Summary: Asset Management: Advanced Investments

Fabian MARBACH, Spring Semester 2016

Useful expressions

Mean, variance and covariances of portfolios

- return of portfolio: $R_{\boldsymbol{w}} = \boldsymbol{w}^\top \boldsymbol{\mu}$
- variance of a portfolio:

$$\sigma^2(\boldsymbol{w}) = \boldsymbol{w}^\top \Sigma \boldsymbol{w} = \sum_{i=1}^{n} w_i^2 \sigma_i^2 + 2 \sum_{i>j} w_i w_j \rho_{ij} \sigma_i \sigma_j$$

- covariance of two portfolios $\boldsymbol{w}_p, \boldsymbol{w}_q$: $\mathrm{Cov}[R_p, R_q] = \boldsymbol{w}_p^\top \Sigma \boldsymbol{w}_q$
- covariance of a single asset i with a portfolio $\boldsymbol{w}$: $\mathrm{Cov}[R_i, R_{\boldsymbol{w}}] = (\Sigma \boldsymbol{w})_i$

Sharpe ratio of a portfolio

$$\mathrm{SR}(\boldsymbol{w}) = \frac{\mu(\boldsymbol{w}) - R_f}{\sigma(\boldsymbol{w})}$$

Linear Algebra

- product rule: $(AB)^\top = B^\top A^\top$
  and: $(A_1 A_2 \cdots A_k)^\top = A_k^\top \cdots A_2^\top A_1^\top$
- transpose of an inverse: $(A^{-1})^\top = (A^\top)^{-1}$
- dot-product of two vectors: $a^\top b = b^\top a \in \mathbb{R}$

1 Mean-variance paradigm

Main advantage of the MV paradigm

- The MV paradigm provides a simple framework to construct and select portfolios, based on *expected performance* of investments and *risk appetite* of investors.

MV critique and approaches to overcome MV shortcomings

- The MV approach relies on *variance as a risk measure.*
  - An investor might not think in terms of utility functions but first wants to make sure that a certain amount of the principal is reserved.

  $\rightsquigarrow$ portfolio optimization based on other risk measures, e.g. based on *VaR* or *ES*
- There is a *high rebalancing activity* / high *instability of optimal portfolio weights $\boldsymbol{w}$* / *high sensitivity to input parameters* (i.e. return estimates).
  - Even small and selected changes in expected returns lead to huge unrealistic shifts/fluctuations in asset weights.
  - Furthermore, MV portfolios often seem counterintuitive and inexplicable.

  This is caused by: *Error maximization* / high *estimation risk:*
  - Most of the estimation risk is due to errors in estimates of expected returns (not estimates of risk).
  - MV optimization exacerbates effect of sampling errors since it takes advantage of unusually high means and low variances.
  - Estimation risk arises naturally, as samples are not sufficiently large and market structures change over time.

  $\rightsquigarrow$ *resampling methods*, *constrained optimization* (e.g. sets constraints on weights), *robust optimization* (considers uncertainty in unknown parameters directly and explicitly), *shrinkage estimators*, *Black-Litterman model*, *Risk-Budgeting* (does not employ explicit forecasts of asset returns)
- There is a *bias in estimated performance.*
  - The MV approach promises far more than it delivers.
  - The actual frontier always lies below the estimated frontier.

  $\rightsquigarrow$ *robust optimization*
- MV optimization may result in *under-diversified* strategies.
  - E.g. during the financial crisis, risk contribution of equities far exceeded their forecast limits — partly due to a realised jump in realised equity correlation.

  $\rightsquigarrow$ *Risk budgeting* (risk factors exhibit by far lower correlation than equity)
- MV is only data-driven, i.e. *subjective views* are not considered.
  $\rightsquigarrow$ the *Black-Litterman approach* incorporates both market data and subjective views.
- There is *no quantification of confidence* in estimated portfolio returns μ_p.
  $\rightsquigarrow$ *robust optimization*, *Black-Litterman*, *Bayesian approaches*
- MV is only a *one-periodical* approach.
  $\rightsquigarrow$ *multi-period approaches* (discrete or continuous time)

1.1 MV without Riskless Asset

Definition

- Given: n risky assets w/ mean vector $\boldsymbol{\mu} \in \mathbb{R}^n$ and covariance matrix $\Sigma \in \mathbb{R}^{n \times n}$
- Chose: target mean return μ_p
- Then: P is the corresponding minimum-variance portfolio (MVP) iff:
  - P has minimal portfolio variance: $\boldsymbol{w}_p^\top \Sigma \boldsymbol{w}_p = \min_{\boldsymbol{w}} \boldsymbol{w}^\top \Sigma \boldsymbol{w}$
  - P has target return μ_p: $\boldsymbol{w}_p^\top \boldsymbol{\mu} = \mu_p$
  - w is a weight vector: $\boldsymbol{w}_p^\top \mathbf{1} = 1$

Optimal weigths

- without any restrictions on the portfolio weights, the weights of the MVF for given μ_p are:

$$\boldsymbol{w}_p = \Sigma^{-1}(\mu_p \boldsymbol{k}_1 + \boldsymbol{k}_2), \quad \boldsymbol{k}_1 = \frac{c\boldsymbol{\mu} - b\mathbf{1}}{d}, \quad \boldsymbol{k}_2 = \frac{a\mathbf{1} - b\boldsymbol{\mu}}{d}$$

  with:

$$a = \boldsymbol{\mu}^\top \Sigma^{-1} \boldsymbol{\mu}, \qquad b = \boldsymbol{\mu}^\top \Sigma^{-1} \mathbf{1}$$
$$c = \mathbf{1}^\top \Sigma^{-1} \mathbf{1}, \qquad d = ac - b^2$$

Black's separation theorem

- Any portfolio of MVPs is also an MVP.
- The MVF can be generated by any two distinct MVPs.

Mean-Variance Frontier

- **mean-variance frontier (MVF):** set of all portfolios with mean return μ_p that solve above's minimization problem, i.e. with:

$$\mu_p = \frac{b}{c} + \sqrt{\frac{d}{c}\left(\sigma_p^2 - \frac{1}{c}\right)}$$

- **efficient frontier:** upper part of the MVF
- **inefficient frontier:** lower part of the MVF

Implications:

- Covariance of a MVP $\boldsymbol{w}_p$ with any asset or portfolio $\boldsymbol{w}_q$ (not necessarily on the MVF) is:

$$\mathrm{Cov}\left[R_p, R_q\right] = \frac{c}{d}\left(\mu_p - \frac{b}{c}\right)\left(\mu_q - \frac{b}{c}\right) + \frac{1}{c}$$

  which can also be written as:

$$\mathrm{Cov}\left[R_p, R_q\right] = e\left(\mu_p - f\right)\left(\mu_q - F\right) + g$$

  with:

$$e = \frac{c}{d}, \qquad f = \frac{b}{c}, \qquad g = \frac{1}{c}$$

- Covariance of two efficient MVPs is at least $\frac{1}{c} > 0$.
- Covariance between two inefficient MVPs is always positive.

Global MVP (gMVP)

- return: $\mu_g = \frac{b}{c}$, variance: $\sigma_g^2 = \frac{1}{c}$
- optimal weights: $\boldsymbol{w}_g = \frac{1}{c}\Sigma^{-1}\mathbf{1}$

Furthermore:

- covariance of any asset or portfolio return R_p with the gMVP is:

$$\mathrm{Cov}\left[R_g, R_p\right] = \frac{1}{c}$$

1.2 Zero Beta Portfolio

Zero Beta Portfolio For each MVP $\boldsymbol{w}_p$, except for the gMVP, $\exists$ a unique MVP, the zero-beta MVP w.r.t. $\boldsymbol{w}_p$, that has zero covariance with $\boldsymbol{w}_p$.
Remarks:

- In the absence of the risk-free rate, the efficient frontier consists of those MVPs with return μ_p equal or higher than the return of the gMVP μ_g.
- The return of the gMVP μ_g is equal or higher than the return of any zero-beta portfolio μ_{0p}.
- Thus it holds: $\mu_p \geq \mu_g \geq \mu_{0p}$

Beta Replication Every portfolio (not necessarily an MVP, μ_q) has a beta representation in terms of a MVP (μ_p) and a portfolio orthogonal to the MVP (μ_{0p}):

$$\mu_q = \mu_{0p} + \beta_{pq}(\mu_p - \mu_{0p}), \qquad \beta_{pq} = \frac{\mathrm{Cov}\left[R_p, R_q\right]}{\sigma_p^2}$$

Remarks:

- All zero-beta portfolios lie on a horizontal line.
- Any feasible mean-variance combination can be constructed from any MVP $\boldsymbol{w}_p$ and some $\boldsymbol{w}_{0q}$, which is orthogonal to $\boldsymbol{w}_p$ (but not necessarily a MVP).

Existence of Zero Beta portfolios

- Zero Beta portfolios only exist if *short-sales* are allowed.
  I.e. if there are short-sales constraints, then the Zero Beta portfolio does generally not exist.
- There exists no Zero Beta portfolio for the *GMV*.

1.3 MV with Risk-Free Asset

In presence of a risk-free asset:

- *Given:* vector of excess returns of the assets over the risk-free rate, i.e. $\boldsymbol{\mu}^e = \boldsymbol{\mu} - \mathbf{1}R_f$, covariance matrix Σ
- *Choose:* target return μ_p
- *Then:* the optimization problem reads:

$$\min_{\boldsymbol{w}_p} \boldsymbol{w}_p^\top \Sigma \boldsymbol{w}_p, \qquad \text{s.t. } \mu_p = R_f + (\boldsymbol{\mu} - \mathbf{1}R_f)^\top \boldsymbol{w}_p$$

  Note that the previous constraint $\mathbf{1}^\top w = 1$ is not anymore necessary.

- *Result:* the weights $\boldsymbol{w}_p$ of the risky assets of the corresponding MVP and its variance are:

$$\boldsymbol{w}_p = \frac{\mu_p - R_f}{(\boldsymbol{\mu}^e)^\top \Sigma^{-1}\boldsymbol{\mu}^e}\Sigma^{-1}\boldsymbol{\mu}^e \quad \sigma_p = \frac{\mu_p - R_f}{(\boldsymbol{\mu}^e)^\top \Sigma^{-1}\boldsymbol{\mu}^e}\sqrt{(\boldsymbol{\mu}^e)^\top \Sigma^{-1}\boldsymbol{\mu}^e}$$

  and the fraction $1 - \mathbf{1}^\top R_f$ is invested in the risk-free asset.

