would allow the portfolio $\mathbf{w} = (1.25, -0.25)$.

Outline

Maximum Sharpe ratio portfolio (MSRP)

- Markowitz's mean-variance framework provides portfolios along the Pareto-optimal frontier and the choice depends on the risk-aversion of the investor.
- But typically one measures an investment with the Sharpe ratio: only one portfolio on the Pareto-optimal frontier achieves the maximum Sharpe ratio.
- Precisely, Sharpe (1966)[8] first proposed the **maximization of the Sharpe ratio**:

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{maximize}} & \dfrac{\mathbf{w}^T\boldsymbol{\mu} - r_f}{\sqrt{\mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w}}} \\ \text{subject to} & \mathbf{1}^T\mathbf{w} = 1, \quad (\mathbf{w} \geq \mathbf{0}) \end{array}$$

  where r_f is the return of a risk-free asset.
- However, this problem is **not convex!**
- This problem belong to the class of **fractional programming (FP)** with many methods available for its resolution.

[8] W. F. Sharpe, "Mutual fund performance," *The Journal of Business*, vol. 39, no. 1, pp. 119–138, 1966.

Interlude: Fractional Programming (FP)

- **Fractional programming (FP)** is a family of optimization problems involving ratios.
- Its history can be traced back to a paper on economic expansion by von Neumann (1937).[9]
- It has since inspired the studies in economics, management science, optics, information theory, communication systems, graph theory, computer science, etc.
- Given functions $f(\mathbf{x}) \geq 0$ and $g(\mathbf{x}) > 0$, a **single-ratio FP** is

$$\begin{array}{ll} \underset{\mathbf{x}}{\text{maximize}} & \dfrac{f(\mathbf{x})}{g(\mathbf{x})} \\ \text{subject to} & \mathbf{x} \in \mathcal{X}. \end{array}$$

- FP has been widely studied and extended to deal with multiple ratios such as

$$\begin{array}{ll} \underset{\mathbf{x}}{\text{maximize}} & \min_i \dfrac{f_i(\mathbf{x})}{g_i(\mathbf{x})} \\ \text{subject to} & \mathbf{x} \in \mathcal{X}. \end{array}$$

---

[9] J. von Neumann, "Uber ein okonomisches gleichgewichtssystem und eine verallgemeinerung des brouwerschen fixpunktsatzes," *Ergebnisse eines Mathematischen Kolloquiums*, vol. 8, pp. 73–83, 1937.

Interlude: How to solve FP

- FPs are nonconvex problems, so in principle they are tough to solve (Stancu-Minasian 1992).[10]
- However, the so-called concave-convex FP can be easibly solved in different ways.
- We will focus on the **concave-convex single-ratio FP**:

$$\begin{array}{ll} \underset{\mathbf{x}}{\text{maximize}} & \dfrac{f(\mathbf{x})}{g(\mathbf{x})} \\ \text{subject to} & \mathbf{x} \in \mathcal{X}. \end{array}$$

  where $f(\mathbf{x}) \geq 0$ is concave and $g(\mathbf{x}) > 0$ is convex.
- Main approaches:
  - via **bisection method (aka sandwich technique)**
  - via **Dinkelbach transform**
  - via **Schaible transform** (**Charnes-Cooper transform** for the linear FP case)

[10] I. M. Stancu-Minasian, *Fractional Programming: Theory, Methods and Applications*. Kluwer Academic Publishers, 1992.

Interlude: Solving FP via bisection

- The idea is to realize that a concave-convex FP, while not convex, is **quasi-convex**.
- This can be easily seen by rewritting the **problem in epigraph form**:

$$\begin{array}{ll} \underset{\mathbf{x},t}{\text{maximize}} & t \\ \text{subject to} & t \leq \dfrac{f(\mathbf{x})}{g(\mathbf{x})} \\ & \mathbf{x} \in \mathcal{X}. \end{array}$$

- If now we fix the variable t to some value (so it is not a variable anymore), then we can rewrite it as a **convex problem**:

$$\begin{array}{ll} \underset{\mathbf{x}}{\text{maximize}} & 0 \\ \text{subject to} & tg(\mathbf{x}) \leq f(\mathbf{x}) \\ & \mathbf{x} \in \mathcal{X}. \end{array}$$

- At this point, one can easily solve this convex problem optimally with a solver and then solve for t via the **bisection algorithm** (aka **sandwich technique**). It converges to the global optimal solution.

