Asset Pricing and Portfolio Theory

Dr. Daniel Jones

jonesdn@maths.ox.ac.uk

1 – Introduction

Topics we will cover:

1. A brief review of CAPM from a practical point of view.
2. Arbitrage Pricing Theory and Multifactor Risk Models.
3. Problems with Forecasting returns and covariances.
4. Asset pricing in practice: The Alpha Model.
5. Portfolio optimisation in practice: Turnover, Constraints and Implementation.
6. Robust Methods.
7. A brief intruction to Behavioural Asset Pricing.

2 – Recommended Reading

These notes are based around the excellent books [2], [4] and [1]. The classic reference for all things practical is [3]; [4] is a more mathematical, up-to-date version of this book. From a quantitative trading perspective all these books are highly recommended.

3 – The Capital Asset Pricing Model

3.1 – Motivation

We need at least 2 inputs for mean-variance optimisation (MVO) of a portfolio:

1. Expected return.
2. Return covariance matrix.

Of course, in reality, we also have additional constraints like position and sector limits; we'll come to that later.

Suppose we have N stocks. Then, we have $N(N+1)/2$ variances/covariances to estimate.

For statistical arbitrage and when we trade the market as a whole, we can have thousands of stocks. The estimation of so many parameters becomes an impossible task.

The Capital Asset Pricing Model (CAPM) (Sharpe (1964), Tobin (1958), Lintner (1965)) provides a simplified covariance structure.

Write

$$r_i = r_f + \beta_i(r_m - r_f) + \epsilon_i, \tag{1}$$

where r_i is the return of stock i, r_m is the market return, r_f is the risk-free rate and ϵ_i is an error term.

Individual stock returns are the sum of a systematic return and specific return.

Beta, β_i, measures the sensitivity of a stock's return to the market return. It is the coefficient of a regression of r_i vs. r_m:

$$\beta_i = \frac{\mathrm{cov}(r_i, r_m)}{\mathrm{var}(r_m)} = \frac{\rho_{i,m}\sigma_i\sigma_m}{\sigma_m^2} = \frac{\rho_{i,m}\sigma_i}{\sigma_m}. \tag{2}$$

$\epsilon_i \sim \mathcal{N}(0, \theta_i)$ is the specific return component.

The volatility θ_i is called the **specific risk**.

CAPM assumes the following:

1. For each stock, systematic and specific returns are independent.
2. Specific returns for different stocks are independent.

The portfolio covariance matrix maps pairwise covariances into linkages via beta.

Consider the covariance matrix $\boldsymbol{\Sigma}$ of a market, or benchmark, portfolio. Let $\boldsymbol{b}$ be the vector of weights. The vector of betas, $\boldsymbol{\beta} := (\beta_1, \ldots, \beta_N)^T$, is then

$$\boldsymbol{\beta} = \frac{\boldsymbol{\Sigma b}}{\boldsymbol{b}^T\boldsymbol{\Sigma b}}. \tag{3}$$

Under CAPM, it can be shown that

$$\boldsymbol{\Sigma} = \boldsymbol{\beta\beta}^T\sigma_m^2 + \boldsymbol{S}, \tag{4}$$

where $\boldsymbol{S} := \mathrm{diag}(\theta_1^2, \ldots, \theta_N^2)$.

Then, for a portfolio with weights $\boldsymbol{w}$, the **portfolio beta** is $\beta_p = \boldsymbol{w}^T\boldsymbol{\beta}$ and consequently,

$$\sigma_p^2 = \beta_p^2\sigma_m^2 + \sum_{i=1}^{N} w_i^2\theta_i^2. \tag{5}$$

We can separate the portfolio variance into systematic and specific components.

So, we have two sources of risk:

Specific Risk. Can be diversified away by increasing the number of stocks. For example, assume we have N stocks with the same specific risk θ_0. An equally-weighted portfolio would have specific variance θ_0^2/N.

Systematic Risk. Can't be diversified away (no dependence on N). Portfolio with a beta of 1 (traditional long-only portfolio) moves in step with the market. Portfolio with a beta of 0 has only specific risk (stat-arb).

3.2 – Optimal Portfolios under CAPM

Consider the mean-variance optimal porfolio (with cash).

Let $\boldsymbol{f} := (f_1, \dots, f_N)^T$ be the expected return vector (forecast), w_0 be the cash weight and $\boldsymbol{w} = (w_1, \dots, w_N)^T$ be the weights for the individual stocks. Let λ be the risk aversion parameter (degree of influence of risk on our portfolio).

We need to solve

$$\begin{aligned} \max \quad & w_0 r_f + \boldsymbol{w}^T \boldsymbol{f} - \frac{1}{2}\lambda \boldsymbol{w}^T \boldsymbol{\Sigma} \boldsymbol{w} \\ \text{s.t.} \quad & w_0 + \boldsymbol{w}^T \boldsymbol{e} = 1, \end{aligned} \tag{6}$$

where $\boldsymbol{e} = (1, \dots, 1)^T$.

This assumes you can borrow unlimited amounts from a low-return asset and invest it in a high-return asset.

To solve, convert to an unconstrained problem by writing $w_0 = 1 - \boldsymbol{w}^T \boldsymbol{e}$ and substituting. The problem is now

$$\max \quad \boldsymbol{w}^T \boldsymbol{f_e} - \frac{1}{2}\lambda \boldsymbol{w}^T \boldsymbol{\Sigma} \boldsymbol{w}, \tag{7}$$

where $\boldsymbol{f_e} = \boldsymbol{f} - r_f \boldsymbol{e}$ (excess return).

Equate partial derivatives to zero to solve:

$$\begin{aligned} \boldsymbol{w}^* &= \frac{1}{\lambda}\boldsymbol{\Sigma}^{-1}\boldsymbol{f_e} \\ w_0^* &= 1 - \frac{1}{\lambda}\boldsymbol{e}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{f_e}. \end{aligned} \tag{8}$$

To solve our problem of optimal portfolios under CAPM, we expect something like this solution, which involves a matrix inversion.

Under CAPM, we can write

$$\boldsymbol{\Sigma}^{-1} = \boldsymbol{S}^{-1} - \frac{\sigma_m^2}{1+\kappa} \boldsymbol{\beta_S} \boldsymbol{\beta_S}^T, \tag{9}$$

where $\kappa = \sum_{i=1}^N \frac{\sigma_m^2 \beta_i^2}{\theta_i^2}$ and $\boldsymbol{\beta_S} := \left(\frac{\beta_1}{\theta_1^2}, \dots, \frac{\beta_N}{\theta_N^2}\right)$.

Then,

$$\boldsymbol{w}^* = \frac{1}{\lambda}\left(\boldsymbol{S}^{-1}\boldsymbol{f_e} - \frac{\sigma_m^2}{1+\kappa}\boldsymbol{\beta_S}\boldsymbol{\beta_S}^T \boldsymbol{f_e}\right). \tag{10}$$

Let $\boldsymbol{w}_0^* := \frac{1}{\lambda}\boldsymbol{S}^{-1}\boldsymbol{f_e}$, which is the partial solution given by the specific risk part of the covariance matrix.

The portfolio beta given by the partial solution is given by

$$\beta_0 := \sum_{i=1}^N \beta_i w_{0,i}^* = \frac{1}{\lambda}\boldsymbol{\beta_S}^T \boldsymbol{f_e}, \tag{11}$$

so we can write

$$\boldsymbol{w}^* = \boldsymbol{w}_0^* - \frac{\sigma_m^2 \beta_0}{1+\kappa}\boldsymbol{\beta_S}. \tag{12}$$

For a particular stock, we have

$$w_i^* = w_{0,i}^* - \frac{1}{1+\kappa}\frac{\sigma_m^2 \beta_0 \beta_i}{\theta_i^2} = w_{0,i}^* - \frac{1}{1+\kappa}\frac{\operatorname{cov}(r_i, r_{\boldsymbol{w}_0^*})}{\theta_i^2}. \tag{13}$$

The beta of the optimal portfolio can be written as

$$\beta^* = \sum_{i=1}^N w_i^* \beta_i = \frac{\beta_0}{1+\kappa} \tag{14}$$

and the specific risk of the portfolio is given by

$$\sum_{i=1}^N (w_i^* \theta_i)^2 = \sum_{i=1}^N \left(w_{0,i}^* \theta_i\right)^2 - \sigma_m^2 \beta_0^2 \left[\frac{1}{1+\kappa} + \frac{1}{(1+\kappa)^2}\right]. \tag{15}$$

The total risk of the portfolio is then

$$\sigma_p^2 = \sum_{i=1}^{N} \left(w_{0,i}^* \theta_i\right)^2 + \left(\beta^*\right)^2 \sigma_m^2 = \sum_{i=1}^{N} \left(w_{0,i}^* \theta_i\right)^2 - \frac{\sigma_m^2 \beta_0^2}{1+\kappa}. \tag{16}$$

3.3 – Beta-neutral Portfolio

Many funds like to control the total beta exposure. Let us assume we would like to add the constraint $\boldsymbol{w}^T \boldsymbol{\beta} = 0$.

This is surprisingly simple under CAPM. A beta-neutral portfolio has only specific risk, by definition. So, we have the following problem to solve:

$$\begin{aligned} \max \quad & \boldsymbol{w}^T \boldsymbol{f} - \frac{1}{2}\lambda \boldsymbol{w}^T \boldsymbol{S} \boldsymbol{w} \\ \text{s.t.} \quad & \boldsymbol{w}^T \boldsymbol{\beta} = 0 \end{aligned} \tag{17}$$

Using the method of Lagrange multipliers, the solution is

$$\boldsymbol{w}^* = \frac{1}{\lambda} \boldsymbol{S}^{-1} \left(\boldsymbol{f} - l\boldsymbol{\beta}\right) \tag{18}$$

with

$$l = \frac{\boldsymbol{f}^T \boldsymbol{S}^{-1} \boldsymbol{\beta}}{\boldsymbol{\beta}^T \boldsymbol{S}^{-1} \boldsymbol{\beta}} \tag{19}$$

As $\boldsymbol{S}$ is diagonal, we have

$$w_i^* = \frac{1}{\lambda} \frac{f_i - l\beta_i}{\theta_i^2}, \quad i = 1, \ldots, N; \quad l = \left(\sum_{i=1}^{N} \frac{\beta_i^2}{\theta_i^2}\right)^{-1} \sum_{i=1}^{N} \frac{f_i \beta_i}{\theta_i^2}. \tag{20}$$

The weights are a function of **beta-adjusted forecasts**.

4 – Arbitrage Pricing Theory

Multifactor risk models (Ross (1976)); natural extension to CAPM.

Very flexible, but we have to decide on the factors that we think explain returns, which can be hard.

Statement of the Arbitrage Pricing Theorem.

Suppose there are N stocks in an economy, and for each stock i, the return is represented as

$$R_i = \mu_i + \boldsymbol{\beta}_i^T \boldsymbol{F} + \epsilon_i, \tag{21}$$

where μ_i is an expected return, $\boldsymbol{\beta}_i$ is an L-dimensional vector of betas to an L-dimensional vector of factors $\boldsymbol{F}$. Assume $\epsilon_i \sim \mathcal{N}(0, \theta_i^2)$ and assume a common variance, $\theta_i^2 = \theta_0^2$.

If there is no arbitrage, and if N is large, we have the approximate relationship

$$\mu_i \approx A + \boldsymbol{\beta}_i^T \boldsymbol{f}, \tag{22}$$

where A is a constant, and $\boldsymbol{f} := (f_1, \ldots, f_L)^T$.

Proof (1-dimensional case)

By linear algebra, we can decompose $\boldsymbol{\mu}$ as

$$\boldsymbol{\mu} = A\boldsymbol{e} + f\boldsymbol{\beta} + \boldsymbol{a}, \tag{23}$$

where A and f are constants and $\boldsymbol{a}$ is a vector orthogonal to $\boldsymbol{e}$ and $\boldsymbol{\beta}$, i.e. $\boldsymbol{a}^T\boldsymbol{e} = 0$ and $\boldsymbol{\beta}^T\boldsymbol{a} = 0$.

We can interpret $\boldsymbol{a}$ as portfolio weights and suppose that all $a_i \neq 0$. For a given number w, consider the alternative portfolio $w\boldsymbol{a}$. This portfolio has expected return

$$\mu_w = \mathbb{E}\left[w\boldsymbol{a}^T\boldsymbol{R}\right] = w\sum_{i=1}^{N} a_i\mu_i = w\sum_{i=1}^{N} a_i^2, \tag{24}$$

by orthogonality of $\boldsymbol{a}$ to $\boldsymbol{e}$ and $\boldsymbol{\beta}$.

The variance of this portfolio is

$$\sigma_w^2 = \mathrm{var}\left[w\sum_{i=1}^{N} a_i R_i\right] = w^2\theta_0^2\sum_{i=1}^{N} a_i^2. \tag{25}$$

We want to argue that

$$\limsup_{N\to\infty} \frac{\sum_{i=1}^N a_i^2}{N} = 0, \tag{26}$$

i.e. that $\sum_{i=1}^N a_i^2$ grows more slowly than N. If this is the case, then most of the a_is have to be small as N gets large so that from

$$\boldsymbol{\mu} = A\boldsymbol{e} + f\boldsymbol{\beta} + \boldsymbol{a}$$

the APT relationship

$$\mu_i \approx A + \boldsymbol{\beta}_i^T \boldsymbol{f} \tag{27}$$

approximately holds.