Tobin's Separation Theorem The relative portfolio fraction is independent of the choice of the targeted portfolio return μ_p.
Implications:

- Every investor's portfolio decision is the same.
- Only difference: relative portion between the risky protfolio and the risk-free rate R_f, which depends on the investor's risk-aversion.

Tangency portfolio The tangency portfolio $\boldsymbol{w}_T$ (with maximal Sharpe ratio) is characterized by:

$$\boldsymbol{w}_T = \frac{\Sigma^{-1}\boldsymbol{\mu}^e}{\mathbf{1}^\top \Sigma^{-1}\boldsymbol{\mu}^e} = \frac{\Sigma^{-1}(\boldsymbol{\mu} - \mathbf{1}R_f)}{b - cR_f}$$

$$\mu_T = R_f + \frac{(\boldsymbol{\mu}^e)^\top \Sigma^{-1}\boldsymbol{\mu}^e}{\mathbf{1}^\top \Sigma^{-1}\boldsymbol{\mu}^e}, \qquad \sigma_T = \frac{\sqrt{(\boldsymbol{\mu}^e)^\top \Sigma^{-1}\boldsymbol{\mu}^e}}{\mathbf{1}^\top \Sigma^{-1}\boldsymbol{\mu}^e}$$

Remarks:

- The tangency portfolio must consist of all assets available to investors, and each asset must be held in proportion to its relative market capitalization ($\equiv$ market portfolio)
- Note that since $\boldsymbol{w}_T$ is independent of μ_p, this proves Tobin's separation theorem.

1.4 CAPM

CAPM

- The CAPM determines a theoretically appropriate required rate of return of an asset, if that asset is to be added to an already well-diversified portfolio, given that asset's non-diversifiable risk.
- The CAPM takes the following aspects into account:
  - asset's sensitivity to non-diversifiable risk (=systematic risk or market risk)
  - expected return of the market
  - expected return of a theoretical risk-free asset.

Assumptions of the CAPM

- variance of returns is an adequate measurement of risk (justified e.g. in case of a quadratic utility function)
- investors are *rational* and *risk-averse*
- investors are price-takers and can lend and borrow any amount under the same risk-free rate
- all investors have access to the *same information*
- investors do not have *preferences* between markets and assets, i.e. investors choose assets soley based on their risk-return profile
- *homogenous expectations assumption:* investors agree on the risk and expected return of all assets
- *no taxes, no transaction costs*
- assets are *infinitely divisible*

Capital market line/security market line

- All optimal portfolios $\boldsymbol{w}^*$ are on the *capital market line:*

  $$\mu^* = R_f + \frac{\mu_m^e}{\sigma_m}\sigma^*$$

  with μ_m^e and σ_m the parameters from the market portfolio.
- This can also be written in terms of the *security market line:*

  $$\mu_i = R_f + \beta_i(\mu_M - R_f)$$

  where R_f is the risk-free rate of return and μ_M is the return of the market portfolio.
  Consequently, the term $\mu_M - R_f$ denotes the *market excess return* or *market risk premium*.

Beta representation For any portfolio or asset i, there is a beta representation, i.e.

$$\mu_i^e = \beta_i \mu_m^e = \frac{\mathrm{Cov}[R_i, R_m]}{\mathrm{Var}[R_m]}\mu_m^e = \frac{\rho_{im}\sigma_i\sigma_m}{\sigma_m^2}\mu_m^e$$

Remarks:

- The graph (μ_i, β_i) is called the **security market line**.
- While *standard deviation* measures risk arising from both systematic and unsystematic sources, the *beta* only measures the risk w.r.t. to the variance from the market portfolio.
- **Low beta anomaly:**
  Historically, low beta stocks have offered a combination of low risks and high returns.

CAPM applied to portfolios

- If an investor wants to achieve a certain return μ_p by investing in the market portfolio and the risk-free asset, the corresponding beta β_p can computed as:

  $$\beta_p = \frac{\mu_p - R_f}{\mu_M - R_f}$$

- The corresponding risk σ_p can then be computed via:

  $$\sigma_p = \beta_p \sigma_M$$

- Since the β of a portfolio is simply the weighted sum of the assets' betas, the β directly determines the ratio that has to be invested in the market portfolio, while the ratio $1-\beta$ determines the investment in the risk-free asset.
- The relationship between the **price of an asset** i and its expected return μ_i is given by:

  $$P_i = \frac{\mathbb{E}[X_i]}{1+\mu_i}$$

  where $\mathbb{E}[X_i]$ is the expected cash-flow of asset i.

Differences between the CAPM and the MV model While MV only looks at the optimization of a *single investor*, the CAPM also considers the impact of this optimization on an aggregate market level by using equilibrium arguments.

2 Downside Risk Measures

2.1 Motivation

Semivariance In the MV approach, variance penalizes over- and underperformance equally.
Semivariance is concerned only with the adverse deviations and is defined as:

$$\sigma_{P,\min}^2 = \mathbb{E}\left[\min\left(\sum_{i=1}^n w_i(R_i - \mu_i), 0\right)\right]$$

Generalization of semivariance: lower partial moment of risk measures.

Lower Partial Moments (LPMs)

$$\sigma_{P,q,R_0} = \mathbb{E}\left[\max(R_0 - R_P, 0)^q\right]^{1/q}$$

with a *power index* q and *target rate of return* R_0.

- Variance and semivariance are consistent with an investor having quadratic utility only. LPMs are consistent with a much wider class of vNM utility functions.
- cases:
  - $0 < q < 1$: risk-seeking investor
  - $q = 1$: risk-neutral investor
  - $1 < q$: risk-averse investor
  - $q = 2$: semivariance / semivolatility

2.2 Value at Risk (VaR) and Expected Shortfall (ES)

2.2.1 Value at Risk (VaR)

Value-at-Risk (VaR)

- $\mathrm{VaR}_\alpha(X, n)$ is the *maximum potential loss* that a portfolio X can suffer in the 100α% *best* cases in n days.
- $\mathrm{VaR}_\alpha(X, n)$ is the *minimum potential loss* that a portfolio X can suffer in the 100$(1-\alpha)$% *worst* cases in n days.
- Given a RV X on $(\Omega, \mathcal{A}, \mathbb{P})$ and a scalar $\alpha \in (0, 1)$,

  $$\mathrm{VaR}_\alpha(X, n) \equiv -\sup\{x : \mathbb{P}[X < x] \leq 1 - \alpha\}, \qquad \alpha \in [0, 1)$$

  i.e. $\mathrm{VaR}_\alpha(X, n)$ defines a quantity such that

  $$\mathbb{P}[X \geq -\mathrm{VaR}_\alpha(X, n)] \geq \alpha$$

  i.e. VaR_α is a α quantile. Usually, $\alpha \in [0.5, 1)$.
- Note that $\mathrm{VaR}_\alpha = -\mathrm{VaR}_{1-\alpha}$.

VaR return

- The 100α%-VaR is the return v s.t. $F(-v) = 1 - \alpha$, $\alpha \in (0.5, 1)$ and F the CDF of the portfolio's return.
- Mathematically:

  $$v(r_w, \alpha) = z_\alpha \sqrt{\boldsymbol{w}^\top \Sigma \boldsymbol{w}} - \boldsymbol{w}^\top \boldsymbol{\mu} = z_\alpha \sigma_{r_w} - \mu_{r_w}$$

  with $z_\alpha = \Phi^{-1}(\alpha)$, Φ the standard normal distribution, and r_w the return of portfolio $\boldsymbol{w}$, $r_w \sim \mathcal{N}(\boldsymbol{w}^\top\boldsymbol{\mu}, \boldsymbol{w}^\top\Sigma\boldsymbol{w})$.

(μ, VaR)-optimization

- Optimization problem:

  $$\min_{\boldsymbol{w}\in\mathcal{W}} z_\alpha \sigma_{r_w} - \mu_{r_w}, \qquad \text{s.t. } \mathbb{E}[r_w] = \bar{r}$$

  with $z_\alpha = \Phi^{-1}(\alpha) > \sqrt{\frac{d}{c}}$ a necessary and sufficient condition for this portfolio.
  Since the return of a MVP is given by $\mu_p = \frac{b}{c} + \sqrt{\frac{d}{c}\left(\sigma_p^2 - \frac{1}{c}\right)}$ and the variance of the GMV portfolio is $\sigma_g = \frac{1}{\sqrt{c}}$, this can be rewritten as:

  $$\min_{\sigma\in[\sigma_g,\infty)} \sigma z_\alpha - \left(\frac{b}{c} + \sqrt{\frac{d}{c}\left(\sigma_p^2 - \frac{1}{c}\right)}\right)$$

- The set of solutions build the *mean-VaR boundary*, which is given as:

  $$\left\{(v(r,\alpha), \mu_r) \quad : \quad \frac{(v(r,\alpha)+\mu_r)^2}{z_\alpha^2/c} - \frac{(\mu_r - b/c)^2}{d/c^2} = 1\right\}$$

- The minimum VaR portfolio existis iff $\alpha > \Phi(\sqrt{d/c})$. The weights are given as:

  $$\boldsymbol{w}_m = \Sigma^{-1}\left(\left(\frac{b}{c} + \sqrt{\frac{d}{c}\left(\frac{z_\alpha^2}{cz_\alpha^2 - d} - \frac{1}{c}\right)}\right)\boldsymbol{k}_1 + \boldsymbol{k}_2\right)$$

  Since the return of the GMV portfolio is given by $\mu_g = \frac{b}{c}$, this can also be written as:

  $$\boldsymbol{w}_m = \boldsymbol{w}_{\mathsf{GMV}} + \Sigma^{-1}\sqrt{\frac{d}{c}\left(\frac{z_\alpha^2}{cz_\alpha^2 - d} - \frac{1}{c}\right)}\boldsymbol{k}_1$$

- *Remark:* since it has to hold that $z_\alpha > \sqrt{\frac{d}{c}}$, the second term in w_m is always positive, i.e. the efficient portfolio minimizing VaR does not coincide with the GMV portfolio.
  Only in the limit ($\alpha \nearrow 1 \Leftrightarrow z_\alpha \nearrow \infty$), the two portfolios coincide.