Interlude: Solving FP via bisection

- Recall the quasi-convex problem we want to solve:

$$\begin{array}{ll} \underset{\mathbf{x},t}{\text{maximize}} & t \\ \text{subject to} & tg(\mathbf{x}) \leq f(\mathbf{x}) \\ & \mathbf{x} \in \mathcal{X}. \end{array}$$

Algorithm 1: Bisection method (aka sandwich technique)

Given upper- and lower-bounds on t: t^{ub} and t^{lb}.

1. compute mid-point: $t = \left(t^{\mathrm{ub}} + t^{\mathrm{lb}}\right)/2$
2. solve the following feasibility problem:

$$\begin{array}{ll} \text{find} & \mathbf{x} \\ \text{subject to} & tg(\mathbf{x}) \leq f(\mathbf{x}) \\ & \mathbf{x} \in \mathcal{X}. \end{array}$$

3. if feasible, then set $t^{\mathrm{lb}} = t$; otherwise set $t^{\mathrm{ub}} = t$
4. if $t^{\mathrm{ub}} - t^{\mathrm{lb}} > \epsilon$ go to step 1; otherwise finish.

Interlude: Solving FP via Dinkelbach tranform

- The **Dinkelbach transform** was proposed in (Dinkelbach 1967)[11].
- It reformulates the original concave-convex FP problem

$$\begin{array}{ll} \underset{\mathbf{x}}{\text{maximize}} & \dfrac{f(\mathbf{x})}{g(\mathbf{x})} \\ \text{subject to} & \mathbf{x} \in \mathcal{X}. \end{array}$$

as the convex problem

$$\begin{array}{ll} \underset{\mathbf{x}}{\text{maximize}} & f(\mathbf{x}) - y g(\mathbf{x}) \\ \text{subject to} & \mathbf{x} \in \mathcal{X} \end{array}$$

with a new auxiliary variable y, which is iteratively updated by

$$y^{(k)} = \frac{f(\mathbf{x}^{(k)})}{g(\mathbf{x}^{(k)})}$$

where k is the iteration index.

[11] W. Dinkelbach, "On nonlinear fractional programming," *Manage. Sci.*, vol. 133, no. 7, pp. 492–498, 1967.

Interlude: Solving FP via Dinkelbach transform

Algorithm 2: Dinkelbach method

Set $k = 0$ and initialize $\mathbf{x}^{(0)}$

repeat

- Set $y^{(k)} = \frac{f(\mathbf{x}^{(k)})}{g(\mathbf{x}^{(k)})}$
- Obtain next point $\mathbf{x}^{(k+1)}$ by solving

$$\begin{array}{ll} \underset{\mathbf{x}}{\text{maximize}} & f(\mathbf{x}) - y^{(k)} g(\mathbf{x}) \\ \text{subject to} & \mathbf{x} \in \mathcal{X} \end{array}$$

- $k \leftarrow k + 1$

until convergence

return $\mathbf{x}^{(k)}$

Dinkelbach method can be shown to converge to the global optimum of the original concave-convex FP by carefully analyzing the increasingness of $y^{(k)}$ and the funtion $F(y) = \arg\max_{\mathbf{x}} f(\mathbf{x}) - yg(\mathbf{x})$.

Interlude: Solving Linear FP via Charnes-Cooper transform

- The **Charnes-Cooper transform** was proposed in (Charnes and Cooper 1962)[12] to solve the linear FP (LFP) case (Bajalinov 2003)[13]:

$$\begin{array}{ll} \underset{\mathbf{x}}{\text{maximize}} & \dfrac{\mathbf{c}^T\mathbf{x}+\alpha}{\mathbf{d}^T\mathbf{x}+\beta} \\ \text{subject to} & \mathbf{A}\mathbf{x}\leq\mathbf{b} \end{array}$$

  over the set $\{\mathbf{x} \mid \mathbf{d}^T\mathbf{x}+\beta>0\}$.

