Risk Analytics
Autumn 2019

Colin Rowat
c.rowat@bham.ac.uk

2019-12-02
(preliminary until end of term)

Contents

Introduction

- financial assets are held out of an interest in their future payoffs
- future payoffs are risky
- hence, asset management necessarily involves risk management
- the job market highly values risk management abilities

	01.10.15	19.09.16	25.09.17	21.09.18	04.10.19
risk jobs	665	512	566	554	539
total jobs	1,257	924	1,062	1,090	1,116

Table: 'Quantitative finance' searches on http://www.indeed.com

Risk managers are in great demand as a result of the troubles the banks have found themselves in (Richard Lipstein, Boyden Global Executive Search)

Banks are failing to implement bonus plans that rein in the types of risks blamed for contributing to the financial crisis, [the Basel Committee on Banking Supervision] said. Bank Bonus Plans Fail to Curb Financial Risks, Regulators Say (Bloomberg, 15 Oct 2010)

Our (ahem!) promise to you

Any graduate of G53 Risk Analytics should be able to beat an index fund over the course of a market cycle.

- a 2015 analysis in the FT of the difficulties of delivering consistent outperformance
- another cautionary 2015 FT article
- 2017: Warren Buffett's $1mn bet that the S&P500 would outperform a portfolio of hedge funds selected by fund manager Ted Seides
- classic statements in favour of passive investment: Malkiel (2016) (too long, but up to date) v Malkiel (2003) (dated, but concise)
  - slides from a talk by John Cochrane on hedge funds
  - Cochrane (2013) addressed a consequence of these results: should we just conclude that "people are stupid"?
- Bookstaber (2007) provides a compelling account of smart, talented, hard-working people losing vast quantities of money
- are G53 techniques the opposite of the 'value' techniques used by investors like Ben Graham and Warren Buffett (Schroeder, 2009), or an augmentation of them?

A little Learning is a dang'rous Thing

If "active" and "passive" management styles are defined in sensible ways, it must be the case that

1. *before costs, the return on the average actively managed dollar will equal the return on the average passively managed dollar and*
2. *after costs, the return on the average actively managed dollar will be less than the return on the average passively managed dollar*

These assertions will hold for any time period. Moreover, they depend only on the laws of addition, subtraction, multiplication and division. Nothing else is required. (Sharpe, 1991)

- market return: "weighted average of the returns on the active and passive segments of the market"
- average passive return: same as the market return
- hence average active return is too
- active management faces higher costs

Market risk

The best known type of risk is probably market risk: the risk of a change in the value of a financial position due to changes in the value of the underlying components on which that portfolio depends, such as stock and bond prices, exchange rates, commodity prices, etc. (McNeil, Frey and Embrechts, 2015, p.5)

- Resti and Sironi (2007, Part II)

Credit risk

The next important category is credit risk: the risk of not receiving promised repayments on outstanding investments such as loans and bonds, because of the "default" of the borrower. (McNeil, Frey and Embrechts, 2015, p.5)

- reflects, more generally, "unexpected changes in the creditworthiness of [a bank's] counterparties" (Resti and Sironi, 2007, Part III); they feel it to be the most important class
- see also Gordy (2000) or Bielecki and Rutkowski (2002)
- Freddie Mac could incur "billions of dollars of losses" for US taxpayers by focussing on mortgages failing in the first two years; while, historically, this is when most have failed, more recent 'teaser' mortgages have low interest rates for three to five years, which then rise sharply (Freddie Mac Loan Deal Defective, Report Says, NYT, 27 Sept 2011)

Operational risk

A further risk category is operational risk: the risk of losses resulting from inadequate or failed internal processes, people and systems, or from external events. (McNeil, Frey and Embrechts, 2015, p.5)

- Resti and Sironi (2007, Part IV)

The loss resulted from unauthorised speculative trading in various S&P 500, Dax and Eurostoxx index futures over the past three months. The positions had been offset in our systems with fictitious, forward-settling, cash ETF positions, allegedly executed by the trader. These fictitious trades concealed the fact that the index futures trades violated UBS's risk limits. (UBS, 18 September 2011)

Model risk

Model risk management has become a board-level process. Now the chief risk officer has to go to the board and not only talk about market risk, credit risk and operational risk, he also has to talk about model risk. It is a huge organisational change. (New York-based model risk manager; Sherif (2016))

- arises from a misspecified model
- e.g. using Black-Scholes when model assumptions don't hold (e.g. normally distributed returns)
- "always present to some degree" (McNeil, Frey and Embrechts, 2015, p.5)
- q.v. Rebonato (2007), *Supervisory Guidance on model risk management* (2011), Morini (2011)

Liquidity risk

When we talk about liquidity risk we are generally referring to price or market liquidity risk, which can be broadly defined as the risk stemming from the lack of marketability of an investment that cannot be bought or sold quickly enough to prevent or minimize a loss. (McNeil, Frey and Embrechts, 2015, p.5)

In banking, there is also the concept of funding liquidity risk, which refers to the easer with which institutions can raise funding to make payments and meet withdrawals as they arise. (McNeil, Frey and Embrechts, 2015, p.5)

The Meucci mantra

1. for each security, identify the iid stochastic terms (§3.1)
2. estimate the distribution of the market invariants (§4)
3. project the invariants to the investment horizon (§3.2)
4. dimension reduce to make the problem more tractable (§3.4)
5. evaluate the portfolio performance at the investment horizon (§5)
   - what is your objective function?
6. pick the portfolio that optimises your objective function (§6)
7. account for estimation risk
   1. replace point parameter estimates with Bayesian distributions (§7)
   2. re-evaluate the portfolio distributions in this light (§8)
   3. robustly re-optimise (§9)

Observation shows that some statistical frequencies are, within narrower or wider limits, stable. But stable frequencies are not very common, and cannot be assumed lightly. Keynes (1921, p.381)

Notational conventions

- τ, investment horizon
- T, time at which the allocation decision is made
  - thus, $T + \tau$ is when the investments are to be evaluated
- $\boldsymbol{P}_t$, the vector of prices at time t
- X_t, a random variable that will realise at time t
- x_t, a realisation of the random variable
- $i_T \equiv \{x_1, \ldots, x_T\}$, a dataset of observed realisations

Random variables

- a number whose realisation is, as yet, unknown
- its distribution may be known
  - a space of events, $\mathfrak{E}$
  - a probability distribution, $\mathbb{P}$
- a function from the space of events to the real line
- thus, $x = X(\mathfrak{e})$ for some event $\mathfrak{e}$ in $\mathfrak{E}$
- the probability of an event giving rise to a realised $x \in [\underline{x}, \bar{x}]$:

$$\mathbb{P}\{X \in [\underline{x}, \bar{x}]\} \equiv \mathbb{P}\{\mathfrak{e} \in \mathfrak{E} \text{ s.t. } X(\mathfrak{e}) \in [\underline{x}, \bar{x}]\}.$$

  reads: the probability that random variable X takes on a value in $[\underline{x}, \bar{x}]$ is the probability of the set of events yielding a realised value of the random variable ...
- going forward, typically suppress dependence on $\mathfrak{e}$, refer just to X

◂ naïve statement

Probability density function (PDF), f_X

1. the probability that the rv X takes on a value within a given interval

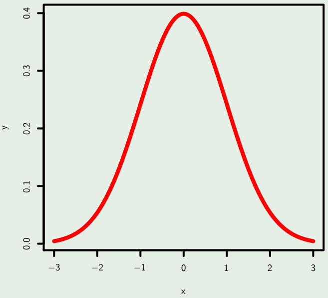
$$\mathbb{P}\left\{X \in [\underline{x}, \bar{x}]\right\} \equiv \int_{\underline{x}}^{\bar{x}} f_X(x)\, dx$$

2. why do the following also hold?

$$f_X(x) \geq 0$$
$$\int_{-\infty}^{\infty} f_X(x)\, dx = 1$$

Example

$$f_X(x) = \frac{1}{\sqrt{\pi}} e^{-x^2}$$



Cumulative distribution function (CDF), F_X

1. the probability that the rv X takes on a value less than x

$$F_X(x) \equiv \mathbb{P}\{X \leq x\} = \int_{-\infty}^{x} f_X(u)\, du$$

2. why do the following also hold?

$$F_X(-\infty) = 0$$
$$F_X(\infty) = 1$$
$$F_X \text{non-decreasing}$$

Example

$$F_X(x) = \frac{1}{2}\left[1 + \operatorname{erf}(x)\right]$$

where erf is the error function

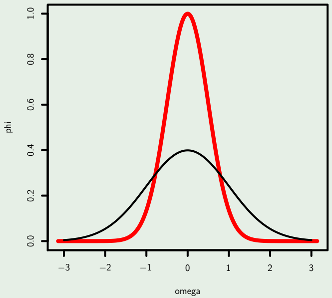
$$\operatorname{erf}(x) \equiv \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$$

Characteristic function (cf), ϕ_X

1. $$\phi_X(\omega) \equiv E\left\{e^{i\omega X}\right\}, \omega \in \mathbb{R}$$
2. its properties are less intuitive (Meucci, 2005, q.v. pp.6-7)
3. particularly useful for handling (weighted) sums of independent rvs

Example

$$\phi_X(\omega) = e^{-\frac{1}{2}\omega^2}$$



Quantile, Q_X

1. the inverse of the CDF

$$Q_X(p) \equiv F_X^{-1}(p)$$

2. the number x such that the probability that X be less than x is p:

$$\mathbb{P}\{X \leq Q_X(p)\} = p$$

Example

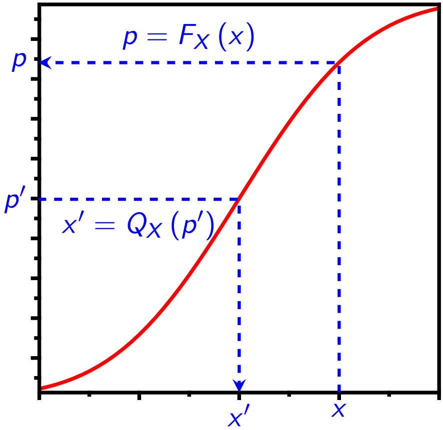
$$Q_X(p) = \mathrm{erf}^{-1}(2p - 1)$$

Example

For the median, $p = \frac{1}{2}$

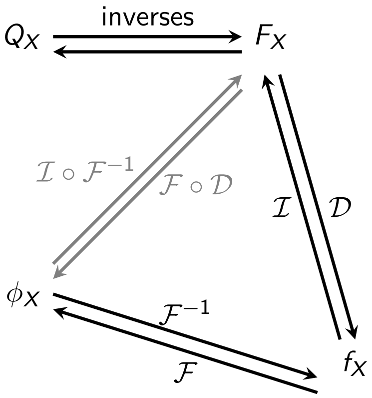
VaR

The quantile and the CDF



- invertibility requires $f_X > 0$
- otherwise, can regularise f_X with $f_{X;\varepsilon}$ Meucci (2005, App. B.4)

Moving between representations of the rv X



- $\mathcal{I}$ is the integration operator
- $\mathcal{D}$ is the derivative operator
- $\mathcal{F}$ is the Fourier transform (FT)
- $\mathcal{F}^{-1}$ is the inverse Fourier transform (IFT)
- all of these are examples of linear operators, $\mathcal{A}[v](\boldsymbol{x})$
  - $\mathcal{A}$, the linear operator
  - v, the function to which it is applied
  - $\boldsymbol{x}$, the function's argument

  q.v. Meucci, Appendix B.3
- (n.b. f_X exists iff F_X is absolutely continuous; ϕ_X always exists)

Lecture 1 exercises

- Meucci exercises
  - pencil-and-paper: 1.1.1; 1.1.2; 1.1.3; 1.1.5; 1.1.6
  - Python: 1.1.4, 1.1.7, 1.1.8
- project
  - pick a six-digit GICS industry (using e.g. Bloomberg, Interactive Brokers' Trader WorkStation) that none of your classmates are using and five US firms within it;
  - enter your pick at

    https://pad.riseup.net/p/rl6GvL7DyTgiHws6fhAR
  - begin to experiment with your Interactive Brokers trading account and the Bloomberg terminals.

Key summary parameters

- full distributions can be expensive to represent
- what summary information helps capture key features?

1. location, $Loc\{X\}$
   - if had one guess as to where X would take its value
   - should satisfy $Loc\{a\} = a$ and affine equivariance

$$Loc\{a + bX\} = a + bLoc\{X\}$$

   to ensure independence of measurement scale/coordinate system

2. dispersion, $Dis\{X\}$
   - how accurate the location guess, above, is
   - affine equivariance is now

$$Dis\{a + bX\} = |b|\, Dis\{X\}$$

   where $|\cdot|$ denotes absolute value

3. z-score normalises a rv, $Z_X \equiv \frac{X - Loc\{X\}}{Dis\{X\}}$: 0 location; 1 dispersion
   - affine equivariance of location & dispersion $\Leftrightarrow (Z_{a+bX})^2 = (Z_X)^2$

Most common location and dispersion measures

	'local'	'semi-local'	'global'
location	mode, $Mod\left\{X\right\}$	median, $Med\left\{X\right\}$	mean / exp'd val, $E\left\{X\right\}$
	$\text{argmax}_{x\in\mathbb{R}}\, f_X\left(x\right)$	$\int_{-\infty}^{Med\left\{X\right\}} f_X\left(x\right)dx = \frac{1}{2}$	$\int_{-\infty}^{\infty} x f_X\left(x\right)dx$
dispersion	modal dispersion	interquantile range	variance
			$\int_{-\infty}^{\infty}\left(x - E\left\{X\right\}\right)^2 f_X\left(x\right)$

- 'global' measures are formed from the whole distribution
- 'semi-local' measures are formed from half (or so) of the distribution
- 'local' measures are driven by individual observations
- generally, we define

$$Dis\left\{X\right\} \equiv \left\|X - Loc\left\{X\right\}\right\|_{X;p}$$

where $\left\|g\right\|_{X;p} \equiv \left(E\left\{\left|g\left(X\right)\right|^p\right\}\right)^{\frac{1}{p}}$ is the norm on the vector space L_X^p

- $p = 1$ is the mean absolute deviation, $MAD\left\{X\right\} \equiv E\left\{\left|X - E\left\{X\right\}\right|\right\}$
- $p = 2$ is the standard deviation, $Sd\left\{X\right\} \equiv \left(E\left\{\left|X - E\left\{X\right\}\right|^2\right\}\right)^{\frac{1}{2}}$

Higher order moments

1. k^{th}-raw moment
$$RM_k^X \equiv E\left\{X^k\right\}$$
is the expectation of the k^{th} power of X
2. k^{th}-central moment is more commonly used
$$CM_k^X \equiv E\left\{(X - E\{X\})^k\right\}$$
de-means the raw moment, making it location-independent
    - skewness, a measure of symmetry, is the normalised 3rd central moment
$$Sk\{X\} \equiv \frac{CM_3^X}{(Sd\{X\})^3}$$
    - kurtosis measures the weights of the distribution's tail relative to its centre
$$Ku\{X\} \equiv \frac{CM_4^X}{(Sd\{X\})^4}$$

Uniform distribution: $X \sim U([a, b])$

- simplest distribution; shall be useful when modelling copulas
- fully described by two parameters, a (lower bound) and b (upper bound)
- any outcome in the $[a, b]$ is equally likely
- closed form representations for $f^U_{a,b}(x), F^U_{a,b}(x), \phi^U_{a,b}(\omega)$ and $Q^U_{a,b}(p)$
- standard uniform distribution is $U([0, 1])$

Normal (Gaussian) distribution: $X \sim N\left(\mu, \sigma^2\right)$

- most widely used, studied distribution
- fully described by two parameters, μ (mean) and σ^2 (variance)
- standard normal distribution when $\mu = 0$ and $\sigma^2 = 1$
- as a stable distribution, the sums of normally distributed rv's are normal
- closed form representations for $f^N_{\mu,\sigma^2}(x)$, $F^N_{\mu,\sigma^2}(x)$, $\phi^N_{\mu,\sigma^2}(\omega)$ and $Q^N_{\mu,\sigma^2}(p)$
- why do we care that $Ku\{X\} = 3$?

Cauchy distribution: $X \sim Ca\left(\mu, \sigma^2\right)$

- 'fat-tailed' distribution: when might this be useful?
- fully described by two parameters, μ and σ^2

$$f^{Ca}_{\mu,\sigma^2}(x) \equiv \frac{1}{\pi\sqrt{\sigma^2}}\left[1 + \frac{(x-\mu)^2}{\sigma^2}\right]^{-1}$$

- what are $E\{X\}, Var\{X\}, Sk\{X\}$ and $Ku\{X\}$?
  - see here for a discussion
- standard Cauchy distribution when $\mu = 0$ and $\sigma^2 = 1$
  - (FYI: if $X, Y \sim NID(0,1)$ then $\frac{X}{Y} \sim Ca(0,1)$)

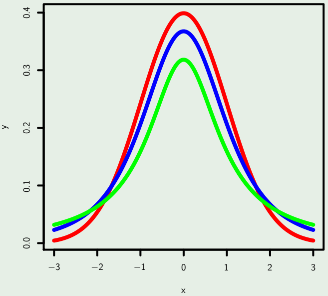
Student t distribution: $X \sim St\left(\nu, \mu, \sigma^2\right)$

- degrees of freedom parameter, ν, determines fatness of tails
- analytical expressions for $f^{St}_{\nu,\mu,\sigma^2}$, $F^{St}_{\nu,\mu,\sigma^2}$ and $\phi^{St}_{\nu,\mu,\sigma^2}$ use the gamma, beta and Bessel functions; none for $Q^{St}_{\nu,\mu,\sigma^2}$
  - limit of analytical expressions quickly reached
- standard Student distribution when $\mu = 0$ and $\sigma^2 = 1$
- when are $E\{X\}, Var\{X\}, Sk\{X\}$ and $Ku\{X\}$ defined?

Example ($\nu = 3$)

$$\nu \to \infty \Rightarrow St\left(\nu, \mu, \sigma^2\right) \to^d N\left(\mu, \sigma^2\right)$$

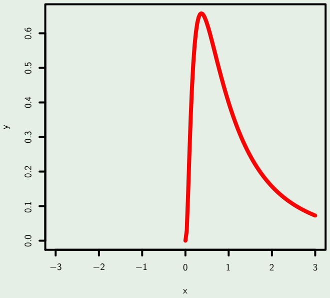
$$\nu \to 1 \Rightarrow St\left(\nu, \mu, \sigma^2\right) \to^d Ca\left(\mu, \sigma^2\right)$$



Log-normal distribution: $X \sim LogN\left(\mu, \sigma^2\right)$

- if $Y \sim N\left(\mu, \sigma^2\right)$ then $X \equiv e^Y \sim LogN\left(\mu, \sigma^2\right)$ (Bailey: should be called 'exp-normal' distribution?)
- now $\phi^{LogN}_{\mu,\sigma^2}$ has no known analytic form
- properties
  - $X > 0$
  - (% changes in X) $\sim N$ ▸ %
  - asymmetric (positively skewed)
- commonly applied to stock prices (Hull (2009, §12.6, §13.1), Stefanica (2011, §4.6))

Example



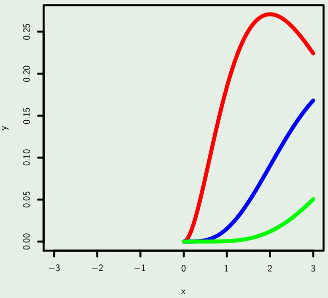
Gamma distribution: $X \sim Ga\left(\nu, \mu, \sigma^2\right)$

let $Y_1, \ldots, Y_\nu \sim IID$ s.t. $Y_t \sim N\left(\mu, \sigma^2\right) \forall t \in \{1, \ldots \nu\}$

- non-central gamma distribution, $X \equiv \sum_{t=1}^{\nu} Y_t^2 \sim Ga\left(\nu, \mu, \sigma^2\right)$
- ν: degrees-of-freedom (shape); μ: non-centrality; σ^2: scale
- Bayesians: each observation is an rv $\Rightarrow$ their variance $\sim Ga$

1. $\mu = 0 \Rightarrow$ central gamma distribution, $X \sim Ga\left(\nu, \sigma^2\right)$ (most common)
2. $\sigma^2 = 1 \Rightarrow$ non-central chi-square distribution
3. $\mu = 0, \sigma^2 = 1 \Rightarrow$ chi-square distribution, $X \sim \chi^2_\nu$

Example $\left(\mu = 0, \sigma^2 = 1\right)$



◂ $X \sim W$

Empirical distribution: $X \sim Em\left(i_T\right)$

- data defines distribution: future occurs with same probability as past

$$f_{i_T}\left(x\right) \equiv \frac{1}{T}\sum_{t=1}^{T}\delta^{\left(x_t\right)}\left(x\right)$$

$$F_{i_T}\left(x\right) \equiv \frac{1}{T}\sum_{t=1}^{T}H^{\left(x_t\right)}\left(x\right)$$

  - $\delta^{\left(x_t\right)}\left(\cdot\right)$ is Dirac's delta function centred at x_t, a generalised function (if wish to treat X as discrete, Kronecker's delta function defines probability mass function)
  - $H^{\left(x_t\right)}\left(\cdot\right)$ is Heaviside's step function, with its step at x_t
- what do these look like? What do regularised versions look like?
- defining $Q_{i_T}\left(p\right)$ obtained by bandwidth techniques of Appendix B: order observations, then count from lowest

◂ $X \sim Em$

Lecture 2 exercises

- Meucci exercises
  - pencil-and-paper: 1.2.5 (not Python), 1.2.6, 1.2.7
  - Python: 1.2.3, 1.2.5 (Python)
- project
  - install and configure Interactive Broker's Python API;
  - can you read historical data via the IB API, or do you receive this error message:

    Requested market data is not subscribed. Historical Market Data Serv error message: No market data permissions for NYSE STK.