Implications of the VaR constraints

- *Highly risk averse investors:* may select a portfolio with larger standard deviation
- *Slightly risk averse investors:* may select a portfolio with smaller standard deviation
- Thus: not a-priori clear whether VaR gives stronger incentives to reduce risk of a portfolio

---

2.2.2 Expected Shortfall (ES)

Expected Shortfall (ES)

- $\mathrm{ES}_\alpha(X,n)$ is the expected value of the loss that portfolio X can suffer in the $100(1-\alpha)\%$ worst cases in n days.
- Mathematically:
$$\mathrm{ES}_\alpha(X,n) = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_m(X,n)dm$$
- Note that:
$$\mathrm{ES}_\alpha(X,n) = -\frac{1}{1-\alpha}\int_0^{1-\alpha} \mathrm{VaR}_m(X,n)dm$$
- e.g. under the *normal distribution*:
  Consider ES for a portfolio X with $X \sim \mathcal{N}(\boldsymbol{w}^\top\boldsymbol{\mu}, \boldsymbol{w}^\top\boldsymbol{\Sigma}\boldsymbol{w})$, i.e. $X \sim \mathcal{N}(\mu_r, \sigma_r^2)$ and $z_\alpha = \Phi^{-1}(\alpha)$, with ϕ and Φ the PDF and CDF of the standard normal distribution:
$$\mathrm{ES}_\alpha(X) = \frac{\sigma_r}{1-\alpha}\int_\alpha^1 z_m dm - \mu_r$$
  and since: $\int_\alpha^1 \Phi^{-1}(m)dm = \phi(\Phi^{-1}(\alpha))$, it follows that:
$$\mathrm{ES}_\alpha(X) = \frac{\sigma_r}{1-\alpha}\phi(z_\alpha) - \mu_r$$

(μ, ES)-optimization

- ES return:
$$\mathrm{ES}(\alpha) = \frac{\phi(z_\alpha)}{1-\alpha}\sigma_{r_w} - \mu_{r_w}$$
  Note that ES is linear in σ_r and μ_r.
- Optimization problem:
$$\min_{\boldsymbol{w}\in\mathcal{W}} \frac{\phi(z_\alpha)}{1-\alpha}\sigma_{r_w} - \mu_{r_w} \qquad \text{s.t.}\mathbb{E}[r_w] = \bar{r}$$

Implications of the ES constraints

- An ES-constraint reduces the possible set of (μ,σ)-efficient portfolios fulfilling the VaR constraint with the same confidence level.
- Even for very low $(1-\alpha$, the difference between $\mathcal{B}_V(\alpha)$ and $\mathcal{B}_E(\alpha)$ is substantial.
- ES-constraints tend to preclude portfoliows with large volatility, but also portfolios with small volatility.
  Depending on the ES-level and the risk-aversion, the investor will move to less risky or riskier portfolios.

2.2.3 Comments on VaR and ES

Economic implications

- The MVP is mean-VaR and mean-ES inefficient for every $1-\alpha > 0$.
- As $1-\alpha \searrow 0$ the MVaRP and MESP converge to the MVP.
- Note that in general: $\mathrm{ES}_\alpha(X) \geq \mathrm{VaR}_\alpha(X)$
  (which can easily be proved)

Leaving normality

- Return distributions are characterized by *fat tails*.
  Now: assume returns follow a T-distribution, $\mathcal{T}(\mu,\Omega,\lambda)$ with λ degrees of freedom.
- When $r_w \sim \mathcal{T}(\boldsymbol{w}^\top\boldsymbol{\mu}, \boldsymbol{w}^\top\Omega\boldsymbol{w}, \lambda)$ and $\alpha > 0.5$, the VaR and ES are linear functions of the portfolio variance $(\boldsymbol{w}^\top\Sigma\boldsymbol{w})$ and expected return $(\boldsymbol{w}^\top\boldsymbol{\mu})$.
  For the VaR and ES to be larger than in the normal case, we need α to be below some critical value.

2.3 Coherent Risk Measures

Coherent risk measures A coherent risk measure ρ for portfolio X has the following properties:

- **Subadditivity:** $\rho(X) + \rho(Y) \geq \rho(X+Y)$
  i.e. the total risk of two separate portfolios is never smaller than the risk of the two portfolios together (*diversification* effect)
- **Homogeneity:** $\rho(\lambda X) = \lambda\rho(X)$
  i.e. increase in leverage of any position leads to a proportional increase in risk (debatable in presence of liquidity risk)
- **Monotonicity:** $Y \succ X \Rightarrow \rho(Y) \leq \rho(X)$
  ($Y \succ X$ means that Y is superior to X in all states of the world)
  i.e. if one portfolio is never worse than another one, then its risk should be smaller or equal to the risk of the second one
- **Translation invariance:** $\rho(X+m) = \rho(X) - m$, w/ cash position m
  i.e. adding cash to the portfolio reduces the risk by this amount

Remarks:

- *Expected Shortfall* ES_α is a coherent risk measure, i.e. it fulfils all four properties.
- But *Value-at-Risk* VaR_α is *not* a coherent risk measure, since it fails to fulfil the *subadditivity* property.
  As a consequence, VaR_α might discourage diversification.

Spectral risk measures Given a non-increasing density function ϕ on $(0,1)$ and a profit and loss distribution F, a *spectral risk measure* is defined by
$$\rho_\phi(X) = -\int_0^1 F^{-1}(m)\phi(m)dm$$
where $\phi(\cdot)$ is called the **risk spectrum**.

Remarks:

- Cases:
  - $\phi =$ const.: risk-neutral investors w.r.t. losses
  - ϕ decreasing: risk-averse investors
- Further **requirements on spectral risk measures** (Balbas et al., 2009):
  - $\phi(m) > 0, \forall m \in (0,1)$
    i.e. positive weights on any possible outcome of the profit and loss distribution
  - For $0 \leq \alpha \leq \beta \leq 1$, we must have $\phi(\alpha) > \phi(\beta)$
    i.e. larger weights on worse outcomes, ϕ is strictly decreasing
  - $\lim_{m\to 0}\phi(m) = \infty$ and $\lim_{m\to 1}\phi(m) = 0$
    i.e. put infinite mass on the worst possible outcome and zero mass on the best possible outcome
- ES does not use information in a large part of a loss distribution, e.g. it fails to properly adjust for *extreme low-probability losses*.
- Although ES fulfills all four properties of a coherent risk measure, it fails to fulfill any of the three additional requirements for spectral risk measures.
- Further spectral risk measures:
  - *Exponential* spectral risk measure:
$$\phi_{\text{exp}}(m) = \frac{ae^{-am}}{1-e^{-a}}, \qquad a \geq 0, m \in (0,1)$$
  - *Power* spectral risk measure:
$$\phi_{\text{pow}}(m) = bm^{b-1}, \qquad b \in (0,1], m \in (0,1)$$

– *Wang transformation:*

$$\phi_W(m) = \frac{n(N^{-1}(m) - N^{-1}(\zeta))}{n(N^{-1}(m))}, \quad \zeta \in \left(0, \frac{1}{2}\right), m \in (0,1)$$

with $n(\cdot)$ and $N(\cdot)$ the PDF and CDF of the standard normal distribution.
Risk aversion is reflected by the fact that gains obtain a lower weight than losses.

– *Remark:* ϕ_{exp} and ϕ_{pow} fail to fulfill the third additional property, but the Wang transformation ϕ_W fulfills all properties.

3 Resampling and Robust Portfolio Optimization

3.1 Resampling Methods

Advantages of resampling

- Resampling is most effective when *correcting for errors in means.*
- Resampled portfolio weights change in a *smooth* way as risk tolerance changes.
- Portfolios based on resampling are close in $\mu-\sigma$ space to standard frontier portfolios (but they are far apart in "weight-space").

Drawbacks of resampling

- The resampling approach misses some aspects of the additional risk that comes from *sampling error*.
- *Short-sales constraints* may return peculiar statics.
- Upward bending frontier not plausible, otherwise linear combination of assts possible that yields a better risk-return trade-off.
- Averaging weights with resampling leads to *greater diversification* than is theoretically optimal.
  As one approaches the higher risk portfolios, the over-diversification may diminish, leading the resampled frontier to be slightly convex.
- Resampling can also change the *maximum Sharpe ratio:*
  Risk averse investors may increase cash; resampling overallocates to volatile assets.
- If there are *short-selling constraints*, the average weight on the assets may go up.

Resampling Method

- Given: estimates $\boldsymbol{\mu}_0^*$ and Σ_0^* based on T observations of excess returns.
- Procedure:
  (i) Generate: $(\boldsymbol{\mu}_1^*, \Sigma_1^*)$, $(\boldsymbol{\mu}_2^*, \Sigma_2^*)$, ..., $(\boldsymbol{\mu}_m^*, \Sigma_m^*)$
  by: drawing m times from the (joint normal) distribution given by $\boldsymbol{\mu}_0^*$ and Σ_0^*.
  (ii) For each $i = 1, \ldots, m$, calculate portfolio weights $\boldsymbol{w}_{ij}$ based on $(\boldsymbol{\mu}_i^*, \Sigma_i^*)$ for set of target μ_j^p, $j = 1, \ldots, P$.
  (iii) Evaluate:
  $$\sigma_{ij}^p = \sqrt{\boldsymbol{w}_{ij}^\top \Sigma_0^* \boldsymbol{w}_{ij}}, \qquad \mu_{ij}^p = \boldsymbol{w}_{ij}^\top \boldsymbol{\mu}_0^*$$
  (iv) Determine resampled weights by averaging $\boldsymbol{w}_{ij}$ for each portfolio on the MVF, i.e. for the l^{th} portfolio:
  $$\tilde{\boldsymbol{w}}_l = \frac{1}{m}\sum_{i=1}^{m} \boldsymbol{w}_{il}$$

Remarks

- Resulting portfolio mean-standard deviation pairs $(\mu_{ij}^p, \sigma_{ij}^p)$ lie below the MVF given by $(\mu_{0j}^p, \sigma_{0j}^p)$.
- Weights w_{ij} and w_{0j} are statistically equivalent.
- Comparing simple MV with resampling:
  – In the *simple MV* optimization: only a few assets play a large role in the portfolio and weights jump as risk tolerance increases.
  – Using *resampling:* resampled portfolio weights change in a smooth way as risk tolerance changes ,i.e. weight graph is smoothened, kinks of simple MV are removed/reduced.
  – While large differences may arise to weights, the efficient frontiers are very close.
- Conclusions:
  – Resampling is most effective when correcting for errors in means.
  – Portfolios based on resampling are close in $\mu-\sigma$ space to the standard MVF, but they are far apart in weight space.
- **Concerns:**
  – Portfolio weight averages may reflect a few extreme observations.
  – If there are *short-selling constraints*, the average weight on the asset may go up.