- The Charnes-Cooper transform introduces two new variables

$$\mathbf{y}=\frac{1}{\mathbf{d}^T\mathbf{x}+\beta}\mathbf{x}, \quad t=\frac{1}{\mathbf{d}^T\mathbf{x}+\beta}.$$

[12] A. Charnes and W. W. Cooper, "Programming with linear fractional functionals," *Naval Research Logistics Quarterly*, vol. 9, no. 3-4, pp. 181–186, 1962.

[13] E. B. Bajalinov, *Linear-Fractional Programming: Theory, Methods, Applications and Software*. Kluwer Academic Publishers, 2003.

Interlude: Solving Linear FP via Charnes-Cooper transform

- The LFP becomes then a linear program (LP):

$$\begin{array}{ll} \underset{\mathbf{y},t}{\text{maximize}} & \mathbf{c}^T\mathbf{y} + \alpha t \\ \text{subject to} & \mathbf{A}\mathbf{y} \leq \mathbf{b}t \\ & \mathbf{d}^T\mathbf{y} + \beta t = 1 \\ & t \geq 0 \end{array}$$

- The solution for $\mathbf{y}$ and t yields the solution of the original problem as

$$\mathbf{x} = \frac{1}{t}\mathbf{y}.$$

- Note that the number of constraints in the LP formulation has increased.

Interlude: Solving Linear FP via Charnes-Cooper transform

Proof:

- Any feasible point $\mathbf{x}$ in the original LFP leads to a feasible point $(\mathbf{y}, t)$ in the LP (via the equations in the Charnes-Cooper transform) with the same objective value. The objective is
$$\frac{\mathbf{c}^T\mathbf{x}+\alpha}{\mathbf{d}^T\mathbf{x}+\beta} = \frac{\mathbf{c}^T\mathbf{y}/t+\alpha}{\mathbf{d}^T\mathbf{y}/t+\beta} = \frac{\mathbf{c}^T\mathbf{y}+\alpha t}{\mathbf{d}^T\mathbf{y}+\beta t}$$
and since it is scale invariant, we can choose to set the denominator equal to 1.
- Conversely, any feasible point $(\mathbf{y}, t)$ in the LP leads to a feasible point $\mathbf{x}$ in the original LFP (via $\mathbf{x} = \frac{1}{t}\mathbf{y}$). From the denominator constraint
$$\mathbf{d}^T\mathbf{y}+\beta t = 1 \Longleftrightarrow \mathbf{d}^T\mathbf{y}/t+\beta = 1/t \Longleftrightarrow 1/(\mathbf{d}^T\mathbf{x}+\beta) = t$$
which leads to the objective
$$\mathbf{c}^T\mathbf{y}+\alpha t = t(\mathbf{c}^T\mathbf{y}/t+\alpha) = t(\mathbf{c}^T\mathbf{x}+\alpha) = (\mathbf{c}^T\mathbf{x}+\alpha)/(\mathbf{d}^T\mathbf{x}+\beta).$$

Interlude: Solving FP via Schaible transform

- The **Schaible transform** is a generalization of the Charnes-Cooper transform proposed in (Schaible 1974)[14] to solve the concave-convex FP:

$$\begin{array}{ll} \underset{\mathbf{x}}{\text{maximize}} & \dfrac{f(\mathbf{x})}{g(\mathbf{x})} \\ \text{subject to} & \mathbf{x} \in \mathcal{X}. \end{array}$$

- The Schaible transform introduces two new variables (note that $\mathbf{x} = \mathbf{y}/t$):

$$\mathbf{y} = \frac{\mathbf{x}}{g(\mathbf{x})}, \quad t = \frac{1}{g(\mathbf{x})}.$$

- The original concave-convex FP is equivalent to the convex problem:

$$\begin{array}{ll} \underset{\mathbf{y},t}{\text{maximize}} & tf\left(\frac{\mathbf{y}}{t}\right) \\ \text{subject to} & tg\left(\frac{\mathbf{y}}{t}\right) \le 1 \\ & t \ge 0 \\ & \mathbf{y}/t \in \mathcal{X}. \end{array}$$

[14] S. Schaible, "Parameter-free convex equivalent and dual programs of fractional programming problems," *Zeitschrift fur Operations Research*, vol. 18, no. 5, pp. 187–196, 1974.