Suppose instead that

$$\limsup_{N\to\infty} \frac{\sum_{i=1}^N a_i^2}{N} = C > 0. \tag{28}$$

Then, with $w = \frac{1}{N}$, $\mu_w \to C > 0$ and $\sigma_w^2 \to 0$. We consider this to be arbitrage, and so impossible. Thus,

$$\limsup_{N\to\infty} \frac{\sum_{i=1}^N a_i^2}{N} = 0$$

holds in the absence of arbitrage. ■

4.1 – Side Note: Equilibrium pricing models versus risk models

CAPM was originally developed as an equilibrium pricing model, not a risk model. As a pricing model, it relates expected returns of individual stocks to the market return, through beta:

$$\mathbb{E}\left[r_i - r_f\right] = \beta_i \left[\mathbb{E}[r_m] - r_f\right] \tag{29}$$

CAPM states that the market will set prices of stocks so that expected returns are proportional to systematic returns. Furthermore, specific risks are diversified away and not rewarded.

This is not how we have used CAPM. We used it as a risk model:

total risk = systematic risk + specific risk.

We left expected returns aside.

Statistically, pricing and risk models come from the same equation. However, the former interprets the equation by expectation, the latter by variance.

When CAPM was first proposed, a long list of anomalies contradicting CAPM followed. This resulted in many variants of CAPM being developed that described how assets are priced in equilibrium.

For example, the Fama-French three-factor model (1992), which has beta, market capitalisation and the book-to-price ratio as factors.

For practical purposes, this also means that there are other priced factors in addition to beta and so the risk model needs extending.

4.2 – APT Models

Very general structure:

$$r_i = b_{i0} + b_{i1} I_1 + \cdots + b_{iK} I_K + \epsilon_i, \tag{30}$$

with K factors $I_1, \ldots, I_K$ and b_{ij} is the sensitivity (or loading) of stock i to factor j.

Assume the ϵ_is are uncorrelated with eachother and the factors.

We can write the covariance as

$$\boldsymbol{\Sigma} = \boldsymbol{B}\boldsymbol{\Sigma}_I \boldsymbol{B}^T + \boldsymbol{S}. \tag{31}$$

$\boldsymbol{B}$ is the exposure matrix:

$$\boldsymbol{B} = \begin{pmatrix} b_{11} & \cdots & b_{1K} \\ \vdots & \ddots & \vdots \\ b_{N1} & \cdots & b_{NK} \end{pmatrix}_{(N \times K)} = \left(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_K\right). \tag{32}$$

The vector $\boldsymbol{b}_i$ contains the exposures of all stocks to factor i.

$\boldsymbol{\Sigma}_I$ is the factor return covariance matrix,

$$\boldsymbol{\Sigma}_I = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1K} \\ \vdots & \ddots & \vdots \\ \sigma_{K1} & \cdots & \sigma_{KK}, \end{pmatrix} \tag{33}$$

and $\boldsymbol{S}$ is the diagonal matrix of specfic risks.

This is a very general framework and it is hard to test empirically in terms of identifying the return-generating process and associated pricing mechanism.

The flexibility has lead to multiple approaches to multifactor risk modelling:

1. Macroeconomic factor models.
2. Fundamental factor models.
3. Statistical factor models.

Macroeconomic Factor Model

The discounted cashflow model states that the stock price is the present value of all future payments received by shareholders.

Macroeconomic factors affect both company earnings and required rate of return for an investment.

Examples:

1. if interest rates are cut, the stock market tends to response favourably.
2. If the oil price is high, this creates a drag on the economy and is bad for the stock market overall. However, individual stocks will respond differently. High prices are good for oil producers and bad for airlines, whereas technology stocks tend to be agnostic.

Typical factors: market return, change in short-term interest rates, change in industrial production, change in inflation, term spreads, default spreads, change in the oil price.

To estimate exposures for each stock to a set of factors, we perform a time-series regression:

$$r_{it} - r_{ft} = \alpha_i + \beta_i(r_{mt} - r_{ft}) + \sum_{k=1}^{K} b_{ki} I_{kt} + \epsilon_{it} \tag{34}$$

In order to estimate this model, we need a very long (rolling) window of many months.

Fundamental Factor Model

One potential avenue for looking for stock returns is through **fundamental research**, which involves looking through balance sheet and cash flow information for companies.

The BARRA US equity model is a widely used risk model. This has a number of fundamental factors: industry group, size, book-to-price, dividend yield, earnings yield, momentum, growth, earnings variability, financial leverage, volatility, trading activity.

Steps to build the model:

1. A stock's exposure to these factors are just the values of the attributes. Normalise factors cross-sectionally so they are mean 0, stdev 1.
2. Run cross-sectional regression of stock returns vs. factor exposures. The coefficients are then the factor returns for the given time period.
3. For a given time period t, regress stock returns for each stock's

subsequent return against factor exposures known at time t, to get the specific risks:

$$r_i^{t+1} = b_0^t + b_1^t I_{i,1}^t + \cdots + b_K^t I_{i,K}^t + \epsilon_i. \tag{35}$$

4. Use the factor returns in a series of cross-sectional regressions for multiple periods to calculate the factor return covariance matrix.

Extensions:

1. The risk model **must** cover a large proportion of the total stock market capitalisation. In practice, we would use a weighted regression, with market caps as the weights.
2. More recent factor returns are more informative. We put higher weight on more recent data, using a weighting decay: $\ldots, \omega^{t-T}, \ldots, \omega^2, \omega, 1$, where $\omega < 1$; half life is $-\log 2/\log\omega$.
3. The estimation of specific risk is very hard. Ideally, for each stock we would get the time series of residuals from a time-series regression and

use the volatility of this as a proxy for specific risk. Newly issued stocks cause problems as we have insufficient data. Specific risks tend to be estimated partially on fundamental characteristics.

It is very time-consuming and fiddly to build fundamental factor models, so a lot of quant funds will just use a third-party product like BARRA.

Statistical Factor Models

Based purely on historical returns; typically derived from principal components.

Tend to be good at exploiting price information and thus explaining risk. They can overfit price data (which may just be noise). They lack economic reasoning and may not be good for longer-term forecasting.

Principal Component Analysis (PCA). What combination of raw data leads to maximum variance among all possible combinations?

Consider the raw $N \times N$ covariance matrix $\boldsymbol{\Sigma}$. PCA decomposes this into

$$\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{P}\boldsymbol{L}^T, \tag{36}$$

with $\boldsymbol{P} = \mathrm{diag}(\lambda_1, \ldots, \lambda_N)$, where $\lambda_1 > \lambda_2 > \cdots > \lambda_N > 0$ are the eigenvalues of $\boldsymbol{\Sigma}$. The matrix $\boldsymbol{L}$ is an orthogonal matrix of eigenvectors, $\boldsymbol{L}\boldsymbol{L}^T = \boldsymbol{I}$.

Recall that under APT, $\boldsymbol{\Sigma} = \boldsymbol{B}\boldsymbol{\Sigma}_I\boldsymbol{B}^T + \boldsymbol{S}$, so $\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{P}\boldsymbol{L}^T$ represents a model of N orthogonal factors with eigenvalues as variances and $\boldsymbol{L}$ being the exposure matrix of each security to these factors.

Let $\boldsymbol{R} := (r_{jt})_{N\times T}$ and $\boldsymbol{Q} := (q_{jt})_{N\times T}$ be returns of N securities and N orthogonal factors over T time periods.

We can write $\boldsymbol{R} = \boldsymbol{L}\boldsymbol{Q}$, because $\boldsymbol{L}\boldsymbol{L}^T = \boldsymbol{I}$ and $\boldsymbol{Q} = \boldsymbol{L}^T\boldsymbol{R}$.

Given $\boldsymbol{Q}$, we can derive the covariance matrix of the N principal component factors as

$$\hat{\boldsymbol{\Sigma}} = \boldsymbol{P}. \tag{37}$$

It is tricky to get the number of factors right: too few and we have a poor description of risk, too many and we risk overfit.

Random Matrix Theory. We compare the distribution of eigenvalues with a random matrix and select only the statistically significant factors.

5 – Forecasting Expected Return and Risk

Key elements to a successful (quantitative) trading program:

1. Good return expectations/forecasts.
2. Risk model and risk controls (portfolio constraints).
3. Trading and transaction cost management (portfolio implementation).
4. Performance monitoring.

We will come to constraints and implementation. The first two elements require estimates of return and covariance.

Most common approach: use estimates from historical data.

5.1 – Sample mean and covariance as estimators

Issues:

1. Expected returns exhibit significant time variation (nonstationarity).
2. Forecasts based on historical data can be influenced by changes in the economic cycle and market conditions (interest rates, political environment, monetary/fiscal policy, investor confidence, business cycles of different industries, etc.).
3. Realised returns are influenced by changed in expected returns (Fama/French study; see [2], p. 222).

Consider historical return series R_{it} and R_{jt} for securities i, j and $t = 1, \ldots, T$.

Sample mean. $\bar{R}_i = \frac{1}{T}\sum_{t=1}^{N} R_{it}$

Sample covariance. $\sigma ij = \frac{1}{T-1}\sum_{t=1}^{T}\left(R_{it} - \bar{R}_i\right)\left(R_{jt} - \bar{R}_j\right)$.

For N securities, we can compute the sample covariance matrix as

$$\mathbf{\Sigma} = \frac{1}{N-1}\boldsymbol{X}\boldsymbol{X}^T; \quad \boldsymbol{X} = \begin{pmatrix} R_{11} & \cdots & R_{1T} \\ \vdots & \ddots & \vdots \\ R_{N1} & \cdots & R_{NT} \end{pmatrix} - \begin{pmatrix} \bar{R}_1 & \cdots & \bar{R}_1 \\ \vdots & \ddots & \vdots \\ \bar{R}_N & \cdots & \bar{R}_N \end{pmatrix}. \tag{38}$$

If security resturns are i.i.d., then $\mathbf{\Sigma}$ is the maximum likelihood estimator of the population covariance. This matrix follows a Wishart distribution with $N-1$ degrees of freedom.

Wishart Distribution.

Suppose $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N$ are i.i.d. random vectors s.t. $\boldsymbol{X}_i \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{V})$, i.e. $\mathbb{E}[\boldsymbol{X}_i] = \mathbf{0}$, where $\mathbf{0}$ is a p-dim vector and $\operatorname{var}(\boldsymbol{X}_i) = \mathbb{E}\left[\boldsymbol{X}_i \boldsymbol{X}_i^T\right] = \boldsymbol{V}$ is a $p \times p$-matrix.

Then, the Wishart Distribution with N degrees of freedom is the probability distribution of the $p \times p$-matrix $\boldsymbol{S} = \sum_{i=1}^N \boldsymbol{X}_i \boldsymbol{X}_i^T$ and we write $\boldsymbol{S} \sim \mathcal{W}_p(\boldsymbol{V}, N)$.

For $p = 1, \boldsymbol{V} = 1$, this is the chi-square distribution.

Note. Over long periods of time, it is common for R_f to change significantly. It is common practice to convert raw returns to excess returns. The expected return is then,

$$\bar{R}_i = R_{f,t} + \frac{1}{N} \sum_{i=1}^{T} (R_{it} - R_{f,t}) . \tag{39}$$

Alternatively, you can use the expected excess return directly in the mean-variance optimisation framework.

5.2 – Sample mean as an estimator

The sample mean is the **Best Linear Unbiased Estimator (BLUE)** of the population mean if the distribution is not heavy-tailed. In this case, increasing the sample size always improves the performance of the estimator.

Financial time series are heavy-tailed and exhibit skewness, so this is not valid. Furthermore, they are nonstationary so the mean is not a good estimator of expected return.

Finally, this estimator has a large estimation error, which significantly influences the mean-variance portfolio optimisation process:

1. Equally-weighted portfolios often outperform MVO portfolios.
2. MVO portfolios are often not well-diversified.
3. Uncertainty of returns can influence MVO portfolios more than risk.

How do you fix this?

1. Create more robust/stable (lower standard error) estimates of expected return. Impose structure on forecasts: impose a factor model; Bayesian modelling; shrinkage estimates.
2. MVO is very sensitive to inputs. When using classical MVO, we are assuming (implicitly) that inputs are known with great accuracy; that they are deterministic in fact. Bad inputs $\implies$ worse outputs.

   Portfolio resampling (Monte Carlo simulations) and Robust techniques can be useful here.

5.3 – Sample covariance as an estimator

The basic estimator can be improved by using weighted data. Let $d < 1$ and use

$$\begin{aligned}\sigma_{ij} &= \sum_{t=1}^{T}\left(\frac{d^{T-t}}{\sum_{t=1}^{T} d^{T-t}}\right)\left(R_{it}-\bar{R}_i\right)\left(R_{jt}-\bar{R}_j\right) \\ &= \frac{1-d}{1-d^T}\sum_{t=1}^{T}\left(R_{it}-\bar{R}_i\right)\left(R_{jt}-\bar{R}_j\right). \qquad (40)\end{aligned}$$

For large enough T, $\frac{1-d}{1-d^T} \approx 1-d$. The decay parameter d can be estimated by MLE or minimising OOS forecast error. This estimator still performs poorly though, and it is usually better to apply shrinkage techniques.

Sample covariance is a nonparametric (unstructured) estimator. An alternative is to make assumptions on what underlying economic factors contribute and to create a factor model.

Other issues:

1. **Serial correlation and heteroskedasticity**. Autocorrelation and time-varying variances/covariances lead to bias in estimators. Simple ways to correct this: Newey-West corrections.
2. **Missing/truncated data**. Data cleaning very important and EM algorithm can be used for handling missing data. Use Stambaugh for handling truncated time series (see [2], p. 228).
3. **Data frequency**. Estimates are improved as sampling frequency increases. Merton: even if expected returns are constant over time, a long history still required to estimate accurately.