  - explore the distribution of returns for your assets

Direct extensions of univariate statistics

- if interested in portfolios (or even arbitrage), must be able to consider how an asset's movements depend on others'
- N-dimensional rv, $\boldsymbol{X} \equiv (X_1, \ldots, X_N)'$, so that $\boldsymbol{x} \in \mathbb{R}^N$
- probability density function

$$\mathbb{P}\{\boldsymbol{X} \in \mathcal{R}\} \equiv \int_{\mathcal{R}} f_{\boldsymbol{X}}(\boldsymbol{x})\, d\boldsymbol{x}, \text{ st } f_{\boldsymbol{X}}(\boldsymbol{x}) \geq \boldsymbol{0}, \int_{\mathbb{R}^N} f_{\boldsymbol{X}}(\boldsymbol{x})\, d\boldsymbol{x} = 1$$

- cumulative or joint distribution function (df, DF, CDF, JDF ...)

$$F_{\boldsymbol{X}}(\boldsymbol{x}) \equiv \mathbb{P}\{\boldsymbol{X} \leq \boldsymbol{x}\} = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_N} f_{\boldsymbol{X}}(u_1, \ldots, u_N)\, du_N \cdots du_1$$

- characteristic function

$$\phi_{\boldsymbol{X}}(\boldsymbol{\omega}) \equiv E\left\{e^{i\boldsymbol{\omega}'\boldsymbol{X}}\right\}, \boldsymbol{\omega} \in \mathbb{R}^N$$

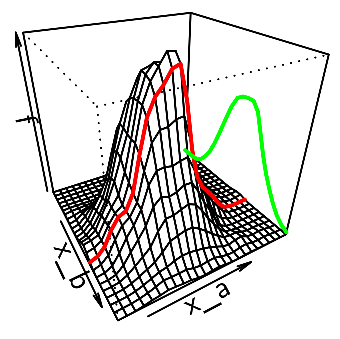
- what about the quantile? (hint: $F_{\boldsymbol{X}} : \mathbb{R}^N \to \mathbb{R}^1$)

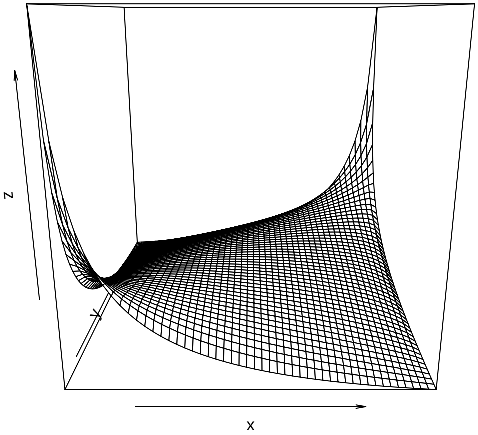
Marginal distribution/density of $\boldsymbol{X}_B$

- partition $\boldsymbol{X}$ into K-dimensional $\boldsymbol{X}_A$ and $(N-K)$-dimensional $\boldsymbol{X}_B$
- distribution of $\boldsymbol{X}_B$ whatever $\boldsymbol{X}_A$'s (technically: integrates out $\boldsymbol{X}_A$)

$$
\begin{aligned}
F_{\boldsymbol{X}_B}(\boldsymbol{x}_B) \equiv& \mathbb{P}\{\boldsymbol{X}_B \leq \boldsymbol{x}_B\} \\
=& \mathbb{P}\{\boldsymbol{X}_A \leq \infty, \boldsymbol{X}_B \leq \boldsymbol{x}_B\} \\
=& F_{\boldsymbol{X}}(\infty, \boldsymbol{x}_B) \\
f_{\boldsymbol{X}_B}(\boldsymbol{x}_B) \equiv& \int_{\mathbb{R}^K} f_{\boldsymbol{X}}(\boldsymbol{x}_A, \boldsymbol{x}_B)\, d\boldsymbol{x}_A \\
\phi_{\boldsymbol{X}_B}(\boldsymbol{\omega}) \equiv& E\left\{e^{i\boldsymbol{\omega}'\boldsymbol{X}_B}\right\} \\
=& E\left\{e^{i\boldsymbol{\psi}'\boldsymbol{X}_A + i\boldsymbol{\omega}'\boldsymbol{X}_B}\right\}|_{\boldsymbol{\psi}=\mathbf{0}} \\
=& \phi_{\boldsymbol{X}}(\mathbf{0}, \boldsymbol{\omega})
\end{aligned}
$$

Example





What, roughly, do the marginals of this pdf look like? copula defined

Conditional distribution/density of $\boldsymbol{X}_A$ given $\boldsymbol{x}_B$

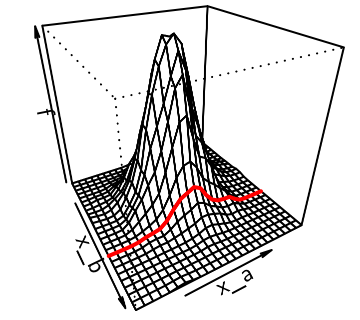
e.g. fix assets B's returns at $\boldsymbol{x}_B$; what is that of assets A?

$$f_{\boldsymbol{X}_A|\boldsymbol{x}_B}(\boldsymbol{x}_A) \equiv \frac{f_{\boldsymbol{X}}(\boldsymbol{x}_A, \boldsymbol{x}_B)}{f_{\boldsymbol{X}_B}(\boldsymbol{x}_B)}$$

- can decompose JDF into product of marginal and conditional
- Bayes' rule for updating beliefs is an immediate consequence

$$f_{\boldsymbol{X}_A|\boldsymbol{x}_B}(\boldsymbol{x}_A) = \frac{f_{\boldsymbol{X}_B|\boldsymbol{x}_A}(\boldsymbol{x}_B)\, f_{\boldsymbol{X}_A}(\boldsymbol{x}_A)}{f_{\boldsymbol{X}_B}(\boldsymbol{x}_B)}$$

Example



Location parameter, $Loc\{\boldsymbol{X}\}$

- desiderata of location extend directly from univariate case
  1. for constant $\boldsymbol{m}$, $Loc\{\boldsymbol{m}\} = \boldsymbol{m}$
  2. for invertible $\boldsymbol{B}$, affine equivariance now

$$Loc\{\boldsymbol{a} + \boldsymbol{BX}\} = \boldsymbol{a} + \boldsymbol{B}Loc\{\boldsymbol{X}\}$$

- expected value
  1. $E\{\boldsymbol{X}\} = (E\{X_1\}, \ldots, E\{X_N\})'$
  2. affine equivariance property holds for any conformable $\boldsymbol{B}$, not just invertible ($Med\{\boldsymbol{X}\}, Mod\{\boldsymbol{X}\}$ require invertible)
  3. relatively easy to calculate when $\phi_{\boldsymbol{X}}$ known, analytical (Meucci, 2005, §T2.10)

Dispersion parameter, $Dis\left\{\boldsymbol{X}\right\}$

- recall: in the univariate case, the z-score normalises a distribution so that it is invariant under affine transformations

$$|Z_{a+bX}| = |Z_X| \equiv \sqrt{\frac{(X - Loc\left\{X\right\})(X - Loc\left\{X\right\})}{Dis\left\{X\right\}^2}}$$

- let $\boldsymbol{\Sigma}$ be a symmetric PD or PSD matrix; then Mahalanobis distance from $\boldsymbol{x}$ to $\boldsymbol{\mu}$, normalised by the metric $\boldsymbol{\Sigma}$, is

$$Ma\left(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}\right) \equiv \sqrt{(\boldsymbol{x} - \boldsymbol{\mu})' \, \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})}$$

- given an ellipsoid centred at $\boldsymbol{\mu}$ whose principal axes' lengths equal the square roots of the eigenvalues of $\boldsymbol{\Sigma}$, all $\boldsymbol{x}$ on its surface have the same Mahalanobis distance from $\boldsymbol{\mu}$ ◂ IID heuristic test 2
- multivariate z-score is then $Ma_{\boldsymbol{X}} \equiv Ma\left(\boldsymbol{X}, Loc\left\{\boldsymbol{X}\right\}, DisSq\left\{\boldsymbol{X}\right\}\right)$
- benchmark (squared) dispersion or scatter parameter: covariance

Correlation

- normalised covariance

$$\rho(X_m, X_n) = Cor\{X_m, X_n\} \equiv \frac{Cov\{X_m, X_n\}}{Sd\{X_m\}\, Sd\{X_n\}} \in [-1, 1]$$

where

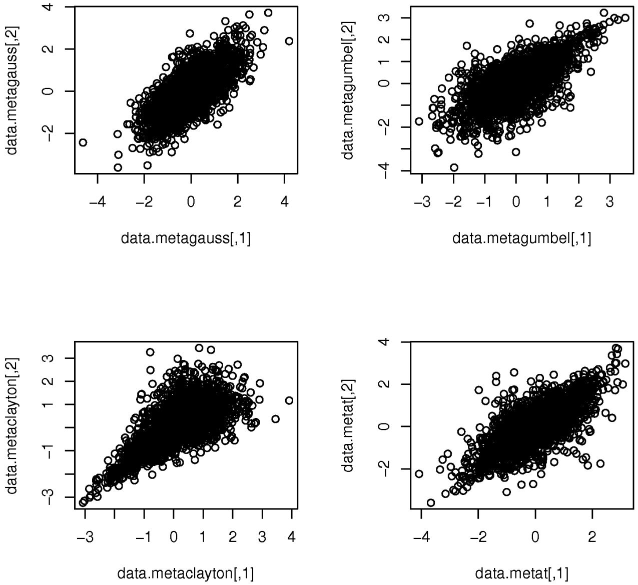
$$Cov\{X_m, X_n\} \equiv E\{(X_m - E\{X_m\})(X_n - E\{X_n\})\}$$

- when is this not defined?
- a measure of linear dependence, invariant under strictly increasing linear transformations

$$\rho(\alpha_m + \beta_m X_m, \alpha_n + \beta_n X_n) = \rho(X_m, X_n)$$

- fallacy (McNeil, Frey and Embrechts, 2015, p.241): given marginal dfs F_1, F_2 and any $\rho \in [-1, 1]$, can always find a JDF F binding them
  - true for elliptical distributions; generally, attainable correlations are a strict subset of $[-1, 1]$ (McNeil, Frey and Embrechts, 2015, Ex. 7.29)
- conventional wisdom: during market stress, all correlations $\Rightarrow 1$

Standard normal marginals, $\rho \approx .7$



fallacy (McNeil, Frey and Embrechts, 2015, p.239): marginal distributions and pairwise correlations of a rv determine its joint distribution

Independence

- information about one variable does not affect distribution of others

$$f_{\boldsymbol{X}_B}(\boldsymbol{x}_B) = f_{\boldsymbol{X}_B|\boldsymbol{x}_A}(\boldsymbol{x}_B)$$

- probability of two independent events: $\mathbb{P}\{\mathfrak{e} \cap \mathfrak{f}\} = \mathbb{P}\{\mathfrak{e}\}\mathbb{P}\{\mathfrak{f}\}$

$$F_{\boldsymbol{X}}(\boldsymbol{x}_A, \boldsymbol{x}_B) = F_{\boldsymbol{X}_A}(\boldsymbol{x}_A) F_{\boldsymbol{X}_B}(\boldsymbol{x}_B)$$

- from definitions of conditional distribution and independence (try it!)

$$f_{\boldsymbol{X}}(\boldsymbol{x}_A, \boldsymbol{x}_B) = f_{\boldsymbol{X}_A}(\boldsymbol{x}_A) f_{\boldsymbol{X}_B}(\boldsymbol{x}_B)$$

- above true if $\boldsymbol{X}_A, \boldsymbol{X}_B$ transformed by arbitrary $g(\cdot)$ and $h(\cdot)$: if $\boldsymbol{x}_A$ doesn't explain $\boldsymbol{X}_B$, transformed versions won't either ◂ linear returns plot
- therefore allows non-linear relations
- independent implies uncorrelated, but not the converse

Example

Given $X^2 + Y^2 = 1$, are the rvs X and Y (un)correlated, (in)dependent?
Hint: if fitting $y_i = mx_i + b + \varepsilon_i$, what are $m, \hat{m}$?

Uniform distribution

- idea is as in univariate case, but domain may be anything
- often elliptical domain, $\mathcal{E}_{\boldsymbol{\mu},\boldsymbol{\Sigma}}$ where $\boldsymbol{\mu}$ is centroid, $\boldsymbol{\Sigma}$ is positive matrix

Example

$$f_{X_1,X_2}(x_1,x_2) = \frac{1}{\pi}\mathbb{I}_{\{x_1^2+x_2^2\le 1\}}(x_1,x_2)$$

where $\mathbb{I}_{\mathcal{S}}$ is the indicator function on the set $\mathcal{S}$

- marginal density: $f_{X_1}(x_1) = \int_{-\sqrt{1-x_1^2}}^{\sqrt{1-x_1^2}} \frac{1}{\pi} dx_2 = \frac{2}{\pi}\sqrt{1-x_1^2}$
- conditional density: $f_{X_1|x_2}(x_1) = \frac{f_{X_1,X_2}(x_1,x_2)}{f_{X_2}(x_2)} = \frac{1}{2\sqrt{1-x_2^2}}$
- are X_1 and X_2 (un)correlated, (in)dependent?

Normal (Gaussian) distribution: $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

- most widely used, studied distribution
- fully described by two parameters, $\boldsymbol{\mu}$ (location) and $\boldsymbol{\Sigma}$ (dispersion)
- standard normal distribution when $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \boldsymbol{I}$ (identity matrix)
- closed form representations for $f^N_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\boldsymbol{x})$, $F^N_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\boldsymbol{x})$, and $\phi^N_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\boldsymbol{\omega})$
- as symmetric and unimodal

$$E\{\boldsymbol{X}\} = Mod\{\boldsymbol{X}\} = Med\{\boldsymbol{X}\} = \boldsymbol{\mu}$$
$$Cov\{\boldsymbol{X}\} = \boldsymbol{\Sigma}$$

- marginal, conditional distributions also normal

Student t distribution: $\boldsymbol{X} \sim St(\nu, \boldsymbol{\mu}, \boldsymbol{\Sigma})$

- again, symmetrically distributed about a peak
- again, three parameters
  - as symmetric and unimodal, $E\{\boldsymbol{X}\} = Mod\{\boldsymbol{X}\} = Med\{\boldsymbol{X}\} = \boldsymbol{\mu}$
  - scatter parameter $\neq$ covariance: $Cov\{\boldsymbol{X}\} = \frac{\nu}{\nu-2}\boldsymbol{\Sigma}$
- standard Student t distribution when $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \boldsymbol{I}$
- Meucci (2005) claims characteristic function depends on whether ν even or odd; Hurst (1995) and Berg and Vignat (2008) do not
- marginal distributions are also t; conditional distributions are not; thus, if $\boldsymbol{X} \sim St$, can't be independent ▸ t dependence

Cauchy distribution: $\boldsymbol{X} \sim Ca\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$

- as in the univariate case, the fat-tailed limit of the Student t-distribution: $Ca\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right) = St\left(1, \boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$
- standard Cauchy distribution when $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \boldsymbol{I}$ (identity matrix)
- same problem with moments as univariate case

Log-distributions

- exponentials of other distributions, applied component-wise
- thus, useful for modelling positive values
- if $\boldsymbol{Y}$ has pdf $f_{\boldsymbol{Y}}$ then $\boldsymbol{X} \equiv e^{\boldsymbol{Y}}$ is log-$\boldsymbol{Y}$ distributed

Example (Log-normal)

Let $\boldsymbol{Y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Then, if $\boldsymbol{X} \equiv e^{\boldsymbol{Y}}$, so that $X_i \equiv e^{Y_i}$ for all $i = 1, \ldots, N$, $\boldsymbol{X} \sim LogN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Wishart distribution: $\boldsymbol{W} \sim W(\nu, \boldsymbol{\Sigma})$

- consider N-dimensionals IID rvs $\boldsymbol{X}_t \sim N(\boldsymbol{0}, \boldsymbol{\Sigma})$ for $t = 1, \ldots, \nu \geq N$
- then Wishart distribution with ν degrees of freedom is the random matrix

$$\boldsymbol{W} \equiv \boldsymbol{X}_1 \boldsymbol{X}_1' + \cdots + \boldsymbol{X}_\nu \boldsymbol{X}_\nu'$$

- as $\boldsymbol{\Sigma}$ is symmetric and PD, so is $\boldsymbol{W}$
- multivariate generalisation of the gamma distribution $X \sim Ga$
- furthermore, given generic $\boldsymbol{a}$,

$$\boldsymbol{W} \sim W(\nu, \boldsymbol{\Sigma}) \Rightarrow \boldsymbol{a}' \boldsymbol{W} \boldsymbol{a} \sim Ga(\nu, \boldsymbol{a}' \boldsymbol{\Sigma} \boldsymbol{a})$$

- as inverse of symmetric, PD matrix is symmetric, PD, inverse Wishart

$$\boldsymbol{Z}^{-1} \sim W(\nu, \Psi^{-1}) \Rightarrow \boldsymbol{Z} \sim IW(\nu, \Psi)$$

- as a random PD matrix, Wishart useful in estimating random $\boldsymbol{\Sigma}$
  - e.g. sample covariance matrix from multivariate normal; Bayesian

priors

Empirical distribution: $\boldsymbol{X} \sim Em\left(i_T\right)$

- direct extension of univariate case ▸ $X \sim Em$

$$f_{i_T}\left(\boldsymbol{x}\right) \equiv \frac{1}{T} \sum_{t=1}^{T} \delta^{\left(\boldsymbol{x}_t\right)}\left(\boldsymbol{x}\right)$$

$$F_{i_T}\left(\boldsymbol{x}\right) \equiv \frac{1}{T} \sum_{t=1}^{T} H^{\left(\boldsymbol{x}_t\right)}\left(\boldsymbol{x}\right)$$

$$\phi_{i_T}\left(\boldsymbol{\omega}\right) \equiv \frac{1}{T} \sum_{t=1}^{T} e^{i \boldsymbol{\omega}' \boldsymbol{x}_t}$$

- moments include
  1. sample mean: $\hat{E}_{i_T} \equiv \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{x}_t$
  2. sample covariance: $\hat{Cov}_{i_T} \equiv \frac{1}{T} \sum_{t=1}^{T} \left(\boldsymbol{x}_t - \hat{E}_{i_T}\right)\left(\boldsymbol{x}_t - \hat{E}_{i_T}\right)'$

Elliptical distributions: $\boldsymbol{X} \sim El\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g_N\right)$

- highly symmetrical, analytically tractable, flexible
- $\boldsymbol{X}$ is elliptically distributed with location parameter $\boldsymbol{\mu}$ and scatter matrix $\boldsymbol{\Sigma}$ if its iso-probability contours form ellipsoids centred at $\boldsymbol{\mu}$ whose principal axes' lengths are proportional to the square roots of $\boldsymbol{\Sigma}$'s eigenvalues
- elliptical pdf must be

$$f_{\boldsymbol{\mu},\boldsymbol{\Sigma}}\left(\boldsymbol{x}\right) = \left|\boldsymbol{\Sigma}\right|^{-\frac{1}{2}} g_N\left(Ma^2\left(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}\right)\right)$$

  where $g_N\left(\cdot\right) \geq 0$ is a generator function rotated to form the distribution.
- examples include: uniform (sometimes), normal, Student t, Cauchy
- affine transformations: for any K-vector $\boldsymbol{a}$, $K \times N$ matrix $\boldsymbol{B}$, and the right generator g_K,

$$\boldsymbol{X} \sim El\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g_N\right) \Rightarrow \boldsymbol{a} + \boldsymbol{BX} \sim El\left(\boldsymbol{a} + \boldsymbol{B\mu}, \boldsymbol{B\Sigma B'}, g_K\right)$$

- correlation captures all dependence structure (copula adds nothing)

Stable distributions

- let $\boldsymbol{X}$, $\boldsymbol{Y}$ and $\boldsymbol{Z}$ be IID rvs; their distribution is stable if a linear combination of them has the same distribution, up to location, scale parameters: for any constants $\alpha, \beta > 0$ there exist constants γ and $\delta > 0$ such that

$$\alpha \boldsymbol{X} + \beta \boldsymbol{Y} \stackrel{d}{=} \gamma + \delta \boldsymbol{Z}$$

- examples: normal, Cauchy (but not lognormal, or generic Student t)
- closed under linear combinations, thus allows easy projection to investment horizons
- stability implies additivity (the sum of two IID rvs belongs to the same family of distributions), but not the reverse

Example

1. stable $\Rightarrow$ additive: $X, Y, Z \sim NID\left(1, \sigma^2\right) \Rightarrow X + Y \stackrel{d}{=} 2 - \sqrt{2} + \sqrt{2} Z$
2. additive $\not\Rightarrow$ stable:

$$\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z} \sim WID\left(\nu, \boldsymbol{\Sigma}\right) \Rightarrow \boldsymbol{X} + \boldsymbol{Y} \sim W\left(2\nu, \boldsymbol{\Sigma}\right) \stackrel{d}{\neq} \gamma + \delta \boldsymbol{Z}$$

Infinitely divisible distributions

- the distribution of rv $\boldsymbol{X}$ is infinitely divisible if it can be expressed as ... the sum of an arbitrary number of IID rvs: for any integer T

$$\boldsymbol{X} \stackrel{d}{=} \boldsymbol{Y}_1 + \cdots + \boldsymbol{Y}_T$$

  for some IID rvs $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_T$
- examples include: all elliptical, gamma, LogN (but not Wishart for $N > 1$)
- shall see: assists in projection to arbitrary investment horizons (e.g. any T)

Lecture 3 exercises

- Meucci exercises
  - pencil-and-paper: 1.3.1, 1.3.4, 2.1.3
  - Python: 1.2.8, 1.3.2, 1.3.3, 2.1.1, 2.1.2
- project
  - can you fit standard distributions to your assets' compound returns (univariate and multivariate)?

Introduction

the copula is a standardized version of the purely joint features of a multivariate distribution, which is obtained by filtering out all the purely one-dimensional features, namely the marginal distribution of each entry X_n. (Meucci, 2005, p.40)

- McNeil, Frey and Embrechts (2015, Ch 7) goes into more detail than (Meucci, 2005, Ch 2) on copulas
  - more material about the book is available at www.qrmtutorial.org
- see Embrechts (2009) for thoughts on the "copula craze", from one of its pioneers, and a "must-read" for context
- the classic text is Nelsen (2006); it contains worked examples and set questions, and has the space to properly develop the basic concepts
- a 2009 wired.com article blamed the Gaussian copula formula for "killing" Wall Street

Copulas defined

Definition

An N-dimensional copula, $\boldsymbol{U}$, is defined on $[0,1]^N$; its JDF, $F_{\boldsymbol{U}}$, has standard uniform marginal distributions.

copula example

- Embrechts (2009, p.640) notes that other standardisations than the copula's to the unit hypercube may sometimes be more useful

Sklar's theorem

Theorem (Sklar, 1959)

Let $F_{\boldsymbol{X}}$ be a JDF with marginals, $F_{X_1}, \ldots, F_{X_N}$. Then there exists a copula, $\boldsymbol{U}$, with JDF $F_{\boldsymbol{U}} : [0,1]^N \to [0,1]$ such that, for all $x_1, \ldots, x_N \in \mathbb{R}$,

$$F_{\boldsymbol{X}}(\boldsymbol{x}) = F_{\boldsymbol{U}}\left(F_{X_1}(x_1), \ldots, F_{X_N}(x_N)\right). \quad (1)$$

If the marginals are continuous, $F_{\boldsymbol{U}}$ is unique.
Conversely, if $\boldsymbol{U}$ is a copula and $F_{X_1}, \ldots, F_{X_N}$ are univariate CDFs, then $F_{\boldsymbol{X}}$, defined in equation 1 is a JDF with marginals $F_{X_1}, \ldots, F_{X_N}$.