Distance of Portfolio Weights

- Define reference portfolio $\boldsymbol{w}_p$.
- Calculate *test statistic* for all resampled portfolio weights.
  Possible test statitistics:
  – Variation in portfolio weigths $\boldsymbol{w} \in \mathbb{R}^{n\times 1}$:
    * Compute covariance matrix Ω from the dataset of weight vectors.
    * Compute quadratic form $(\boldsymbol{w}_p - \boldsymbol{w}_i)^\top \Omega^{-1} (\boldsymbol{w}_p - \boldsymbol{w}_i)$, which should be distributed χ_2 with k-degress of freedom.
    * Intuition: small weight differences for highly correlated assets might be of greater significance than large weight differences for assets with negative correlation.
    * Issue: long-only constraint invalidates the normality assumption.
  – Distance to return distribution:
  $$(\boldsymbol{w}_p - \boldsymbol{w}_i)^\top \Sigma_0^* (\boldsymbol{w}_p - \boldsymbol{w}_i)$$
  which is equivalent to the *squared tracking error* (volatility of return differences between portfolios $\boldsymbol{w}_i$ and $\boldsymbol{w}_p$).
- See if the test statistic based on an optimal set of weights (assuming no sampling error) is greater than the α-quantile of the test statistic distribution.
- Result: statement whether the optimized weights are statistically significantly different from the current portfolio.

3.2 Constrained Optimization

Motivation for constrained optimization: estimation, legal, preferences, ...

Formulation of constrainted optimization

$$\max_{\boldsymbol{w}} \boldsymbol{w}^\top \boldsymbol{\mu} - \frac{\lambda}{2}\boldsymbol{w}^\top \Sigma_0 \boldsymbol{w}, \qquad \text{subject to: } \boldsymbol{w}_{\text{lower}} \le A\boldsymbol{w} \le \boldsymbol{w}_{\text{upper}}$$

Lagrangian

- Standard MV setting:
  $$\mathcal{L}(\boldsymbol{w}, \lambda_0) = \frac{1}{2}\boldsymbol{w}^\top \Sigma \boldsymbol{w} - \lambda_0(\mathbf{1}^\top \boldsymbol{w} - 1)$$
  which yields the solution:
  $$\boldsymbol{w}^* = \Sigma^{-1}\mathbf{1}/(\mathbf{1}^\top \Sigma \mathbf{1}) = \boldsymbol{w}^*(\Sigma)$$

- With *constraints* $\mathcal{C}(\boldsymbol{w}^-, \boldsymbol{w}^+)$, with $\boldsymbol{w}^- \leq \boldsymbol{0} \leq \boldsymbol{w}^+$:

$$\mathcal{L}(\boldsymbol{w}, \lambda_0, \boldsymbol{\lambda}^-, \boldsymbol{\lambda}^+) = \mathcal{L}(\boldsymbol{w}, \lambda_0) - (\boldsymbol{\lambda}^-)^\top (\boldsymbol{w} - \boldsymbol{w}^-) - (\boldsymbol{\lambda}^+)^\top (\boldsymbol{w}^+ - \boldsymbol{w})$$

  for which we get the *Kuhn-Tucker conditions:*

$$\begin{aligned} \Sigma \boldsymbol{w} - \lambda_0 \boldsymbol{1} - \boldsymbol{\lambda}^- + \boldsymbol{\lambda}^+ &= \boldsymbol{0} \\ \boldsymbol{1}^\top \boldsymbol{w} - 1 &= 0 && \text{(full investment)} \\ \min(\lambda_i^-, w_i - w_i^-) &= 0 && \text{(lower boundaries)} \\ \min(\lambda_i^+, w_i^+ - w_i) &= 0 && \text{(upper boundaries)} \end{aligned}$$

  with $\lambda_i^+, \lambda_i^-, \lambda_0 \geq 0 \quad \forall i$.

Effect of constraints

- Given a constrained portfolio $\tilde{\boldsymbol{w}}$, we can find a covariance matrix $\tilde{\Sigma}$ s.t. $\tilde{\boldsymbol{w}}$ is the solution of a GMV portfolio.
- Let $\tilde{\Sigma} = \Sigma + (\boldsymbol{\lambda}^+ - \boldsymbol{\lambda}^-)\boldsymbol{1}^\top + \boldsymbol{1}(\boldsymbol{\lambda}^+ - \boldsymbol{\lambda}^-)$.
  Then $\tilde{\Sigma} > 0$ and $\tilde{\boldsymbol{w}} = \boldsymbol{w}^*(\tilde{\Sigma})$.
- Interpretation: the implied voariance matrix can be interpreted as *perturbed matrix* of the form:

$$\tilde{\Sigma}_{ij} = \Sigma_{ij} + \Delta_{ij}$$

  Pertubation is:
  - null if no constraint is violated,
  - positive if one asset reaches the upper bound,
  - negative if one asset reaches the lower bound.
- Impact on *volatility:* $\tilde{\sigma}_i = \sqrt{\sigma_i^2 + \Delta_{ii}}$.
  Impact on *correlation:* direction of the impact less obvious.

3.3 Robust Optimization

Advantage of robust optimization

- Robust optimization *considers uncertainty* in unknown parameters directly and explicity in the optimization problem.
- Robust optimization ensures a certain performance/limits underperformance in *worst cases*. It is generally concerned with ensuring that decision are adequate even if estimates of input parameters are incorrect.

Drawbacks of robust optimization

- Interesting at first sight, but actually does not add much within the MV framework.
- Practitioners feel that robust optimization is a *conservative* and hence *prudent* (if not even *overly pessimistic*) form of portfolio construction.
  - e.g. if *cash* is included:
    We will end up with a 100% cash holding for any formulation as long as we look deep enough into the estimation error tail.
- *Estimation error* is still built into the optimization process.
- It is formally *equivalent to shrinkage estimators* / very narrow *Bayesian priors*.
- *Computational difficulties* arise already when linear constraints (e.g. long-only) are added.

Three different frontiers i.e.

Frontier	returns $\boldsymbol{\mu}$	covariance Σ	weights $\boldsymbol{w}$
True frontier (TF)	true	true	implicit
Estimated frontier (EF)	estimates	true	implicit
Actual frontier (AF)	true	true	from EF

Remarks:

- Both true expected returns and true covariance matrix are unobservable.
- The AF will always lie below the TF.

3.3.1 First model

Model

- Assumption: investors are ambiguous about correct covariance matrix and correct mean vector.
  Thus: $\exists$ a set of *candidate mean vectors* $\boldsymbol{\mu} \in \mathcal{S}_\mu$ and *candidate covariance matrices* $\Sigma \in \mathcal{S}_\Sigma$.
- All matrices are given equal importance.
- Then, the optimization problem reads:

$$\max_{\boldsymbol{w}} \left(\min_{\boldsymbol{\mu} \in \mathcal{S}_\mu, \Sigma \in \mathcal{S}_\Sigma} \left(\boldsymbol{w}^\top \boldsymbol{\mu} - \frac{\lambda}{2} \boldsymbol{w}^\top \Sigma \boldsymbol{w} \right) \right)$$

- Interpretation:
  Maximization of the worst case (quadratic) utility for all combinations of $\boldsymbol{\mu} \in \mathcal{S}_\mu$ and $\Sigma \in \mathcal{S}_\Sigma$.
- In reality: this means being very pessimistic (since the solution even has a good outcome in the worst case).

Robust optimization under long-only constraints

- Under long-only constraints, the max-min optimization is equivalent to:

$$\max_{\boldsymbol{w} \geq \boldsymbol{0}} \boldsymbol{w}^\top \boldsymbol{\mu}_l - \frac{\lambda}{2} \boldsymbol{w}^\top \Sigma_h \boldsymbol{w}$$

  where $\boldsymbol{\mu}_l$ the worst-case return vector (i.e. the smallest element in $\mathcal{S}_\mu$) and Σ_h the worst case covariance matrix (i.e. the largest element in $\mathcal{S}_\Sigma$).
- Construction: apply bootstrapping
  - Simulate m mean return vectors and covariance matrices from the original $\boldsymbol{\mu}_0$ and Σ_0.
  - Construction of Σ_h: select for each element in Σ_h e.g. the highest 5% entry across all m matrices, and ensure that $\Sigma_h \geq \Sigma_0$.
    It follows: $\boldsymbol{w}^\top \Sigma_h \boldsymbol{w} - \boldsymbol{w}^\top \Sigma_0 \boldsymbol{w} \geq 0$, i.e. Σ_h is riskier.
  - Construction of μ_h: idem, i.e. select e.g. the lowest 5% entries across all m return vectors.

3.3.2 Second model (more general approach)

Model

- If estimated returns $\bar{\boldsymbol{\mu}} \in \mathbb{R}^n$ have covariance $\Omega \in \mathbb{R}^{n \times n}$, then the 100η%-confidence region is defined as

$$(\boldsymbol{\mu} - \bar{\boldsymbol{\mu}})^\top \Omega^{-1} (\boldsymbol{\mu} - \bar{\boldsymbol{\mu}}) \leq \kappa^2$$

with $\kappa^2 = \chi_n^2(1-\eta)$, where χ^2 is the inverse cumulative distribution function of the chi-squared distribution with n degrees of freedom (n the numbers of assets).

- Difference between true returns and actual returns: $\bar{\boldsymbol{\mu}}^\top \bar{\boldsymbol{w}} - \boldsymbol{\mu}^\top \bar{\boldsymbol{w}}$
- The lowest possible value of the actual expected return of the portfolio over the given confidence region (100η%) of true expected return can be found via the following maximization problem:

$$\max_{\boldsymbol{\mu}} \underbrace{\bar{\boldsymbol{\mu}}^\top \bar{\boldsymbol{w}}}_{\text{est. front.}} - \underbrace{\boldsymbol{\mu}^\top \bar{\boldsymbol{w}}}_{\text{actual front.}}, \qquad \text{s.t. } (\boldsymbol{\mu} - \bar{\boldsymbol{\mu}})^\top \Omega^{-1} (\boldsymbol{\mu} - \bar{\boldsymbol{\mu}}) \leq \kappa^2$$

with optimal solution:

$$\boldsymbol{\mu} = \bar{\boldsymbol{\mu}} - \sqrt{\frac{\kappa^2}{\bar{\boldsymbol{w}}^\top \Omega \bar{\boldsymbol{w}}}} \Omega \bar{\boldsymbol{w}}$$

and we thus get:

$$\boldsymbol{\mu}^\top \bar{\boldsymbol{w}} = \bar{\boldsymbol{\mu}}^\top \bar{\boldsymbol{w}} - \underbrace{\kappa \sqrt{\bar{\boldsymbol{w}}^\top \Omega \bar{\boldsymbol{w}}}}_{\text{maximum difference}}$$

- New portfolio optimization problem:

$$\max_{\boldsymbol{w}} \underbrace{\bar{\boldsymbol{\mu}}^\top \bar{\boldsymbol{w}}}_{\text{est. front.}} - \underbrace{\kappa \sqrt{\bar{\boldsymbol{w}}^\top \Omega \bar{\boldsymbol{w}}}}_{\text{maximum difference}},$$

$$\text{s.t. } \underbrace{\mathbf{1}^\top \boldsymbol{w} = 1}_{\text{full inv.}}, \underbrace{\boldsymbol{w}^\top \Sigma \boldsymbol{w} \leq \sigma_P^2}_{\text{given risk}}, \underbrace{\boldsymbol{w} \geq 0}_{\text{no short-sales}}$$

Impact of robust optimization Consider the actual return and the maximum difference between actual and estimated portfolio return:

$$\boldsymbol{\mu} = \bar{\boldsymbol{\mu}} - \sqrt{\frac{\kappa^2}{\bar{\boldsymbol{w}}^\top \Omega \bar{\boldsymbol{w}}}} \Omega \bar{\boldsymbol{w}}, \qquad \bar{\boldsymbol{\mu}}^\top \bar{\boldsymbol{w}} - \boldsymbol{\mu}^\top \bar{\boldsymbol{w}} = \kappa \frac{\bar{\boldsymbol{w}}^\top \Omega \bar{\boldsymbol{w}}}{\sqrt{\bar{\boldsymbol{w}}^\top \Omega \bar{\boldsymbol{w}}}}$$

to find:

- Expected returns of assets with positive weights will be adjusted downwards.
- Expected returns of assets with negative weights (i.e. short holdings) will be adjusted upwards.
- Size of the adjustment: controlled by κ, i.e. the size of the confidence region.