5.4 – Random Matrix Theory

Developed in the 1950s by quantum physicists.

Sample covariance for large number of assets is unstable through time. Look at SP500 stocks: matrix fluctuates in a nearly random way (but the average correlation remains high).

Random Matrix. Covariance matrix of a set of independent random walks. Entries are a set of zero-mean, i.i.d. variables. The mean of random correlation coeffs is zero as they are distributed symmetrically in the interval $[-1,+1]$.

Suppose $T, N \to \infty$ s.t. $Q := T/N \geq 1$ remains fixed. It can be shown that the density of eigenvalues of the random matrix tends to

$$\rho(\lambda) = \frac{Q}{2\pi\sigma^2}\frac{\sqrt{\left(\lambda_{max}-\lambda\right)\left(\lambda_{min}-\lambda\right)}}{\lambda}, \qquad (41)$$

where $\lambda_{max}, \lambda_{min} = \sigma^2\left(1+\frac{1}{Q} \pm 2\sqrt{\frac{1}{Q}}\right)$ and σ^2 is the average eigenvalue of the matrix.

The shape of the distribution of eigenvalues is a signature of the randomness.

SP500.

- Fairly close to a random matrix distribution, with the exception of a few eigenvalues that have significantly higher values.
- Little information in the full sample covariance of a large portfolio.
- Filter out the smaller eigenvalues and eigenvectors $\implies$ reduce the estimation error in the covariance matrix in portfolio optimisation.

Other approaches: Higham (nearest covariance in Frobenius norm), Factor models, shrinkage estimation.

6 – Evaluation of Alpha Factors.

6.1 – Performance measures

Central to our endeavour is the forecasting model: how do we assess **alpha factors** and combine optimally?

1. **Sharpe ratio,** $SR = (\bar{\mu} - r_f)/\sigma$**.** Ratio of excess return to stdev of excess return; σ is total risk of portfolio. Derived from CAPM.
2. **Information ratio,** $IR = \bar{\alpha}/\sigma(\alpha)$**.** Average of active portfolio return (relative to a passive benchmark), relative to volatility of active portfolio (relative to passive benchmark).

IR allows you to compare long-only to active managers to passive benchmark.

Tracking error. Periodic deviation from passive benchmark (or active risk).

Long-only. Alpha is the portfolio excess return over benchmark.

Long-short market-neutral. Alpha is the excess return over cash.

Note that IR is a multiperiod measure; you need a sequence of α_t. This is easy to compute ex post, but it is not easy to estimate ex ante. IR is very close to a t-stat:

$$t_\alpha = \frac{\sqrt{T-1}\bar{\alpha}}{\sigma(\alpha)} = IR\sqrt{T-1}, \tag{42}$$

where T is the number of sample points.

6.2 – Information coefficient (IC)

The IC is key for measuring the effective power of an alpha factor (see [3]). It is a linear statistic measuring the cross-sectional correlation between forecasts and subsequent returns. As we shall see, we can write

$$IR = \frac{\overline{IC_t}}{\mathrm{std}(IC_t)}, \tag{43}$$

so IC translates directly to IR.

6.3 – Raw IC

Consider a single-period portfolio optimisation problem, with active weights $\boldsymbol{w} = (w_1, \dots, w_N)^T$ and subsequent returns $\boldsymbol{r} = (r_1, \dots, r_N)^T$. The total realised, excess return is

$$\alpha_t = \sum_{i=1}^{N} w_i r_i = \boldsymbol{w}^T \boldsymbol{r}. \tag{44}$$

For a dollar-neutral, long-short portfolio, or a long-only portfolio against a benchmark, $\boldsymbol{w}^T \boldsymbol{e} = 0$. So, we can simply write

$$\alpha_t = \boldsymbol{w}^T (\boldsymbol{r} - \bar{r}\boldsymbol{e}) \tag{45}$$

and hence

$$\alpha_t = (N-1)\,\mathrm{corr}(\boldsymbol{w}, \boldsymbol{r})\mathrm{dis}(\boldsymbol{w})\mathrm{dis}(\boldsymbol{r}), \tag{46}$$

where $\mathrm{dis}(\cdot)$ is the dispersion (cross-sectional standard deviation).

Note. $\text{dis}(\boldsymbol{w}), \text{dis}(\boldsymbol{r}) \geq 0$, so α_t is the same sign as $\text{corr}(\boldsymbol{w}, \boldsymbol{r})$.

This makes sense: if we upweight high-return stocks and downweight low-return stocks, we expect positive excess return.

Key relationship. Assume (unrealistically) $\boldsymbol{w} = c\boldsymbol{f}$, where c is a constant and $\boldsymbol{f}$ is a vector of forecasts with zero mean.

It can be shown that

$$\alpha_t = c(N-1)IC\text{dis}(\boldsymbol{f})\text{dis}(\boldsymbol{r});\ IC = \text{corr}(\boldsymbol{f}, \boldsymbol{r}), \tag{47}$$

i.e.

$$\alpha \propto \textbf{skill} \times \textbf{conviction} \times \textbf{opportunity}.$$

6.4 – Risk-adjusted IC

Raw IC is intuitive and simple, but seriously flawed: $\boldsymbol{w} = c\boldsymbol{f}$.

We need a **risk-adjusted IC**, consistent with a realistic portfolio we want to hold: need to remove the systematic risk and incorporate individual stock specific risks.

Proceed as follows:

1. Solve MVO to get weights of market-neutral portfolio.
2. Derive single-period α using these weights and subsequent returns.
3. Derive the risk-adjusted IC.

Step 1. Solve

$$\begin{aligned} \max \quad & \boldsymbol{f}^T\boldsymbol{w} - \frac{1}{2}\lambda \boldsymbol{w}^T\boldsymbol{\Sigma}\boldsymbol{w}, \\ \text{s.t.} \quad & \boldsymbol{w}^T\boldsymbol{e} = 0 \text{ and } \boldsymbol{w}^T\boldsymbol{B} = 0. \end{aligned} \tag{48}$$

Recall that $\boldsymbol{\Sigma} = \boldsymbol{B}\boldsymbol{\Sigma}_I\boldsymbol{B}^T + \boldsymbol{S}$, where $\boldsymbol{B}$ is the factor exposure matrix and $\boldsymbol{S}$ is the diagonal matrix of specific volatilities.

Actual weights are dollar-neutral and neutral to risk factors, so there is no systematic risk in the final portfolio. Therefore, provided we keep the constraints, the objective function reduces to $\boldsymbol{f}^T\boldsymbol{w} - \frac{1}{2}\lambda\boldsymbol{w}^T\boldsymbol{S}\boldsymbol{w}$.

Solve analytically using Lagrange multipliers: new objective function is

$$\sum_{i=1}^N w_i f_i - \frac{1}{2}\lambda\sum_{i=1}^N w_i^2\sigma_i^2 - l_0\sum_{i=1}^N w_i - l_1\sum_{i=1}^N w_i B_{1i} - \dots l_K\sum_{i=1}^N w_i B_{Ki}$$

$K+1$ lagrange multipliers: 1 for the dollar-neutral constraint, K for the risk factors.

Taking partial derivatives,

$$w_i = \frac{f_i - l_0 - l_1\beta_{1i} - \dots - l_K\beta_{Ki}}{\lambda\sigma_i^2}. \tag{49}$$

Let $\langle \boldsymbol{x}, \boldsymbol{y}\rangle := \boldsymbol{x}^T\boldsymbol{S}^{-1}\boldsymbol{y} = \sum_{i=1}^N \frac{x_i y_i}{\sigma_i^2}$. From the constraints,

$$\begin{aligned} l_0\langle \boldsymbol{e}, \boldsymbol{e}\rangle + l_1\langle \boldsymbol{e}, \boldsymbol{b}_1\rangle + \dots + l_K\langle \boldsymbol{e}, \boldsymbol{b}_K\rangle &= \langle \boldsymbol{e}, \boldsymbol{f}\rangle \\ l_0\langle \boldsymbol{b}_1, \boldsymbol{e}\rangle + l_1\langle \boldsymbol{b}_1, \boldsymbol{b}_1\rangle + \dots + l_K\langle \boldsymbol{b}_1, \boldsymbol{b}_K\rangle &= \langle \boldsymbol{b}_1, \boldsymbol{f}\rangle \\ &\vdots \\ l_0\langle \boldsymbol{b}_K, \boldsymbol{e}\rangle + l_1\langle \boldsymbol{b}_K, \boldsymbol{b}_1\rangle + \dots + l_K\langle \boldsymbol{b}_K, \boldsymbol{b}_K\rangle &= \langle \boldsymbol{b}_K, \boldsymbol{f}\rangle. \end{aligned} \tag{50}$$

(51)

The solution is

$$\begin{pmatrix} l_0 \\ l_1 \\ \vdots \\ l_K \end{pmatrix} = \begin{pmatrix} \langle \boldsymbol{e}, \boldsymbol{e} \rangle & \langle \boldsymbol{e}, \boldsymbol{b}_1 \rangle & \cdots & \langle \boldsymbol{e}, \boldsymbol{b}_K \rangle \\ \langle \boldsymbol{b}_1, \boldsymbol{e} \rangle & \langle \boldsymbol{b}_1, \boldsymbol{b}_1 \rangle & \cdots & \langle \boldsymbol{b}_1, \boldsymbol{b}_K \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \boldsymbol{b}_K, \boldsymbol{e} \rangle & \langle \boldsymbol{b}_K, \boldsymbol{b}_1 \rangle & \cdots & \langle \boldsymbol{b}_K, \boldsymbol{b}_K \rangle \end{pmatrix}^{-1} \begin{pmatrix} \langle \boldsymbol{e}, \boldsymbol{f} \rangle \\ \langle \boldsymbol{b}_1, \boldsymbol{f} \rangle \\ \vdots \\ \langle \boldsymbol{b}_K, \boldsymbol{f} \rangle \end{pmatrix}. \tag{52}$$

Now,

$$\alpha_t = \sum_{i=1}^{N} w_i r_i = \frac{1}{\lambda} \sum_{i=1}^{N} \left(\frac{f_i - l_0 - l_1 \beta_{1i} - \cdots - l_K \beta_{Ki}}{\sigma_i^2} \right) r_i. \tag{53}$$

Because of the contraints placed on the active weights, we can replace r_i by $r_i - m_0 - m_1 \beta_{1i} - \cdots - m_K \beta_{Ki}$, where $m_1, \ldots, m_K$ are returns to the K risk factors (derived from a cross-sectional regression), and the equation is unchanged.

Define risk-adjusted forecasts and returns:

$$\begin{aligned} F_i &= \frac{f_i - l_0 - l_1 \beta_{1i} - \cdots - l_K \beta_{Ki}}{\sigma_i} \\ R_i &= \frac{r_i - m_0 - m_1 \beta_{1i} - \cdots - m_K \beta_{Ki}}{\sigma_i}. \end{aligned} \tag{54}$$

m_0 is still undetermined. Then,

$$\alpha_t = \frac{1}{\lambda} \sum_{i=1}^{N} F_i R_i, \tag{55}$$

which we can write as

$$\alpha_t = (N-1) \lambda_t^{-1} \mathrm{corr}(\boldsymbol{F}_t, \boldsymbol{R}_t) \mathrm{dis}(\boldsymbol{F}_t) \mathrm{dis}(\boldsymbol{R}_t), \tag{56}$$

provided that $\bar{\boldsymbol{R}}_t = 0$; choose m_0 such that this is true.

$\mathrm{corr}(\boldsymbol{F}_t, \boldsymbol{R}_t)$ is the risk-adjusted information coefficient. It relates directly to the excess return of a risk-managed portfolio.

Equation (56) is central to the assessment of predictors in the context of a market-neutral, cash-neutral trading strategy.

However, tight neutrality constraints are rather restrictive in practice. Treat the risk-adjusted IC as indicative of performance where we have portfolio constraints such as limited exposures.

6.5 – Tracking Error and Risk Aversion

No systematic risk above, so tracking error is just residual variance:

$$\sigma_{model}^2 = \sum_{i=1}^{N} w_i \sigma_i^2 = \lambda_t^{-2} \sum_{i=1}^{N} F_i^2 \tag{57}$$

It can be shown that

$$\sigma_{model} \approx \lambda_t^{-1} \sqrt{N-1} \mathrm{dis}(\boldsymbol{F}_t), \tag{58}$$

so that

$$\lambda_t \approx \frac{\sqrt{N-1} \mathrm{dis}(\boldsymbol{F}_t)}{\sigma_{model}}. \tag{59}$$

Scaling forecasts and the risk aversion by the same amount has not effect on weights and tracking error at all.

Substituting back,

$$\alpha_t = IC_t\sqrt{N-1}\sigma_{model}\text{dis}(\boldsymbol{R}_t) \approx IC_t\sqrt{N}\sigma_{model}\text{dis}(\boldsymbol{R}_t). \tag{60}$$

The single-period excess return is the product of the risk-adjusted IC (skill), square-root of N (breadth), targeted tracking error (risk budget) and dispersion of risk-adjusted returns (opportunity).

If the risk model is adequate, the dispersion of risk-adjusted return should show little time-series variation, and $R_i \sim \mathcal{N}(0,1)$ for each i. The variance of N of these (stocks), is a scaled chi-squared distribution. If N is sufficiently large, the dispersion is close to 1 (Keeping 1995). Then,

$$\alpha_t \approx IC_t\sqrt{N}\sigma_{model}. \tag{61}$$

6.6 – Ex Ante Information Ratio

Recall that we wanted to work out our expected information ratio from expected return and risk.