Useful to decompose rv into marginals and copula:

1. may have more confidence in marginals than JDF
   - e.g. multivariate t with differing tail-thickness parameters
   - can modify joint distributions of extreme values
2. can run shock experiments: idiosyncratic via marginals, common via copula

Meucci (2005, (2.30)) relates $f_{\boldsymbol{X}}$ to $f_{\boldsymbol{U}}$: sometimes more useful

Probability and quantile transformations

If want to stochastically simulate Z, but X is easier to generate, and can calculate/approximate Q_Z:

Theorem (Proposition 7.2 McNeil, Frey and Embrechts (2015); Meucci 2.25 - 2.27)

Let F_X be a CDF and let Q_X denote its inverse. Then

1. *if X has a continuous univariate CDF, F_X, then $F_X(X) \sim U([0,1])$*
   proof
2. *if $U \equiv F_X(X) \stackrel{d}{=} F_Z(Z) \sim U([0,1])$, then $Z \stackrel{d}{=} Q_Z(U)$*

- the new rv, U is the grade of X
- now have 3rd representation for copulas: $\boldsymbol{U}$, the copula of a multivariate rv, $\boldsymbol{X}$, is the joint distribution of its grades

$$(U_1, \ldots, U_N)' \equiv (F_{X_1}(X_1), \ldots, F_{X_N}(X_N))'$$

Independence copula

- independence of rvs $\Leftrightarrow$ JDF is the product of their univariate CDFs
- applying Sklar's theorem to independent rvs, $X_1, \ldots, X_N$

$$F_{\boldsymbol{X}}(\boldsymbol{x}) = \prod_{n=1}^{N} F_{X_n}(x_n) = F_{\boldsymbol{U}}\left(F_{X_1}(x_1), \ldots, F_{X_N}(x_N)\right)$$

- thus, substituting $F_{X_n}(x_n) = u_n$, provides the independence copula

$$\Pi(\boldsymbol{u}) \equiv F_{\boldsymbol{U}}(u_1, \ldots, u_N) = \prod_{n=1}^{N} u_n$$

  which is uniformly distributed on the unit hyper-cube, with a horizontal pdf, $\pi(\boldsymbol{u}) = 1$
- Schweizer-Wolf measures of dependence (indexed by p in L_p-norm): distance between a copula and the independence copula

Strictly increasing transformations of the marginals

- recall: correlation only invariant under linear transformations

Theorem (Proposition 7.7 McNeil, Frey and Embrechts (2015))

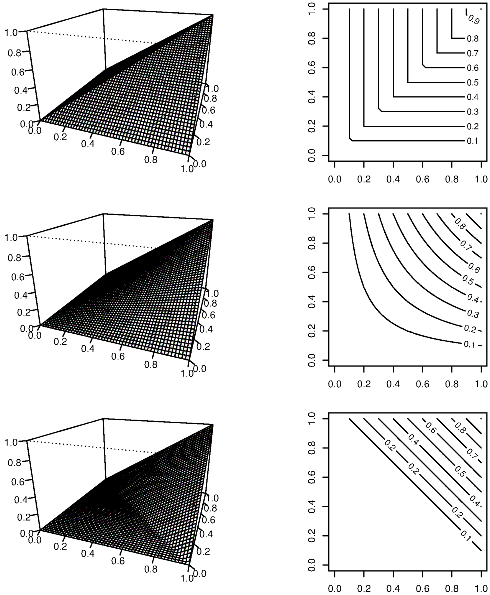
Let $(X_1, \ldots, X_N)$ be a rv with continuous marginals and copula $\boldsymbol{U}$, and let $g_1, \ldots, g_N$ be strictly increasing functions. Then $(g_1(X_1), \ldots, g_N(X_N))$ also has copula $\boldsymbol{U}$.

- a special case of this is the co-monotonicity copula
  - let the rvs $X_1, \ldots, X_N$ have continuous dfs that are perfectly positively dependent, so that $X_n = g_n(X_1)$ almost surely for all $n \in \{2, \ldots, N\}$ for strictly increasing $g_n(\cdot)$
  - co-monotonicity copula is then

$$M(\boldsymbol{u}) \equiv \min\{u_1, \ldots, u_N\}$$

where the JDF of the rv $(U, \ldots, U)$ is s.t. $U \sim U([0,1])$ (McNeil, Frey and Embrechts, 2015, p.226)

Fréchet-Hoeffding bounds



- co-monotonicity copula, M, is Fréchet-Hoeffding upper bound
- Fréchet-Hoeffding lower bound, W, isn't copula for $N > 2$:

$$W(\boldsymbol{u}) \equiv \max\left\{1 - N + \sum_{n=1}^{N} u_n, 0\right\}$$

- any copula's CDF fits between these

$$W(\boldsymbol{u}) \leq F_{\boldsymbol{U}}(\boldsymbol{u}) \leq M(\boldsymbol{u})$$

- which copula is 2nd figure?
- R code: Härdle and Okhrin (2010)

A call option

Example

Consider two stock prices, the rvs $\boldsymbol{X} = (X_1, X_2)$, and a European call option on the first with strike price K. The payoff on this option is therefore also a rv, $C_1 \equiv \max\{X_1 - K, 0\}$.
Thus, C_1 and X_1 are co-monotonic; their copula is M, the co-monotonicity copula. Further, (X_1, X_2) and (C_1, X_2) are also co-monotonic; the copula of (X_1, X_2) is the same as that of (C_1, X_2).
What technical detail is the above missing? How is this overcome?

◂ co-monotonic additivity

Conceptual overview

Meucci (2005) identifies the following steps for building the link between historical performance and future distributions

1. detecting the invariants
   - what market variables can be modelled as IID rvs?
   - Meucci (2017): risk drivers are time-homogenous variables driving P&L; invariants are their IID shocks
2. determining the distribution of the invariants
   - how frequently do these change (q.v. Bauer and Braun (2010))?
3. projecting the invariants into the future
4. mapping the invariants into the market prices

As the dimension of 'most' randomness may be much less than that of the portfolio space, dimension reduction techniques will enhance tractability

Univariate stylised facts

Given an asset price P_t, let its compound return at time t for horizon τ be

$$C_{t,\tau} \equiv \ln \frac{P_t}{P_{t-\tau}}$$

Then, following McNeil, Frey and Embrechts (2015, §3.1):

1. series of compound returns are not IID, but show little serial correlation across different lags
   - if not IID, then prices don't follow random walk
   - if neither IID nor normal, Black-Scholes-Merton pricing is in trouble
2. volatility clustering: series of $|C_{t,\tau}|$ or $C_{t,\tau}^2$ show profound serial correlation
3. conditional (on any history) expected returns are close to zero
4. volatility appears to vary over time
5. extreme returns appear in clusters
6. returns series are leptokurtic (heavy-tailed)

- as horizon increases, returns more IID, less heavy-tailed

Multivariate stylised facts

Given a vector of asset prices $\boldsymbol{P}_t$, let its compound return at time t for horizon τ be defined component-wise as

$$\boldsymbol{C}_{t,\tau} \equiv \ln \frac{\boldsymbol{P}_t}{\boldsymbol{P}_{t-\tau}}$$

Following McNeil, Frey and Embrechts (2015, §3.2)

1. $\boldsymbol{C}_{t,\tau}$ series show little evidence of (serial) cross-correlation, except for contemporaneous returns
2. $|\boldsymbol{C}_{t,\tau}|$ series show profound evidence of (serial) cross-correlation
3. correlations between contemporaneous returns vary over time
4. extreme returns in one series often coincide with extreme returns in several other series

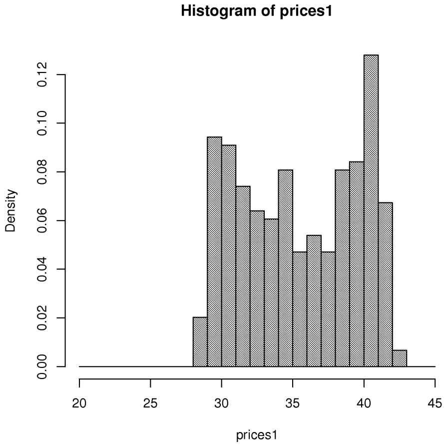
Market invariants

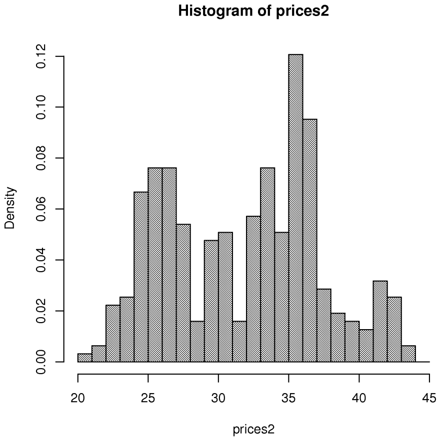
- market invariants/risk drivers, X_t
  1. takes on realised values x_t at time t
  2. behave like random walks
- they are time homogeneous if the IID distribution does not depend on a reference date, $\tilde{t}$
- risk drivers like this make it 'easy' to forecast
- how test for IID (Campbell, Lo and MacKinlay, 1997, Chapter 2)?
  - in particular, how posit the right H_1?
  - tests against particular H_1's often missed non-linear deterministic relationships
  - e.g. logistic map, $x_{t+1} = rx_t(1 - x_t)$ and tent map,
    $$x_{t+1} = \left\{ \begin{array}{ll} \mu x_t & \text{if } x_t < \frac{1}{2} \\ \mu(1 - x_t) & \text{otherwise} \end{array} \right\}$$
  - BDS(L) test (Brock et al., 1996) designed to capture this, but fails in the presence of real noise; not often used due to strong theoretical priors on H_1
- we therefore present two heuristic tests (q.v. Meucci, 2009, §2)

Heuristic test 1: compare split sample histograms

- by the Glivenko-Cantelli theorem, empirical pdf $\rightarrow$ true pdf as the number of IID observations grows
- split the time series in half and compare the two histograms
- what should the two histograms look like if IID?

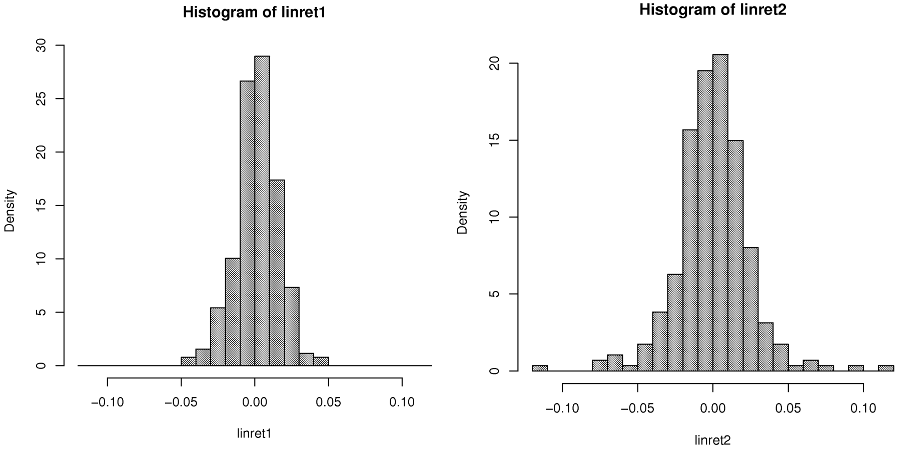
Do stock prices, P_t, pass the histogram test?





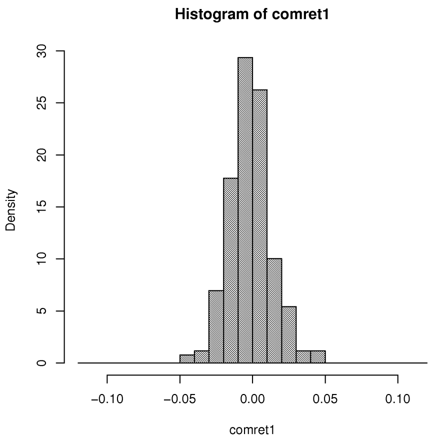
Caveat: apparent similarity changes with bin size choice
All data: THARGES:ID 01/01/07 – 10/09/09

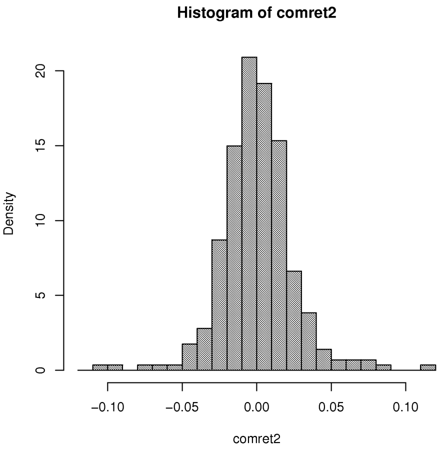
Do linear stock returns, $L_{t,\tau}$, pass the histogram test?



Linear returns are $L_{t,\tau} \equiv \frac{P_t}{P_{t-\tau}} - 1$

Do compound stock returns, $C_{t,\tau}$, pass the histogram test?



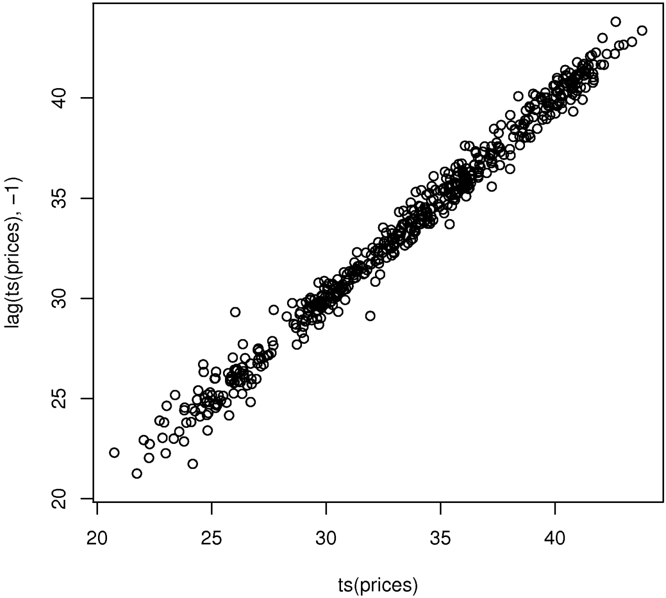


Compound returns are $C_{t,\tau} \equiv \ln \frac{P_t}{P_{t-\tau}}$

Heuristic test 2: plot x_t v $x_{t-\tilde{\tau}}$

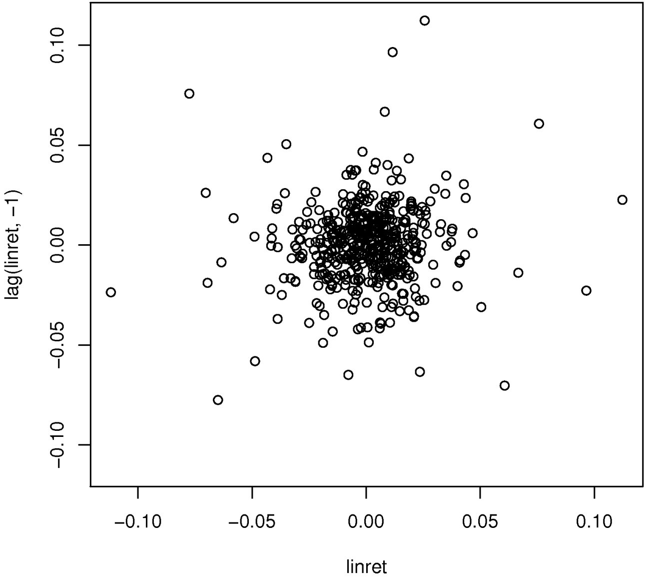
- plot x_t v $x_{t-\tilde{\tau}}$, where $\tilde{\tau}$ is the estimation interval
- what should the plot look like if IID?
  - symmetric about the diagonal: if IID, doesn't matter if plot x_t v $x_{t-\tilde{\tau}}$ or $x_{t-\tilde{\tau}}$ v x_t
  - circular: mean-variance ellipsoid with location (μ, μ), dispersion same in each direction, aligned with coordinate axes as covariance zero (due to independence) (Meucci, 2005, p.55) hint

Do stock prices, P_t, pass the lagged plot test?



What does this tell us about stock prices?

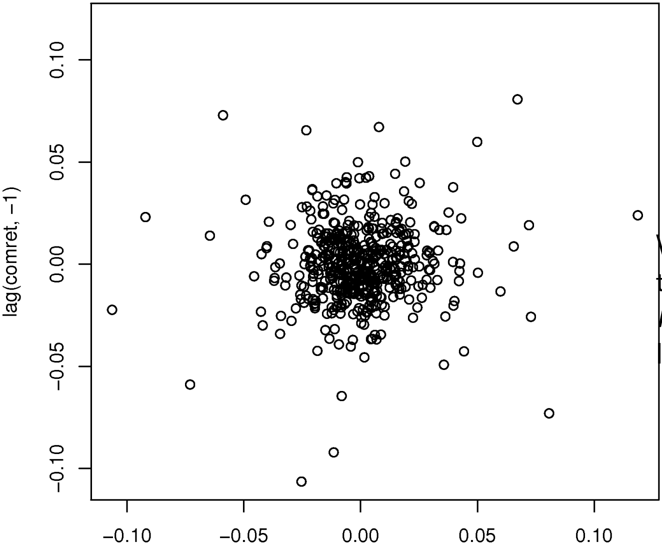
Do linear stock returns, $L_{t,\tau}$, pass the lagged plot test?



What do we expect compound returns to look like, as a result?

▸ independence

Do compound stock returns, $C_{t,\tau}$, pass the lagged plot test?



What do we expect total returns, $H_{t,\tau} \equiv \frac{P_t}{P_{t-\tau}}$ to look like?

Risk drivers for equities, commodities and exchange rates

- THARGES equity fund: do linear, compound, total returns pass the heuristic tests?
- prefer to use compound returns as
  1. shall see that can more easily project distributions to investment horizon
  2. greater symmetry facilitates modelling by elliptical distributions

◂ Δ YTM

- individual equities, commodities, exchange rates have similar properties: no time horizons
- key assumptions
  1. equities: either no dividends, or dividends ploughed back in
  2. generally, non-overlapping - see W_t in Meucci's online exercise 3.2.1 (Oct 2009) as a counter-example
- accept compound returns as IID as expositional device (recall stylised facts); see Meucci (2009) for more discussion

Lecture 4 exercises

- Nelsen (2006, Exercise 2.12) Let X and Y be rvs with JDF

$$H(x,y) = \left(1 + e^{-x} + e^{-y}\right)^{-1}$$

for all $x, y \in \bar{\mathbb{R}}$, the extended reals.

1. show that X and Y have standard (univariate) logistic distributions

$$F(x) = \left(1 + e^{-x}\right)^{-1} \text{ and } G(y) = \left(1 + e^{-y}\right)^{-1}.$$

2. show that the copula of X and Y is $C(u,v) = \frac{uv}{u+v-uv}$.

- Meucci exercises
  - pencil-and-paper: 3.2.1
  - Python: 2.2.1, 2.2.3, 2.2.4, 2.2.6, 3.1.3
- project
  - do your assets' compound returns appear invariant, or do they display GARCH properties?
  - fit an Archimedean copula to the assets' univariate returns

Fixed income: zero-coupon bonds

- make no termly payments
- as simplest form of bond, form basis for analysis of bonds
- fixed income as certain [?] payout at face or redemption value
  - (see Brigo, Morini and Pallavicini (2013) for richer risk modelling)
- bond price then $Z_t^{(E)}$, where $t \leq E$ is date, and E is maturity date
- normalise $Z_E^{(E)} = 1$
- are bond prices invariants?
  1. stock prices weren't
  2. time homogeneity violated
- are returns (total, simple, compound) invariants?

Fixed income: a time homogeneous framework

- construct a synthetic series of bond prices with the same time to maturity, v:
  1. $Z_t^{(E)}$ (e.g. Nov 2019 price of a bond that matures in Feb 2024)
  2. $Z_{t-\tilde{\tau}}^{(E-\tilde{\tau})}$ (e.g. Nov 2018 price of a bond that matures in Feb 2023)
  3. $Z_{t-2\tilde{\tau}}^{(E-2\tilde{\tau})}$ (e.g. Nov 2017 price of a bond that matures in Feb 2022)
  4. $\vdots$
- target duration funds: an established fixed income strategy (Langetieg, Leibowitz and Kogelman, 1990)
- can now define pseudo-returns, or rolling (total) returns to maturity

$$R_{t,\tilde{\tau}}^{(v)} \equiv \frac{Z_t^{(t+v)}}{Z_{t-\tilde{\tau}}^{(t-\tilde{\tau}+v)}}$$

  where $\tilde{\tau}$ is the estimation interval (e.g. a year)
- candidates for passing the two heuristic tests (Meucci, 2005, Figure 3.5)

Fixed income: yield to maturity

- what is the most convenient fixed income invariant to work with?
- define $Y_t^{(v)} \equiv -\frac{1}{v} \ln Z_t^{(t+v)}$ and manipulate to obtain a compound return:

$$vY_t^{(v)} = -\ln Z_t^{(t+v)} = \ln 1 - \ln Z_t^{(t+v)} = \ln \frac{1}{Z_t^{(t+v)}} = \ln \frac{Z_{t+v}^{(t+v)}}{Z_t^{(t+v)}}$$

- $Y_t^{(v)}$ is yield to maturity v; yield curve graphs $Y_t^{(v)}$ as a function of v
- if $\tilde{\tau}$ is a year (standard), then YTM is like an annualised yield
- changes in yield to maturity can be expressed in terms of rolling returns to maturity,

$$X_{t,\tilde{\tau}}^{(v)} \equiv Y_t^{(v)} - Y_{t-\tilde{\tau}}^{(v)} = -\frac{1}{v} \ln \frac{Z_t^{(t+v)}}{Z_{t-\tilde{\tau}}^{(t-\tilde{\tau}+v)}} = -\frac{1}{v} \ln R_{t,\tilde{\tau}}^{(v)}$$

usually pass the heuristics, have similarly desirable properties to compound returns for equities compound returns

Derivatives

- derived from underlying raw securities (e.g. stocks, zero-coupon bonds, . . .)
  - or see here for Senator Trent Lott's views, via Webster's dictionary
- vanilla European options are the most liquid derivatives (why?)
  - the right, but not the obligation, to buy or sell . . .
  - on expiry date E . . .
  - an underlying security trading at price U_t at time t . . .
  - for strike price K

Example (European call option)

The price of a European call option at time $t \leq E$ is often expressed as

$$C_t^{(K,E)} \equiv C^{BSM}\left(E - t, K, U_t, Z_t^{(E)}, \sigma_t^{(K,E)}\right) \text{ s.t. } C_E^{(K,E)} = \max\left\{U_E - K, 0\right\}$$

where $E - t$ is the time remaining, and $\sigma_t^{(K,E)}$ is the volatility of U_t.
The option is in the money when $U_t > K$, at the money when $U_t = K$ and out of the money otherwise.