Solution (with short-sales) If only the full investment constraint $\mathbf{1}^\top \boldsymbol{w} = 1$ is applied, we get:

$$\boldsymbol{w}^*_{\text{rob}} = \underbrace{\frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}^\top \Sigma^{-1} \mathbf{1}}}_{\boldsymbol{w}_{\text{GMV}}} + \underbrace{\left(1 - \frac{T^{-1/2}\kappa}{\lambda \sigma_p^* + T^{-1/2}\kappa}\right)}_{<1 \Rightarrow \text{ shrinking}} \underbrace{\frac{1}{\lambda} \Sigma^{-1} \left(\bar{\boldsymbol{\mu}} - \frac{\bar{\boldsymbol{\mu}}^\top \Sigma^{-1} \mathbf{1}}{\mathbf{1}^\top \Sigma^{-1} \mathbf{1}} \mathbf{1}\right)}_{\boldsymbol{w}_{\text{spec}}}$$

with T the number of observations and σ_p^* the standard deviation of the optimal robust portfolio.

Remarks:

- For low required confidence levels $\kappa \to 0$ and many data points $T \to \infty$, we get $\boldsymbol{w}^*_{\text{rob}} = \boldsymbol{w}^*_{\text{MV}}$.
- One could say that the portfolio weigths are robustified by shrinking them to the glboal minimum variance portfolio $\boldsymbol{w}_{\text{GMV}}$

4 Risk Budgeting

Advantages/Motivation

- During the financial crisis: risk contribution of equities far exceeded their forecast limits. Part of this surprise was due to the jump in realised equity correlation during the crisis.
- **Risk budgeting (RB):**
  - is a purely heuristic approach
  - treats equity as carrier of risk premia
  - decideds only on the risk to take ($\leadsto$ diversification) — return is achieved via leverage
- While asset returns can be highly correlated, *risk factors* are not likely to be nearly as highly correlated.

Drawbacks of risk budgeting

- Risk budgeting does not satisfy *duplication invariance* and the *polico invariance property*.

4.1 Risk Budgeting Portfolio

Risk measures/risk contributions

- Let $\mathrm{R}(\boldsymbol{w})$ be a homogeneous *risk measure* for portfolio $\boldsymbol{w}$ with assets $i = 1, \ldots, n$
  Denote by $\mathrm{RC}_i(\boldsymbol{w})$ the corresponding *risk contribution* of asset i.
  This can be written as:

$$\mathrm{R}(\boldsymbol{w}) = \sum_{i=1}^n \mathrm{RC}_i(\boldsymbol{w}) = \sum_{i=1}^n \underbrace{w_i \frac{\partial \mathrm{R}(\boldsymbol{w})}{\partial w_i}}_{=\mathrm{RC}_i(\boldsymbol{w})}$$

- Then a set of given *risk budgets* $\{b_i\}_{i=1,\ldots,n}$ can be expressed as:

$$\mathrm{RC}_i(\boldsymbol{w}) = b_i \, \mathrm{R}(\boldsymbol{w}), \qquad \mathbf{1}^\top \boldsymbol{b} = 1, \qquad \text{i.e. } b_i = \frac{\mathrm{RC}_i(\boldsymbol{w})}{\mathrm{R}(\boldsymbol{w})}$$

  Often the constraint $b_i > 0, \forall i$ is imposed IOT avoid some common issues.

Quadratic optimization problem To solve for a *long-only* portfolio that fulfills the risk budgeting constraint, one can solve:

$$\boldsymbol{w}^* = \arg\min f(\boldsymbol{w}, \boldsymbol{b}), \qquad \text{s.t. } \mathbf{1}^\top \boldsymbol{w} = 1, w_i \in [0,1], \forall i$$

with

$$f(\boldsymbol{w}, \boldsymbol{b}) = \sum_{i=1}^n \left(\underbrace{w_i \partial_{w_i} \mathrm{R}(\boldsymbol{w})}_{\mathrm{RC}_i(\boldsymbol{w})} - \underbrace{b_i \, \mathrm{R}(\boldsymbol{w})}_{\approx \mathrm{RC}_i(\boldsymbol{w})} \right)^2$$

or alternatively:

$$f(\boldsymbol{w}, \boldsymbol{b}) = \sum_{j=1}^n \sum_{i=1}^n \left(\underbrace{\frac{w_i \partial_{w_i} \mathrm{R}(\boldsymbol{w})}{b_i}}_{\approx \mathrm{R}(\boldsymbol{w})} - \underbrace{\frac{b_i \, \mathrm{R}(\boldsymbol{w})}{b_j}}_{\approx \mathrm{R}(\boldsymbol{w})} \right)^2$$

4.2 Equal Risk Contribution

Equal risk contribution (ERC) portfolio

- The ERC portfolio is a special case of the RB portfolio:

$$b_i = \frac{1}{n}, \forall i$$

  i.e. each asset gets assigned the same risk budget/the manager has a neutral view on the risk budgets.
- Consider special cases from the slides ...
- Comparison between MVP, the equally weighted portfolio (EW) and the ERC portfolio:

$$\sigma(\boldsymbol{w}_{\text{MV}}) \leq \sigma(\boldsymbol{w}_{\text{ERC}}) \leq \sigma(\boldsymbol{w}_{\text{EW}})$$

- Optimality of the ERC portfolio: iff the Sharpe ratio of all assets is the same, then the ERC portfolio is optimal.

4.3 Diversification Revisited

Diversification

- *Diversification index:*

$$\mathrm{D}(\boldsymbol{w}) = \frac{\mathrm{R}(\boldsymbol{w})}{\sum_{i=1}^{n} w_i \, \mathrm{R}(\boldsymbol{I}_i)} \leq 1$$

- *Diversification ratio:*

$$\mathrm{DR}(\boldsymbol{w}) = \frac{\sum_{i=1}^{n} w_i \sigma_i}{\sigma(\boldsymbol{w})} = \frac{\boldsymbol{w}^\top \boldsymbol{\sigma}}{\sqrt{\boldsymbol{w}^\top \Sigma \boldsymbol{w}}} \geq 1$$

  Interpretation: The DR equals the weighted-average volatility of the individual assets divided by the actual risk of a portfolio. Due to diversification, the denominator (actual portfolio risk) will always be less or equal to the numerator.

- For a general risk measure $R(\boldsymbol{w})$, where the risk associated to the i^{th} asset is denoted as R_i, the diversification ratio is defined as:

$$\mathrm{DR}(\boldsymbol{w}) = \frac{\sum_{i=1}^{n} w_i \, \mathrm{R}_i}{\mathrm{R}(\boldsymbol{w})}$$

- The upper bound w.r.t. a general risk measure $\mathrm{R}(\boldsymbol{w})$ is given by:

$$\mathrm{DR}(\boldsymbol{w}) = \frac{\sup \mathrm{R}_i}{\mathrm{R}(\boldsymbol{w}_{\mathsf{mr}})}$$

  where $\boldsymbol{w}_{\mathsf{mr}}$ is the portfolio that minimizes the risk measure.

Long-only most diversified portfolio (MDP)

- Long-only MDP:

$$\boldsymbol{w}^* = \arg\max \log \mathrm{DR}(\boldsymbol{w}), \qquad \text{s.t. } \mathbf{1}^\top \boldsymbol{w} = 1, \mathbf{0} \leq w \leq \mathbf{1}$$

- The long-only MDP is the long-only portfolio s.t. the correlation between any other portfolio and itself is at least as high as the ratio of their diversification ratios.
- All stock belonging to the MDP have the same correlation to it.
- If all assets have the same Sharpe ratio, then $\mathrm{DR}(\boldsymbol{w})$ equals the Sharpe Ratio of the portfolio divided by the Sharpe Ratio of the assets.

Correlations with the MDP

- An arbitrary portfolio $\boldsymbol{w}$ has the following correlation with the *long-only* MDP $\boldsymbol{w}^*_{\mathsf{LO}}$ and the *unconstrained* MDP $\boldsymbol{w}^*_{\mathsf{UC}}$:

$$\rho(\boldsymbol{w}, \boldsymbol{w}^*_{\mathsf{LO}}) \geq \frac{\mathrm{DR}(\boldsymbol{w})}{\mathrm{DR}(\boldsymbol{w}^*_{\mathsf{LO}})}, \qquad \rho(\boldsymbol{w}, \boldsymbol{w}^*_{\mathsf{UC}}) = \frac{\mathrm{DR}(\boldsymbol{w})}{\mathrm{DR}(\boldsymbol{w}^*_{\mathsf{UC}})}$$

  Hence, the long-only MDP is the long-only portfolio s.t. the correlation between any other long-only portfolio and itself is at least as high as the ratio of their diversification ratios.

Other concentration indexes If π are the portfolio weights of a long-only portfolio:

- *Herfindahl Index:*

$$H(\pi) = \sum_{i=1}^{n} \pi_i^2, \qquad H^*(\pi) = \frac{nH(\pi) - 1}{n - 1} \quad \text{(normalized)}$$

  - minimum: $H(\pi) = \frac{1}{n}$ for the equally weighted portfolio $\pi_i = \frac{1}{n}$
  - maximum: $H(\pi) = 1$ if the portfolio π is concentrated in only one asset

- *Gini coefficient*

$$G(\pi) = \frac{A}{A + B}, \qquad G = 1 - 2\int_0^1 L(x)dx$$

  with $L(x)$ the Lorenz curve.