Assume N and σ_{model} are both constant. Also, assume $\text{dis}(\boldsymbol{R}_t) = \overline{\text{dis}(\boldsymbol{R}_t)}$. Then,

$$\overline{\alpha}_t = \overline{IC}_t\sqrt{N}\sigma_{model}\overline{\text{dis}(\boldsymbol{R}_t)}. \tag{62}$$

Also, our expected active risk is

$$\sigma = \text{std}(IC_t)\sqrt{N}\sigma_{model}\overline{\text{dis}(\boldsymbol{R}_t)}. \tag{63}$$

Consequently,

$$IR = \frac{\overline{IC}_t}{\text{std}(IC_t)}, \tag{64}$$

as expected.

6.7 – Fundamental Law of Active Management (FLAM)

This is a term coined by Grinold and Kahn [3] and relates to the relationship $IR = \overline{IC}_t \times \sqrt{breadth} = \overline{IC}_t\sqrt{N}$.

From the previous section, this is true only if $\text{std}(IC_t) = 1/\sqrt{N}$.

Furthermore, if this holds, then $\sigma = \sigma_{model}\overline{\text{dis}(\boldsymbol{R}_t)}$, i.e. the target tracking error accurately predicts the active risk.

When is $\text{std}(IC_t) = 1/\sqrt{N}$ true?

This is approximately true if the underlying population correlation coefficient is constant over time and the stdev of IC over time is purely due to sampling error.

If $\text{corr}(\boldsymbol{F}_t, \boldsymbol{R}_t) = \rho$, then $\text{stderr}(IC_t) \approx \sqrt{1-\rho^2}/\sqrt{N}$. Most ICs are less than 0.1 on a quarterly basis, so $\text{stderr}(IC_t) \approx 1/\sqrt{N}$.

7 – Multifactor Alpha Model.

Aim. To Maximise the IR of a multifactor model (using mean-variance optimisation techniques).

The correlations between factor (or predictor) ICs will be important in providing diversification.

Note. The correlation between the ICs is not the same as the correlation between predictor values.

Consider M factors $(\boldsymbol{F}_1, \ldots, \boldsymbol{F}_M)$, and a weight vector $\boldsymbol{v} = (v_1, \ldots, v_M)^T$. Once selected, $\boldsymbol{v}$ will remain constant through time.

Assume that all factors are risk-adjusted. The composite factor model is written as

$$\boldsymbol{F}_c = \sum_{i=1}^{M} v_i \boldsymbol{F}_i. \tag{65}$$

The composite model is risk-adjusted (the portfolio is neutral to all risk factors) and is mean-variance optimal.

We can write

$$\begin{aligned} \alpha_t &= \frac{(N-1)}{\lambda_t}\mathrm{cov}(\boldsymbol{F}_{c,t}, \boldsymbol{R}_t) \\ &= \frac{(N-1)}{\lambda_t}\mathrm{corr}(\boldsymbol{F}_{c,t}, \boldsymbol{R}_t)\mathrm{dis}(\boldsymbol{F}_{c,t})\mathrm{dis}(\boldsymbol{R}_t). \end{aligned} \tag{66}$$

Let $\phi_{ij,t} := \mathrm{cov}(\boldsymbol{F}_{i,t}, \boldsymbol{F}_{j,t})$ and $\boldsymbol{\Phi}_t := (\phi_{ij,t})_{i,j=1}^{M}$ be the factor covariance matrix. Then,

$$\alpha_t = IC_{c,t}\sqrt{N-1}\sigma_{model}\mathrm{dis}(\boldsymbol{R}_t), \tag{67}$$

where

$$IC_t = \mathrm{corr}(\boldsymbol{F}_{c,t}, \boldsymbol{R}_t) = \frac{\sum_{i=1}^{M} v_i IC_{i,t}\mathrm{dis}(\boldsymbol{F}_{i,t})}{\sqrt{\boldsymbol{v}^T\boldsymbol{\Phi}_t\boldsymbol{v}}}. \tag{68}$$

Notes.

1. There are many time-varying components to $IC_{c,t}$, e.g. ICs of factors, dispersion of forecasts and their covariance matrix.
2. As $\boldsymbol{v}$ appears in quadratic form in the denominator it is unlikely that we can find an analytic solution to maximise IR.

   Couple of approaches:

   (a) Approximation (assume factor correlations are constant).

   (b) Transformation of factors to orthogonal factors.

7.1 – Optimal Alpha Model.

Standardise factors so that $\mathrm{dis}(\boldsymbol{F}_{i,t}) = 1 \, \forall i, t$.

This is common practice: equalises contribution of each factor, reduces turnover (reduces exposure to single factor changing). Also induces a naturally time-varying alpha model.

With standardised factors, the covariance matrix Φ reduces to a correlation matrix. The composite IC becomes a linear combination of factor ICs, scaled by the dispersion of the composite factor, τ:

$$IC_{c,t} = \frac{1}{\tau}\sum_{i=1}^{M} v_i IC_{i,t}. \tag{69}$$

Let $\overline{\boldsymbol{IC}} := (\overline{IC}_1, \overline{IC}_2, \ldots, \overline{IC}_M)^T$ and $\boldsymbol{\Sigma}_{IC} := (\rho_{ij,IC})_{i,j=1}^{M}$, i.e. the mean and covariance of ICs. Then,

$$\overline{IC}_c = \frac{1}{\tau}\boldsymbol{v}^T\overline{\boldsymbol{IC}} \tag{70}$$

and

$$\text{std}(IC_c) = \frac{1}{\tau}\sqrt{\boldsymbol{v}\boldsymbol{\Sigma}_{IC}\boldsymbol{v}}. \tag{71}$$

Hence,

$$IR = \frac{\boldsymbol{v}^T\overline{\boldsymbol{IC}}}{\sqrt{\boldsymbol{v}^T\boldsymbol{\Sigma}_{\boldsymbol{IC}}\boldsymbol{v}}}, \tag{72}$$

i.e. τ drops out and the determining factor of active risk is thus $\boldsymbol{\Sigma}_{IC}$.

It can be shown that the optimal weight vector is

$$\boldsymbol{v}^* = s\boldsymbol{\Sigma}_{IC}^{-1}\overline{\boldsymbol{IC}}, \tag{73}$$

for arbitrary constant s. We choose s such that the weights sum to 1. Substituting in,

$$IR^* = \sqrt{\overline{\boldsymbol{IC}}^T\boldsymbol{\Sigma}_{IC}^{-1}\overline{\boldsymbol{IC}}} \tag{74}$$

7.2 – Composite Alpha Model with Orthogonalised Factors.

In the last section we assumed constant factor correlations through time. Factor correlations are actually time-varying (although they are small compared to IC volatilities, so the approximation was somewhat justified).

Let's relax this assumption. We orthogonalise the factors, so that the correlations are zero by construction and the composite IC becomes

$$IC_{c,t} = \frac{1}{\sqrt{\boldsymbol{v}^T\boldsymbol{v}}}\sum_{i=1}^{M} v_i IC_{i,t}. \tag{75}$$

Then, the ICs are the only terms that vary in time and the IR is precisely

$$IR_c = \frac{\boldsymbol{v}^T\overline{\boldsymbol{IC}}}{\sqrt{\boldsymbol{v}^T\Sigma_{IC}\boldsymbol{v}}} \tag{76}$$

and the optimal weights solution holds without approximation.

Gram-Schmidt Procedure.

Consider M standardised factors, $(\boldsymbol{F}_1, \ldots, \boldsymbol{F}_M)$.

1st orthogonal factor: $\boldsymbol{F}_1^o = F_1$.

2nd orthogonal factor: $\boldsymbol{F}_2^o = \frac{1}{\sqrt{1-\rho_{21}^2}}(\boldsymbol{F}_2 - \rho_{21}\boldsymbol{F}_1^o)$, where $\rho_{21} = \text{corr}(\boldsymbol{F}_2, \boldsymbol{F}_1^o)$.

3rd orthogonal factor: $\boldsymbol{F}_3^o = \frac{1}{\sqrt{1-\rho_{32}^2-\rho_{31}^2}}(\boldsymbol{F}_3 - \rho_{32}\boldsymbol{F}_2^o - \rho_{31}\boldsymbol{F}_1^o)$, where $\rho_{31} = \text{corr}(\boldsymbol{F}_3, \boldsymbol{F}_1^o)$ and $\rho_{32} = \text{corr}(\boldsymbol{F}_3, \boldsymbol{F}_2^o)$.

And so on.

The optimal model with orthogonalised factors follows the standard form, $\boldsymbol{v}^* = s\boldsymbol{\Sigma}_{IC}^{-1}\overline{\boldsymbol{IC}}$.

Another standard method is **principal component analysis**. The principal components of $(\boldsymbol{F}_1, \ldots, \boldsymbol{F}_M)$ are linear combinations; the first component has the largest cross-sectional dispersion across all linear combinations, second component has next largest, etc.

8 – Portfolio Turnover and the Optimal Alpha Model.

There are two parts to effectively making money:

1. Theoeretical value of alpha skill (gross profit).
2. Cost of implementation (leads to net profit).

Ideally, $1 \gg 2$!

The total **assets under management** (AUM) affects 2. Lots of strategies make money with small AUM but scale badly. As the AUM increases we need to reduce turnover to reduce the cost of trading.

Typically, managers build an alpha model and put this into an optimiser with turnover constraints. This decouples the turnover problem from the alpha model so it can be hard to assess the model's effectiveness. Also, it can make it hard to change the model as AUM grows. It would be better if turnover was part of the model selection process.

Transaction Costs.

Sources include exchanges fees, broker commissions, bid/ask spread, market impact.

Costs vary by stock (liquidity). Overall costs are related to portfolio turnover.

Short-term alphas incur larger cost (more frequent rebalancing): factor autocorrelation is key.

Passive portfolio drift.

Suppose we start with weights $\boldsymbol{w} = (w_1, \ldots, w_N)^T$, where $\boldsymbol{e}^T \boldsymbol{w} = 1$.

If returns are $\boldsymbol{r} = (r_1, \ldots, r_N)^T$, the portfolio return is $r_p = \boldsymbol{w}^T \boldsymbol{r}$.

The new portfolio weights are $w_i(1+r_i)/(1+r_p)$; $i = 1, \ldots, N$, so the change in weight for stock i is

$$\Delta w_i = \frac{w_i\,(r_i - r_p)}{1 + r_p}. \tag{77}$$

Weights drift purely due to changing prices.

8.1 – Turnover of Fixed-weight portfolio

We would like to maintain fixed weights (i.e. correct for drift).

Definition (Turnover). In moving from the portfolio $\boldsymbol{w}^0 := (w_1^0, \ldots, w_N^0)^T$ to the portfolio $\boldsymbol{w}^1 := (w_1^1, \ldots, w_N^1)^T$, we induce a turnover of

$$T := \frac{1}{2} \sum_{i=1}^{N} \left| w_i^1 - w_i^0 \right|. \tag{78}$$

Note that the amount of buying offsets the amount of selling. We divide by 2 so that this is one-way turnover.

For a long-only portfolio, if we entirely replaced the portfolio, we have $T = \frac{1}{2} \sum_{i=1}^{N} \left(w_i^1 + w_i^0 \right) = 1$, i.e. 100% turnover.

Turnover is normally quoted on an annualised basis: a portfolio with 100% turnover in 1 year implies that the average holding period for a position is 1 year.

Turnover due to drift.

We have

$$T = \frac{1}{2} \sum_{i=1}^{N} |\Delta w_i| = \frac{1}{2(1+r_p)} \sum_{i=1}^{N} |w_i(r_i - r_p)|. \tag{79}$$

For the equally-weighted portfolio,

$$T = \frac{1}{2(1+r_p)N} \sum_{i=1}^{N} |r_i - r_p|. \tag{80}$$

Assume that $r \sim \mathcal{N}(\bar{r}, d^2)$; $\bar{r}$ is the average return, d is the dispersion. Then,

$$T \approx \frac{1}{2(1+r_p)} \mathbb{E}\left[|r - r_p|\right]. \tag{81}$$

For an equally-weighted portfolio, $r_p = \bar{r}$ and given that $\mathbb{E}\left[|r - \bar{r}|\right] = \sqrt{\frac{2}{\pi}} d$, we have

$$T \approx \frac{d}{(1+\bar{r})\sqrt{2\pi}}. \tag{82}$$

The turnover is directly related to dispersion and the average return. This is a simplified example though, and weights are generally uneven.

Assume weights and returns are independent. This is incorrect for active portfolios that outperform though; there we would have positive correlation. Now,

$$\begin{aligned} T &= \frac{N}{2(1+r_p)} \frac{1}{N} \sum_{i=1}^{N} |w_i| \, |r_i - r_p| \approx \frac{N}{2(1+r_p)} \mathbb{E}\left[|w| \, |r - r_p|\right] \\ &= \frac{N}{2(1+r_p)} \mathbb{E}[|w|]\mathbb{E}[|r - r_p|] = \frac{1}{2(1+r_p)} \sum_{i=1}^{N} |w_i| \mathbb{E}[|r - r_p|]. \end{aligned} \tag{83}$$

For a long-only portfolio, we have $\sum_{i=1}^{N} |w_i| = 1$. For long-short portfolios,

we have $\sum_{i=1}^{N} |w_i| = L$, where L is the portfolio leverage. So,

$$T \approx \frac{L}{2(1+r_p)} \mathbb{E}(|r - r_p|). \tag{84}$$

If $\bar{r}$ is different from r_p, then

$$T \approx \frac{Ld}{\sqrt{2\pi}(1+r_p)} \left[1 + \frac{1}{2}\left(\frac{\Delta r}{d}\right)^2\right]; \; \Delta r := r_p - \bar{r}. \tag{85}$$

The difference between the portfolio return and the average return contributes to turnover; the ratio of this difference to dispersion is key and large values can impact turnover significantly.