Derivatives: volatility

- pricing options requires a measure of volatility
  1. historical or realised volatility: determined from historical values of U_t (esp. ARCH models); backward looking but model-free
  2. implied volatility: as the call option's price increases in σ_t, the BSM pricing formula has an inverse, allowing volatility to be implied from option prices; forward looking, but model-dependent; e.g. VXO
  3. model-free volatility expectations: risk-neutral expectation of OTM option prices; forward looking, less model-dependent (but assumes stochastic process doesn't jump); e.g. VIX
- Taylor, Yadav and Zhang (2010) compare the three volatility measures
- at-the-money-forward (ATMF) implied percentage volatility of the underlying: "implied percentage volatility of an option whose strike is equal to the forward price of the underlying at expiry" (Meucci, 2005)
  - $K_t = \frac{U_t}{Z_t^{(E)}} = e^{r_t(E-t)} U_t$, where latter rearranges the no-arbitrage forward price formula (Stefanica, 2011, §1.10), $Z_t^{(E)} e^{r_t(E-t)} = 1$

▸ why ATMF?

Derivatives: a time homogeneous framework

- as with $Z_t^{(E)}$ for fixed income, $\sigma_t^{(K,E)}$ converges as $t \to E$
- consider set of rolling implied percentage volatilities with same time to maturity v, $\sigma_t^{(K_t,t+v)}$
- substitute ATMF definition for K_t into C^{BSM} pricing formula for

$$\sigma_t^{(K_t,E)} = \sqrt{\frac{8}{E-t}} \operatorname{erf}^{-1}\left(\frac{C_t^{(K_t,E)}}{U_t}\right) \approx \sqrt{\frac{2\pi}{v}} \frac{C_t^{(K_t,t+v)}}{U_t}$$

  by first order Taylor expansion of erf^{-1} (q.v. Technical Appendix §3.1)
- normalisation by U_t should remove non-stationarity of $\sigma_t^{(K_t,E)}$
- as $C_t^{(K_t,t+v)}$, U_t not invariant, ratio usually not (Meucci, 2005, p.118), but changes in rolling ATMF implied volatility pass heuristic tests (like differencing $I(1)$ series?)

Projecting invariants to the investment horizon

- have identified invariants, $\boldsymbol{X}_{t,\tilde{\tau}}$ given estimation interval $\tilde{\tau}$
- want to know distribution of $\boldsymbol{X}_{T+\tau,\tau}$, rv at investment horizon, τ
- our preferred invariants are specified in terms of differences
  1. compounds returns for equities, commodities, FX

$$\boldsymbol{X}_{T+\tau,\tau} = \ln P_{T+\tau} - \ln P_T$$

  2. changes in YTM for fixed income

$$\boldsymbol{X}_{T+\tau,\tau} = Y_{T+\tau} - Y_T$$

  3. changes in implied volatility for derivatives

$$\boldsymbol{X}_{T+\tau,\tau} = \boldsymbol{\sigma}_{T+\tau} - \boldsymbol{\sigma}_T$$

all of which are additive, so that they satisfy

$$\boldsymbol{X}_{T+\tau,\tau} = \boldsymbol{X}_{T+\tau,\tilde{\tau}} + \boldsymbol{X}_{T+\tau-\tilde{\tau},\tilde{\tau}} + \cdots + \boldsymbol{X}_{T+\tilde{\tau},\tilde{\tau}}$$

Distributions at the investment horizon

- for expositional simplicity, assume that $\tau = k\tilde{\tau}$, where $k \in \mathbb{Z}_{++}$
  - no problem if not as long as distribution is infinitely divisible (why?)
- as all of the invariants in

$$\boldsymbol{X}_{T+\tau,\tau} = \boldsymbol{X}_{T+\tau,\tilde{\tau}} + \boldsymbol{X}_{T+\tau-\tilde{\tau},\tilde{\tau}} + \cdots + \boldsymbol{X}_{T+\tilde{\tau},\tilde{\tau}}$$

are IID, the projection formula is

$$\phi_{\boldsymbol{X}_{T+\tau,\tau}} = \left(\phi_{\boldsymbol{X}_{t,\tilde{\tau}}}\right)^{\frac{\tau}{\tilde{\tau}}}$$

proof

- can translate back and forth between cf and pdf with Fourier and inverse Fourier transforms

$$\phi_{\boldsymbol{X}} = \mathcal{F}\left[f_{\boldsymbol{X}}\right] \text{ and } f_{\boldsymbol{X}} = \mathcal{F}^{-1}\left[\phi_{\boldsymbol{X}}\right]$$

- by contrast, linear return projections yield

$$\boldsymbol{L}_{T+\tau,\tau} = diag\left(\boldsymbol{1} + \boldsymbol{L}_{T+\tau,\tilde{\tau}}\right) \times \cdots \times diag\left(\boldsymbol{1} + \boldsymbol{L}_{T+\tilde{\tau},\tilde{\tau}}\right) - \boldsymbol{1}$$

where the diagonal entries in the $N \times N$ *diag* matrix are those in its vector-valued argument; its off-diagonal entries are zero

Joint normal distributions

Example

Let the weekly compound returns on a stock and the weekly yield changes for three-year bonds be normally distributed. Thus, the invariants are

$$\mathbf{X}_{t,\tilde{\tau}} = \begin{pmatrix} C_{t,\tilde{\tau}} \\ X_{t,\tilde{\tau}}^{(v)} \end{pmatrix} \equiv \begin{pmatrix} \ln P_t - \ln P_{t-\tilde{\tau}} \\ Y_t^{(v)} - Y_{t-\tilde{\tau}}^{(v)} \end{pmatrix}.$$

Bind these marginals so that their joint distribution is also normal, $\mathbf{X}_{t,\tilde{\tau}} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. By joint normality, the cf is $\phi_{\mathbf{X}_{t,\tilde{\tau}}}(\boldsymbol{\omega}) = e^{i\boldsymbol{\omega}'\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{\omega}'\boldsymbol{\Sigma}\boldsymbol{\omega}}$.
From the previous slide, $\mathbf{X}_{T+\tau,\tau}$ has cf $\phi_{\mathbf{X}_{T+\tau,\tau}}(\boldsymbol{\omega}) = e^{i\boldsymbol{\omega}'\frac{\tau}{\tilde{\tau}}\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{\omega}'\frac{\tau}{\tilde{\tau}}\boldsymbol{\Sigma}\boldsymbol{\omega}}$.
Thus,

$$\mathbf{X}_{T+\tau,\tau} \sim N\left(\frac{\tau}{\tilde{\tau}}\boldsymbol{\mu}, \frac{\tau}{\tilde{\tau}}\boldsymbol{\Sigma}\right).$$

Properties of the horizon distribution

- the projection formula allows derivation of moments (when they are defined)
  1. expected values sum

$$E\left\{\boldsymbol{X}_{T+\tau,\tau}\right\} = \frac{\tau}{\tilde{\tau}} E\left\{\boldsymbol{X}_{t,\tilde{\tau}}\right\}$$

  2. square-root of time rule of risk propagation

$$Cov\left\{\boldsymbol{X}_{T+\tau,\tau}\right\} = \frac{\tau}{\tilde{\tau}} Cov\left\{\boldsymbol{X}_{t,\tilde{\tau}}\right\} \Leftrightarrow Sd\left\{\boldsymbol{X}_{T+\tau,\tau}\right\} = \sqrt{\frac{\tau}{\tilde{\tau}}} Sd\left\{\boldsymbol{X}_{t,\tilde{\tau}}\right\}$$

  Normalising $\tilde{\tau} = 1$ year: standard deviation of the horizon invariant is the square root of the horizon times the standard deviation of the annualised invariant

  - intuition? Portfolio diversifies itself by receiving IID shocks over time
  - see Danielsson and Zigrand (2006) for warnings about non-robustness

Raw securities: horizon prices

- prices depend on invariants through some pricing function, $\boldsymbol{P}_{T+\tau} = \boldsymbol{g}\left(\boldsymbol{X}_{T+\tau,\tau}\right)$

1. for equities, manipulating the compound returns formula yields

$$\boldsymbol{P}_{T+\tau} = \boldsymbol{P}_T e^{\boldsymbol{X}_{T+\tau,\tau}}$$

2. for zero coupon bounds, manipulating the definitions of $\boldsymbol{R}_{T+\tau,\tau}^{(E-T-\tau)}$ and $\boldsymbol{X}_{T+\tau,\tau}^{(E-T-\tau)}$ yields

$$\boldsymbol{Z}_{T+\tau}^{(E)} = \boldsymbol{Z}_T^{(E-\tau)} e^{-(E-T-\tau)\boldsymbol{X}_{T+\tau,\tau}^{(E-T-\tau)}}$$

n.b. could use $v \equiv E - (T + \tau)$

Raw securities: horizon price distribution

- for both equities and fixed income, $\boldsymbol{P}_{T+\tau} = e^{\boldsymbol{Y}_{T+\tau,\tau}}$, where

$$\boldsymbol{Y}_{T+\tau,\tau} \equiv \gamma + diag\left(\varepsilon\right)\boldsymbol{X}_{T+\tau,\tau}$$

  an affine transformation
- thus, they have a $\log - \boldsymbol{Y}$ distribution
- this can be represented as

$$\phi_{\boldsymbol{Y}_{T+\tau,\tau}}\left(\boldsymbol{\omega}\right) = e^{i\boldsymbol{\omega}'\gamma}\phi_{\boldsymbol{X}_{T+\tau,\tau}}\left(diag\left(\varepsilon\right)\boldsymbol{\omega}\right)$$

- usually impossible to compute closed form for full distribution
- may suffice just to compute first few moments
- e.g. can compute $E\left\{P_n\right\}$ and $Cov\left\{P_m, P_n\right\}$ from cf

Derivatives: horizon prices

- prices are still functions of invariants, $\boldsymbol{P}_{T+\tau} = \boldsymbol{g}\left(\boldsymbol{X}_{T+\tau,\tau}\right)$
- as prices reflect multiple invariants, no longer simple $\log - \boldsymbol{Y}$ structure

Example

Again: price of a European call option at horizon $T + \tau \leq E$ is

$$C_{T+\tau}^{(K,E)} \equiv C^{BSM}\left(E - T - \tau, K, U_{T+\tau}, Z_{T+\tau}^{(E)}, \sigma_{T+\tau}^{(K,E)}\right).$$

The horizon distributions of the three invariants are then

$$\begin{aligned} U_{T+\tau} &= U_T e^{X_1} \\ Z_{T+\tau}^{(E)} &= Z_T^{(E-\tau)} e^{-X_2 v} \\ \sigma_{T+\tau}^{(K,E)} &= \sigma_T^{(K_T,E-\tau)} + X_3 \end{aligned}$$

for $v \equiv E - T - \tau$ and suitably defined K_T and invariants, X_1 to X_3.

Derivatives: approximating horizon prices

- options pricing formula is already complicated, non-linear
- adding in possibly complicated horizon projections almost certainly prevents exact solutions
- but can approximate $P_{T+\tau} = g\left(\boldsymbol{X}_{T+\tau,\tau}\right)$ with Taylor expansion

$$P_{T+\tau} \approx g\left(\boldsymbol{m}\right) + \left(\boldsymbol{X} - \boldsymbol{m}\right) \nabla g\left(\boldsymbol{m}\right) + \frac{1}{2} \left(\boldsymbol{X} - \boldsymbol{m}\right)' H\left(g\left(\boldsymbol{m}\right)\right) \left(\boldsymbol{X} - \boldsymbol{m}\right)$$

  where $\nabla g\left(\boldsymbol{m}\right)$ is gradient, $H\left(g\left(\boldsymbol{m}\right)\right)$ Hessian and $\boldsymbol{m}$ some significant value of the invariants $\boldsymbol{X}_{T+\tau,\tau}$
- this approximation produces the Greeks

Example (BetOnMarkets)

BetOnMarkets has to price custom options in less than 15 seconds. Monte Carlo is far too slow; even Black-Scholes may be. They use Vanna-Volga.

Lecture 5 exercises

- Meucci exercises
  - pencil-and-paper: 5.3
  - Python: 3.2.2, 3.2.3, 5.1 (modify code to display one-period and horizon distributions; contrast to Meucci (2005) equations 3.95, 3.96), 5.5.1, 5.5.2, 5.6
- project
  - produce horizon price distributions for your assets (ideally using fully multivariate techniques) by means of one of the techniques mentioned in Daníelsson (2015) *and* a technique in scikit-learn.
  - use the IB API to execute trades algorithmically.

Why dimension reduction?

1. actual dimension of the market is less than the number of securities

Example

Consider a stock whose price is U_t and a European call option on it with strike K and expiry date $T+\tau$. Their horizon prices are

$$\boldsymbol{P}_{T+\tau} = \begin{pmatrix} U_{T+\tau} \\ \max\{U_{T+\tau} - K, 0\} \end{pmatrix}.$$

These are perfectly positively dependent.

2. randomness in the market can be well approximated with fewer than N dimensions (that of the market invariants, $\boldsymbol{X}$)
    - this is the possibility considered in what follows
    - can considerably reduce computational complexity

Common factors

- would like to express N-vector $\boldsymbol{X}_{t,\tilde{\tau}}$ in terms of
  1. a K-vector of common factors, $\boldsymbol{F}_{t,\tilde{\tau}}$;
     1. explicit factors are measurable market invariants
     2. hidden factors are synthetic invariants extracted from the market invariants
  2. an N-vector of residual perturbations, $\boldsymbol{U}_{t,\tilde{\tau}}$

  as follows

$$\boldsymbol{X}_{t,\tilde{\tau}} = \boldsymbol{h}\left(\boldsymbol{F}_{t,\tilde{\tau}}\right) + \boldsymbol{U}_{t,\tilde{\tau}}$$

- for tractability, usually use linear factor model (first order Taylor approximation),

$$\boldsymbol{X}_{t,\tilde{\tau}} = \boldsymbol{B}\boldsymbol{F}_{t,\tilde{\tau}} + \boldsymbol{U}_{t,\tilde{\tau}}$$

  with an $N \times K$ factor loading matrix, $\boldsymbol{B}$

Common factors: desiderata

1. substantial dimension reduction, $K \ll N$
2. independence of $\boldsymbol{F}_{t,\tilde{\tau}}$ and $\boldsymbol{U}_{t,\tilde{\tau}}$ (why?)
   - hard to attain, so often relax to $Cor\left\{\boldsymbol{F}_{t,\tilde{\tau}}, \boldsymbol{U}_{t,\tilde{\tau}}\right\} = \mathbf{0}_{K\times N}$
3. goodness of fit
   - want recovered invariants to be close, $\tilde{\boldsymbol{X}} \equiv \boldsymbol{h}(\boldsymbol{F}) \approx \boldsymbol{X}$
   - use generalised R^2

$$R^2\left\{\boldsymbol{X}, \tilde{\boldsymbol{X}}\right\} \equiv 1 - \frac{E\left\{\left(\boldsymbol{X} - \tilde{\boldsymbol{X}}\right)'\left(\boldsymbol{X} - \tilde{\boldsymbol{X}}\right)\right\}}{tr\left\{Cov\left\{\boldsymbol{X}\right\}\right\}}$$

where the trace of $\boldsymbol{Y}$, $tr\left\{\boldsymbol{Y}\right\}$, is the sum of its diagonal entries

1. what is in the numerator?
2. what is in the denominator?
3. how does this differ from the usual coefficient of determination, R^2?

Explicit factors

- suppose that theory provides a list of explicit market variables as factors, $\boldsymbol{F}$
- how does one determine the loadings matrix, $\boldsymbol{B}$?
- with linear factor model, $\boldsymbol{X} = \boldsymbol{B}\boldsymbol{F} + \boldsymbol{U}$, pick $\boldsymbol{B}$ to maximise generalised R^2

$$\boldsymbol{B}_r \equiv \underset{\boldsymbol{B}}{\operatorname{argmax}}\, R^2 \left\{\boldsymbol{X}, \boldsymbol{B}\boldsymbol{F}\right\}$$

  where the subscript indicates that these are determined by regression
- this is solved by

$$\boldsymbol{B}_r = E\left\{\boldsymbol{X}\boldsymbol{F}'\right\} E\left\{\boldsymbol{F}\boldsymbol{F}'\right\}^{-1}$$

  - how does this differ from OLS?
- even weak version of second desideratum, $Cor\left\{\boldsymbol{F}, \boldsymbol{U}\right\} = \mathbf{0}_{K\times N}$ not generally satisfied; but:
  1. $E\left\{\boldsymbol{F}\right\} = \mathbf{0} \Rightarrow Cor\left\{\boldsymbol{F}, \boldsymbol{U}\right\} = \mathbf{0}_{K\times N}$
  2. adding constant factor to $\boldsymbol{F} \Rightarrow E\left\{\boldsymbol{U}_r\right\} = \mathbf{0}, Cor\left\{\boldsymbol{F}, \boldsymbol{U}_r\right\} = \mathbf{0}_{K\times N}$
     - cf. including constant term in OLS regression

Explicit factors: picking factors

1. want the set of factors to be as highly correlated as possible with the market invariants
    - maximises explanatory power of the factors
    - if do principal components decomposition on $\boldsymbol{F}$, so that $Cov\{\boldsymbol{F}\} = \boldsymbol{E}\Lambda\boldsymbol{E}'$ and $\boldsymbol{C}_{\boldsymbol{XF}} \equiv Cor\{\boldsymbol{X}, \boldsymbol{E}'\boldsymbol{F}\}$ ($\boldsymbol{E}'\boldsymbol{F}$ are rotated factors) then

$$R^2\left\{\boldsymbol{X}, \tilde{\boldsymbol{X}}_r\right\} = \frac{tr\left(\boldsymbol{C}_{\boldsymbol{XF}}\boldsymbol{C}'_{\boldsymbol{XF}}\right)}{N}$$

2. want the set of factors to be as uncorrelated with each other as possible
    - extreme version of correlation is multicollinearity
    - in this case, adding additional factors doesn't add explanatory power, and leaves regression plane ill conditioned
3. more generally, trade-off between more accuracy and more computational intensivity when adding factors

Example (Capital assets pricing model (CAPM))

The linear returns (invariants) of N stocks are $L_{t,\tilde{\tau}}^{(n)} \equiv \frac{P_t^{(n)}}{P_{t-\tilde{\tau}}^{(n)}} - 1$. If the price of the market index is M_t, the linear return on the market index, $F_{t,\tilde{\tau}}^{M} \equiv \frac{M_t}{M_{t-\tilde{\tau}}} - 1$, is a linear factor. The general regression result (3.127) then reduces, in this special case, to:

$$\tilde{L}_{t,\tilde{\tau}}^{(n)} = E\left\{L_{t,\tilde{\tau}}^{(n)}\right\} + \beta_{\tilde{\tau}}^{(n)}\left(F_{t,\tilde{\tau}}^{M} - E\left\{F_{t,\tilde{\tau}}^{M}\right\}\right).$$

Assuming mean-variance utility and efficient markets, linear returns lie on the security market line

$$E\left\{L_{t,\tilde{\tau}}^{(n)}\right\} = \beta_{\tilde{\tau}}^{(n)} E\left\{F_{t,\tilde{\tau}}^{M}\right\} + \left(1 - \beta_{\tilde{\tau}}^{(n)}\right) R_{t,\tilde{\tau}}^{f}$$

where $R_{t,\tilde{\tau}}^{f}$ are risk-free returns (q.v. Dybvig and Ross, 1985). The CAPM then follows

$$\tilde{L}_{t,\tilde{\tau}}^{(n)} = R_{t,\tilde{\tau}}^{f} + \beta_{\tilde{\tau}}^{(n)}\left(F_{t,\tilde{\tau}}^{M} - R_{t,\tilde{\tau}}^{f}\right).$$

Example (Fama and French (1993) three factor model)

The Fama and French (1993) three factor model reduces the compound returns, $C^{(n)}_{t,\tilde{\tau}}$ of N stocks to three explicit linear factors and a constant:

1. C^M, the compound return to a broad stock index
2. SmB, size (small minus big), the difference between the compound return to a small-cap stock index and a large-cap stock index
3. HmL, value (high minus low), the difference between the compound return to a high book-to-market stock index and a low book-to-market stock index

Hidden factors

- now let factors, $\boldsymbol{F}\left(\boldsymbol{X}_{t,\tilde{\tau}}\right)$ be synthetic invariants extracted from market invariants
- thus, the affine model is

$$\boldsymbol{X}_{t,\tilde{\tau}} = \boldsymbol{q} + \boldsymbol{BF}\left(\boldsymbol{X}_{t,\tilde{\tau}}\right) + \boldsymbol{U}_{t,\tilde{\tau}}$$

- what is the trivial way of maximising generalised R^2?
  - what is the weakness of doing so?
- otherwise, main approach taken is principal component analysis (PCA)
  - Matlab's pca
  - Python's sklearn.decomposition.PCA
  - R's prcomp

Principal component analysis (PCA)

- assume the hidden factors are affine transformations of the $\boldsymbol{X}_{t,\tilde{\tau}}$

$$\boldsymbol{F}_p\left(\boldsymbol{X}_{t,\tilde{\tau}}\right) = \boldsymbol{d}_p + \boldsymbol{A}_p'\boldsymbol{X}_{t,\tilde{\tau}}$$

- given these affine assumptions, the optimally recovered invariants are

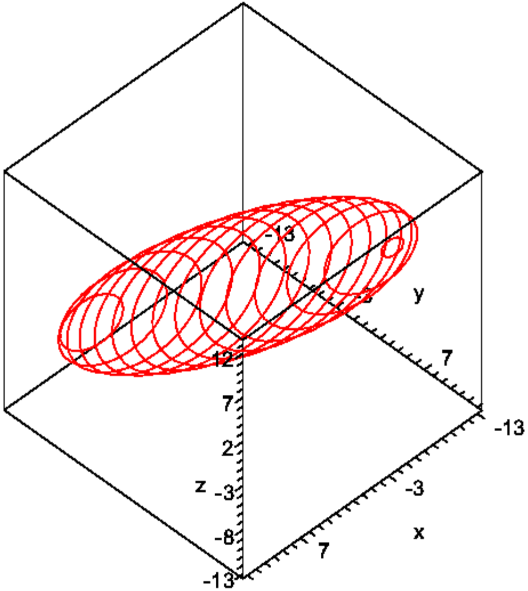
$$\tilde{\boldsymbol{X}}_p = \boldsymbol{m}_p + \boldsymbol{B}_p\boldsymbol{A}_p'\boldsymbol{X}_{t,\tilde{\tau}}$$

  where

$$\left(\boldsymbol{B}_p, \boldsymbol{A}_p, \boldsymbol{m}_p\right) \equiv \underset{\boldsymbol{B},\boldsymbol{A},\boldsymbol{m}}{\operatorname{argmax}} R^2\left\{\boldsymbol{X}, \boldsymbol{m} + \boldsymbol{B}\boldsymbol{A}'\boldsymbol{X}_{t,\tilde{\tau}}\right\}$$

- heuristically
  - want orthogonal factors
  - consider location-dispersion ellipsoid generated by $\boldsymbol{X}_{t,\tilde{\tau}}$
  - asking what its longest principal axes are

Location-dispersion ellipsoid



- consider rv $\boldsymbol{X}$ in $\mathbb{R}^3$
- given location and dispersion parameters, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, can form location-dispersion ellipsoid
- if $K = 1$, which factor would you choose? What would $\tilde{\boldsymbol{X}}_p$ look like?
- what if $K = 2$?
- what if $K = 3$?