Solving the Maximum Sharpe ratio portfolio (MSRP)

- Recall the **maximization of the Sharpe ratio** proposed in (Sharpe 1966)[15]:

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{maximize}} & \dfrac{\mathbf{w}^T\boldsymbol{\mu} - r_f}{\sqrt{\mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w}}} \\ \text{subject to} & \mathbf{1}^T\mathbf{w} = 1, \quad (\mathbf{w} \geq \mathbf{0}). \end{array}$$

- This problem is **nonconvex**, but upon recognition as a **fractional program (FP)**, we can consider its resolution via:
  - **bisection method (aka sandwich technique)**,
  - **Dinkelbach transform**, and
  - **Schaible transform** (aka **Charnes-Cooper transform** for the linear FP case).

[15] W. F. Sharpe, "Mutual fund performance," *The Journal of Business*, vol. 39, no. 1, pp. 119–138, 1966.

Maximum Sharpe ratio portfolio via bisection

- The idea is to realize that the problem, while not convex, is **quasi-convex**.
- This can be easily seen by rewritting the **problem in epigraph form**:

$$
\begin{array}{ll}
\underset{\mathbf{w},t}{\text{maximize}} & t \\
\text{subject to} & t \leq \dfrac{\mathbf{w}^T\boldsymbol{\mu} - r_f}{\sqrt{\mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w}}} \\
& \mathbf{1}^T\mathbf{w} = 1, \quad (\mathbf{w} \geq \mathbf{0})\,.
\end{array}
$$

- If now we fix the variable t to some value (so it is not a variable anymore), the problem is easily recognized as a **(convex) second order cone program (SOCP)**:

$$
\begin{array}{ll}
\text{find} & \mathbf{w} \\
\text{subject to} & t\left\|\boldsymbol{\Sigma}^{1/2}\mathbf{w}\right\|_2 \leq \mathbf{w}^T\boldsymbol{\mu} - r_f \\
& \mathbf{1}^T\mathbf{w} = 1, \quad (\mathbf{w} \geq \mathbf{0})\,.
\end{array}
$$

- At this point, one can easily solve the convex problem with an SOCP solver and then solve for t via yhe **bisection algorithm** (aka **sandwich technique**).

Maximum Sharpe ratio portfolio via Dinkelbach

- The **Dinkelbach transform** proposed in (Dinkelbach 1967)[16] reformulates the original concave-convex FP problem

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{maximize}} & \dfrac{\mathbf{w}^T\boldsymbol{\mu} - r_f}{\sqrt{\mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w}}} \\ \text{subject to} & \mathbf{1}^T\mathbf{w} = 1, \quad (\mathbf{w} \geq \mathbf{0}) \end{array}$$

  as the convex problem

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{maximize}} & \mathbf{w}^T\boldsymbol{\mu} - y\sqrt{\mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w}} \\ \text{subject to} & \mathbf{1}^T\mathbf{w} = 1, \quad (\mathbf{w} \geq \mathbf{0}) \end{array}$$

  with a new auxiliary variable y, which is iteratively updated by

$$y^{(k)} = \frac{\mathbf{w}^{(k)T}\boldsymbol{\mu} - r_f}{\sqrt{\mathbf{w}^{(k)T}\boldsymbol{\Sigma}\mathbf{w}^{(k)}}}$$

  where k is the iteration index.

[16] W. Dinkelbach, "On nonlinear fractional programming," *Manage. Sci.*, vol. 133, no. 7, pp. 492–498, 1967.

Maximum Sharpe ratio portfolio via Dinkelbach

- By noting that $\sqrt{\mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w}} = \left\|\boldsymbol{\Sigma}^{1/2}\mathbf{w}\right\|_2$, where $\boldsymbol{\Sigma}^{1/2}$ satisfies $\boldsymbol{\Sigma}^{T/2}\boldsymbol{\Sigma}^{1/2} = \boldsymbol{\Sigma}$, we can finally rewrite the problem as a SOCP.
- The iterative Dinkelbach-based method obtains $\mathbf{w}^{(k+1)}$ by solving the following problem at the kth iteration:

$$\begin{array}{ll} \underset{\mathbf{w},t}{\text{maximize}} & \mathbf{w}^T\boldsymbol{\mu} - y^{(k)}t \\ \text{subject to} & t \geq \left\|\boldsymbol{\Sigma}^{1/2}\mathbf{w}\right\|_2 \\ & \mathbf{1}^T\mathbf{w} = 1, \quad (\mathbf{w} \geq \mathbf{0}) \end{array}$$

where

$$y^{(k)} = \frac{\mathbf{w}^{(k)T}\boldsymbol{\mu} - r_f}{\sqrt{\mathbf{w}^{(k)T}\boldsymbol{\Sigma}\mathbf{w}^{(k)}}}.$$