Turnover induced by changing forecasts.

Most turnover is due to forecast change for active portfolios. We have to balance benefit of going to a new portfolio with a higher expectation with the cost of doing so.

Assume we solve an unconstrained MVO at each time t:

$$w_i^t = \frac{1}{\lambda_t} \frac{F_i^t}{\sigma_i}. \tag{86}$$

Assume that N, σ_i and σ_{model} are all constant through time. Then,

$$\lambda_t = \frac{\sqrt{N-1}\,\mathrm{dis}(\boldsymbol{F}^t)}{\sigma_{model}} \tag{87}$$

and

$$w_i^t = \frac{\sigma_{model}}{\sqrt{N-1}} \frac{\tilde{F}_i^t}{\sigma_i}, \tag{88}$$

where $\tilde{\boldsymbol{F}}_t := \boldsymbol{F}_t / \mathrm{dis}(\boldsymbol{F}_t)$.

Just focusing on the turnover due to forecast change (i.e. ignore weight drift), we can write

$$\begin{aligned} T &= \frac{1}{2} \sum_{i=1}^{N} |w_i^{t+1} - w_i^t| = \frac{\sigma_{model}}{2\sqrt{N-1}} \sum_{i=1}^{N} \frac{|\tilde{F}_i^{t+1} - \tilde{F}_i^t|}{\sigma_i} \\ &\approx \frac{\sigma_{model}\sqrt{N}}{2} \mathbb{E}\left[\frac{|\tilde{F}^{t+1} - \tilde{F}^t|}{\sigma}\right]. \end{aligned} \tag{89}$$

So, assuming independence,

$$T \approx \frac{\sigma_{model}\sqrt{N}}{2} \mathbb{E}(|\tilde{F}_i^{t+1} - \tilde{F}_i^t|)\mathbb{E}\left(\frac{1}{\sigma}\right). \tag{90}$$

We'd expect that if our forecast has a higher autocorrelation, we would have lower turnover.

Let's make this concrete. Assume that $\tilde{F}_i^t$ and $\tilde{F}_i^{t+1}$ have a bivariate Normal distribution (with stdev 1) and correlation ρ_f (lag 1 autocorrelation). Both forecasts are Normally-distributed, so

$$\tilde{F}^{t+1} - \tilde{F}^{t} \sim \mathcal{N}\left(0, 2(1-\rho_f)\right). \tag{91}$$

Then,

$$\mathbb{E}(|\tilde{F}^{t+1} - \tilde{F}^{t}|) = 2\sqrt{\frac{1-\rho_f}{\pi}} \tag{92}$$

and so

$$T = \sqrt{\frac{N}{\pi}}\sigma_{model}\sqrt{1-\rho_f}\,\mathbb{E}\left(\frac{1}{\sigma}\right). \tag{93}$$

So, the forecast-induced turnover for an unconstrained long-short portfolio can be expressed in terms of the targeted tracking error, forecast autocorrelation and distribution of specific risks.

Turnover and leverage.

Leverage, L, can be written as

$$L = \sum_{i=1}^{N} |w_i| \approx \sigma_{model}\sqrt{N}\,\mathbb{E}\left[|\tilde{F}^{t}|\right]\mathbb{E}\left[\frac{1}{\sigma}\right], \tag{94}$$

assuming independence between forecasts and specific risks. As $\tilde{F}^{t} \sim \mathcal{N}(0,1)$, we have $\mathbb{E}\left[|\tilde{F}^{t}|\right] = \sqrt{2/\pi}$, and so

$$L = \sqrt{\frac{2N}{\pi}}\sigma_{model}\,\mathbb{E}\left[\frac{1}{\sigma}\right]. \tag{95}$$

Hence, using previous results,

$$T = \frac{L\sqrt{1-\rho_f}}{\sqrt{2}}. \tag{96}$$

Turnover (hence costs) increase linearly with leverage.

8.2 – Optimal Turnover-constrained Model.

To reduce turnover, include lagged forecasts in the model (even if these are weaker predictors) to increase the autocorrelation of the composite forecast.

Alternatively, include other (slower-varying) forecasts. Value factors have little decay, so we can give a high weight to lags of these factors. Momentum factors decay more quickly.

Consider $F^{t}_{c,ma} = v_{01}F^{t}_{1} + v_{02}F^{t}_{2} + v_{11}F^{t-1}_{1} + v_{12}F^{t-1}_{2}$. Augment the weight vector, $\boldsymbol{v} = (v_{01}, v_{02}, v_{11}, v_{12})^T$ and create the stacked vector $(\boldsymbol{F}^{t+1}_1, \boldsymbol{F}^{t+1}_2, \boldsymbol{F}^{t}_1, \boldsymbol{F}^{t}_2, \boldsymbol{F}^{t-1}_1, \boldsymbol{F}^{t-1}_2)$.

Label elements of the correlation matrix as $\rho^{l,k}_{ij} = \mathrm{corr}(\boldsymbol{F}^{t+l}_i, \boldsymbol{F}^{t+k}_j)$. If you write this out, you will see that

$$\mathrm{var}(\boldsymbol{F}^{t}_{c,ma}) = \boldsymbol{v}^T\boldsymbol{C}_4\boldsymbol{v} \text{ and } \mathrm{cov}(\boldsymbol{F}^{t}_{c,ma}, \boldsymbol{F}^{t-1}_{c,ma}) = \boldsymbol{v}^T\boldsymbol{D}_4\boldsymbol{v}, \tag{97}$$

where $\boldsymbol{C}_4$ and $\boldsymbol{D}_4$ are the 4×4-matrices in the upper left and upper right of the correlation matrix, respectively.

Therefore, we can write the autocorrelation of the composite model as

$$\rho_{f_{c,ma}} = \frac{\boldsymbol{v}^T\boldsymbol{D}_4\boldsymbol{v}}{\boldsymbol{v}^T\boldsymbol{C}_4\boldsymbol{v}}, \tag{98}$$

which is the general form for a composite forecasting model with multiple factors and lags. If we wish to find the IR-optimal portfolio with a turnover constraint, we must solve the problem

$$\begin{aligned} \max \quad & \frac{\boldsymbol{v}^T\overline{\boldsymbol{IC}}}{\sqrt{\boldsymbol{v}^T\boldsymbol{\Sigma}_{IC}\boldsymbol{v}}} \\ \text{s.t.} \quad & \frac{\boldsymbol{v}^T\boldsymbol{D}_4\boldsymbol{v}}{\boldsymbol{v}^T\boldsymbol{C}_4\boldsymbol{v}} = \rho_t, \end{aligned} \tag{99}$$

where ρ_t is the target autocorrelation. This is a nonlinear optimisation problem with a quadratic constraint, so we need to use numerical techniques to solve it.

9 – Portfolio Constraints.

9.1 – Typical constraints.

For effective risk and cost control, we need to add constraints to the optimisation problem. Some typical constraints:

1. Risk exposure (market, size, growth, statistical): reduce systematic risk.
2. Size of holding: no single stock exposure can be greater than some percent of the portfolio.
3. Sector bounds: only allow ±3% for sector/country bets.
4. Long-only: Impose $w_i > 0 \; \forall i$, "no-short rule".
5. Long-short: constrain leverage.
6. Gross book constraint.

Constraints generally boil down to equality and inequality constraints. For the former, we can usually find exact solutions for the optimal weights. For the latter, numerical techniques are required.

9.2 – Sector-neutral constraint.

Some sectors look more expensive than others for some fundamental factors. For example, using a growth metric, tech stocks are overpriced relative to utilities. Without sector neutrality, you risk being long utilities and short tech in this example, so this is an important constraint.

Recall that

$$\alpha_t = \lambda_t^{-1} \sum_{i=1}^{N} F_i R_i, \tag{100}$$

where F_i and R_i are risk-adjusted forecasts and returns, respectively.

Suppose we have S sectors, such that $N = \sum_{s=1}^{S} N_s$. Write

$$\alpha_t = \lambda_t^{-1} \sum_{s=1}^{S} \sum_{i=1}^{N_s} F_{si} R_{si}, \tag{101}$$

where F_{si} is the forecast for stock i in sector s, etc.

Letting $\bar{F}_s = \frac{1}{N_s} \sum_{i=1}^{N_s} F_{si}$, we have $\bar{F} = \sum_{s=1}^{S} \frac{N_s}{N} \bar{F}_s$.

We have equivalent expressions for $\bar{R}_s$ and $\bar{R}$.

Assume that $\bar{F}$ and $\bar{R}$ are close to zero. Then, it can be shown that

$$\alpha_t = \lambda_t^{-1} \sum_{s=1}^{S} \sum_{i=1}^{N_s} \left[\left(F_{si} - \bar{F}_s\right) \left(R_{si} - \bar{R}_s\right) \right] + \lambda_t^{-1} \sum_{s=1}^{S} N_s \bar{F}_s \bar{R}_s. \tag{102}$$

The first term is the excess return generated by the sector-relative forecast. The second term is related to the excess return of the sector itself:

$$\sum_{s=1}^{S} N_s \bar{F}_s \bar{R}_s = N \sum_{s=1}^{S} \frac{N_s}{N} \bar{F}_s \bar{R}_s \approx N \sum_{s=1}^{S} \frac{N_s}{N} \left(\bar{F}_s - \bar{F}\right) \left(\bar{R}_s - \bar{R}\right), \tag{103}$$

which is the covariance of the aggregated sector forecast and the aggregated sector return.

9.3 – Mean-variance optimisation with linear and range constraints.

For all typical constraint types we encounter, we are faced with a quadratic programming problem with linear and range constraints. These problems have to be solved numerically for a general covariance matrix. However, in the case where the covariance matrix is diagonal (e.g. we have neutralised all systematic factor exposures and are optimising with residual alphas and risks), there is an efficient algorithm.

Our problem:

$$\begin{aligned}
\max \quad & \boldsymbol{f}^T \boldsymbol{w} \\
\text{s.t.} \quad & \boldsymbol{w}^t \boldsymbol{\Sigma} \boldsymbol{w} = \sigma^2_{target}, \\
& \boldsymbol{w}^t \boldsymbol{e} = 0, \\
& \boldsymbol{w}^T \boldsymbol{B} = 0, \\
& w_i - U_i \le 0,\ L_i - w_i \le 0 \text{ for } i = 1, \dots, N, \qquad (104)
\end{aligned}$$

where $\boldsymbol{f}$ is the forecast vector, $\boldsymbol{\Sigma} = \boldsymbol{B}\boldsymbol{\Sigma}_I\boldsymbol{B}^T + \boldsymbol{S}$ is the covariance matrix and σ_{target} is the target tracking error.

$\boldsymbol{w}^T\boldsymbol{e} = 0$ is the cash-neutral constraint and $\boldsymbol{w}^T\boldsymbol{B} = 0$ is the market neutral contraint. Finally, we have the range constraints (note that we have broken these into separate inequality constraints.

Karusch-Kuhn-Tucker (KKT) conditions for optimality.

Suppose our problem is to max $p(\boldsymbol{w})$ s.t. $g_j(\boldsymbol{w}) \le 0$ for $j = 1, \dots, m$. Define the Lagrangian function as

$$L(\boldsymbol{w}) = p(\boldsymbol{w}) - \sum_{j=1}^{m} l_j g_j(\boldsymbol{w}), \qquad (105)$$

where l_i are Lagrange multipliers. The KKT conditions for optimality are

$$\frac{\partial L(\boldsymbol{w})}{\partial w_i} = \frac{\partial p(\boldsymbol{w})}{\partial w_i} - \sum_{j=1}^{m} l_j \frac{\partial g_j(\boldsymbol{w})}{\partial w_i} = 0 \text{ for } i = 1, \dots, N. \qquad (106)$$

and

$$g_j(\boldsymbol{w}) \le 0,\ l_j \ge 0 \text{ and } l_i g_i(\boldsymbol{w}) = 0 \text{ for } i = 1, \dots, m. \qquad (107)$$

Back to our problem. The Lagrangian is

$$\begin{aligned}
L(\boldsymbol{w}) \quad = \quad & \boldsymbol{f}^T \boldsymbol{w} \\
& - \lambda(\boldsymbol{w}^T \boldsymbol{\Sigma} \boldsymbol{w} - \sigma^2_{target}) \\
& - l_0 \boldsymbol{w}^t \boldsymbol{e} \\
& - \sum_{i=1}^{K} l_i \boldsymbol{w}^T \boldsymbol{b}_i \\
& - \sum_{j=1}^{N} \left[\tilde{l}_{j1}(w_j - U_j) + \tilde{l}_{j2}(L_j - w_j) \right], \qquad (108)
\end{aligned}$$

where λ, l_0, $\tilde{l}_{j1}$ and $\tilde{l}_{j2}$ are Lagrangian multipliers associated with separate constraints.

Next, the KKT conditions. We have

$$\frac{\partial L(\boldsymbol{w})}{\partial \boldsymbol{w}} = \boldsymbol{f} - 2\lambda \boldsymbol{\Sigma} \boldsymbol{w} - l_0 \boldsymbol{e} - \sum_{i=1}^{K} l_i \boldsymbol{b}_i - (\tilde{\boldsymbol{l}}_1 - \tilde{\boldsymbol{l}}_2) = 0, \tag{109}$$

where $\tilde{\boldsymbol{l}}_1 := (\tilde{l}_{11}, \ldots, \tilde{l}_{1N})^T$; similarly for $\tilde{\boldsymbol{l}}_2$.