Optimal factors in PCA

- optimal factors rotate, translate and collapse the location-dispersion ellipsoid's co-ordinates (q.v. Meucci, 2005, App A.5)
- thus

$$\left(\boldsymbol{B}_p, \boldsymbol{A}_p, \boldsymbol{m}_p\right) = \left(\boldsymbol{E}_K, \boldsymbol{E}_K, \left(\boldsymbol{I}_N - \boldsymbol{E}_K \boldsymbol{E}_K'\right) E\left\{\boldsymbol{X}_{t,\tilde{\tau}}\right\}\right)$$

where

$$\boldsymbol{E}_K \equiv \left(\boldsymbol{e}^{(1)}, \ldots, \boldsymbol{e}^{(K)}\right)$$

with $\boldsymbol{e}^{(k)}$ being the eigenvector of $Cov\left\{\boldsymbol{X}_{t,\tilde{\tau}}\right\}$ corresponding to λ_k, the k^{th} largest eigenvalue.

- $\boldsymbol{m}_p$ translates, and $\boldsymbol{B}_p \boldsymbol{A}_p'$ rotates and collapses, for

$$\begin{aligned}\tilde{\boldsymbol{X}}_p = \boldsymbol{m}_p + \boldsymbol{B}_p \boldsymbol{A}_p' \boldsymbol{X}_{t,\tilde{\tau}} = & \left(\boldsymbol{I}_N - \boldsymbol{E}_K \boldsymbol{E}_K'\right) E\left\{\boldsymbol{X}_{t,\tilde{\tau}}\right\} + \boldsymbol{E}_K \boldsymbol{E}_K' \boldsymbol{X}_{t,\tilde{\tau}} \\ = & E\left\{\boldsymbol{X}_{t,\tilde{\tau}}\right\} + \boldsymbol{E}_K \boldsymbol{E}_K' \left(\boldsymbol{X}_{t,\tilde{\tau}} - E\left\{\boldsymbol{X}_{t,\tilde{\tau}}\right\}\right)\end{aligned}$$

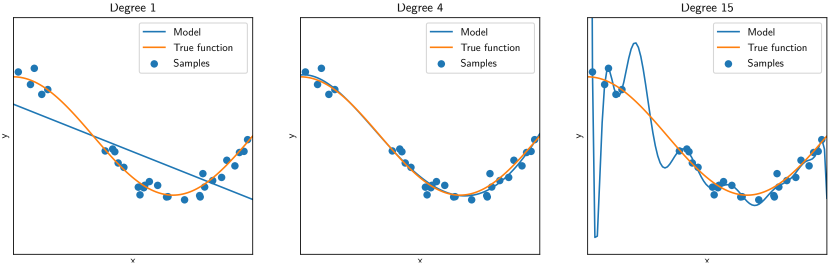
- why are $E\left\{\boldsymbol{U}_p\right\} = \boldsymbol{0}$ and $Cor\left\{\boldsymbol{F}_p, \boldsymbol{U}_p\right\} = \boldsymbol{0}_{K \times N}$? ▸ $E\left\{\boldsymbol{U}_p\right\} = \boldsymbol{0}$
- as $R^2\left\{\boldsymbol{X}_{t,\tilde{\tau}}, \tilde{\boldsymbol{X}}_p\right\} = \frac{\sum_{k=1}^{K} \lambda_k}{\sum_{n=1}^{N} \lambda_n}$, can see effect of each further factor

Explicit factors v PCA?

- as PCA projects onto the most informative K dimensions, it yields a higher R^2 than any K-factor explicit factor model
- however, the synthetic dimensions of PCA are harder to interpret, and therefore perhaps to understand
  - but see Meucci (2005, Fig 3.19, p.158) for a decomposition of the swap market yield curve into level, slope and curvature factors
- are PCA factors less stable out of sample?
- see pp.67- of Smith and Fuertes' Panel Time Series notes for a discussion of how to use and interpret PCA models

The variance-bias tradeoff: what should an estimator do?

- estimator: a function mapping from $\boldsymbol{i}_T$ to a number, the estimate
- bias: distance between estimate (over all data given DGP) and true
- inefficiency: dispersion of estimate over all data given DGP
- error: $Err^2 = Bias^2 + Inef^2$ (Meucci, 2005, p.176)



Over- and underfitting example from Python's scikit-learn

1. underfitting & bias: model forces too many parameters to zero
2. overfitting & inefficiency: "memorises the training set"/"fits to noise"

Example (Estimating the mean in small samples)

Let X correspond to an independent throw of a fair die, so that $E\{X\} = \mu = \frac{7}{2}$. Suppose we do not know μ, but wish to estimate it. The sample mean, $\hat{s} \equiv \frac{1}{T}\sum_{t=1}^{T} x_t$ is unbiased, $E\{\hat{s} - \mu\} = 0$, but may be inefficient as $Var\{\hat{s}\} \equiv E\left\{(\hat{s} - \mu)^2\right\}$ may be high. When $T = 1$, for example,

$$Var\{\hat{s}\} = \frac{1}{6}\left[\frac{25}{4} + \frac{9}{4} + \frac{1}{4}\right] \times 2 = \frac{35}{12} < 3.$$

When $T = 2$,

$$Var\{\hat{s}\} = \cdots = \frac{105}{72} < \frac{35}{12}.$$

Thus, when $T = 2$, $\hat{s}$ has an error of $\sqrt{0 + \frac{105}{72}} = \sqrt{\frac{35}{24}}$. When $T = 1$, the error is $\sqrt{3}$.
The fixed estimator $\tilde{s} \equiv 2$ has a bias of $\frac{3}{2}$, but $Var\{\tilde{s}\} = E\left\{(\tilde{s} - 2)^2\right\} = 0$, for an error of $\frac{3}{2}$ – better than $\hat{s}$ for $T = 1$, but worse when $T \geq 2$.

A rose by any other name

1. classical econometric parlance
   1. non-parametric estimators no identifying restrictions on the empirical distribution ⇒ good estimates require large samples
   2. parametric estimators restrict distributions, so can estimate well on small samples (unless bad parametric restrictions have been made)
   3. shrinkage estimators: for the smallest samples, perform Bayesian averages of estimated values with a constant, reducing error by improving efficiency at the cost of bias ◂ Bayes-Stein ◂ sample-based allocation
2. machine learning parlance (Kolanovic and Krishnamachari, 2017)
   1. supervised ML
      1. regressions (continuous DV): parametric (e.g. ridge, lasso) & non-parametric (e.g. k-NN)
      2. classifications (discrete DV): inc. logit, probit, SVM, decision tree, random forest, HMM
   2. unsupervised ML: clustering (e.g. k-means, ward, birch) & factor analyses (e.g. PCA)
   3. deep/reinforcement learning: 'neurons' provide input to the next layer if linear scores exceeed threshold; most useful for images, text so far

Weighted estimates

- if $i_T \equiv \{\mathbf{x}_1, \ldots, \mathbf{x}_T\}$ truly generated by IID invariants, then can work with empirical distributions, without attention to order of realisation
- if think more recent observations are more informative, may fudge, weighting the empirical distribution by w_t

$$f_{i_T} \equiv \frac{1}{\sum_{t=1}^{T} w_t} \sum_{t=1}^{T} w_t \delta^{\mathbf{x}_t}$$

1. Rolling window treats last W observations equally, discarding all earlier

$$w_t = 1 \text{ if } t > (T - W)$$
$$w_t = 0 \text{ if } t \leq (T - W)$$

2. Exponential smoothing picks a decay factor, $\lambda \in [0, 1]$, and weights by $w_t = (1 - \lambda)^{T-t}$
   - approach used by RiskMetrics, special case of the Kalman filter (Meinhold and Singpurwalla, 1983), (Kolanovic and Krishnamachari, 2017, p.73)

Lecture 6 exercises

- Meucci exercises
  - pencil-and-paper: 6.2.1, 6.4.1, 6.4.2, 6.4.4
  - Python: 6.1, 6.4.3, 6.4.6
- project
  - experiment with dimension reduction: what percentage of the variance in the full five dimensional distribution can you explain using one to four dimensions?

Evaluating allocations

- let $\boldsymbol{\alpha}$ be a portfolio or allocation, an N-vector of asset holdings, and $\boldsymbol{P}_{T+\tau,\tau}$ the investment horizon price distribution

1. investors care about their portfolio's performance at the horizon
    - e.g. absolute wealth, relative wealth, net profits
    - call this their objective, Ψ_α, a random variable
2. need to convert this random variable into a real number
    - call this an index of satisfaction, $\mathcal{S}(\alpha)$ (suppressing dependence on Ψ)

    1. 'economist': certainty-equivalence associated with expected utility
    2. 'practitioners': Value at Risk based on evaluating quantiles of the objective at given confidence levels
    3. 'finance': coherent indices, and spectral indices as a subset, including expected shortfall (aka conditional Value at Risk)

Typical objectives, $\Psi_{\boldsymbol{\alpha}}$

1. absolute wealth

$$\Psi_{\boldsymbol{\alpha}} = W_{T+\tau}(\boldsymbol{\alpha}) = \boldsymbol{\alpha}'\mathbf{P}_{T+\tau}$$

   - e.g. investor concerned about her wealth at retirement

2. relative wealth

$$\Psi_{\boldsymbol{\alpha}} = W_{T+\tau}(\boldsymbol{\alpha}) - \gamma(\boldsymbol{\alpha}) W_{T+\tau}(\boldsymbol{\beta}) = \boldsymbol{\alpha}'\mathbf{K}\mathbf{P}_{T+\tau}$$

   where $\gamma(\boldsymbol{\alpha}) \equiv \frac{w_T(\boldsymbol{\alpha})}{w_T(\boldsymbol{\beta})}$ and $\mathbf{K} \equiv \mathbf{I}_N - \frac{\mathbf{p}_T\boldsymbol{\beta}'}{\boldsymbol{\beta}'\mathbf{p}_T}$

   - e.g. mutual fund manager evaluated annually against a benchmark

3. net profits

$$\Psi_{\boldsymbol{\alpha}} = W_{T+\tau}(\boldsymbol{\alpha}) - w_T(\boldsymbol{\alpha}) = \boldsymbol{\alpha}'(\mathbf{P}_{T+\tau} - \mathbf{p}_T)$$

   - e.g. trader concerned with daily profit and loss (P & L); prospect theory

By non-satiation, more of an objective is preferred.

Benchmarking: relative wealth objectives

- given a relative wealth objective,

$$\Psi_{\boldsymbol{\alpha}} \equiv \boldsymbol{\alpha}' \boldsymbol{P}_{T+\tau} - \gamma \boldsymbol{\beta}' \boldsymbol{P}_{T+\tau}$$

  where $\boldsymbol{\beta}$ is a benchmark portfolio and $\gamma \equiv \frac{\boldsymbol{\alpha}' \boldsymbol{P}_T}{\boldsymbol{\beta}' \boldsymbol{P}_T}$ equalises portfolio costs
- expected overperformance is $EOP(\boldsymbol{\alpha}) \equiv E\{\Psi_{\boldsymbol{\alpha}}\}$
- tracking error is $TE(\boldsymbol{\alpha}) \equiv Sd\{\Psi_{\boldsymbol{\alpha}}\}$
- the information ratio normalises outperformance by tracking error:

$$IR(\boldsymbol{\alpha}) \equiv \frac{EOP(\boldsymbol{\alpha})}{TE(\boldsymbol{\alpha})}$$

- see Baker, Bradley and Wurgler (2011) for dangers of benchmarking in long-only portfolios

Objectives, in general

- in all the objectives considered

$$\Psi_{\boldsymbol{\alpha}} = \boldsymbol{\alpha}' \boldsymbol{M}$$

  where $\boldsymbol{M} \equiv \boldsymbol{a} + \boldsymbol{B}\boldsymbol{P}_{T+\tau}$ is the market vector, the relevant affine transformation of horizon prices, and $\boldsymbol{B}$ is invertible
- (what are $\boldsymbol{a}$ and $\boldsymbol{B}$ for the previous examples?)
- the distribution of $\boldsymbol{M}$ is easily computed from that of $\boldsymbol{P}_{T+\tau}$

$$\begin{aligned}\phi_{\boldsymbol{M}}(\boldsymbol{\omega}) \equiv & E\left\{e^{i\boldsymbol{\omega}'\boldsymbol{M}}\right\} = E\left\{e^{i\boldsymbol{\omega}'(\boldsymbol{a}+\boldsymbol{B}\boldsymbol{P}_{T+\tau})}\right\} = E\left\{e^{i\boldsymbol{\omega}'\boldsymbol{a}} e^{i\boldsymbol{\omega}'\boldsymbol{B}\boldsymbol{P}_{T+\tau}}\right\} \\ = & e^{i\boldsymbol{\omega}'\boldsymbol{a}} \phi_{\boldsymbol{P}_{T+\tau}}\left(\boldsymbol{B}'\boldsymbol{\omega}\right)\end{aligned}$$

- can easily show that $\Psi_{\boldsymbol{\alpha}}$ is
  1. homogeneous of degree one: $\Psi_{\lambda\boldsymbol{\alpha}} = \lambda \Psi_{\boldsymbol{\alpha}}$
  2. additive: $\Psi_{\boldsymbol{\alpha}+\boldsymbol{\beta}} = \Psi_{\boldsymbol{\alpha}} + \Psi_{\boldsymbol{\beta}}$
- as objective is a rv, how compare two portfolios, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$?

Stochastic dominance

1. allocation $\boldsymbol{\alpha}$ strongly dominates allocation $\boldsymbol{\beta}$ iff

$$\forall \mathfrak{e} \in \mathfrak{E}, \Psi_{\boldsymbol{\alpha}} > \Psi_{\boldsymbol{\beta}}$$

   - also known as zero order dominance
   - how often can we expect this?

2. allocation $\boldsymbol{\alpha}$ weakly dominates allocation $\boldsymbol{\beta}$ iff

$$\forall \psi \in (-\infty, \infty), F_{\Psi_{\boldsymbol{\alpha}}}(\psi) \leq F_{\Psi_{\boldsymbol{\beta}}}(\psi) \Leftrightarrow Q_{\Psi_{\boldsymbol{\alpha}}}(p) \geq Q_{\Psi_{\boldsymbol{\beta}}}(p) \forall p \in (0,1)$$

   - aka first order stochastic dominance (FOSD)
   - very rare

3. allocation $\boldsymbol{\alpha}$ second-order stochastically dominates allocation $\boldsymbol{\beta}$ iff

$$\forall \psi \in (-\infty, \infty), \int_{-\infty}^{\psi} (\psi - s) f_{\Psi_{\boldsymbol{\alpha}}}(s)\, ds \geq \int_{-\infty}^{\psi} (\psi - s) f_{\Psi_{\boldsymbol{\beta}}}(s)\, ds$$

   so that lower partial expectation for $\psi_{\boldsymbol{\alpha}}$ exceeds that of $\psi_{\boldsymbol{\beta}}$ for all ψ

- see Levy (1992) for a full treatment of stochastic dominance

Measures of satisfaction

- stochastic dominance does not generate complete order
- wish, therefore, to have one-dimensional index of satisfaction

$$\boldsymbol{\alpha} \mapsto \mathcal{S}(\boldsymbol{\alpha})$$

- risk measure is $-\mathcal{S}$; operationalised via risk capital interpretation
- what features would be desirable for such summary statistics to have?

1. four axioms define coherent measures (Artzner et al., 1999)
2. two more define spectral measures (Acerbi, 2002)

Coherence axiom 1: translation invariance

- if allocation $\boldsymbol{b}$ yields deterministic, $\psi_{\boldsymbol{b}}$ translation invariance requires

$$\mathcal{S}(\boldsymbol{\alpha}+\boldsymbol{b})=\mathcal{S}(\boldsymbol{\alpha})+\mathcal{S}(\boldsymbol{b})=\mathcal{S}(\boldsymbol{\alpha})+\psi_{\boldsymbol{b}}$$

- this, in turn, implies
  1. constancy: $\boldsymbol{\alpha}=\mathbf{0}\Rightarrow\mathcal{S}(\boldsymbol{\alpha})=\psi_{\boldsymbol{\alpha}}=0$, so that $\mathcal{S}(\boldsymbol{b})=\psi_{\boldsymbol{b}}$ (satisfaction of the deterministic outcome is the outcome itself)
  2. if unit of measurement is money, money-equivalence: receiving extra £1mn increases satisfaction (*resp* decreases risk capital) by £1mn

- n.b. additive objectives do not imply additive satisfaction:

$$\Psi_{\boldsymbol{\alpha}+\boldsymbol{\beta}}=\Psi_{\boldsymbol{\alpha}}+\Psi_{\boldsymbol{\beta}}\nRightarrow\mathcal{S}(\boldsymbol{\alpha}+\boldsymbol{\beta})=\mathcal{S}(\boldsymbol{\alpha})+\mathcal{S}(\boldsymbol{\beta})$$

◂ certainty-equivalence | ◂ quantile | ◂ coherent indices | ◂ expected shortfall

Money-equivalence v scale-invariance

Example

1. expected value: $\mathcal{S}(\boldsymbol{\alpha}) = E\{\Psi_{\boldsymbol{\alpha}}\}$
2. Sharpe ratio: $SR(\boldsymbol{\alpha}) = \frac{E\{\Psi_{\boldsymbol{\alpha}}\}}{Sd\{\Psi_{\boldsymbol{\alpha}}\}}$
   When have we seen a Sharpe ratio previously?

- which of the above are money-equivalent?
- by contrast, dimensionless scale-invariance (homogeneity of degree zero)

$$\mathcal{S}(\lambda\boldsymbol{\alpha}) = \mathcal{S}(\boldsymbol{\alpha})\,\forall\lambda > 0$$

  normalises size of portfolio away
- which of the above are scale-invariant?

certainty-equivalence | quantile | expected shortfall

Coherence axiom 2: super-additivity

- an index of satisfaction is super-additive if two portfolios yield a higher index of satisfaction than the indices of the portfolios individually

$$\mathcal{S}(\boldsymbol{\alpha}+\boldsymbol{\beta}) \geq \mathcal{S}(\boldsymbol{\alpha})+\mathcal{S}(\boldsymbol{\beta})$$

- this is desirable as the summed portfolio is at least as diversified as the individual portfolios
- super-additive satisfaction measure implies what sort of risk measure?