  - A is the area between the line indicating no concentration (i.e. the diagonal, $L_{\pi^-}(x)$) and the Lorenz curve (i.e. the line indicating the actual concentration, $L(x)$)
  - B is the area below the Lorenz curve (i.e. between the Lorenz curve and the line indicating perfect concentration, $L_{\pi^+_\infty}(x)$).
  - No concentration: $G(\pi) = 0$.
    Perfect concentration: $G(\pi) = 1$
    $\rightsquigarrow$ i.e. the lower the Gini coefficient the better the diversification/equality — the higher the Gini coefficient the higher the inequality

- *Shannon entropy:*

$$I(\pi) = -\sum_{i=1}^{n} \pi_i \log \pi_i$$

5 Black-Litterman Model

5.1 Introduction

Advantages/Motivation

- Its main practical value is the *"view-expressing scheme"*.
- The BL model is related to *Bayesian* analysis since it provides a sensible way to dampen down over-fitting and to incorporate information that the investor may possesses/investors' priors.
- The starting point is NOT the historical sample average returns but returns *implicit in market allocations* $\rightarrow$ *CAPM equilibrium market portfolio* $\rightarrow$ direct connection to the market
- The BL model can also be extended to *non-normal* returns.

- **Intuition:**
  - in case of no views: (CAPM) market portfolio results again
  - investors' views tilt the resulting weights away from the market portfolio, depending on the confidence in views

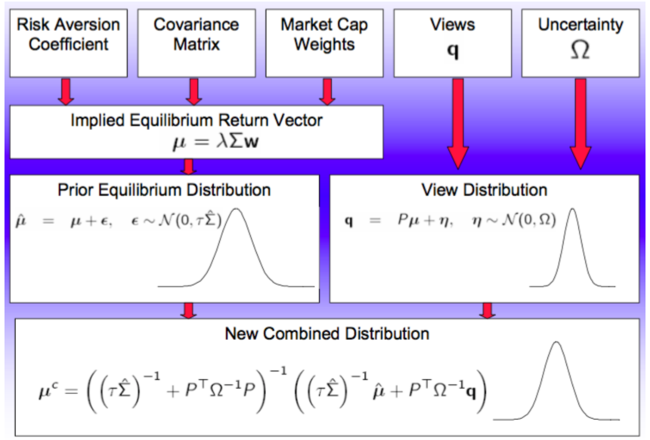
Drawback of the BL model

- It ignores information in realized returns (since it depends only on equilibrium returns and views)
  $\rightsquigarrow$ e.g. choose a starting point different from CAPM

Outcome of BL approach

- A single view causes the return of every asset in the portfolio to change from its implied equilibrium return
  (since each individual return is linked to the other returns via the covariance matrix of excess returns Σ).
  But: the weights of the assets on which no views were expressed remain unchanged.
- If further *investment constraints* are added: then the BL model becomes less intuitive.

BL roadmap The full approach:



5.2 The Math Behind Black-Litterman

Main assumptions

- **Market structure:**
  - n assets
  - expected return vector $\boldsymbol{\mu} \in \mathbb{R}^{n\times 1}$
  - expected covariance matrix $\Sigma \in \mathbb{R}^{n\times n}$
- **Investors:**
  - quadratic utility function / static optimization problem:

    $$\max_{\boldsymbol{w}} \left(\boldsymbol{w}^\top \boldsymbol{\mu} + (1-\boldsymbol{w}^\top \mathbf{1})R_f - \frac{\lambda}{2}\boldsymbol{w}^\top \Sigma \boldsymbol{w}\right)$$

    where λ is the risk aversion coefficient.

5.2.1 Step 1: Equilibrium as Reference Point

CAPM equilibrium return

- CAPM equilibrium returns are the current market collective forecasts of next period returns ($\rightsquigarrow$ "prediction markets").
- The market capitalization $\boldsymbol{w}$ (since we assume being in equilibrium) can be computed via:
  - Approach with excess returns $\boldsymbol{\mu} - \mathbf{1}R_f$:

    $$\boldsymbol{w}^* = \arg\max_{\boldsymbol{w}} \left(\boldsymbol{w}^\top \boldsymbol{\mu} + (1-\boldsymbol{w}^\top \mathbf{1})R_f - \frac{\lambda}{2}\boldsymbol{w}^\top \Sigma \boldsymbol{w}\right)$$

    which leads to:

    $$\boldsymbol{\mu} - \mathbf{1}R_f = \lambda\Sigma\boldsymbol{w}^* \quad \Rightarrow \quad \boldsymbol{w}^* = (\lambda\Sigma)^{-1}(\boldsymbol{\mu} - \mathbf{1}R_f)$$

  - Approach with returns $\boldsymbol{\mu}$ (cf. application section):

    $$\boldsymbol{w}^* = \arg\max_{\boldsymbol{w}} \left(\boldsymbol{w}^\top \boldsymbol{\mu} - \frac{\lambda}{2}\boldsymbol{w}^\top \Sigma \boldsymbol{w}\right), \qquad \lambda \approx 3$$

    which leads to:

    $$\boldsymbol{\mu} = \lambda\Sigma\boldsymbol{w}^* \quad \Rightarrow \quad \boldsymbol{w}^* = (\lambda\Sigma)^{-1}\boldsymbol{\mu}$$

- The risk aversion coefficient λ:
  (i) can be computed via the market's excess return μ_m and its variance σ_m^2:

  $$\lambda = \frac{\mu_m}{\sigma_m^2}$$

  (ii) or can be calibrated to the *historical Sharpe ratio*, e.g. often $\lambda \approx 3$.

Note on estimation errors

- Estimation cannot directly be derived (since equilibrium returns are not actually estimated).
- But the following is known:
  The estimation error of the means of returns should be less than the covariance fo the returns.
- Practical approach:
  Define estimation error proportional to the covariance matrix of returns via a scalar τ, e.g. $\tau\Sigma$ with $\tau < 1$.
- Setting of τ: (very different approaches)
  - $\tau \in [0.01, 0.05]$
    since: The uncertainty in the mean is less than the uncertainty in the variance, thus τ should be close to zero.
  - $\tau = 1$ (missing justification ...)
  - often $\tau = 0.3$ (another arbitrary choice)
  - $\tau = 1/\text{number of observations}$
    since: $\tau\Sigma$ is interpreted as the standard error of the estimate of the implied equilibrium return vector.

Prior distribution The prior distribution of expected returns is $\mathcal{N}(\boldsymbol{\mu}, \tau\hat{\Sigma})$, where $\hat{\Sigma}$ is the estimated covariance matrix.

5.2.2 Step 2: Our Views

Characteristics of views

- Each view is unique and uncorrelated with other views.
  ($\rightsquigarrow$ improved stability and simplification of the problem)
- A view on every asset is NOT required — views may even conflict.
- two types of views:
  - *absolute view:* sum of weights is one
  - *relative view:* sum of weights is zero
    Note: relative view weights on a group of assets can either be equal ($\rightsquigarrow \frac{1}{n}$) or account for relative market capitalization.

Formulation of views

- *view vector:*

  $$\boldsymbol{q} = P\boldsymbol{\mu} + \boldsymbol{\eta}, \qquad \eta \sim \mathcal{N}(0, \Omega), \qquad P \in \mathbb{R}^{k\times n}, \quad \Omega \in \mathbb{R}^{k\times k}$$

  i.e. there are k views for a market with n assets.
- *view portfolios $P\mu$:* each row represents a weight vector of n assets, i.e. $P\mu$ expresses our views via k view portfolios.
  Note that P is not required to be invertible.
- *confidence/covariance of view portfolios Ω:*
  there appear to be two approaches presented in the lecture:
  - directly based on *investors' confidences:*

    $$\Omega = \operatorname{diag}(\omega_1, \ldots, \omega_k)$$

  - based on *historical estimates:*

    $$\Omega = \tau \operatorname{diag}(\boldsymbol{p}_1\hat{\Sigma}\boldsymbol{p}_1^\top, \ldots, \boldsymbol{p}_k\hat{\Sigma}\boldsymbol{p}_k^\top)$$

    with $P = (\boldsymbol{p}_1^\top, \ldots, \boldsymbol{p}_k^\top)^\top$ and $\hat{\Sigma}$ the historical estimate of the real covariance matrix Σ.

5.2.3 Step 3: Combining Equilibrium and View

Derivation

- Given:
  - expected returns:

    $$\hat{\boldsymbol{\mu}} = \boldsymbol{\mu} + \boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(0, \tau\hat{\Sigma})$$

  - views:

    $$\boldsymbol{q} = P\boldsymbol{\mu} + \boldsymbol{\eta}, \qquad \boldsymbol{\eta} \sim \mathcal{N}(0, \Omega)$$

- Define the following:

  $$\boldsymbol{y} = (\hat{\boldsymbol{\mu}}, \boldsymbol{q})^\top \in \mathbb{R}^{n+k} \qquad X = (I, P^\top)^\top \in \mathbb{R}^{(k+n)\times n}$$
  $$\boldsymbol{u} = (\boldsymbol{\epsilon}, \boldsymbol{\eta})^\top \in \mathbb{R}^{n+k} \qquad \psi = \operatorname{diag}(\tau\hat{\Sigma}, \Omega) \in \mathbb{R}^{(n+k)\times(n+k)}$$

  with $\boldsymbol{u} \sim \mathcal{N}(0, \psi)$
- *Regression equation:*

  $$\boldsymbol{y} = X\boldsymbol{\mu} + \boldsymbol{u}$$

- *Generalized least-square estimator:*

  $$\boldsymbol{\mu}^c = (X^\top \psi^{-1} X)^{-1} X^\top \psi^{-1} \boldsymbol{y} \quad \in \mathbb{R}^n$$

BL Master Formula

$$\boldsymbol{\mu}^c = \underbrace{\left((\tau\hat{\Sigma})^{-1} + P^\top \Omega^{-1} P\right)^{-1}}_{\text{normalization factor}} \cdot \underbrace{\left((\tau\hat{\Sigma})^{-1}\hat{\boldsymbol{\mu}} + P^\top \Omega^{-1} \boldsymbol{q}\right)}_{\text{weighting factors: eq. returns \& views}}$$
$$= \hat{\boldsymbol{\mu}} + \tau\hat{\Sigma}P^\top \left(\tau P\hat{\Sigma}P^\top + \Omega\right)^{-1} (\boldsymbol{q} - P\hat{\boldsymbol{\mu}})$$

Variance of μ^c

$$(\sigma^c)^2 = \left((\tau\hat{\Sigma})^{-1} + P^\top \Omega^{-1} P\right)^{-1}$$

(which corresponds to the normalization factor in the BL master formula)

Limiting cases consider:

- case $\boldsymbol{\mu}^c = \hat{\boldsymbol{\mu}}$:
  - no view: $P = 0$
  - no confidence: $\Omega \to \infty$
  - no estimation error: $\tau \to 0$
- case $\boldsymbol{\mu}^c = P^{-1}\boldsymbol{q}$:
  - absolute confidence: $\Omega \to 0$
  - infinite estimation error: $\tau \to \infty$

5.2.4 Step 4: Optimization

Markowitz optimization with adjusted mean μ^c and given estimated covariance matrix $\hat{\Sigma}$, which gives the whole efficient frontier.

$$\begin{aligned}\boldsymbol{w}^c &= \frac{1}{\lambda}\hat{\Sigma}^{-1}\boldsymbol{\mu}^c \\ &= \boldsymbol{w} + P^\top\left(P\hat{\Sigma}P^\top + \frac{1}{\tau}\Omega\right)^{-1}\left(\frac{1}{\lambda}\boldsymbol{q} - P\hat{\Sigma}\boldsymbol{w}\right)\end{aligned}$$

Often, Ω is specified as $\Omega = \operatorname{diag}\left[P\hat{\Sigma}P^\top\right]\tau$ to make w^c independent of τ.