Equality constraints must satisfy

$$\boldsymbol{w}^T \boldsymbol{\Sigma} \boldsymbol{w} = \sigma^2_{target}, \boldsymbol{w}^T \boldsymbol{e} = 0, \boldsymbol{w}^T \boldsymbol{B} = 0. \tag{110}$$

The range constraints must satisy

$$\begin{aligned} \tilde{l}_{j1} \geq 0, w_j - U_j \leq 0 \text{ and } \tilde{l}_{j1}(w_j - U_j) &= 0; \\ \tilde{l}_{j2} \geq 0, L_j - w_j \leq 0 \text{ and } \tilde{l}_{j2}(L_j - w_j) &= 0. \end{aligned} \tag{111}$$

Equation (109) can be solved to give

$$\boldsymbol{w} = \frac{1}{2\lambda} \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{f} - l_0 \boldsymbol{e} - \sum_{i=1}^{K} l_i \boldsymbol{b}_i - \tilde{\boldsymbol{l}}_1 + \tilde{\boldsymbol{l}}_2 \right) =: \frac{1}{2\lambda} \boldsymbol{\Sigma}^{-1} \boldsymbol{f}_{adj}. \tag{112}$$

This is a solution to an unconstrained optimisation problem, but with the forecast adjusted for constraints, then scaled by λ to give the required tracking error.

If $L_i < w_i < U_i$ then $\tilde{l}_{i1} = \tilde{l}_{i2} = 0$ (non-binding constraints).

If $w_i = U_i$ then $\tilde{l}_{i1} \geq 0, \tilde{l}_{i2} = 0$.

If $w_i = L_i$ then $\tilde{l}_{i1} = 0, \tilde{l}_{i2} \geq 0$.

Only one of $\tilde{l}_{i1}$ and $\tilde{l}_{i2}$ can be nonzero. So, let $\tilde{l}_i = \tilde{l}_{i1} - \tilde{l}_{i2}$.

If $\boldsymbol{\Sigma} = \boldsymbol{B} \boldsymbol{\Sigma}_I \boldsymbol{B}^T + \boldsymbol{S}$, then

$$\boldsymbol{w} = \frac{1}{2\lambda} \boldsymbol{S}^{-1} \boldsymbol{f}_{adj}, \tag{113}$$

as we are factor-neutral. For each stock i we have

$$w_i = \frac{f_i - l_0 - l_1 b_{1i} - \cdots - l_K b_{Ki} - \tilde{l}_i}{2\lambda \sigma_i^2} \tag{114}$$

The numerical procedure to solve this is as follows:

At each step n, we have w_i^n and $l_0^n, l_1^n, \ldots, l_K^n, \tilde{l}_i^n, \lambda^n$. If any weights violate the range constraint, do the following:

1. Update weights: $w_i^{new} = \max\left(\min\left(w_i^n, U_i\right), L_i\right)$.
2. Update range multipliers:
   $\tilde{l}_i^{n+1} = f_i - l_0^n - l_1^n b_{1i} - \cdots - l_K^n b_{Ki} - 2\lambda^n \sigma_i^2 w_i^{new}$
3. Update $l_0^{n+1}, \ldots, l_K^{n+1}$ from the solution of the set of linear equations.
4. Calculate $\sigma^{new} = \sqrt{\sum_{i=1}^{N} \left(w_i^{new}\right)^2 \sigma_i^2}$ and then $\lambda^{n+1} = \lambda^n \sigma^{new} / \sigma_{target}$.
5. Calculate new weights w_i^{n+1} using updated multipliers in equation (114).

Repeat this procedure until there are no range violations.

10 – Portfolio Implementation.

10.1 – Introduction

Turnover is only a rough proxy for trading costs. What we actually need in practice are detailed costs, i.e. different costs for different stocks.

Given an initial portfolio $\boldsymbol{w}_0$, we consider the move to a new portfolio $\boldsymbol{w}$ and the associated costs, assumed to be a function $c(\Delta \boldsymbol{w})$; $\Delta w := \boldsymbol{w} - \boldsymbol{w}_0$.

The function $c(\cdot)$ depends on how we trade and the stock liquidity. Once we have defined $c(\cdot)$, it must be added to the objective function.

You can't separate implementation from optimisation.

10.2 – Components of $c(\cdot)$.

Fixed costs. Commissions, bid/ask spread, taxes, borrow fees. These are (generally) on a per-share basis, and so are a linear function of the traded value. So, model these as a constant (vector) × abs weight change:

$$c(\Delta \boldsymbol{w}) = \boldsymbol{\theta}^T |\Delta \boldsymbol{w}|, \boldsymbol{\theta} > \boldsymbol{0}. \tag{115}$$

For bid/ask spread, we assume we pay half on the way in and half on the way out.

Variable costs. Market impact (price change due to our trading; temporary/permanent), opportunity cost (what we lose by not getting filled; limit order placement). Market impact is not linear: increases dramatically as trade size increases. Model as a quadratic cost:

$$c(\Delta w_i) = \psi_i \left(w_i\right)^2, \psi_i \geq 0 \tag{116}$$

10.3 – Examples with a single asset.

Quadratic cost. Solve the following problem (risky asset with cash):

$$\max f w - \frac{1}{2}\lambda\sigma^2 w^2 - \psi(w - w_0)^2 =: U(w) \tag{117}$$

We have

$$U'(w) = f - \lambda\sigma^2 w - 2\psi(w - w_0) = 0 \implies w^* = \frac{f + 2\psi w_0}{\lambda\sigma^2 + 2\psi}. \tag{118}$$

As ψ increases from 0, $w^* \to w_0$ (slowly).

Let $\Delta w^* := w^* - w_0$ be the optimal trade with cost and $\Delta\tilde{w} := \tilde{w} - w_0$, where $\tilde{w} = f/(\lambda\sigma^2)$, be the optimal cost-free trade.

Note that

$$\Delta w^* = \frac{\Delta\tilde{w}}{1 + 2\psi/(\lambda\sigma^2)}. \tag{119}$$

The costed trade is a fraction of the cost-free trade; scaling is a ratio of cost to risk.

There is always some trading with quadratic costs, no matter how large ψ is. The higher ψ is, the closer to w_0 we trade.

Linear cost. Solve the following problem:

$$\max f w - \frac{1}{2}\lambda\sigma^2 w^2 - \theta|w - w_0| =: U(w). \tag{120}$$

$|x|$ is not differentiable at $x = 0$, so this is trickier. With no costs, $\theta = 0$, optimal weight is $\tilde{w} = f/(\lambda\sigma^2)$ as before.

When $\theta > 0$, formulate the problem in terms of Δw:

$$\begin{aligned} U(\Delta w) &= f(\Delta w + w_0) - \frac{1}{2}\lambda^2(\Delta w + w_0)^2 - \theta|\Delta w| \\ &= U(w_0) + \left[\lambda\sigma^2(\tilde{w} - w_0)\Delta w - \theta|\Delta w| - \frac{1}{2}\lambda\sigma^2(\Delta w)^2\right], \end{aligned} \tag{121}$$

where $U(w_0) = fw_0 - \frac{1}{2}\lambda\sigma^2 w_0^2$.

Then,

$$\Delta U = U(\Delta w) - U(w_0) = \lambda\sigma^2 \Delta\tilde{w}\Delta w - \theta|\Delta w| - \frac{1}{2}\lambda\sigma^2(\Delta w)^2, \qquad (122)$$

with $\Delta\tilde{w} = \tilde{w} - w_0$.

We have three cases:

1. $\tilde{w} = w_0 \implies \Delta w = 0$ is optimal and $U(\Delta w) = U(w_0)$.
2. $\tilde{w} > w_0$. The last two terms in equation (122) are < 0 so the only way to improve utility is in the first term. We need $\Delta\tilde{w}$ and Δw to have the same sign. So, when $\tilde{w} > 0$, look for a buy solution, $\Delta w \geq 0$.

   For $\Delta w > 0$, $|\Delta w| = \Delta w$, so we have

$$U'(\Delta w) = \lambda\sigma^2\Delta\tilde{w} - \theta - \lambda\sigma^2\Delta w = 0 \qquad (123)$$

so that

$$\Delta w^* = \Delta\tilde{w} - w_c;\ w_c = \frac{\theta}{\lambda\sigma^2}. \qquad (124)$$

3. Similar analysis follows for $\tilde{w} < w_0$.

In the final solution, we have a zone in the middle where we don't trade:

$$\Delta w^* = \begin{cases} \Delta\tilde{w} - w_c, & \Delta\tilde{w} > w_c \\ 0, & |\Delta\tilde{w}| \leq w_c \\ \Delta\tilde{w} + w_c, & \Delta\tilde{w} < -w_c. \end{cases} \qquad (125)$$

11 – Robust methods.

Problems with MVO.

1. MVO portfolios are often outperformed by simple strategies like equally-weighted portfolios.
2. Portfolio weights are unstable over time. Leads to excessive turnover and associated costs.
3. Weights are often extreme holdings ("corner solutions") in a few stocks, zero in the rest. Leads to undiversified portfolio (and excessive risk).
4. Classical MVO assumes inputs are deterministic (and hence known with certainty).

Sensitivity to Estimation Error.

Stocks with large forecasts and low specific variance are overweighted. Large errors in the forecast/variance introduce large errors in the weights. Optimisers are often called "error maximisers".

Errors in expected returns have greater influence than errors in variances/covariances. The relative magnitudes depend on the risk aversion, but as a rule of thumb: errors in $\boldsymbol{f}$ are $10\times$ more important than errors in $\boldsymbol{\Sigma}$. Further, errors in variances are $2\times$ more important than errors in covariances [2].

Approaches to reduce the influence of errors:

1. Constrain portfolio weights.
2. Improve accuracy of inputs using a robust approach: shrinkage/Bayesian estimators.
3. Incorporate errors directly in optimisation process.

11.1 – Shrinkage estimation.

Stein. Biased estimators often give better results than unbiased estimators.

Consider $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{X}$ is N-dimensional, $N > 2$. Assume we are trying to estimate $\boldsymbol{\mu}$, with $\boldsymbol{\Sigma}$ known.

Sample mean $\hat{\boldsymbol{\mu}}$ is not the best estimator in terms of a quadratic loss function, $L(\boldsymbol{\mu}, \hat{\boldsymbol{\mu}}) := (\mu - \hat{\boldsymbol{\mu}})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \hat{\boldsymbol{\mu}})$.

The James-Stein shrinkage estimator,

$$\hat{\boldsymbol{\mu}}_{JS} := (1-w)\hat{\boldsymbol{\mu}} + w\mu_0 \boldsymbol{e} \tag{126}$$

with

$$w := \min\left(1, \frac{N-2}{T(\hat{\boldsymbol{\mu}} - \mu_0\boldsymbol{e})^T \boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\mu}} - \mu_0\boldsymbol{e})}\right), \tag{127}$$

has a lower quadratic loss than the sample mean. Here T is the number of observations and μ_0 can be *any* number (Stein paradox!). $\mu_0\boldsymbol{e}$ is called the shrinkage target and w is the shrinkage intensity.

Definition (Shrinkage). Form of averaging different estimators where the shrinkage estimator has three components:

1. Estimator with little/no structure (e.g. sample mean).
2. Estimator with a lot of structure (e.g. shrinkage target).
3. Shrinkage intensity.

We have the following requirements for the shrinkage target:

1. Small number of free parameters (robust, lots of structure).
2. Some properties in common with the unknown quantity being estimated.

The shrinkage intensity is chosen considering theory or by numerical simulation.

Best known shrinkage estimator for expected returns is Jorion's. The target is $\mu_g\boldsymbol{e}$, where

$$\mu_g = \frac{\boldsymbol{e}^T\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\mu}}}{\boldsymbol{e}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{e}} \tag{128}$$

and

$$w = \frac{N+2}{N+2+T(\hat{\boldsymbol{\mu}} - \mu_g\boldsymbol{e})^T\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\mu}} - \mu_g\boldsymbol{e})}. \tag{129}$$

μ_g is the return on a minimum variance portfolio.

Several studies have shown using that when using this estimator in MVO problems we see:

1. A decrease in portfolio weights variability from one period to the next.
2. A significant improvement in out-of-sample risk-adjusted performance.

Shrinkage techniques can also be applied to covariance matrix estimation.

Ledoit/Wolf. Shrink to $\mathbf{\Sigma}$ from a single factor (CAPM) model or a constant-correlation $\mathbf{\Sigma}$. In practice, very similar results obtained using either, but a constant-correlation covariance structure is easier to handle:

$$\hat{\mathbf{\Sigma}}_{LW} = w\hat{\mathbf{\Sigma}}_{CC} + (1-w)\hat{\mathbf{\Sigma}}, \tag{130}$$

where $\hat{\mathbf{\Sigma}}$ is the sample covariance matrix, $\hat{\mathbf{\Sigma}}_{CC}$ is a covariance matrix with constant correlation.