◂ certainty-equivalence ◂ quantile ◂ coherent indices ◂ expected shortfall

Example (Expected value)

$$\mathcal{S}(\boldsymbol{\alpha}+\boldsymbol{\beta}) \equiv E\{\Psi_{\boldsymbol{\alpha}+\boldsymbol{\beta}}\} = E\{\Psi_{\boldsymbol{\alpha}}\} + E\{\Psi_{\boldsymbol{\beta}}\} = \mathcal{S}(\boldsymbol{\alpha})+\mathcal{S}(\boldsymbol{\beta})$$

Coherence axiom 3: positive homogeneity

- we know rescaling an allocation rescales the objective identically

$$\Psi_{\lambda\boldsymbol{\alpha}} = \lambda\Psi_{\boldsymbol{\alpha}} \forall\lambda \geq 0$$

- if an index of satisfaction rescales similarly, it is homogeneous with degree one or positive homogenous

$$\mathcal{S}\left(\lambda\boldsymbol{\alpha}\right) = \lambda\mathcal{S}\left(\boldsymbol{\alpha}\right) \forall\lambda \geq 0$$

- Euler's homogeneous function theorem allows satisfaction to be decomposed into hotspots, contributions from each security

$$\mathcal{S}\left(\boldsymbol{\alpha}\right) = \sum_{n=1}^{N} \alpha_n \frac{\partial\mathcal{S}\left(\boldsymbol{\alpha}\right)}{\partial\alpha_n}$$

Positive homogeneity + super-additivity ⇒ concavity

- an index of satisfaction is concave iff

$$\mathcal{S}\left(\lambda\boldsymbol{\alpha}+(1-\lambda)\boldsymbol{\beta}\right)\geq\lambda\mathcal{S}\left(\boldsymbol{\alpha}\right)+(1-\lambda)\mathcal{S}\left(\boldsymbol{\beta}\right)\forall\lambda\in[0,1]$$

- relevance: diversification via convex combinations of two portfolios (e.g. budget constrained) increase satisfaction
- positive homogeneity and super-additivity imply concavity (*resp.* convexity for risk measures)

$$\begin{aligned}\mathcal{S}\left(\lambda\boldsymbol{\alpha}+(1-\lambda)\boldsymbol{\beta}\right)\geq&\mathcal{S}\left(\lambda\boldsymbol{\alpha}\right)+\mathcal{S}\left((1-\lambda)\boldsymbol{\beta}\right)\\=&\lambda\mathcal{S}\left(\boldsymbol{\alpha}\right)+(1-\lambda)\mathcal{S}\left(\boldsymbol{\beta}\right)\end{aligned}$$

by super-additivity, positive homogeneity respectively

◂ certainty-equivalence ◂ quantile ◂ expected shortfall

Coherence axiom 4: monotonicity

- by non-satiation, a satisfaction index satisfies monotonicity iff

$$\Psi_{\alpha} \geq \Psi_{\beta} \forall \mathfrak{e} \in \mathfrak{E} \Rightarrow \mathcal{S}(\alpha) \geq \mathcal{S}(\beta)$$

- thus, monotonicity requires consistency with strong dominance
  - α strongly dominates $\beta \Rightarrow \mathcal{S}(\alpha) \geq \mathcal{S}(\beta)$
- again, seems a sensible requirement

Counterexample: 2006 Swiss Solvency Test (SST)

a framework for determining "the solvency capital required for an insurance company . . . There are situations where the company is allowed to give away a profitable non-risky part of its asset-liability portfolio while reducing its target capital." (Filipovi and Vogelpoth, 2008)
Reduces Ψ_{α} to $\Psi_{\beta} \leq \Psi_{\alpha}$, but $-\mathcal{S}(\beta) \leq -\mathcal{S}(\alpha) \Leftrightarrow \mathcal{R}(\beta) \leq \mathcal{R}(\alpha)$.

certainty-equivalence quantile coherent indices expected shortfall

Spectral axiom 5: law invariance

- law invariant: $\mathcal{S}$ depends only on distribution of Ψ_α (e.g. $f_{\Psi_\alpha}, F_{\Psi_\alpha}, \phi_{\Psi_\alpha}, Q_{\Psi_\alpha}$)
- equivalent to estimable from empirical data: by Glivenko-Cantelli, as samples become large, identically distributed rvs yield the same $\mathcal{S}$

Counterexample: general equilibrium measures of risk

> *"the risk of a portfolio depends on the other assets . . . in the economy (the market portfolio) . . . The corresponding measure of risk of a portfolio [is the] cash needed to sell the risk . . . in the portfolio to the market" (Csóka, Herings and Kóczy, 2007)*

Market impact: thin markets (illiquid or emerging) or correlated behaviour (Shin, 2010), inc. algorithmic crowding

Worst conditional expectation (Artzner et al., 1999, Definition 5.2) not law invariant as conditions on state space (Acerbi, 2002)

Hanson, Kashyap and Stein (2011) on macroprudential regulation

Spectral axiom 6: co-monotonic additivity

- allocations $\boldsymbol{\alpha}$ and $\boldsymbol{\delta}$ are co-monotonic if their objectives are co-monotonic ▸ co-monotonicity
- combining co-monotonic allocations does not provide genuine diversification
- thus, index of satisfaction is co-monotonically additive iff

$$(\boldsymbol{\alpha}, \boldsymbol{\delta}) \text{ co-monotonic } \Rightarrow \mathcal{S}(\boldsymbol{\alpha} + \boldsymbol{\delta}) = \mathcal{S}(\boldsymbol{\alpha}) + \mathcal{S}(\boldsymbol{\delta})$$

- such indices are "derivative-proof"

◂ certainty-equivalence ◂ quantile ◂ expected shortfall

law invariant + monotonic ⇒ consistent with stochastic dominance

- have applied non-satiation to stochastic dominance: can also apply to weaker concepts of dominance
- spectral measure ⇒ consistence with weak / first order dominance

$$Q_{\Psi_{\boldsymbol{\alpha}}}(p) \geq Q_{\Psi_{\boldsymbol{\beta}}}(p)\,\forall p \in (0,1) \Rightarrow \mathcal{S}(\boldsymbol{\alpha}) \geq \mathcal{S}(\boldsymbol{\beta})$$

  (Meucci, 2005, p.291, www.5.2)

- monotonicity's $\Psi \boldsymbol{\alpha} \geq \Psi_{\boldsymbol{\beta}} \forall \mathfrak{e} \in \mathfrak{E}$ is stronger than FOSD's $F_{\Psi_{\boldsymbol{\alpha}}}(\psi) \leq F_{\Psi_{\boldsymbol{\beta}}}(\psi)\,\forall \psi \in \mathbb{R}$; law invariance prevents any other factors

◂ certainty-equivalence ◂ quantile ◂ expected shortfall

Desideratum: risk-aversion

- let $\boldsymbol{b}$ be an allocation yielding a deterministic objective, $\psi_{\boldsymbol{b}}$
- let $\boldsymbol{f}$ be a 'fair game' allocation whose objective has $E\left\{\Psi_{\boldsymbol{f}}\right\} = 0$
- an index of satisfaction displays risk-aversion iff

$$\mathcal{S}\left(\boldsymbol{b}\right) \geq \mathcal{S}\left(\boldsymbol{b} + \boldsymbol{f}\right)$$

- the risk-premium is the dissatisfaction associated with the risky $\boldsymbol{f}$

$$RP \equiv \mathcal{S}\left(\boldsymbol{b}\right) - \mathcal{S}\left(\boldsymbol{b} + \boldsymbol{f}\right)$$

  (if money-equivalent, how interpret?)
- if $\mathcal{S}$ satisfies constancy, and $E\left\{\Psi_{\boldsymbol{\alpha}}\right\}$ exists, can factor into deterministic and 'fair game' components

$$RP\left(\boldsymbol{\alpha}\right) \equiv E\left\{\Psi_{\boldsymbol{\alpha}}\right\} - \mathcal{S}\left(\boldsymbol{\alpha}\right)$$

  (why?)
- risk-aversion $\Leftrightarrow RP\left(\boldsymbol{\alpha}\right) \geq 0$
- (relationship to concavity?)

◂ certainty-equivalence ◂ quantile ◂ expected shortfall

Lecture 7 exercises

- Meucci exercises
  - pencil-and-paper: 7.2.1, 7.2.2, 7.3.2 (how do we “notice that normal marginals [bound together by] a normal copula give rise to a normal joint distribution”?), 7.3.3
  - Python: 7.1.1 (why is equation 440 not a typo?)
- project: given a portfolio composed of your assets, write code to calculate its wealth relative to the MSCI World benchmark.

Expected utility

- recall, a measure of satisfaction maps from an allocation to a number: $\boldsymbol{\alpha} \mapsto \mathcal{S}(\boldsymbol{\alpha})$
- utility function associated with each realisation, ψ, some utility, $u(\psi)$
- expected utility is therefore

$$\boldsymbol{\alpha} \mapsto E\{u(\Psi_{\boldsymbol{\alpha}})\} \equiv \int_{\mathbb{R}} u(\psi) f_{\Psi_{\boldsymbol{\alpha}}}(\psi)\, d\psi$$

- (why not just use the expected value of the objective, $E\{\Psi_{\boldsymbol{\alpha}}\}$?)
- as utility has no meaningful units, invert to obtain certainty-equivalent

$$\boldsymbol{\alpha} \mapsto CE(\boldsymbol{\alpha}) \equiv u^{-1}(E\{u(\Psi_{\boldsymbol{\alpha}})\})$$

Properties of certainty-equivalence

1. translation invariance? ▸ translation invariance
   - only for $u(\psi) = -e^{-\frac{1}{\zeta}\psi}$ (Meucci, 2005, www.5.3)
2. super-additivity? ▸ super-additivity
   - (Meucci, 2005, p.267): only holds for linear utility, $u(\psi) \equiv \psi$
   - (what do Hennessy and Lapan (2006) results say?)
3. positive homogeneity? ▸ positive homogeneity
   - only for $u(\psi) = \psi^{1-\frac{1}{\gamma}}, \gamma \geq 1 \Rightarrow$ (Meucci, 2005, www.5.3)
4. monotonicity? ▸ monotonicity
   - what condition is required?
5. law-invariance? ▸ law-invariance
6. co-monotonic additivity? ▸ co-monotonic additivity
   - (Meucci, 2005, p.267): only holds for linear utility, $u(\psi) \equiv \psi$

Properties of certainty-equivalence

7. concavity? ▸ concavity
   - as sum of concave functions is concave, $E\{u(\cdot)\}$ concave if $u(\cdot)$ is
   - but this implies that u^{-1} is convex, so $u^{-1}(E\{u(\cdot)\})$ needn't be
8. risk-aversion? ▸ risk-aversion
   - as CE satisfies constancy

$$RP(\alpha) \equiv E\{\Psi_{\alpha}\} - CE(\alpha)$$

   - $u(\cdot)$ concave $\Leftrightarrow RP(\alpha) \geq 0$ (Meucci, 2005, www.5.3)

Computing $CE(\boldsymbol{\alpha}) \equiv u^{-1}(E\{u(\boldsymbol{\alpha}'\boldsymbol{M})\})$

Example (Exponential utility; normally distributed markets)

- exponential utility: $u(\psi) \equiv -e^{-\frac{1}{\zeta}\psi} \Rightarrow CE(\boldsymbol{\alpha}) = -\zeta \ln\left(\phi_{\boldsymbol{M}}\left(\frac{i}{\zeta}\boldsymbol{\alpha}\right)\right)$
- normally distributed markets:
  $\boldsymbol{M} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \Rightarrow CE(\boldsymbol{\alpha}) = \boldsymbol{\alpha}'\boldsymbol{\mu} - \frac{\boldsymbol{\alpha}'\boldsymbol{\Sigma}\boldsymbol{\alpha}}{2\zeta}$

- usually must approximate, e.g. second-order Taylor series expansion

$$CE(\boldsymbol{\alpha}) \equiv E\{\Psi_{\boldsymbol{\alpha}}\} - RP(\boldsymbol{\alpha}) \approx E\{\Psi_{\boldsymbol{\alpha}}\} - \frac{1}{2}A(E\{\Psi_{\boldsymbol{\alpha}}\})\, Var\{\Psi_{\boldsymbol{\alpha}}\}$$

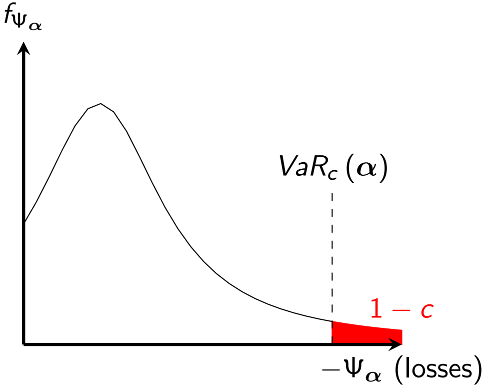
where $A(\psi) \equiv -\frac{u''(\psi)}{u'(\psi)}$ is the Arrow-Pratt measure of absolute risk-aversion

Introduction to VaR

how much can we lose on our trading portfolio by tomorrow's close? (Attributed to Dennis Weatherstone, motivating his famous 4:15 reports (Allen, Boudoukh and Saunders, 2004))

- given an investment horizon, and a confidence level, c, VaR is the maximum loss over that period $c\%$ of the time
- popularity grew after 1996, when J.P. Morgan published its VaR methodology
- in 1998, J.P. Morgan spun off the RiskMetrics group
- preferred measure of market risk adopted since Basel II
- can control bankruptcy risk (Shin, 2010)
- credit risk version called potential future exposure

VaR illustrated



- if the objective is net profits

$$\Psi_\alpha \equiv W_{T+\tau}(\alpha) - w_T$$

then, from our verbal definition,

$$\mathbb{P}\{-\Psi_\alpha \geq VaR_c(\alpha)\} = 1-c$$
$$\mathbb{P}\{\Psi_\alpha \leq -VaR_c(\alpha)\} = 1-c$$
$$F_{\Psi_\alpha}(-VaR_c(\alpha)) = 1-c$$
$$-VaR_c(\alpha) = Q_{\Psi_\alpha}(1-c)$$
$$VaR_c(\alpha) \equiv -Q_{\Psi_\alpha}(1-c)$$

quantile

- applies equally to any other Ψ_α

$$Q_c(\alpha) \equiv Q_{\Psi_\alpha}(1-c)$$

Properties of quantile measures

1. translation invariance? ▸ translation invariance
   - intuition? (Meucci, 2005, www.5.4)
2. super-additivity? ▸ super-additivity
   - fully concentrated portfolios can have lower VaR than fully diversified ones (McNeil, Frey and Embrechts, 2015, Example 2.25) ▸ VaR fail
   - this failure prompted search for alternatives
   - but holds for elliptical markets (McNeil, Frey and Embrechts, 2015, Theorem 8.28(2))
   - Embrechts, Lambrigger and Wüthrich (2009) for detailed discussion of the importance of super-additivity failures for VaR ◂ expected value
3. positive homogeneity? ▸ positive homogeneity
   - intuition? (Meucci, 2005, www.5.4) $\therefore$ Euler condition holds
4. monotonicity? ▸ monotonicity
5. law-invariance? ▸ law-invariance
6. co-monotonic additivity? ▸ co-monotonic additivity
   - intuition? (Meucci, 2005, www.5.4)
   - thus, consistent with first order stochastic dominance (counter-examples for second and higher orders Meucci (2005, p.279))

Properties of quantile measures

7. concavity? ▸ concavity
   - failure related to that of super-additivity, above?
8. risk-aversion? ▸ risk-aversion
   - $RP\left(\alpha\right)$ can take on any sign

Computing $Q_c(\boldsymbol{\alpha}) \equiv Q_{\boldsymbol{\alpha}'\boldsymbol{M}}(1-c)$

Example (Net profits and normally distributed markets)

$$\boldsymbol{P}_{T+\tau} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \text{ and } \Psi_{\boldsymbol{\alpha}} \equiv \boldsymbol{\alpha}'\boldsymbol{M} \Rightarrow \Psi_{\boldsymbol{\alpha}} \sim N\left(\mu_{\boldsymbol{\alpha}}, \sigma^2_{\boldsymbol{\alpha}}\right)$$

$$Q_c(\boldsymbol{\alpha}) = \mu_{\boldsymbol{\alpha}} + \sqrt{2}\sigma_{\boldsymbol{\alpha}} \operatorname{erf}^{-1}(1-2c)$$

- usually must approximate
  1. delta-gamma approximation: second order Taylor series expansion
  2. Cornish-Fisher expansion: expansion whose terms are the rv's moments
  3. extreme value theory as $c \to 1$: just fit the tail (e.g. using a generalised Pareto distribution)
- simulated data: sort by $\Psi_{\boldsymbol{\alpha}}$ and pick scenario nearest desired quantile
- Gourier, Farkas and Abbate (2009) applies VaR to Italian bank data; see Kritzman (2011) on thoughtful v. naïve use

Spectral indices (Acerbi, 2002)

- existing indices either satisfied or failed to satisfy certain properties
- both expected utility (in general) and quantile measures fail to satisfy super-additivity, concavity
- both may therefore fail to understand motives for diversification
- coherent indices designed to satisfy these properties
- given a coherent index, how can others be generated?
- question gave rise to spectral indices, a subclass of coherent indices
  - in satisfying additional two axioms, also satisfy risk-aversion

Expected value as a spectral measure of satisfaction

Theorem

The expected value, $E\left\{\Psi_{\boldsymbol{\alpha}}\right\}$, is a spectral measure of satisfaction.

Proof.

1. translation invariance:
   $E\left\{\Psi_{\boldsymbol{\alpha}}+\psi_{\boldsymbol{b}}\right\}=E\left\{\Psi_{\boldsymbol{\alpha}}\right\}+E\left\{\psi_{\boldsymbol{b}}\right\}=E\left\{\Psi_{\boldsymbol{\alpha}}\right\}+\psi_{\boldsymbol{b}}$
2. super-additivity: $E\left\{\Psi_{\boldsymbol{\alpha}+\boldsymbol{\beta}}\right\}=E\left\{\Psi_{\boldsymbol{\alpha}}\right\}+E\left\{\Psi_{\boldsymbol{\beta}}\right\}$
3. positive homogeneity: $E\left\{\Psi_{\lambda\boldsymbol{\alpha}}\right\}=E\left\{\lambda\Psi_{\boldsymbol{\alpha}}\right\}=\lambda E\left\{\Psi_{\boldsymbol{\alpha}}\right\}$
4. monotonicity: $\Psi_{\boldsymbol{\alpha}}\geq\Psi_{\boldsymbol{\beta}}\forall\mathfrak{e}\in\mathfrak{E}\Rightarrow E\left\{\Psi_{\boldsymbol{\alpha}}\right\}\geq E\left\{\Psi_{\boldsymbol{\beta}}\right\}$
5. law invariance: $E\left\{\Psi_{\boldsymbol{\alpha}}\right\}\equiv\int_{\mathbb{R}}\psi f_{\Psi_{\boldsymbol{\alpha}}}(\psi)\,d\psi$
6. co-monotonic additivity: additive for any $\boldsymbol{\alpha},\boldsymbol{\beta}$, not just co-monotonic

□

Expected value as an average of quantiles

Lemma

$E\{\Psi_\alpha\}$ can be written as the unweighted average of the quantiles

$$E\{\Psi_\alpha\} \equiv \int_{\mathbb{R}} \psi f_{\Psi_\alpha}(\psi)\, d\psi = \int_0^1 Q_{\Psi_\alpha}(p)\, dp$$

proof

- recall: the quantile itself is not super–additive quantile
  - but expected value, as the average of all quantiles, is
- what about an average over the worst scenarios?

Expected shortfall

- expected value averages over all scenarios

$$E\{\Psi_{\alpha}\} \equiv \int_0^1 Q_{\Psi_{\alpha}}(p)\, dp$$

- now define expected shortfall to average over the worst scenarios,

$$ES_c\{\alpha\} \equiv \frac{1}{1-c}\int_0^{1-c} Q_{\Psi_{\alpha}}(p)\, dp = E\{\Psi_{\alpha} \,|\Psi_{\alpha} \leq Q_c(\alpha)\}$$

  where $c \in [0,1]$ indexes the confidence level sought
- why is the $(1-c)^{-1}$ term present?
- when $f_{\Psi_{\alpha}}$ is smooth (Acerbi and Tasche, 2002), equivalent to
  - tail conditional expectation (TCE)
  - conditional value at risk (CVaR)

Properties of expected shortfall

1. translation invariance? ▸ translation invariance
   - from linearity of integral and translation invariance of quantile
2. super-additivity? ▸ super-additivity
   - as averaging over tail, can't 'bomb' it as can VaR
   - see Acerbi and Tasche (2002, Proposition A.1)
3. positive homogeneity? ▸ positive homogeneity
   - integral is linear; quantile is positively homogeneous; again, Euler condition holds
4. monotonicity? ▸ monotonicity
5. law-invariance? ▸ law-invariance
6. co-monotonic additivity? ▸ co-monotonic additivity
   - from linearity of integral and co-monotonic additivity of quantile

Properties of expected shortfall

7. concavity? ▸ concavity
   - from positive homogeneity and super–additivity
8. risk-aversion? ▸ risk-aversion
   - from the other properties of spectral indices

- thus $ES_c(\alpha)$ is a spectral measure of satisfaction for any $c \in [0, 1]$

Building spectral indices of satisfaction

- to generate family of spectral indices, begin with spectral basis
- use $ES_c(\boldsymbol{\alpha})$ to generate class (Acerbi (2002), Meucci (2005, www.5.5))

$$Spc_{\varphi}(\boldsymbol{\alpha}) \equiv \int_0^1 \varphi(p)\, Q_{\Psi_{\boldsymbol{\alpha}}}(p)\, dp$$

  where the spectrum, φ, (weakly) decreases to $\varphi(1) = 0$, and sets $\int_0^1 \varphi(p)\, dp = 1$
  - φ gives more weight to the lowest quantiles (the worst outcomes)
- any spectral index can be defined by a φ satisfying the above

Example

$\varphi_{ES_c}(p) \equiv \frac{1}{1-c} H^{(c-1)}(-p)$, where $H^{(x)}$ is the Heaviside step function.

- draw φ for expected value
- can you draw φ for $Q_c(\boldsymbol{\alpha})$? Why or why not?
- for other applications, q.v. Ellison and Sargent (2012)

Computing $Spc_{\varphi}(\boldsymbol{\alpha}) \equiv \int_0^1 \varphi(p)\, Q_{\boldsymbol{\alpha}'\mathbf{M}}(p)\, dp$

Example (Net profits and normally distributed markets)

$$\mathbf{P}_{T+\tau} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \text{ and } \Psi_{\boldsymbol{\alpha}} \equiv \boldsymbol{\alpha}'\mathbf{M} \Rightarrow \Psi_{\boldsymbol{\alpha}} \sim N\left(\mu_{\boldsymbol{\alpha}}, \sigma^2_{\boldsymbol{\alpha}}\right)$$

$$Q_{\boldsymbol{\alpha}'\mathbf{M}}(\boldsymbol{\alpha}) = \mu_{\boldsymbol{\alpha}} + \sqrt{2}\sigma_{\boldsymbol{\alpha}} \operatorname{erf}^{-1}(1-2c)$$

$$\Rightarrow Spc_{\boldsymbol{\alpha}} = \mu_{\boldsymbol{\alpha}} + \sqrt{2}\sigma_{\boldsymbol{\alpha}} \int_0^1 \varphi(p) \operatorname{erf}^{-1}(2p-1)\, dp$$

- usually approximate, e.g. delta-gamma approximation or Cornish-Fisher expansion
- for $ES_c(\boldsymbol{\alpha})$, can also use extreme value theory as $c \to 1$
- simulated data: sort by $\Psi_{\boldsymbol{\alpha}}$ and average scenarios below $Q_c(\boldsymbol{\alpha})$

15 Jan 2015 CHFEUR depreciation (Daníelsson, 2015)

regulator-approved standard risk models . . . under-forecast risk before the announcement and over-forecast risk after the announcement, getting it wrong in all states of the world.