6 Bayesian Mean Variance Analysis and Shrinkage

Overview

- As the BL model, Bayesian analysis provides a method to incorporate an investor's prior information into the estimation of mean returns.
- In principle, incorporating prior information about the covariance matrix of returns is also possible.
- Including prior information reduces over-fitting and smoothes out the influence of the particular sample available.

6.1 Basic Bayes

Likelihood function

- given: data series Y and model with unknown parameter θ
- joint density function of Y for a given value of θ: $f(y_1, \dots, y_m|\theta)$
- vs. likelihood function: $L(\theta|y_1, \dots, y_m) = f(y_1, \dots, y_m|\theta)$

Bayes theorem

- Bayes rule:

  $$\mathbb{P}[E|D] = \frac{\mathbb{P}[E \cap D]}{\mathbb{P}[D]} = \frac{\mathbb{P}[D|E]}{\mathbb{P}[D]} \cdot \mathbb{P}[E]$$

  i.e.

  $$\mathbb{P}[E|D] \cdot \mathbb{P}[D] = \mathbb{P}[D|E] \cdot \mathbb{P}[E]$$

- Bayes theorem consists of the following elements:
  - $\mathbb{P}[D|E]$: *likelihood*
    i.e. the conditional probability of the new data given that the prior evidende E is true
  - $\mathbb{P}[D]$: *evidence*
    i.e. the unconditional probability of the additional data (new observation)
  - $\mathbb{P}[E]$: *prior* probability
    i.e. the prior belief
    i.e. the probability of the evidence before the additional data (new observation)
  - $\mathbb{P}[E|D]$: *posterior* probability
    i.e. probability of the evidence after the additional data (new observation)

  which can be summarized as:

  $$\text{posterior} = \frac{\text{likelihood}}{\text{evidence}}\text{prior}$$

- *Remarks:*
  - Updated posterior beliefs are the result of a tradeoff between prior and data distributions.
  - The degree of the tradeoff is determined by the strength of the prior and the amount of available data.

Priors

- *Informative* prior elicitation
  - modify sustantially the information contained in the sample
  - also enclose information about the spread of the distribution
- *Noninformative* prior distribution
  - vague or diffuse priors are often modeled via uniform distribution or Jeffrey's prior
  - may be improper (i.e. do not integrate to one) — although resulting densities are usually proper
- *Conjugate* prior distribution
  - choice of prior often governed by aim to obtain analytically tractable solution
  - $\rightsquigarrow$ conjugate prior distributions guarantee that the posterior distribution is of the same class as the prior distribution

6.2 A Simple Bayesian Model

- case: unknown mean $\mu \sim \mathcal{N}(\hat{\mu}, \hat{\sigma}_\mu^2)$, known variance σ_0^2
- aim: find mean $\hat{\mu}$ and variance $\hat{\sigma}_\mu^2$ of the unknown mean μ
- data distribution:

  $$X|\mu, \sigma_0^2 \sim \mathcal{N}(\mu, \sigma_0^2) = \frac{1}{\sqrt{2\pi\sigma_0^2}}\exp\left(-\frac{1}{2\sigma_0^2}(X-\mu)^2\right)$$

- normal prior distribution (for unknown mean):

  $$\mu|r, s^2 \sim \mathcal{N}(r, s^2)$$

  with r, s known
- after computing the posterior distribution $\mathbb{P}[\mu|\boldsymbol{x}] \propto \mathbb{P}[\boldsymbol{x}|\mu]\mathbb{P}[\mu]$, the following *point estimates* are obtained:

  $$\hat{\mu} = \frac{\frac{r}{s^2} + \frac{m\bar{x}}{\sigma_0^2}}{\frac{1}{s^2} + \frac{m}{\sigma_0^2}}, \qquad \hat{\sigma}_\mu^2 = \left(\frac{1}{s^2} + \frac{m}{\sigma_0^2}\right)^{-1}$$

  where r, s, σ_0^2 are known, $\bar{x}$ is the mean of the data set and m denotes the number of data points.
- Characteristics:
  - prior precision: $1/s^2$
  - data precision: m/σ_0^2
  - posterior precision: $1/s^2 + m/\sigma_0^2$
- Note that for an infinite amount of data, i.e. $m \to \infty$:

  $$\lim_{m\to\infty}\hat{\mu} = \bar{x}, \qquad \lim_{m\to\infty}\hat{\sigma}_\mu^2 = 0$$

- *Remarks:*
  - non-Gaussian prior could have been adopted as well
  - uniform prior: possible, but does not change the posterior from the sample mean
  - prior could also be placed on variance, but this is not important in practice

6.3 Bayesian Portfolio Selection

Excess returns

- Predictive return density of the yet unobserved next-period excess return R_{T+1}:

$$p(R_{T+1}|R) = \int p(R_{T+1}|\boldsymbol{\mu},\Sigma)p(\boldsymbol{\mu},\Sigma|R)d\boldsymbol{\mu}d\Sigma$$

where $R \in \mathbb{R}^{T\times n}$, $p(\boldsymbol{\mu},\Sigma|R)$ the joint posterior density of the two parameters of the multivariate normal and $p(R_{T+1}|\boldsymbol{\mu},\Sigma)$ is the multivariate normal density.

- *Remark:* averaging over the posterior distribution accounts for estimation risk.
- In the multivariate normal setup:

$$L(\boldsymbol{\mu},\Sigma|R) \propto |\Sigma|^{-T/2}\exp\left(-\frac{1}{2}\sum_{t=1}^{T}(R_t-\boldsymbol{\mu})^\top\Sigma^{-1}(R_t-\boldsymbol{\mu})\right)$$

Scenario 1: MV with diffuse (improper) priors

- Assume: investor has *no particular prior knowledge* of the distribution parameters $\boldsymbol{\mu}$ and Σ.
- typical choice: **Jeffreys' prior**

$$p(\boldsymbol{\mu},\Sigma) \propto |\Sigma|^{-(n+1)/2}$$

- Predicitive distribution of the excess return is a multivariate Student's t-distribution with $T-n$ degrees of freedom.
- Then: predictive mean and covariance matrix of returns:

$$\tilde{\boldsymbol{\mu}} = \hat{\boldsymbol{\mu}}, \qquad \tilde{\Sigma} = \frac{(1+1/T)(T-1))}{T-n-2}\hat{\Sigma}$$

with the sample estimate:

$$\hat{\Sigma} = \frac{1}{T-1}\sum_{t=1}^{T}(R_t-\hat{\boldsymbol{\mu}})(R_t-\hat{\boldsymbol{\mu}})^\top$$

- **Interpretation:**
Since the predictive expected return is not shrunk towards the prior mean, the full amount of any sampling error is transferred to the posterior mean.
Thus, scenario 1 is more appropriate when we do NOT suspect that the sample mean contains substaintial estimation errors.

Scenario 2: MV with proper priors

- Assume: investor has *informative beliefs* about the mean vector and the covariance matrix of return.
Here: *conjugate priors:*
  - conjugate prior of the mean vector of the normal distribution: multivariate normal:

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

  where τ determines the strength of the confidence in $\boldsymbol{\eta}$.
  (e.g. if $\tau = 0$, then the investor has no knowledge and the variance of μ becomes infinite, thus the means becomes uniformly distributed on $\mathbb{R}$)
  - conjugate prior of the unknown covariance matrix of the normal distribution: inverted Whishart distribution:

  $$\Sigma \sim \mathrm{IW}(\Omega,\nu)$$

  where ν refelcts the confidence in Ω.
- Then: predicitve distribution of next-period's excess returns are multivariate Student's t with:

$$\tilde{\boldsymbol{\mu}} = \frac{\tau}{T+\tau}\eta + \frac{T}{T+\tau}\hat{\boldsymbol{\mu}}$$
$$\tilde{\Sigma} = \frac{T+1}{T(\nu+n-1)}\left(\Omega+(T-1)\hat{\Sigma}+\frac{T\tau}{T+\tau}(\boldsymbol{\eta}-\hat{\boldsymbol{\mu}})(\boldsymbol{\eta}-\hat{\boldsymbol{\mu}})^\top\right)$$

- *Remark:* In contrast to scenario 1, the predictive mean and covariance matrix are not proportional to the sample estimates $\hat{\boldsymbol{\mu}}$ and $\hat{\Sigma}$.
- **Interpretation:**
The predictive mean is a weighted average for the prior mean $\boldsymbol{\eta}$ and the sample mean $\hat{\boldsymbol{\mu}}$.
Thus, the sample mean is shrunk towards the prior mean. The stronger the belief in the prior mean, the larger the degreee to which the prior mean influences the predictive mean.

6.4 Shrinking μ

Admissible estimator Let X be a RV with distribution depending on an unknown parameter μ lying in a parameter space Θ. Let δ denote an estimator of μ.

- An estimator δ_1 is *as good as* an estimator δ_2 if:

$$R(\mu,\delta_1) \le R(\mu,\delta_2), \forall\mu\in\Theta \tag{1}$$

- An estimator δ_1 is *better than* an estimator δ_2 if eq. (1) is satisfied and:

$$R(\mu,\delta_1) < R(\mu,\delta_2) \quad \text{for at least one } \mu\in\Theta.$$

- An estimator is said to be *admissible* if there exists no estimator which is better than that.
Otherwise, it is an *inadmissible* estimator.