Compute $\hat{\mathbf{\Sigma}}_{CC}$ as follows:

1. Decompose $\hat{\mathbf{\Sigma}} = \mathbf{\Lambda}\boldsymbol{C}\mathbf{\Lambda}^T$, where $\mathbf{\Lambda}$ is a diagonal matrix of volatilities and

Asset Pricing and Portfolio Theory

$\boldsymbol{C}$ is the sample correlation matrix:

$$\boldsymbol{C} = \begin{pmatrix} 1 & \hat{\rho}_{1,2} & \cdots & \hat{\rho}_{1,N} \\ \hat{\rho}_{2,1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \hat{\rho}_{N-1,N} \\ \hat{\rho}_{N,1} & \cdots & \hat{\rho}_{N,N-1} & 1 \end{pmatrix}. \tag{131}$$

Replace $\boldsymbol{C}$ with $\boldsymbol{C}_{CC}$,

$$\boldsymbol{C}_{CC} = \begin{pmatrix} 1 & \hat{\rho} & \cdots & \hat{\rho} \\ \hat{\rho} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \hat{\rho} \\ \hat{\rho} & \cdots & \hat{\rho} & 1 \end{pmatrix}, \tag{132}$$

where $\hat{\rho}$ is the average of all sample correlations.

Asset Pricing and Portfolio Theory

The optimal shrinkage intensity is proportional to A/T, where A is a constant and T is the length of history (see [2] for details).

Ledoit and Wolf have shown that this estimator performs better than

1. sample covariance;
2. a statistical factor model with the first five principal components; and
3. a factor model based on 48 industry factors.

The shrinkage intensity for their single-factor model is fairly constant through time at 0.8.

Asset Pricing and Portfolio Theory

11.2 – Robust optimisation

MVO assumes that expected returns, $\boldsymbol{\mu}$, and covariances, $\mathbf{\Sigma}$, are known with certainty. How can we incorporate estimation errors directly into the portfolio management process?

Portfolio Resampling. Monte Carlo: repeatedly draw from the $\boldsymbol{\mu}, \mathbf{\Sigma}$ distribution and average to get the portfolio weights. Slow, cumbersome.

Robust allocation. Distributions are put directly into the optimisation process. Fast, efficient. Not part of the mainstream, but becoming increasingly popular.

We will focus on the second, far more elegant approach.

Basic Idea.

Let $\hat{\boldsymbol{\mu}}$ be expected return; assume that we observe the true covariance $\boldsymbol{\Sigma}$.

If we knew the true return $\boldsymbol{\mu}_{true}$, we could solve the problem

$$\begin{aligned} \max \quad & \boldsymbol{w}^T \boldsymbol{\mu}_{true} - \lambda \boldsymbol{w}^T \boldsymbol{\Sigma} \boldsymbol{w}, \\ \text{s.t.} \quad & \boldsymbol{w}^T \boldsymbol{e} = 1. \end{aligned} \tag{133}$$

Instead, we end up solving

$$\begin{aligned} \max \quad & \boldsymbol{w}^T \hat{\boldsymbol{\mu}} - \lambda \boldsymbol{w}^T \boldsymbol{\Sigma} \boldsymbol{w}, \\ \text{s.t.} \quad & \boldsymbol{w}^T \boldsymbol{e} = 1, \end{aligned} \tag{134}$$

so that we are exposed to estimation error in $\hat{\boldsymbol{\mu}}$.

Assume $\hat{\boldsymbol{\mu}} - \boldsymbol{\epsilon} \leq \boldsymbol{\mu}_{true} \leq \hat{\boldsymbol{\mu}} + \boldsymbol{\epsilon}$, for $\boldsymbol{\epsilon} := (\epsilon_1, \ldots, \epsilon_N)^T$.

For example, we could take $\epsilon_i = 1.96\sigma_i/\sqrt{T}$, where T is the sample size, so that we have a 95% confidence interval for $\boldsymbol{\mu}_{true}$.

What is the "worst" estimate of return and how would we allocate the portfolio weights in this scenario?

Set up this optimisation problem:

$$\begin{aligned} \max_{\boldsymbol{w}} \min_{\boldsymbol{\mu}} \quad & \boldsymbol{w}^T \boldsymbol{\mu} - \lambda \boldsymbol{w}^T \boldsymbol{\Sigma} \boldsymbol{w}, \\ \text{s.t.} \quad & \boldsymbol{w}^T \boldsymbol{e} = 1, \\ & (\mu_i - \hat{\mu}_i)^2 \leq \epsilon_i^2, \quad i = 1, \ldots, N. \end{aligned} \tag{135}$$

This is a **max-min problem**: maximises expected utility (the minimum w.r.t. $\boldsymbol{\mu}$) for a given $\lambda > 0$ in the worst case realisation of $\boldsymbol{\mu}$.

If we have low confidence in a particular security return, the magnitude of the weight is reduced.

If we have high confidence, the weight is closer to the MVO weight.

The set $U_\epsilon(\boldsymbol{\mu}) := \left\{ \boldsymbol{\mu} | \, (\mu_i - \hat{\mu}_i)^2 \leq \epsilon_i^2, i = 1, \ldots N \right\}$ is called the **uncertainty set**.

We can write the problem as

$$\begin{aligned} \max_{\boldsymbol{w}} \min_{\boldsymbol{\mu} \in U_\epsilon(\boldsymbol{\mu})} \quad & \boldsymbol{w}^T \boldsymbol{\mu} - \lambda \boldsymbol{w}^T \boldsymbol{\Sigma} \boldsymbol{w}, \\ \text{s.t.} \quad & \boldsymbol{w}^T \boldsymbol{e} = 1. \end{aligned} \tag{136}$$

How do we solve this?

For a fixed $\boldsymbol{w}$, solve the inner problem:

$$\begin{aligned} \min_{\boldsymbol{\mu}} \quad & \boldsymbol{w}^T \boldsymbol{\mu} - \lambda \boldsymbol{w}^T \boldsymbol{\Sigma} \boldsymbol{w}, \\ \text{s.t.} \quad & (\mu_i - \hat{\mu}_i)^2 \leq \epsilon_i^2, \quad i = 1, \ldots, N. \end{aligned} \tag{137}$$

For optimal $\boldsymbol{\mu}^*$, the KKT conditions are

$$\begin{aligned} \nabla_{\boldsymbol{\mu}} L(\boldsymbol{\mu}^*; \boldsymbol{\gamma}) &= 0, \\ (\mu_i^* - \hat{\mu}_i)^2 &\leq \epsilon_i^2, \quad i = 1, \ldots, N, \\ \gamma_i &\geq 0, \quad i = 1, \ldots, N, \\ \gamma_i \left[(\mu_i^* - \hat{\mu}_i)^2 - \epsilon_i^2 \right] &= 0, \quad i = 1, \ldots, N, \end{aligned} \tag{138}$$

with Lagrangian

$$L(\boldsymbol{\mu}; \boldsymbol{\gamma}) := \boldsymbol{w}^T \boldsymbol{\mu} - \lambda \boldsymbol{w}^T \boldsymbol{\Sigma} \boldsymbol{w} - \frac{1}{2} \sum_{i=1}^{N} \gamma_i \left[(\mu_i - \hat{\mu}_i)^2 - \epsilon_i^2 \right]. \tag{139}$$

From $\nabla_{\boldsymbol{\mu}} L = 0$, so $w_i - \gamma_i(\mu_i^* - \hat{\mu}_i) = 0$ and then $\mu_i^* = w_i/\gamma_i + \hat{\mu}_i$.

Substitute $(\mu_i^* - \hat{\mu}_i)^2 \leq \epsilon_i^2$ into this to get $\gamma_i \geq \pm w_i/\epsilon_i$ and finally

$$\mu_i^* = \hat{\mu}_i - \operatorname{sgn}(w_i)\epsilon_i, \quad i = 1, \ldots, N. \tag{140}$$

This is the worst-case expected return as a function of weight and error.

Substitute this into the max-min problem:

$$\max_{\boldsymbol{w}} \boldsymbol{w}^T \left(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_{\boldsymbol{\epsilon},\boldsymbol{w}}\right) - \lambda \boldsymbol{w}^T \boldsymbol{\Sigma} \boldsymbol{w}, \tag{141}$$

where $\boldsymbol{\mu}_{\boldsymbol{\epsilon},\boldsymbol{w}} = \left(\mathrm{sgn}(w_1)\epsilon_1, \ldots, \mathrm{sgn}(w_N)\epsilon_N\right)^T$.

This is a modification to classical MVO, where $\boldsymbol{\mu}$ is adjusted down to account for uncertainty: this induces **shrinkage** of portfolio weights.

Write

$$w_i \mathrm{sgn}(w_i)\epsilon_i = w_i \frac{w_i}{|w_i|}\epsilon_i = \frac{w_i}{\sqrt{|w_i|}}\epsilon_i \frac{w_i}{\sqrt{|w_i|}}. \tag{142}$$

Then, we can write the problem as

$$max_{\boldsymbol{w}} \boldsymbol{w}^T \boldsymbol{\mu} - \lambda \boldsymbol{w}^T \boldsymbol{\Sigma} \boldsymbol{w} - \tilde{\boldsymbol{w}}^T \boldsymbol{E} \tilde{\boldsymbol{w}}, \tag{143}$$

where $\tilde{\boldsymbol{w}} = \left(\frac{w_1}{\sqrt{|w_1|}}, \ldots, \frac{w_N}{\sqrt{|w_N|}}\right)^T$ and $\boldsymbol{E} = \mathrm{diag}\left(\epsilon_1, \ldots, \epsilon_N\right)$.

This can also be viewed as a modification to classical MVO: a risk-like term $\tilde{\boldsymbol{w}}^T \boldsymbol{E} \tilde{\boldsymbol{w}}$ is added. This term represents a risk adjustment for the investor who is averse to estimation risk.

Note that these are not quadratic programming problems in general. They can be converted to second-order cone problems (SOCPs), which can be solved using interior point methods in about the same time as classical MVO.

Generally, robust solutions are more stable and have superior performance to classical MVO, but you have to careful as sometimes the worst-case solution is too pessimistic (and you end up holding no portfolio).

Extensions. Uncertainty in other inputs (e.g. factor returns). Different utility functions and risk measures. Different uncertainty sets. Multiperiod problems.

12 – Behavioural Asset Pricing

12.1 – Introduction

Behavioural Finance is the study of how psychological biases impact investor behaviour and subsequent market prices and returns. Investors do not behave as if they have rational preferences and do not form judgements along Bayesian lines. We see systematic deviations from rational behaviour both in experimental and real markets. There is a tendency to rely on "heuristics" when trading and to commit particular kinds of error, e.g. Representativeness, Anchoring, Gambler's Fallacy.

Behavioural Asset Pricing attempts to provide a concrete theoretical framework that incorporates investor sentiment and heterogeneity into the asset pricing problem.

Key text for this section is [5].

12.2 – A Simple Model with Two Investors

At time $t = 0$, assume there is an aggregrate amount ω_0 available for consumption.

For times $t > 0$ (assume $t = 0, \ldots, T$), assume the aggregate amount available for consumption unfolds through a Binomial process, growing by either $u > 1$ or $d < 1$ at each time, i.e. we either have $\omega_1 = \omega_0 u$ or $\omega_1 = \omega_0 d$, etc.

Denote the sequences of u, d by x_t (date-event pairs).

Define the cumulative growth between dates 0 and t as $g(x_t) = \omega(x_t)/\omega(x_0)$.

Probabilities. Assume we have two investors, indexed by $j = 1, 2$. At time 0, trader j attaches a probability $P_j(x_t)$ to the occurrence of x_t at time t. Only one x_t occurs at each time t, so for a specific t, $\sum_{x_t} P_j(x_t) = 1$.

Utility. Assume the traders only derive utility from consumption. Let $c_j(x_t)$ be the number of units that j consumes at the end of date-event pair x_t.

Discounting. Assume the near-future is more important than the distant-future: discount $c_j(x_t)$ by δ^t, $\delta \leq 1$.

Assume log-utility: investor j associates utility $\delta^t \ln(c_j(x_t))$ to consumption level $c_j(x_t)$ in date-event pair x_t. Consider the vector $\boldsymbol{c}_j$ whose components are $c_j(x_t)$ as the "consumption plan". The investor assesses alternative consumption plans based on expected utilities.

The expected utility associated with plan $\boldsymbol{c}_j$ is

$$\mathbb{E}(u_j) = \sum_{t,x_t} P_j(x_t) \delta^t \ln(c_j(x_t)). \tag{144}$$

State prices. Let $\nu(x_t)$ be the price of a contract that promises delivery of 1 unit of consumption if x_t occurs at t. The key attribute of prices is not the level, but the relative value, so we set $\nu(x_0) = 1$ wlog (i.e. the numeraire is time 0 consumption).

Budget constraint. Let $\omega(x_t)$ be the amount of aggregate consumption available in date-event pair, which is held jointly by the two investors; let $\omega_j(x_t)$ be the amount held by investor j.

The value of holding an initial amount $\omega_j(x_t)$ is $\nu(x_t)\omega_j(x_t)$. So, the initial wealth of investor j is

$$W_j = \sum_{t,x_t} \nu(x_t) \omega_j(x_t). \tag{145}$$

The budget constraint is

$$\sum_{t,x_t} \nu(x_t) c_j(x_t) \leq W_j, \tag{146}$$

i.e. the value of claims must be less than or equal to the initial wealth.

For simplicity, assume that $\omega_j(x_t) = w_j \omega(x_t)$, so that the initial wealth of investor 1 relative to investor 2 is w_1/w_2.

Expected utility maximisation. The problem investor j faces is

$$\begin{aligned} \max \qquad & \mathbb{E}(u_j) = \sum_{t,x_t} P_j(x_t) \delta^t \ln(c_j(x_t)), \\ \text{s.t.} \qquad & \sum_{t,x_t} \nu_j(x_t) c_j(x_t) \leq W_j. \end{aligned} \tag{147}$$

The solution is

$$c_j(x_t) = \frac{\delta^t}{\sum_{\tau=0}^{T} \delta^\tau} \frac{P_j(x_t)}{\nu(x_t)} W_j. \tag{148}$$

Left as exercise for the reader.