	HS	MA	EWMA	GARCH	t-GARCH	EVT
VaR						
15/01/15	€14	€11	€1.6	€1.7	€2.1	€14
19/01/15	€15	€16	€89	€123	€218	€16
ES						
15/01/15	€20	€13	€1.8	€2.0	€2.9	€24
19/01/15	€35	€19	€102	€141	€301	€31
frequency (yrs)						
		2×10^{215}	∞	∞	2,079,405	109

Lecture 8 exercises

- Meucci exercises
  - pencil-and-paper: 7.4.1, 7.5.1 (not Python)
  - Python: 7.4.1, 7.4.2, 7.4.3, 7.5.1 (Python), 7.5.2
- project: given your assets and a one hour horizon, calculate the 1% ES

The ingredients

1. collecting information on the investor's profile, $\mathcal{P}$
   1. existing portfolio, $\boldsymbol{\alpha}^{(0)}$
   2. investment horizon, $T+\tau$
   3. markets of interest (e.g. alternatives, mutual funds, etc.)
   4. objective, $\Psi_{\boldsymbol{\alpha}}$
   5. risk/satisfaction index, $\mathcal{S}(\boldsymbol{\alpha})$
2. collecting information on the market, i_T
   1. current securities prices, $\boldsymbol{p}_T$
   2. horizon securities prices, $\boldsymbol{P}_{T+\tau}$ (how?)
   3. transaction costs, $\mathcal{T}\left(\boldsymbol{\alpha}^{(0)}, \boldsymbol{\alpha}\right)$

Optimal allocations

- an allocation is therefore $\boldsymbol{\alpha} : [\mathcal{P}, i_T] \mapsto \mathbb{R}^N$
- an optimal allocation is

$$\boldsymbol{\alpha}^* \equiv \operatorname{argmax} \mathcal{S}(\boldsymbol{\alpha}) \text{ s.t. } \boldsymbol{\alpha} \in \mathcal{C}$$

  where $\mathcal{C}$ defines the constraint set
  - $\boldsymbol{p}_T' \boldsymbol{\alpha} + \mathcal{T}\left(\boldsymbol{\alpha}^{(0)}, \boldsymbol{\alpha}\right) - b \leq 0$ where b is a budget constraint
  - secondary objectives (e.g. VaR targets)
- these are not generally possible to solve analytically

Constrained optimisation problems

- the general programming problem

$$\boldsymbol{z}^* \equiv \operatorname{argmin} \mathcal{Q}(\boldsymbol{z}) \text{ s.t. } \boldsymbol{z} \in \mathbb{R}^n, f_i(\boldsymbol{z}) \le \mathbf{0}, \text{ for } i = 1, \ldots, m$$

  where
  - $\mathcal{Q}(\boldsymbol{z})$ is an arbitrary objective function
  - the $f_i(\boldsymbol{z})$ are arbitrary constraints
  - $\boldsymbol{z}$ are the choice variables

  is a global optimisation problem, hence $\mathcal{NP}$-hard
- convex programming is a subset such that
  - $\mathcal{Q}(\boldsymbol{z})$ is a convex function
  - the $f_i(\boldsymbol{z})$ are also convex
- convex programming problems can be efficiently solved (e.g. in $\mathcal{P}$), have known uniqueness
- Boyd and Vandenberghe (2004) is a standard, well supported text (q.v. the open courses, here and here)

Cone programming problems

- cone programming is a subset of convex programming such that
  - $\mathcal{Q}(\boldsymbol{z})$ is a linear function, $\boldsymbol{c}'\boldsymbol{z}$
  - the $f_i(\boldsymbol{z})$ define a cone, $\mathcal{K}$

Definition (Cone)

1. closed under positive multiplication: $\boldsymbol{y} \in \mathcal{K}, \lambda \geq 0 \Rightarrow \lambda\boldsymbol{y} \in \mathcal{K}$
2. closed under addition: $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{K} \Rightarrow \boldsymbol{x} + \boldsymbol{y} \in \mathcal{K}$
3. 'pointed': $(\boldsymbol{y} \neq \mathbf{0}) \in \mathcal{K} \Rightarrow -\boldsymbol{y} \notin \mathcal{K}$

- software packages using interior-point methods (Boyd and Vandenberghe, 2004, ch.11) can efficiently solve these
- includes well-known classes as special cases

Cone programming problems

1. linear programming
   - sets $\boldsymbol{Bz} - \boldsymbol{b} \geq \boldsymbol{0}$ so that $\mathcal{K} \equiv \mathbb{R}_+^M$, the non-negative orthant
   - simplex, interior point methods both perform well
2. quadratically constrained quadratic programming (QCQP) includes LP
   - has quadratic objective

   $$\mathcal{Q} = \boldsymbol{z}'\boldsymbol{S}_{(0)}\boldsymbol{z} + 2\boldsymbol{u}'_{(0)}\boldsymbol{z} + v_{(0)}$$

   but can introduce auxiliary variable to transform $\mathcal{Q}$ into linear
3. second-order cone programming (SOCP) includes QCQP
   - auxiliary variable in QCQP transforms constraints to conic
4. semidefinite programming (SDP) includes SOCP
   - semidefinite matrix constraints generalise second-order conic

Example (Mean-variance optimisation)

Boyd and Vandenberghe (2004, pp.155-156) discuss Markowitz (1952) portfolio selection as a quadratic programming problem.

The non-negative orthant cone

$\mathbb{R}_{+}^{M} \equiv \left\{ \boldsymbol{y} \in \mathbb{R}^{M} \,|y_i \geq 0 \forall i = 1, \ldots, M \right\}$
Is this a cone?

1. closed under positive multiplication:
   $\boldsymbol{y} \in \mathbb{R}_{+}^{M} \Rightarrow y_i \geq 0 \Rightarrow \lambda y_i \geq 0 \forall \lambda \geq 0$;
2. closed under addition: $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}_{+}^{M} \Rightarrow x_i, y_i \geq 0 \Rightarrow x_i + y_i \geq 0$;
3. 'pointed': $(\boldsymbol{y} \neq \boldsymbol{0}) \in \mathbb{R}_{+}^{M} \Rightarrow \exists y_i > 0 \Rightarrow -y_i < 0$.

- while the cone itself is unique, the variables can be translated and rotated

$$\boldsymbol{Bz} + \boldsymbol{b} \geq \boldsymbol{0}$$

The Lorentz (second-order) cone

$\mathbb{K}^M \equiv \left\{ \boldsymbol{y} \in \mathbb{R}^M \,\middle|\, \| (y_1, \ldots, y_{M-1})' \|_2 \le y_M \right\}$ so that

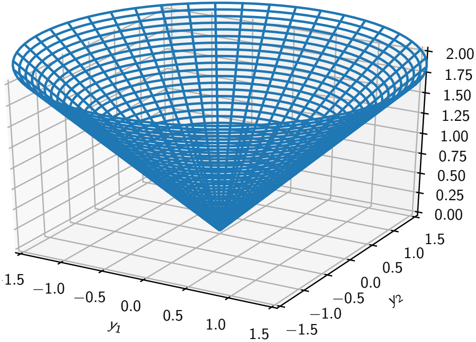
$$\sqrt{y_1^2 + \cdots + y_{M-1}^2} \le y_M$$

▸ is a cone?

- $\mathbb{K}^M$ is the Lorentz, ice-cream, norm or second-order cone
- while the cone itself is unique, constraints can be flexibly posed as

$$\| \boldsymbol{A}_i \boldsymbol{z} + \boldsymbol{b}_i \| \le \boldsymbol{f}_i' \boldsymbol{z} + d_i$$

for $i = 1, \ldots, m$



The semidefinite cone

$\mathbb{S}_+^M \equiv \{\boldsymbol{S} \succcurlyeq \boldsymbol{0}\}$, $\boldsymbol{S} \in \mathbb{R}^{M \times M}$ and $\succcurlyeq \boldsymbol{0}$ denotes PSD ▸ is a cone?

- e.g. when $M = 2$, represent

$$\boldsymbol{S} \equiv \begin{bmatrix} x_1 & x_2 \\ x_2 & x_3 \end{bmatrix}$$

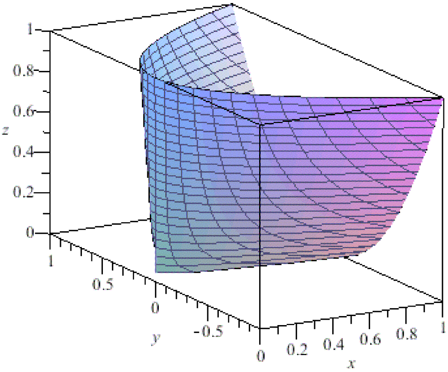
  as $(x_1, x_2, x_3) \in \mathbb{R}^3$

- PSD
  $\Leftrightarrow x_1 \geq 0, x_3 \geq 0, x_1 x_3 \geq x_2^2$

- can flexibly write constraints

$$\boldsymbol{F}_0 + \sum_{i=1}^{n} \boldsymbol{F}_i z_i \succcurlyeq \boldsymbol{0}$$

- some SDP solvers are listed here



The mean-variance approach

- Meucci (2005, §6.1) contains perhaps only non–trivial portfolio optimisation example that can be analytically solved
- more generally, even numerical solutions cannot be guaranteed outside of convex programming
    - $\mathcal{S}$ or constraints are not concave
    - quantile or certainty equivalent fail; spectral pass
- even when the problem is one of a convex programming, the computational cost may be prohibitive
- here, present mean–variance approach
    - use dates back to Markowitz (1952)
    - extremely popular
    - computationally tractable

The geometry of allocation optimisation

- the indices of satisfaction considered here are law invariant
- thus, they can be represented in terms of the distribution of $\Psi_{\boldsymbol{\alpha}}$
- distribution of $\Psi_{\boldsymbol{\alpha}}$, in turn, can be represented in terms of moments
  - certainty equivalent: Taylor expanding $u(\cdot)$ yields moments
  - quantiles or spectral indices: Cornish-Fisher expansion
- therefore, if $\mathcal{S}(\boldsymbol{\alpha})$ is an analytic function,

$$\mathcal{S}(\boldsymbol{\alpha}) \equiv \mathcal{H}\left(E\{\Psi_{\boldsymbol{\alpha}}\}, CM_2\{\Psi_{\boldsymbol{\alpha}}\}, CM_3\{\Psi_{\boldsymbol{\alpha}}\}, \ldots\right)$$

  where CM_k is the k^{th} central moment of $\Psi_{\boldsymbol{\alpha}}$
- iso-satisfaction surfaces therefore live in $(\infty - 1)$-dimensional subspace of moments
- within this, though, $\boldsymbol{\alpha} \in \mathbb{R}^N$ spans a subspace
- within that subspace, the constraint that $\boldsymbol{\alpha} \in \mathcal{C}$ is further restrictive
- solving the problem is finding the point in that final subspace corresponding to the highest level of satisfaction

Dimension reduction: the mean-variance framework

- instead of the infinite-dimensional version of $\mathcal{S}(\boldsymbol{\alpha})$, consider

$$\mathcal{S}(\boldsymbol{\alpha}) \approx \tilde{\mathcal{H}}\left(E\left\{\Psi_{\boldsymbol{\alpha}}\right\}, Var\left\{\Psi_{\boldsymbol{\alpha}}\right\}\right)$$

- this would clearly be easier to solve, when the approximation is good
- all $\mathcal{S}(\boldsymbol{\alpha})$ considered above are consistent with stochastic dominance
  - given some fixed $Var\left\{\Psi_{\boldsymbol{\alpha}}\right\}$, higher $E\left\{\Psi_{\boldsymbol{\alpha}}\right\}$ preferred for any $\tilde{\mathcal{H}}$
  - cannot assume dual: given some fixed $E\left\{\Psi_{\boldsymbol{\alpha}}\right\}$, lower $Var\left\{\Psi_{\boldsymbol{\alpha}}\right\}$ preferred for any $\tilde{\mathcal{H}}$?

$\Rightarrow$ optimal allocation $\boldsymbol{\alpha}^*$ belongs to 1-parameter family $\boldsymbol{\alpha}(v)$, where

$$\boldsymbol{\alpha}(v) \equiv \underset{\substack{\boldsymbol{\alpha} \in \mathcal{C} \\ Var\left\{\Psi_{\boldsymbol{\alpha}}\right\} = v}}{\operatorname{argmax}} E\left\{\Psi_{\boldsymbol{\alpha}}\right\}$$

- solution is Markowitz' mean-variance efficient frontier

A two-step approach to the mean-variance frontier

1. compute the mean-variance efficient frontier,

$$\boldsymbol{\alpha}(v) \equiv \underset{\substack{\boldsymbol{\alpha} \in \mathcal{C} \\ Var\{\Psi_{\boldsymbol{\alpha}}\} = v}}{\operatorname{argmax}} E\{\Psi_{\boldsymbol{\alpha}}\}$$

2. perform the one-dimensional search,

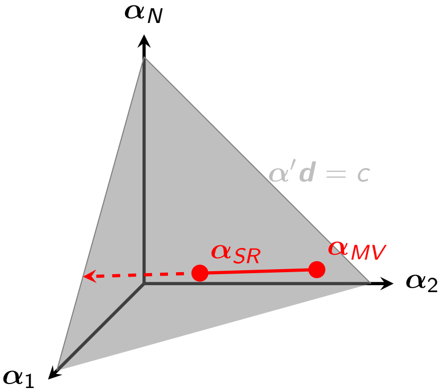
$$\boldsymbol{\alpha}^* \equiv \boldsymbol{\alpha}(v^*) \equiv \underset{v \geq 0}{\operatorname{argmax}} \mathcal{S}(\boldsymbol{\alpha}(v))$$

- $\boldsymbol{\Psi}_{\boldsymbol{\alpha}} \equiv \boldsymbol{\alpha}' \boldsymbol{M} \Rightarrow E\{\Psi\} = \boldsymbol{\alpha}' E\{\boldsymbol{M}\}, Var\{\Psi\} = \boldsymbol{\alpha}' Cov\{\boldsymbol{M}\} \boldsymbol{\alpha}$,
- thus, can write first step in terms of the horizon market vector, $\boldsymbol{M}$:

$$\boldsymbol{\alpha}(v) \equiv \underset{\substack{\boldsymbol{\alpha} \in \mathcal{C} \\ \boldsymbol{\alpha}' Cov\{\boldsymbol{M}\} \boldsymbol{\alpha} = v \geq 0}}{\operatorname{argmax}} \boldsymbol{\alpha}' E\{\boldsymbol{M}\}$$

(why do we do this?)

A single affine constraint: allocation space



Example

$\mathcal{C} : \boldsymbol{\alpha}' \boldsymbol{p}_T = w_T$

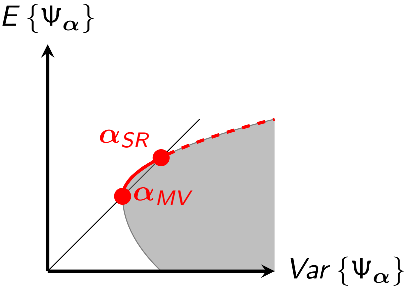
- let $\mathcal{C}$ be an affine constraint

$$\mathcal{C} : \boldsymbol{\alpha}' \boldsymbol{d} = c$$

  s.t. $\boldsymbol{d}, E\{\boldsymbol{M}\}$ not collinear
- frontier then a semi-line on $(N-1)$-d constraint
- two-fund separation theorem: frontier is linear combination of MV, SR portfolios

1. why is frontier linear?
2. why begin at $\boldsymbol{\alpha}_{MV}$?
3. why dotted above $\boldsymbol{\alpha}_{SR}$?

A single affine constraint: mean-variance space



Example

$\mathcal{C}:\boldsymbol{\alpha}'\boldsymbol{p}_T = w_T$

- approximating $\mathcal{S}(\boldsymbol{\alpha})$ by 1st two moments

1. why is frontier a parabola?
2. why begin at $\boldsymbol{\alpha}_{MV}$?
   - what is $\boldsymbol{\alpha}_{MV}$?
3. why dotted above $\boldsymbol{\alpha}_{SR}$?
   - what is $\boldsymbol{\alpha}_{SR}$?
   - Fig 6.11 depicts in (E, Sd) space

$$SR(\boldsymbol{\alpha}) \equiv \frac{E\{\Psi_\alpha\}}{Sd\{\Psi_\alpha\}}$$

- can analyse market neutral special case by $w_T = 0$

MV as an approximation

- remember, are approximating $\mathcal{S}(\boldsymbol{\alpha})$ by

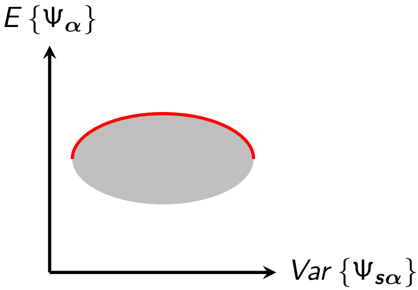
$$\mathcal{S}(\boldsymbol{\alpha}) \approx \tilde{\mathcal{H}}\left(E\left\{\Psi_{\boldsymbol{\alpha}}\right\}, Var\left\{\Psi_{\boldsymbol{\alpha}}\right\}\right)$$

  where $\Psi_{\boldsymbol{\alpha}} \equiv \boldsymbol{\alpha}' \boldsymbol{M}$
- approximation exact iff $\mathcal{S}(\boldsymbol{\alpha})$ only depends on 1st two moments
  - preferences: iff $\mathcal{S}(\boldsymbol{\alpha})$ is certainty equivalent with quadratic utility

$$u(\psi) \equiv \psi - \frac{1}{2\gamma}\psi^2,$$

    then $\tilde{\mathcal{H}} = \mathcal{H}$ for any $\boldsymbol{M}$
  - markets: iff $\boldsymbol{M} \sim El(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g_N)$, then $\tilde{\mathcal{H}} = \mathcal{H}$ for any $\mathcal{S}(\cdot)$
- how well do the first two moments capture your particular problem?
  - can we extend this methodology to first three, four moments?

The dual formulation: check case-by-case



when can we write

$$\alpha(v) \equiv \underset{\substack{\alpha \in \mathcal{C} \\ Var\{\Psi_{\alpha}\} = v}}{\operatorname{argmax}} E\{\Psi_{\alpha}\} \ldots$$

1. as the inequality constrained

$$\alpha(v) \equiv \underset{\substack{\alpha \in \mathcal{C} \\ Var\{\Psi_{\alpha}\} \leq v}}{\operatorname{argmax}} E\{\Psi_{\alpha}\}?$$

2. as the dual formulation

$$\alpha(e) \equiv \underset{\substack{\alpha \in \mathcal{C} \\ E\{\Psi_{\alpha}\} \geq e}}{\operatorname{argmin}} Var\{\Psi_{\alpha}\}?$$

Example

Net profits?

◂ robust optimisation

Lecture 9 exercises

- Meucci exercises
  - pencil-and-paper: 8.1.3 (non-Python component)
  - Python: 8.1.1, 8.1.2, 8.1.3 (Python component), 8.1.4
- project:
  1. for your assets (or reduced dimension version), implement the two-step mean-variance search procedure, first identifying the locus of $\alpha(v)$, and then α^*
  2. think about and explain why the resulting α^* is the optimal portfolio (given a particular estimation/forecast method)

Bayesian estimation

- classical estimation: from data to parameters

$$i_T \equiv \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T\} \mapsto \hat{\boldsymbol{\theta}}$$

  - but different realisations from the same DGP $\rightarrow$ different $\hat{\boldsymbol{\theta}}$
- Bayesian estimation
  - experience, e_C (and confidence in it) imply prior beliefs, $f_{pr}(\boldsymbol{\theta}_0)$
  - combining this with $\hat{\boldsymbol{\theta}}$ from i_T yields a posterior distribution

$$i_T, e_C \mapsto f_{po}(\boldsymbol{\theta})$$

  - can view prior beliefs as reflecting C pseudo-observations

$$C \rightarrow \infty \Rightarrow f_{po}(\boldsymbol{\theta}) \rightarrow \boldsymbol{\theta}_0;\ T \rightarrow \infty \Rightarrow f_{po}(\boldsymbol{\theta}) \rightarrow \hat{\boldsymbol{\theta}};$$

- classical-equivalent estimators use location parameter (e.g. expected value, mode) to summarise f_{po}
  - Bayes-Stein shrinkage estimators when shrink onto prior beliefs
    shrinkage estimators
- often computationally expensive due to integration
- modelling priors as normal-inverse-Wishart may aid tractability

Determining the prior

- theory nice, but who can convert their beliefs into a distribution?

1. peak then tweak
   - often specify location parameter from 'peak' of prior beliefs
   - then 'tweak' dispersion parameter to vary confidence levels
2. allocation implied parameters
   - investors often have a better idea of their preferred portfolio, $\boldsymbol{\alpha}$, than of the underlying market parameters, $\boldsymbol{\theta}$
   - view preferred portfolio as solving

$$\boldsymbol{\alpha}(\boldsymbol{\theta}) \equiv \underset{\boldsymbol{\alpha} \in \mathcal{C}}{\operatorname{argmax}} \, \mathcal{S}^{\boldsymbol{\theta}}(\boldsymbol{\alpha})$$

   - if $\boldsymbol{\theta}$'s dimension more than N, need further restrictions to invert for $\boldsymbol{\theta}$
3. prior constrained likelihood maximisation
   - existing $\boldsymbol{\alpha}$ implies $\tilde{\boldsymbol{\Theta}} \subset \boldsymbol{\Theta}$, subset of parameters consistent with $\boldsymbol{\alpha} \in \mathcal{C}$
   - use ML on i_T to estimate priors within $\tilde{\boldsymbol{\Theta}}$

Opportunity cost of suboptimal allocations

- an optimal allocation solves

$$\boldsymbol{\alpha}^{*} \equiv \underset{\boldsymbol{\alpha}\in\mathcal{C}}{\operatorname{argmax}}\, \mathcal{S}\left(\boldsymbol{\alpha}\right)$$

- the opportunity cost of a generic allocation $\boldsymbol{\alpha}$ is

$$OC\left(\boldsymbol{\alpha}\right) \equiv \mathcal{S}\left(\boldsymbol{\alpha}^{*}\right) - \mathcal{S}\left(\boldsymbol{\alpha}\right) \geq 0$$

(for expositional clarity, ignoring costs of constraint violation)

- as satisfaction from any $\boldsymbol{\alpha}$ depends on the unknown parameters

$$\boldsymbol{\theta} \mapsto \boldsymbol{X}^{\boldsymbol{\theta}}_{T+\tau} \mapsto \boldsymbol{P}^{\boldsymbol{\theta}}_{T+\tau} \Rightarrow \left(\boldsymbol{\alpha}, \boldsymbol{P}^{\boldsymbol{\theta}}_{T+\tau}\right) \mapsto \Psi^{\boldsymbol{\theta}}_{\boldsymbol{\alpha}} \mapsto \mathcal{S}^{\boldsymbol{\theta}}\left(\boldsymbol{\alpha}\right)$$

so does the opportunity cost

$$OC^{\boldsymbol{\theta}}\left(\boldsymbol{\alpha}\right) \equiv \mathcal{S}^{\boldsymbol{\theta}}\left(\boldsymbol{\alpha}^{*}\left(\boldsymbol{\theta}\right)\right) - \mathcal{S}^{\boldsymbol{\theta}}\left(\boldsymbol{\alpha}\right) \geq 0$$

- as generic allocation depends on data, i_T, so does opportunity cost

$$OC^{\boldsymbol{\theta}}\left(\boldsymbol{\alpha}\left[i_T\right]\right) \equiv \mathcal{S}^{\boldsymbol{\theta}}\left(\boldsymbol{\alpha}^{*}\left(\boldsymbol{\theta}\right)\right) - \mathcal{S}^{\boldsymbol{\theta}}\left(\boldsymbol{\alpha}\left[i_T\right]\right) \geq 0$$

but unobtainable $\mathcal{S}^{\boldsymbol{\theta}}\left(\boldsymbol{\alpha}^{*}\left(\boldsymbol{\theta}\right)\right)$ does not (why?)