Stein-James shrinkage estimator

- Consider $X_t \sim \mathcal{N}(\boldsymbol{\mu},\Sigma)$ with $X_t \in \mathbb{R}^n$, $n>2$, $t=1,\dots,T$.
- Stein-James shrinkage estimator:

$$\hat{\delta}_a = (1-w)\hat{\boldsymbol{\mu}} + w\boldsymbol{b}, \qquad w = \frac{a}{(\hat{\boldsymbol{\mu}}-\boldsymbol{b})^\top(\hat{\boldsymbol{\mu}}-\boldsymbol{b})}$$

with $\hat{\boldsymbol{\mu}} = \frac{1}{T}\sum_t X_t \sim \mathcal{N}(\boldsymbol{\mu},\Sigma/T)$ the sample mean, $\boldsymbol{b}$ any constant vector and a any scalar s.t.

$$0 < a < \frac{2}{T}(\mathrm{tr}(\Sigma)-2\lambda_1)$$

where λ_1 is the largest eigenvalue fo the matrix Σ, for which it must hold $(\mathrm{tr}(\Sigma)-2\lambda_1) > 0$.

- The MSE using $\hat{\boldsymbol{\delta}}_a$ is smaller than the MSE from using $\hat{\boldsymbol{\mu}}$.

7 Regime Swiching and Asset Allocation

Motivation

- A RS model allos the data to be drawn from two or more possible distributions (regimes).
- e.g. two regimes for international equity returns:
a *normal regime* and a *bear regime* with:
  - lower average returns
  - higher volatility
  - higher correlation

7.1 A Simple RS Model

Serially uncorrelated data

- Regression model *without* switching:

$$y_t = \beta x_t + e_t, \qquad e_t \sim \text{i.i.d.}\mathcal{N}(0,\sigma^2)$$

To estimate this model, simply maximize the log-likelihood function w.r.t. $beta$ and σ^2:

$$\log L = \sum_{t=1}^{T}\log f(y_t), \qquad f(y_t) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(y_t-\beta x_t)^2}{2\sigma^2}\right)$$

- Now: regression with *structural breaks:*

$$y_t = \beta_{S_t} x_t + e_t, \qquad e_t \sim \text{i.i.d.}\mathcal{N}(0, \sigma^2_{St})$$

with:

$$\beta_{S_t} = \beta_0(1 - S_t) + \beta_1 S_t, \qquad \sigma^2_{S_t} = \sigma^2_0(1 - S_t) + \sigma^2_1 S_t$$

and $S_t \in \{0, 1\}$.
If S_t is known for $t = 1, 2, \ldots, T$, the same log-likelihood function as before can be maximized w.r.t. $\beta_0, \beta_1, \sigma_0, \sigma_1$.

Markov switching

- If S_t is not known a priori, then the following two-step procedure has to be applied:
- *Step 1:* decompose the joint density of y_t and unobserved S_t:

$$f(y_t, S_t|\mathcal{F}_{t-1}) = f(y_t|S_t, \mathcal{F}_{t-1}) f(S_t|\mathcal{F}_{t-1})$$

After integrating S_t our of the joint density $f(y_t|\mathcal{F}_{t-1})$, the log-likelihood function is then given by:

$$\log L = \sum_{t=1}^{T} \log f(y_t|\mathcal{F}_{t-1})$$
$$= \sum_{t=1}^{T} \log \left(\sum_{S_t=0}^{1} f(y_t|S_t, \mathcal{F}_{t-1}) \mathbb{P}[S_t|\mathcal{F}_{t-1}] \right)$$

which is a weighted average of the conditional densities given $S_t = 0$ and $S_t = 1$.
To calculate $\mathbb{P}[S_t|\mathcal{F}_{t-1}]$, a priori assumptions need be made about the behaviour of S_t.

- **Independent Switching**
  - Assume: S_t evolves independently of its own past values.
  - possible specification:

$$\mathbb{P}[S_t = 1] = p = \frac{\exp(p_0)}{1 + \exp(p_0)}, \qquad \mathbb{P}[S_t = 0] = 1 - p$$

    where p_0 is an unconstrained parameter.
  - If S_t does not depend on any other exogenous parameter, then the log-likelihood function can simply be maximized w.r.t. $\beta_0, \beta_1, \sigma_0, \sigma_1$ and p_0.
  - If S_t evolves independently of its own past values, but depends on some exogenous variable Z_{t-1}, then e.g.

$$\mathbb{P}[S_t = 1|\mathcal{F}_{t-1}] = p_t = \frac{\exp(p_0 + Z_{t-1}p_1)}{1 + \exp(p_0 + Z_{t-1}p_1)}$$
$$\mathbb{P}[S_t = 0|\mathcal{F}_{t-1}] = 1 - p_t$$

    and the log-likelihood function can be maximized additionally w.r.t. p_1.

- **Markov Switching**
  - Assume: S_t depends on past values of S_t.
    Consider here: the simplest case of an r^{th} order Markov, i.e. a first-order Markov switching process for S_t.
  - Then: transition probabilities:

$$\mathbb{P}[S_t = 1|S_{t-1} = 1] = p = \frac{\exp(p_0)}{1 + \exp(p_0)}$$
$$\mathbb{P}[S_t = 0|S_{t-1} = 0] = q = \frac{\exp(q_0)}{1 + \exp(q_0)}$$

  - Apply the following **filter:**
    (i) Given $\mathbb{P}[S_{t-1} = i|\mathcal{F}_{t-1}]$, $i = 0, 1$ at the beginning of time t, the weighting terms are calculated as:

$$\mathbb{P}[S_t = j|\mathcal{F}_{t-1}] = \sum_{i=0}^{1} \mathbb{P}[S_t = j|S_{t-1} = i]\mathbb{P}[S_{t-1} = i|\mathcal{F}_{t-1}]$$

    (ii) Once y_t is observed at the end of time t, update the probability term:

$$\mathbb{P}[S_t = j|\mathcal{F}_t] = \frac{f(y_t|S_t = j, \mathcal{F}_{t-1})\mathbb{P}(S_t = j|\mathcal{F}_{t-1})}{\sum_{k=0}^{1} f(y_t|S_t = k, \mathcal{F}_{t-1})\mathbb{P}(S_t = k|\mathcal{F}_{t-1})}$$

  - Repeat the above steps to get $\mathbb{P}[S_t = j|\mathcal{F}_{t-1}]$, $t = 1, \ldots, T$.
  - To start the filter, $\mathbb{P}[S_0^{|}\mathcal{F}_0]$ is required. Thus, emply the steady-state or unconditional probabilities of S_t:

$$\pi_0 = \frac{1 - p}{2 - p - q}, \qquad \pi_1 = \frac{1 - q}{2 - p - q}$$

  - Finally, optimize the log-likelihood function:

$$\log L = \sum_{t=1}^{T} \log \left(\sum_{S_t=0}^{1} f(y_t|S_t, \mathcal{F}_{t-1}) \mathbb{P}[S_t|\mathcal{F}_{t-1}] \right)$$

    w.r.t. $\beta_0, \beta_1, \sigma_0, \sigma_1, p$ and q.

Steady-state probabilities

- Transition probabilities of a first-order, M-state Markov switching process:

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1M} \\ p_{21} & p_{22} & \cdots & p_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ p_{M1} & p_{M2} & \cdots & p_{MM} \end{bmatrix}, \qquad P\mathbf{1} = \mathbf{1}$$

- Then the steady-state probabilitie can be obtained by:

$$\boldsymbol{\pi}_t = (A^\top A)^{-1} A^\top \begin{bmatrix} 0_M \\ 1 \end{bmatrix}$$

  i.e. the steady-state probabilities are the last column of the matrix $(A^\top A)^{-1} A^\top$.

7.2 Application: International CAPM

The Model

- International CAPM model with world market return:

$$y^w_t = \mu^w + \sigma^w \epsilon^w_t, \qquad \epsilon^w_t \sim \mathcal{N}(0, 1)$$

- Now: introduce two states, $s = 1, 2$, with conditional world mean world volatility μ^w_s and σ^w_s (no country-specific regimes).

Regime dynamics

- For each point in time, the portfolio managers knows the realized regime, but does not know which regime will be realized in the next time period.
- The regime follows a Markov process with constant transition probabilities q and p.
  Note that usually $p = 1 - q$ is not realistic. Empirical studies found that both p and q are well above 0.5, indicating *persistent states*.

Expected excess returns

- Expected excess return of country j is given by:

$$e_{j,i} = (1 - \beta^j)\mu^z + \beta^j e^w_i$$

  where i is the prevailing regime and e^w_i is the world's expected excess return, $i = 1, 2$.
- Expected excess returns differ across individual equity indices only through their different betas w.r.t. the world market.

Covariance matrix

- Add *idiosyncratic* part $V = \mathrm{diag}(\bar{\sigma}^2_j) \in \mathbb{R}^{J \times J}$
- Add a *regime-dependent systematic* part.
- Then:

$$\Omega_i = (\boldsymbol{\beta}\boldsymbol{\beta}^\top)(\sigma^w(S_{t+1} = i))^2 + V, \qquad i = 1, 2$$

- Take current regimes into account.
  Depending on which regime we are at the current time, we get different covariance matrices:

$$\Sigma_1 = p\Omega_1 + (1 - p)\Omega_2 + p(1 - p)(\boldsymbol{e}_1 - \boldsymbol{e}_2)(\boldsymbol{e}_1 - \boldsymbol{e}_2)^\top$$
$$\Sigma_2 = q\Omega_2 + (1 - q)\Omega_1 + q(1 - q)(\boldsymbol{e}_1 - \boldsymbol{e}_2)(\boldsymbol{e}_1 - \boldsymbol{e}_2)^\top$$

intentionally left blank

Notations

Unless otherwise specified, the following notations were used:

ϕ	standard normal PDF
Φ	standard normal CDF

Abbreviations

CDF	cumulative distribution function
cf.	conferre
e.g.	exempli gratia
i.e.	id est
iff	if and only if
IOT	in order to
PDF	probability density function
RV	random variable
s.t.	such that
w.r.t.	with respect to

Disclaimer

- This summary is work in progress, i.e. neither completeness nor correctness of the content are guaranteed by the author.
- This summary may be extended or modified at the discretion of the readers.
- Source: Lecture Asset Management: Advanced Investments, Spring Semester 2016, UZH (lecture notes, script, exercises and literature). Copyright of the content is with the lecturers.
- The layout of this summary is based on the summaries of several courses of the BSc ETH ME from Jonas LIECHTI.