Equilibrium prices. For x_t, the demand is $c_1(x_t) + c_2(x_t)$ and the supply is $\omega(x_t)$. From (148), we have

$$\omega(x_t) = c_1(x_t) + c_2(x_t) = \frac{\delta^t}{\sum_{\tau=0}^{T} \delta^\tau} \frac{P_1(x_t)W_1 + P_2(x_t)W_2}{\nu(x_t)} \tag{149}$$

where we have used the equilibrium condition, $c_1(x_t) + c_2(x_t) = \omega(x_t)$. We can rearrange this to yield

$$\nu(x_t) = \frac{\delta^t}{\sum_{\tau=0}^{T} \delta^\tau \omega(x_t)} \left(P_1(x_t)W_1 + P_2(x_t)W_2\right). \tag{150}$$

Define aggregate wealth at $t = 0$ by $W = W_1 + W_2$ and relative wealth at $t = 0$ by $w_j = W_j/W$. Define the wealth-weighted probability

$$P_R(x_t) = w_1 P_1(x_t) + w_2 P_2(x_t). \tag{151}$$

Note that $P_R(x_t) \geq 0$ and for fixed t, $\sum_{x_t} P_R(x_t) = 1$. Then,

$$\nu(x_t) = \frac{\delta^t}{\sum_{\tau=0}^{T} \delta^\tau \omega(x_t)} P_R(x_t) W. \tag{152}$$

By assumption, $\nu(x_0) = 1$. Also, $P_j(x_0) = 1$, i.e. no uncertainty at $t = 0$. So, we have

$$c_j(x_0) = \frac{1}{\sum_{\tau=0}^{T} \delta^\tau} W_j. \tag{153}$$

From $c_1(x_0) + c_2(x_0) = \omega(x_0)$ and $W = W_1 + W_2$, and rearranging, we have

$$W = \left(\sum_{\tau=0}^{T} \delta^\tau\right) \omega(x_0) \tag{154}$$

so that

$$\nu(x_t) = \frac{\delta^t P_R(x_t) \omega(x_0)}{\omega(x_t)} = \frac{\delta^t P_R(x_t)}{g(x_t)}, \tag{155}$$

where we recall that $g(x_t) = \omega(x_t)/\omega(x_0)$.

12.3 – Market Efficiency

Let $\Pi(x_t)$ be the *real* pdf associated with x_t. Given that $\nu(x_t) = \delta^t P_R(x_t)/g(x_t)$, in order for state prices to be efficient prices, we need $P_R(x_t) = \Pi(x_t)$.

We can define **market error** either through the P_R/Π (relative error) or $P_R - \Pi$ (absolute error).

For each x_t, define the discounted investor error as

$$\epsilon_j(x_t) = \delta'(t)(P_j(x_t) - \Pi(x_t)), \tag{156}$$

where $\delta'(x_t) := \delta^t / \sum_{t=0}^{T} \delta^\tau$.

Let $w_j = W_j / \sum_{k=1}^{J} W_k$. We have

$$\mathrm{cov}(w_j, \epsilon_j(x_t)) = \sum_{k=1}^{J} (w_k - 1/J) \left(\epsilon_j(x_t) - \bar{\epsilon}(x_t)\right)/J, \tag{157}$$

where J is the number of investors and $\bar{\epsilon}(x_t) = \sum_{k=1}^{J} \epsilon_k(x_t)/J$.

Theorem. Prices are efficient if and only if

$$\mathrm{cov}(W_j, \epsilon_j(x_t)) + \bar{\epsilon}(x_t) \left(\sum_j W_j/J\right) = 0. \tag{158}$$

The condition can also be expressed as

$$\sum_j w_j \epsilon_j(x_t) = 0. \tag{159}$$

Proof. See Shefrin [5].

When are markets efficient?

Natural efficiency (nonsystemmatic errors, $\bar{\epsilon} = 0$):

1. All $\epsilon_j = 0$ (for all t and x_t). Error-free case.
2. Some investors make errors, but $\bar{\epsilon} = 0$ for all x_t and the wealth is distributed uniformly among all investors. Then, there is positive variation in ϵ, but zero deviation in W_j and the covariance is zero.

Knife-edge efficiency:

1. $\bar{\epsilon} \neq 0$ but the covariance offsets $\bar{\epsilon}$ so that the condition holds. A slight change in wealth or errors triggers a violation of efficiency.
2. If the wealth shifts to one investor, error-wealth covariance is zero and efficiency is regained.

12.4 – Investors with CRRA utility

Consider an investor with CRRA utility, $u(x) = x^{\gamma-1}/(1-\gamma)$, i.e. with a coefficient of relative risk aversion γ.

Let $D_j(x_t) := \delta_j^t P_j(x_t)$ (discounted probability).

Recall that every investor chooses a consumption plan $\boldsymbol{c}_j$ by maximising the sum of probability-weighted discounted utilities:

$$\begin{aligned} \max \quad & \mathbb{E}(u_j) = \sum_{t=1}^{T} D_{x_t j}(x_t) u_j(c_j(x_t)), \\ \text{s.t.} \quad & \sum_{t,x_t} c_j(x_t) \leq W_j. \end{aligned} \tag{160}$$

We can solve this to give

$$c_j(x_t) = \frac{(D_j(x_t)/\nu(x_t))^{1/\gamma} W_j}{\sum_\tau \nu(x_\tau) (D_j(x_\tau)/\nu(x_\tau))^{1/\gamma}}, \tag{161}$$

which specifies the fraction of wealth to be consumed in each date-event pair x_t.

It is useful to consider this relative to initial consumption $c_j(x_0)$:

$$\frac{c_j(x_t)}{c_j(x_0)} = (D_j(x_t)/\nu(x_t))^{1/\gamma} \tag{162}$$

Representative Investor. When all investors have log-utility, equilibrium prices are set by a representative investor. This generalises to the case when alll investors have CRRA utility with the same coefficient of relative risk aversion γ.

Suppose there is one investor, R. Then,

$$g(x_t) = (D_R(x_t)/\nu(x_t))^{1/\gamma}. \tag{163}$$

Given that $g(x_t) = \sum_{j=1}^J c_j(x_t) / \sum_{j=1}^J c_j(x_0)$, we have

$$g(x_t) = \sum_{j=1}^{J} \frac{c_j(x_0)}{\sum_{k=1}^J c_k(x_0)} (D_j(x_t)/v(x_t))^{1/\gamma} \tag{164}$$

Aggregate consumption growth is a convex combination of the different investors' rates of consumption growth.

Comparing equations (163) and (164), in the case when there are exactly J investors, prices ν are set as if there is a representative investor R for which

$$D_R(x_t) = \left(\sum_{j=1}^{J} \frac{c_j(x_0)}{\sum_{k=1}^J c_k(x_0)} D_j(x_t)^{1/\gamma} \right)^\gamma. \tag{165}$$

This equation is central to both the probability density functions P_R and the discount factor δ_R. For fixed t, we have $\delta^t = \sum_{x_t} D_R(x_t)$ and $P_R(x_t) = D_R(x_t)/\delta_R(t)$.

12.5 – Representative Investor in a Heterogeneous CRRA model.

The above results generalise to an economy where investors have CRRA utility with different risk coefficients as well as beliefs and time discount factors. What does this representative investor look like?

Rpresentative Investor Characterization Theorem. Let ν be an equilibrium state price vector.

1) ν satisfies

$$\nu(x_t) = \delta^t_{R,t} P_R(x_t) g(x_t)^{-\gamma_R(x_t)}, \tag{166}$$

where γ_R, δ_R, P_R are as follows.

$$\frac{1}{\gamma_R(x_t)} = \sum_j \theta_j(x_t)/\gamma_j, \tag{167}$$

where

$$\theta_j(x_t) = \frac{c_{j,\pi}(x_0)}{\omega(x_t)} \left[\delta^t_j \Pi(x_t)/\nu_\pi(x_t)\right]^{1/\gamma_j}. \tag{168}$$

θ_j is investor j's share of consumption in date-event pair x_t for the case $P_j = \Pi$; ν_π and $c_{j,\pi}$ are the state price vector and equilibrium value of c_j when this holds.

$$\delta^t_{R,t} = \sum_{x_t} \nu(x_t)\zeta(x_t)^{\gamma_R(x_t)}, \tag{169}$$

where this summation is over all investors and x_t events at time t, and

$$P_R(x_t) = \frac{\nu(x_t)\zeta(x_t)^{\gamma_R(x_t)}}{\delta^t_{R,t}} \tag{170}$$

with

$$\zeta(x_t) = \sum_{j=1}^{J} \frac{c_j(x_0)\left[D_j(x_t)/\nu(x_t)\right]^{1/\gamma_j}}{\sum_{k=1}^{J} c_k(x_0)}. \tag{171}$$

2) The representative investor is not unique. Any two representative investors, R_1 and R_2 are related through

$$\frac{\delta^t_{R_1}}{\delta^t_{R_2}} \frac{P_{R_1}}{P_{R_2}} = g^{\gamma_{R_1} - \gamma_{R_2}}. \tag{172}$$

Proof. See Shefrin [5].

Moment-like structure: discounted probabilities are based on weighted averages of individual investors' terms raised to powers.

Is the nonuniqueness of the representative investor a problem?

12.6 – Sentiment

Sentiment relates to aggregate investor errors, which then impact prices. A lot of the Behavioural Finance literature only focus on the first two moments of investors' probability density functions, but higher moments are also relevant:

Second moment. Errors in risk perception.

Third Moment. Investors are optimistic but concerned about a downturn.

Fourth Moment. Investors attach high probabilities to extreme events.

It makes sense then to define sentiment in terms of entire probability distributions. Let

$$\Phi(x_t) = \frac{P_R(x_t)}{\Pi(x_t)} \frac{\delta_R(t)}{\delta_{R,\Pi}(t)}. \tag{173}$$

$\delta_{R,\Pi}$ is the value of δ_R in the last theorem when all investors hold correct beliefs. The process of aggregation can cause deviations in δ_R, hence the inclusion of the discount factor ratio. When all investors hold correct beliefs, $\Phi = 1$.

Define the sentiment function

$$\Lambda = \ln(\Phi) = \ln(P_R/\Pi) + \ln(\delta_R/\delta_{R,\Pi}). \tag{174}$$

Notes.

- Encapsulates aggregate investor errors in the market.
- Time-varying; reflects evolution of error-wealth covariance.
- Magnitude of Λ varies as wealth shifts between investors who have taken sides in trades.
- Weight attached to investors' beliefs is an increasing function of trading success.

12.7 – The Behavioural Stochastic Discount Factor

In a discrete-time, discrete-state model, the SDF (pricing kernel), M_t, is the state price per unit probability: $\mathrm{M}_t = \nu/\Pi$.

Consider a risk-free security F with return $r_{t+1}(F) = i_1$, priced at 1 unit at t. Then, $\mathbb{E}_t(M_{t+1}i_1) = 1$, so that $\mathbb{E}_t(M_{t+1}) = 1/i_1$.

Consider also a risky security Z, priced at 1 unit at t, i.e. $\mathbb{E}_t(M_{t+1}r_{t+1}(Z)) = 1$. Then, $\mathbb{E}_t(M_{t+1}(r_{t+1}(Z) - r_{t+1}(F))) = 0$.

Using $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y) + \mathrm{cov}(X, Y)$, it can be shown that

$$\mathbb{E}_t(r_{t+1}(Z) - r_{t+1}(F)) = -i_1 \mathrm{cov}(M_{t+1}, r_{t+1}(Z)). \tag{175}$$

The risk premium is determined by covariance with a stochastic discount factor. The existence of an SDF is equivalent to the law of one price. The existence of a strictly positive SDF is equivalent to the absence of arbitrage.

Sentiment and the SDF.

The state price vector, ν, gives the present value at time $t = 0$ of a contingent claim to one x_t-dollar. We have

$$M_1 := M(x_1) = M(x_1|x_0) = \delta_R(P_R(x_1)/\Pi(x_1))g(x_1)^{-\gamma_R}. \tag{176}$$

Define the log-SDF, $m := \ln(M)$. Then,

$$m = \Lambda - \gamma_R \ln(g) + \ln(\delta_{R,\Pi}). \tag{177}$$

The log-SDF is the sum of a sentiment process and a fundamental process which is based on aggregate consumption growth.

Prices are efficient only when the sentiment Λ is uniformly zero (i.e. zero at every node). In this case, fundamentals alone drive prices. When sentiment is non-zero, we see deviations in prices away from those associated with the traditional SDF.

References

[1] J. Cvitanić and F. Zapatero. *Introduction to the Economics and Mathematics of Finacial Markets.* MIT Press, 2004.

[2] F. J. Fabozzi, S. M. Focardi, and P. N.i Kolm. *Financial Modeling of the Equity Market: From CAPM to Cointegration.* John Wiley & Sons, 2006.

[3] R. C. Grinold and R. N. Kahn. *Active Portfolio Management: A quantitative approach for producing superior returns and selecting superior money managers.* McGraw-Hill.

[4] E. E. Qian, R. H. Hua, and E. H. Sorensen. *Quantitative Equity Portfolio Manangement: Modern Techniques and Applications.* Chapman & Hall/CRC, 2007.

[5] H. Shefrin. *A Behavioural Approach to Asset Pricing.* Elsevier, 2005.