Maximum Sharpe ratio portfolio via Schaible

- The **Schaible transform** is a generalization of the Charnes-Cooper transform proposed in (Schaible 1974)[17] to solve a concave-convex FP:

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{maximize}} & \dfrac{\mathbf{w}^T\boldsymbol{\mu} - r_f}{\sqrt{\mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w}}} \\ \text{subject to} & \mathbf{1}^T\mathbf{w} = 1, \quad (\mathbf{w} \geq \mathbf{0}). \end{array}$$

- The Schaible transform introduces two new variables (note that $\mathbf{w} = \tilde{\mathbf{w}}/t$):

$$\tilde{\mathbf{w}} = \frac{\mathbf{w}}{\sqrt{\mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w}}}, \quad t = \frac{1}{\sqrt{\mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w}}}.$$

- The original maximum Sharpe ratio portfolio is equivalent to the convex quadratic problem (QP):

$$\begin{array}{ll} \underset{\tilde{\mathbf{w}}, t}{\text{maximize}} & \tilde{\mathbf{w}}^T\boldsymbol{\mu} - r_f t \\ \text{subject to} & \tilde{\mathbf{w}}^T\boldsymbol{\Sigma}\tilde{\mathbf{w}} \leq 1 \\ & t \geq 0 \\ & \mathbf{1}^T\tilde{\mathbf{w}} = t, \quad (\tilde{\mathbf{w}} \geq \mathbf{0}). \end{array}$$

[17] S. Schaible, "Parameter-free convex equivalent and dual programs of fractional programming problems," *Zeitschrift fur Operations Research*, vol. 18, no. 5, pp. 187–196, 1974.

Maximum Sharpe ratio portfolio via Schaible

- The previous problem can be simplified by eliminating the variable $t = \mathbf{1}^T\tilde{\mathbf{w}}$:

$$\begin{array}{ll} \underset{\tilde{\mathbf{w}}}{\text{maximize}} & \tilde{\mathbf{w}}^T(\boldsymbol{\mu} - r_f\mathbf{1}) \\ \text{subject to} & \tilde{\mathbf{w}}^T\boldsymbol{\Sigma}\tilde{\mathbf{w}} \leq 1 \\ & \mathbf{1}^T\tilde{\mathbf{w}} \geq 0, \quad (\tilde{\mathbf{w}} \geq \mathbf{0}) \end{array}$$

  from which we can recover the original solution as $\mathbf{w} = \tilde{\mathbf{w}}/t = \tilde{\mathbf{w}}/\left(\mathbf{1}^T\tilde{\mathbf{w}}\right)$.

- Interestingly, we can get the following slightly different formulation by starting with a minimization of a ratio:

$$\begin{array}{ll} \underset{\tilde{\mathbf{w}}}{\text{minimize}} & \tilde{\mathbf{w}}^T\boldsymbol{\Sigma}\tilde{\mathbf{w}} \\ \text{subject to} & \tilde{\mathbf{w}}^T(\boldsymbol{\mu} - r_f\mathbf{1}) = 1 \\ & \mathbf{1}^T\tilde{\mathbf{w}} \geq 0, \quad (\tilde{\mathbf{w}} \geq \mathbf{0})\,. \end{array}$$

Proof of convex reformulation of MSRP

- Start with the original problem formulation:

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{maximize}} & \frac{\mathbf{w}^T\boldsymbol{\mu}-r_f}{\sqrt{\mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w}}} \\ \text{subject to} & \mathbf{1}^T\mathbf{w}=1 \\ & (\mathbf{w}\geq\mathbf{0}) \end{array} \iff \begin{array}{ll} \underset{\mathbf{w}}{\text{minimize}} & \frac{\sqrt{\mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w}}}{\mathbf{w}^T(\boldsymbol{\mu}-r_f\mathbf{1})} \\ \text{subject to} & \mathbf{1}^T\mathbf{w}=1 \\ & (\mathbf{w}\geq\mathbf{0}) \end{array}$$

- Now, since the objective is scale invariant w.r.t. $\mathbf{w}$, we can choose the proper scaling factor for our convenience. We define $\tilde{\mathbf{w}}=t\mathbf{w}$ with the scaling factor $t=1/\left(\mathbf{w}^T\boldsymbol{\mu}-r_f\right)>0$, so that the objective becomes $\tilde{\mathbf{w}}^T\boldsymbol{\Sigma}\tilde{\mathbf{w}}$, the sum constraint $\mathbf{1}^T\tilde{\mathbf{w}}=t$, and the problem is

$$\begin{array}{ll} \underset{\mathbf{w},\tilde{\mathbf{w}},t}{\text{minimize}} & \sqrt{\tilde{\mathbf{w}}^T\boldsymbol{\Sigma}\tilde{\mathbf{w}}} \\ \text{subject to} & t=1/\mathbf{w}^T(\boldsymbol{\mu}-r_f\mathbf{1})>0 \\ & \tilde{\mathbf{w}}=t\mathbf{w} \\ & \mathbf{1}^T\tilde{\mathbf{w}}=t>0 \\ & (\tilde{\mathbf{w}}\geq\mathbf{0}). \end{array}$$

Proof of convex reformulation of MSRP

- The constraint $t = 1/\mathbf{w}^T(\boldsymbol{\mu} - r_f\mathbf{1})$ can be rewritten in terms of $\tilde{\mathbf{w}}$ as $1 = \tilde{\mathbf{w}}^T(\boldsymbol{\mu} - r_f\mathbf{1})$. So the problem becomes

$$\begin{array}{ll} \underset{\mathbf{w},\tilde{\mathbf{w}},t}{\text{minimize}} & \tilde{\mathbf{w}}^T\boldsymbol{\Sigma}\tilde{\mathbf{w}} \\ \text{subject to} & \tilde{\mathbf{w}}^T(\boldsymbol{\mu} - r_f\mathbf{1}) = 1 \\ & \tilde{\mathbf{w}} = t\mathbf{w} \\ & \mathbf{1}^T\tilde{\mathbf{w}} = t > 0 \\ & (\tilde{\mathbf{w}} \geq \mathbf{0}). \end{array}$$

- Now, note that the strict inequality $t > 0$ is equivalent to $t \geq 0$ because $t = 0$ can never happen as $\tilde{\mathbf{w}}$ would would be zero and the first constraint would not be satisfied.

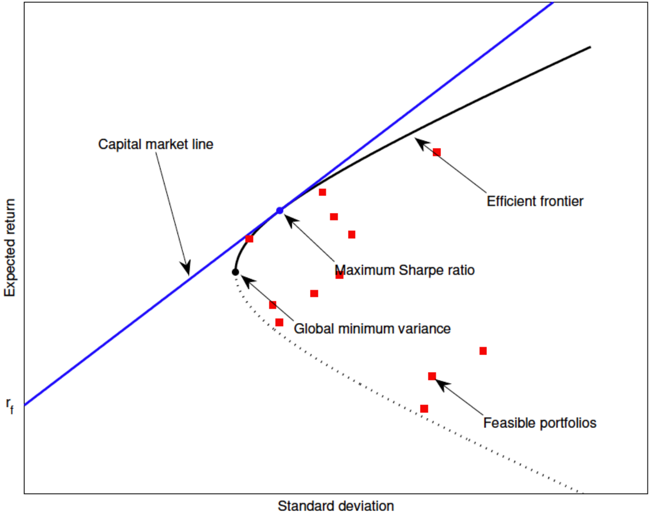
Proof of convex reformulation of MSRP

- Finally, we can now get rid of $\mathbf{w}$ and t in the formulation as they can be directly obtained as $t = \mathbf{1}^T\tilde{\mathbf{w}}$ and $\mathbf{w} = \tilde{\mathbf{w}}/t$:

$$\begin{array}{ll} \underset{\tilde{\mathbf{w}}}{\text{minimize}} & \tilde{\mathbf{w}}^T\boldsymbol{\Sigma}\tilde{\mathbf{w}} \\ \text{subject to} & \tilde{\mathbf{w}}^T(\boldsymbol{\mu} - r_f\mathbf{1}) = 1 \\ & \mathbf{1}^T\tilde{\mathbf{w}} \geq 0 \\ & (\tilde{\mathbf{w}} \geq \mathbf{0})\,. \end{array}$$

- QED!
- Recall that the portfolio is then obtained with the correct scaling factor as $\mathbf{w} = \tilde{\mathbf{w}}/(\mathbf{1}^T\tilde{\mathbf{w}})$.

Sharpe ratio portfolio in the efficient frontier



Basic references

- Textbook on financial data (Ruppert and Matteson 2015):

  D. Ruppert and D. Matteson. Statistics and Data Analysis for Financial Engineering: With R Examples. Springer, 2015.

- Textbooks on portfolio optimization (Cornuejols and Tütüncü 2006; Fabozzi 2007; Feng and Palomar 2016):

  G. Cornuejols and R. Tutuncu. Optimization Methods in Finance. Cambridge University Press, 2006.

  F. J. Fabozzi. Robust Portfolio Optimization and Management. Wiley, 2007.

  Y. Feng and D. P. Palomar. A Signal Processing Perspective on Financial Engineering. Foundations and Trends in Signal Processing, Now Publishers, 2016.

Thanks

For more information visit:

https://www.danielppalomar.com

References I

Bajalinov, E. B. (2003). *Linear-fractional programming: Theory, methods, applications and software*. Kluwer Academic Publishers.

Charnes, A., & Cooper, W. W. (1962). Programming with linear fractional functionals. *Naval Research Logistics Quarterly*, *9*(3-4), 181–186.

Chopra, V., & Ziemba, W. (1993). The effect of errors in means, variances and covariances on optimal portfolio choice. *Journal of Portfolio Management*.

Cornuejols, G., & Tütüncü, R. (2006). *Optimization methods in finance*. Cambridge University Press.

DeMiguel, V., Garlappi, L., & Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *The Review of Financial Studies*.

Dinkelbach, W. (1967). On nonlinear fractional programming. *Manage. Sci.*, *133*(7), 492–498.

References II

Duchin, R., & Levy, H. (2009). Markowitz versus the talmudic portfolio diversification strategies. *Journal of Portfolio Management*, *35*(2), 71.

Fabozzi, F. J. (2007). *Robust portfolio optimization and management*. Wiley.

Feng, Y., & Palomar, D. P. (2016). *A Signal Processing Perspective on Financial Engineering*. Foundations; Trends in Signal Processing, Now Publishers.

Jolliffe, I. (2002). *Principal component analysis*. Springer-Verlag.

Lütkepohl, H. (2007). *New introduction to multiple time series analysis*. Springer.

Markowitz, H. (1952). Portfolio selection. *J. Financ.*, *7*(1), 77–91.

Ruppert, D. (2010). *Statistics and data analysis for financial engineering*. Springer.

Ruppert, D., & Matteson, D. (2015). *Statistics and data analysis for financial engineering* (2nd Ed.). Springer.

References III

Schaible, S. (1974). Parameter-free convex equivalent and dual programs of fractional programming problems. *Zeitschrift fur Operations Research*, *18*(5), 187–196.

Sharpe, W. F. (1966). Mutual fund performance. *The Journal of Business*, *39*(1), 119–138.

Stancu-Minasian, I. M. (1992). *Fractional programming: Theory, methods and applications*. Kluwer Academic Publishers.

Tsay, R. S. (2005). *Analysis of financial time series* (3rd ed.). John Wiley & Sons.

Tsay, R. S. (2013). *Multivariate time series analysis: With r and financial applications*. John Wiley & Sons.

von Neumann, J. (1937). Uber ein okonomisches gleichgewichtssystem und eine verallgemeinerung des brouwerschen fixpunktsatzes. *Ergebnisse eines Mathematischen Kolloquiums*, *8*, 73–83.