Opportunity cost as a random variable

- i_T is a realisation of random $I_T^{\boldsymbol{\theta}} \equiv \left\{ \boldsymbol{X}_1^{\boldsymbol{\theta}}, \ldots, \boldsymbol{X}_T^{\boldsymbol{\theta}} \right\}$
- thus, if an allocation is result of a decision rule, $\boldsymbol{\alpha}\left[I_T^{\boldsymbol{\theta}}\right]$ is an rv
- in turn, the opportunity cost itself is a random variable

$$OC^{\boldsymbol{\theta}}\left(\boldsymbol{\alpha}\left[I_T^{\boldsymbol{\theta}}\right]\right) \equiv \mathcal{S}^{\boldsymbol{\theta}}\left(\boldsymbol{\alpha}^*\left(\boldsymbol{\theta}\right)\right) - \mathcal{S}^{\boldsymbol{\theta}}\left(\boldsymbol{\alpha}\left[I_T^{\boldsymbol{\theta}}\right]\right) \geq 0$$

- can now stress test an allocation (mapping from the i_T)
  - how does OC vary over $\boldsymbol{\Theta}$, subset expected to contain the true $\boldsymbol{\theta}$?
  - ideally, want OC low for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$
- Meucci does not aggregate this into a single number
  - $\mathcal{S}\left(\boldsymbol{\alpha}\right)$ aggregates $\Psi_{\boldsymbol{\alpha}}$, but ...
  - "modeling the investor's attitude toward estimation risk is an even harder task than modeling his attitude toward risk"
  - is this what the ambiguity literature (Epstein and Schneider, 2010) seeks to do?

Allocating on the basis of priors only

- consider the prior allocation rule

$$\boldsymbol{\alpha}_p \left[i_T \right] \equiv \boldsymbol{\alpha}$$

where $\boldsymbol{\alpha}$ is a fixed portfolio

- then the opportunity cost becomes deterministic, given $\boldsymbol{\theta}$

$$OC^{\boldsymbol{\theta}} \left(\boldsymbol{\alpha}_p \left[I_T^{\boldsymbol{\theta}} \right] \right) \equiv \mathcal{S}^{\boldsymbol{\theta}} \left(\boldsymbol{\alpha}^* \left(\boldsymbol{\theta} \right) \right) - \mathcal{S}^{\boldsymbol{\theta}} \left(\boldsymbol{\alpha}_p \right) \geq 0$$

- further, $OC^{\boldsymbol{\theta}} \left(\boldsymbol{\alpha}_p \right)$ is generally large
  - q.v. bias of constant estimator

Example (Equally-weighted portfolio)

The equally-weighted portfolio is determined exclusively by prior views.

Sample-based allocation

- until this week, have estimated parameters from data, $\hat{\boldsymbol{\theta}}\left[i_T\right]$
- let the sample allocation be

$$\boldsymbol{\alpha}_s\left[i_T\right] \equiv \boldsymbol{\alpha}\left(\hat{\boldsymbol{\theta}}\left[i_T\right]\right) \equiv \underset{\boldsymbol{\alpha}\in\mathcal{C}^{\hat{\theta}\left[i_T\right]}}{\operatorname{argmax}}\, \mathcal{S}^{\hat{\theta}\left[i_T\right]}\left(\boldsymbol{\alpha}\right)$$

- to evaluate performance, do the following $\forall\boldsymbol{\theta}\in\boldsymbol{\Theta}$ (the stress test set)
  1. compute the deterministic $\mathcal{S}^{\boldsymbol{\theta}}\left(\boldsymbol{\alpha}^*\left(\boldsymbol{\theta}\right)\right)$
  2. generate a distribution of i_T's $\Rightarrow$ distribution for $\hat{\boldsymbol{\theta}}\left[I_T^{\boldsymbol{\theta}}\right]$
  3. produce distribution of 'optimal' allocations, indexed by $\hat{\boldsymbol{\theta}}$

$$\boldsymbol{\alpha}_s\left[I_T^{\boldsymbol{\theta}}\right] \equiv \boldsymbol{\alpha}\left(\hat{\boldsymbol{\theta}}\left[I_T^{\boldsymbol{\theta}}\right]\right) \equiv \underset{\boldsymbol{\alpha}\in\mathcal{C}^{\hat{\theta}\left[I_T^{\boldsymbol{\theta}}\right]}}{\operatorname{argmax}}\, \mathcal{S}^{\hat{\theta}\left[I_T^{\boldsymbol{\theta}}\right]}\left(\boldsymbol{\alpha}\right)$$

  4. compute distribution of $\mathcal{S}^{\boldsymbol{\theta}}\left(\boldsymbol{\alpha}_s\left[I_T^{\boldsymbol{\theta}}\right]\right)$, given indexed $\boldsymbol{\theta}$
  5. compute the distribution of $OC^{\boldsymbol{\theta}}\left(\boldsymbol{\alpha}_s\left[I_T^{\boldsymbol{\theta}}\right]\right) \equiv \mathcal{S}^{\boldsymbol{\theta}}\left(\boldsymbol{\alpha}^*\left(\boldsymbol{\theta}\right)\right) - \mathcal{S}^{\boldsymbol{\theta}}\left(\boldsymbol{\alpha}_s\left[I_T^{\boldsymbol{\theta}}\right]\right)$
- thus, for each $\boldsymbol{\theta}\in\boldsymbol{\Theta}$ have a distribution of $OC^{\boldsymbol{\theta}}\left(\boldsymbol{\alpha}_s\left[I_T^{\boldsymbol{\theta}}\right]\right)$

Evaluating the sample-based allocation approach

- if $\hat{\boldsymbol{\theta}}$ is an unbiased estimator of $\boldsymbol{\theta}$, then bulk of $OC^{\boldsymbol{\theta}}\left(\boldsymbol{\alpha}_s\left[I_T^{\boldsymbol{\theta}}\right]\right)$ distribution is close to zero
- however, $\boldsymbol{\alpha}_s$ is inefficient due to sensitivity of optimal allocation to inefficiency in $\hat{\boldsymbol{\theta}}$
  1. as sample-based estimators are inefficient, $\hat{\boldsymbol{\theta}}\left[I_T^{\boldsymbol{\theta}}\right]$ is
  2. inefficient estimates of $\hat{\boldsymbol{\theta}}\left[I_T^{\boldsymbol{\theta}}\right]$ propagate estimation error into estimates of satisfaction, $\mathcal{S}^{\hat{\boldsymbol{\theta}}}$, the constraints, $\mathcal{C}^{\hat{\theta}}$, and ...
  3. the computed optimal allocation itself, $\boldsymbol{\alpha}_s$
- thus, allocations can vary greatly with the particular history used
- can trade off bias against efficiency by using shrinkage estimators

▸ shrinkage estimators

Overview

- approach described previously
  1. estimates market distribution
  2. inputs these estimates into a classical optimiser

  so that the parameter estimation inefficiency propagates through
- now consider alternatives that limit sensitivity
  1. Bayesian allocation that shrinks parameters estimates to priors on $\boldsymbol{\theta}$
  2. Black–Litterman allocation that shrinks with respect to market views
     - this 2014 article discusses roboadvisors' use of Black-Litterman
  3. Michaud resampling . . . ?
- as well as approaches that don't just try to limit sensitivity
  4. robust allocation doesn't limit sensitivity, but picks to ensure against bad outcome
  5. robust Bayesian blends robust with Bayesian

Bayesian allocation

- as don't know true $\boldsymbol{\theta}$, can never implement optimal allocation

$$\boldsymbol{\alpha}(\boldsymbol{\theta}) \equiv \underset{\boldsymbol{\alpha} \in \mathcal{C}^{\boldsymbol{\theta}}}{\operatorname{argmax}} \mathcal{S}^{\boldsymbol{\theta}}(\boldsymbol{\alpha})$$

- but can implement classical–equivalent Bayesian allocation decision

$$\boldsymbol{\alpha}_{ce}[i_T, e_C] \equiv \boldsymbol{\alpha}\left(\hat{\boldsymbol{\theta}}_{ce}[i_T, e_C]\right) \equiv \underset{\boldsymbol{\alpha} \in \mathcal{C}^{\hat{\boldsymbol{\theta}}_{ce}[i_T, e_C]}}{\operatorname{argmax}} \mathcal{S}^{\hat{\boldsymbol{\theta}}_{ce}[i_T, e_C]}(\boldsymbol{\alpha})$$

  where $\hat{\boldsymbol{\theta}}_{ce}[i_T, e_C]$ is posterior location parameter
- as with sample–based, evaluate CEBA over all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ (domain of f_{po})
  1. compute the deterministic $\mathcal{S}^{\boldsymbol{\theta}}(\boldsymbol{\alpha}^*(\boldsymbol{\theta}))$
  2. generate a distribution of i_T's $\Rightarrow$ distribution for $\hat{\boldsymbol{\theta}}_{ce}\left[I_T^{\boldsymbol{\theta}}, e_C\right]$
  3. produce $\boldsymbol{\alpha}_{ce}\left[I_T^{\boldsymbol{\theta}}, e_C\right]$, distribution of 'optimal' allocations given $\hat{\boldsymbol{\theta}}_{ce}$
  4. compute distribution of $\mathcal{S}^{\boldsymbol{\theta}}\left(\boldsymbol{\alpha}_{ce}\left[I_T^{\boldsymbol{\theta}}, e_C\right]\right)$, given indexed $\boldsymbol{\theta}$
  5. compute the distribution of $OC^{\boldsymbol{\theta}}\left(\boldsymbol{\alpha}_{ce}\left[I_T^{\boldsymbol{\theta}}, e_C\right]\right)$
- relative to $\boldsymbol{\alpha}_s$, minimises, tightens OC especially where prior strongest

Black-Litterman allocation

- as don't know true $\boldsymbol{\theta}$, can never implement optimal allocation

$$\boldsymbol{\alpha}(\boldsymbol{\theta}) \equiv \underset{\boldsymbol{\alpha}\in\mathcal{C}^{\boldsymbol{\theta}}}{\operatorname{argmax}}\, \mathcal{S}^{\boldsymbol{\theta}}(\boldsymbol{\alpha})$$

- as CEBA, BL uses Bayes to limit sensitivity to $\boldsymbol{\theta}$'s inefficient estimation
- CEBA shrinks estimates of market parameters, $\boldsymbol{\theta}$, to their priors
- BL shrinks estimates of market distribution, say $\boldsymbol{X}$, to their priors
  1. given some market rv $\boldsymbol{X}$, quants determine a distribution $f_{\boldsymbol{X}}$
  2. experienced investor provides view, $\boldsymbol{v}$
     - $\boldsymbol{v}$ seen as realisation of rv $\boldsymbol{V}$ (else complete shrinkage)
     - $\boldsymbol{V}|g(\boldsymbol{x})$ is investor's view given model prediction, e.g. $\boldsymbol{V}|\boldsymbol{x} \sim N(\boldsymbol{x}, \boldsymbol{\Omega}')$
     - $g(\boldsymbol{x})$ allows the investor's views to depend on a function of the market
  3. Bayes' rule computes posterior distribution

$$f_{\boldsymbol{X}|\boldsymbol{v}}(\boldsymbol{x}) = \frac{f_{\boldsymbol{V}|g(\boldsymbol{x})}(\boldsymbol{v})\, f_{\boldsymbol{X}}(\boldsymbol{x})}{\int f_{\boldsymbol{V}|g(\boldsymbol{x}')}(\boldsymbol{v})\, f_{\boldsymbol{X}}(\boldsymbol{x}')\, d\boldsymbol{x}'}$$

  4. Black-Litterman allocation decision depends on i_T iff quant model does

$$\boldsymbol{\alpha}_{BL}[\boldsymbol{v}] \equiv \operatorname{argmax} \mathcal{S}^{\boldsymbol{v}}(\boldsymbol{\alpha})$$

Example (Linear expertise on normal markets (Meucci, 2010))

1. stock indices for Italy, Spain, Switzerland, Canada, USA, Germany are modelled as normal, $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \Sigma)$
2. investor provides point estimates on areas of expertise, $\boldsymbol{v}$
   - two views: Spanish index to gain 12% annualised, and German index to outperform US index by 10% annualised
   - investor's expertise is linear, $g(\boldsymbol{x}) = \boldsymbol{P}\boldsymbol{x}$
     - $\boldsymbol{P}$ is $K \times N$ pick matrix, representing K views
     - k^{th} row is N-vector corresponding to k^{th} view
     - a view on Spain, and one on US-German relative performance are

$$\boldsymbol{P} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix}$$

   - conditional distribution of views is normal, $\boldsymbol{V}|f(\boldsymbol{x}) = \boldsymbol{V}|\boldsymbol{P}\boldsymbol{x} \sim N(\boldsymbol{P}\boldsymbol{x}, \Omega)$
3. the posterior market vector given the view is $\boldsymbol{X}|\boldsymbol{v} \sim N(\boldsymbol{\mu}_{BL}, \Sigma_{BL})$.

Robust allocation

- now, don't try to reduce sensitivity to inefficient estimation of $\boldsymbol{\theta}$
- instead, most conservative approach: minimise the maximum OC over the stress–test set

1. use i_T to define robustness set, $\hat{\boldsymbol{\Theta}}[i_T]$, smallest $\boldsymbol{\Theta}$ that contains true $\boldsymbol{\theta}$
2. define constraint set to ensure allocation feasible for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$

$$\mathcal{C}^{\hat{\boldsymbol{\Theta}}[i_T]} \equiv \left\{ \boldsymbol{\alpha} \in \mathcal{C}^{\boldsymbol{\theta}} \forall \boldsymbol{\theta} \in \hat{\boldsymbol{\Theta}}[i_T] \right\}$$

3. the robust allocation decision then maps from i_T to solve

$$\boldsymbol{\alpha}_r[i_T] \equiv \underset{\boldsymbol{\alpha} \in \mathcal{C}^{\hat{\boldsymbol{\Theta}}[i_T]}}{\operatorname{argmin}} \max_{\boldsymbol{\theta} \in \hat{\boldsymbol{\Theta}}[i_T]} \left[\mathcal{S}^{\boldsymbol{\theta}}(\boldsymbol{\alpha}^*(\boldsymbol{\theta})) - \mathcal{S}^{\boldsymbol{\theta}}(\boldsymbol{\alpha}) \right]$$

- e.g. zero-sum game against evil demon picking the worst $\boldsymbol{\theta} \in \boldsymbol{\Theta}$

Mean-variance framework for robust allocation

- prohibitively expensive to implement minmax computationally
  - therefore, use two-step mean-variance framework again
- further simplifying assumptions
  1. constraints don't depend on $\mathbf{\Theta}$
  2. can write variance constraint as $Var\{\Psi_{\boldsymbol{\alpha}}\} \leq v$ (Meucci, 2005, §6.5.3)

dual

- given $\mathbf{\Theta}$, first step then becomes

$$\boldsymbol{\alpha}_r(v) \equiv \underset{\boldsymbol{\alpha}}{\operatorname{argmax}} \min_{\boldsymbol{\mu} \in \hat{\mathbf{\Theta}}_{\boldsymbol{\mu}}} \boldsymbol{\alpha}'\boldsymbol{\mu} \text{ s.t. } \begin{cases} \boldsymbol{\alpha} & \in \mathcal{C} \\ \max_{\mathbf{\Sigma} \in \hat{\mathbf{\Theta}}_{\mathbf{\Sigma}}} \boldsymbol{\alpha}'\mathbf{\Sigma}\boldsymbol{\alpha} & \leq v \end{cases}$$

- careful choice of $\hat{\mathbf{\Theta}}$ allows problem to be cast as SOCP

Example (Elliptical expectations, known covariances)

1. elliptical expectations: $\hat{\mathbf{\Theta}}_{\boldsymbol{\mu}} \equiv \left\{\boldsymbol{\mu} \text{ s.t. } Ma^2(\boldsymbol{\mu}, \boldsymbol{m}, \boldsymbol{T}) \leq q^2\right\}$
2. known covariances: $\hat{\mathbf{\Theta}}_{\mathbf{\Sigma}} \equiv \hat{\mathbf{\Sigma}}$

The Meucci mantra

1. for each security, identify the iid stochastic terms (§3.1)
2. estimate the distribution of the market invariants (§4)
3. project the invariants to the investment horizon (§3.2)
4. dimension reduce to make the problem more tractable (§3.4)
5. evaluate the portfolio performance at the investment horizon (§5)
    - what is your objective function?
6. pick the portfolio that optimises your objective function (§6)
7. account for estimation risk
    1. replace point parameter estimates with Bayesian distributions (§7)
    2. re-evaluate the portfolio distributions in this light (§8)
    3. robustly re-optimise (§9)

Observation shows that some statistical frequencies are, within narrower or wider limits, stable. But stable frequencies are not very common, and cannot be assumed lightly. Keynes (1921, p.381)

- Resti and Sironi (2007, Part V) is extensive

References I

Acerbi, Carlo (July 2002). "Spectral measures of risk: a coherent representation of subjective risk aversion". *Journal of Banking and Finance* 26.7, pp. 1505–1518.

Acerbi, Carlo and Dirk Tasche (July 2002). "On the coherence of expected shortfall". *Journal of Banking and Finance* 26.7, pp. 1487–1503.

Allen, Linda, Jacob Boudoukh and Anthony Saunders (2004). *Understanding market, credit, and operational risk: the Value at Risk approach*. Blackwell Publishing.

Artzner, Philippe et al. (July 1999). "Coherent measures of risk". *Mathematical Finance* 9.3, pp. 203–228.

Baker, Malcolm, Brendan Bradley and Jeffrey Wurgler (Jan. 2011). "Benchmarks as limits to arbitrage: understanding the low-volatility anomaly". *Financial Analysts Journal* 67.1, pp. 40–54.

References II

Bauer, Rob and Robin Braun (Nov. 2010). "Misdeeds matter: long-term stock price performance after the filing of class-action lawsuits". *Financial Analysts Journal* 66.6, pp. 74–92.

Berg, Christian and Christophe Vignat (Jan. 2008). "Linearization coefficients of Bessel polynomials and properties of Student t-distributions". *Constructive Approximation* 27.1, pp. 15–32.

Bielecki, Tomasz R. and Marek Rutkowski (2002). *Credit risk: modeling, valuation and hedging*. Springer Finance. Springer.

Bookstaber, Richard (2007). *A Demon of our own Design*. John Wiley & Sons.

Boyd, Stephen P. and Lieven Vandenberghe (2004). *Convex optimization*. Cambridge University Press.

Brigo, Damiano, Massimo Morini and Andrea Pallavicini (2013). *Counterparty credit risk, collateral and funding*. Finance. Wiley.

References III

- Brock, William A. et al. (1996). "A Test for Independence Based on the Correlation Dimension". *Econometric Reviews* 15.3, pp. 197–235.
- Campbell, John Y., Andrew W. Lo and A. Craig MacKinlay (1997). *The Econometrics of Financial Markets*. Princeton University Press.
- Cochrane, John H. (Spring 2013). "Finance: function matters, not size". *Journal of Economic Perspectives* 27.2, pp. 29–50.
- Csóka, Péter, P. Jean-Jacques Herings and László Á. Kóczy (Aug. 2007). "Coherent measures of risk from a general equilibrium perspective". *Journal of Banking and Finance* 31.8, pp. 2517–2534.
- Daníelsson, Jón (18 January 2015). *What the Swiss FX shock says about risk models*. Tech. rep. URL: http://voxeu.org/article/what-swiss-fx-shock-says-about-risk-models.
- Danielsson, Jon and Jean-Pierre Zigrand (Oct. 2006). "On time-scaling of risk and the square-root-of-time rule". *Journal of Banking and Finance* 30.10, pp. 2701–2713.

References IV

Dybvig, Philip H. and Stephen A. Ross (June 1985). "Differential Information and Performance Measurement Using a Security Market Line". *Journal of Finance* 40.2, pp. 383–399.

Ellison, Martin and Thomas J. Sargent (Nov. 2012). "A Defence of the FOMC". *International Economic Review* 53.4, pp. 1047–65.

Embrechts, Paul (Sept. 2009). "Copulas: a personal view". *Journal of Risk and Insurance* 76.3, pp. 639–650.

Embrechts, Paul, Dominik D. Lambrigger and Mario V. Wüthrich (June 2009). "Multivariate extremes and the aggregation of dependent risks: examples and counter-examples". *Extremes* 12.2, pp. 107–127.

Epstein, Larry G. and Martin Schneider (Dec. 2010). "Ambiguity and asset markets". *Annual Review of Financial Economics* 2, pp. 315–346.

Fama, Eugene F. and Kenneth French (Feb. 1993). "Common Risk Factors in the Returns on Stocks and Bonds". *Journal of Financial Economics* 33.1, pp. 3–56.

References V

Filipovi, Damir and Nicolas Vogelpoth (June 2008). "A note on the Swiss Solvency Test risk measure". *Insurance: Mathematics and Economics* 42.3, pp. 897–902.

Gordy, Michael B. (Jan. 2000). "A comparative anatomy of credit risk models". *Journal of Banking and Finance* 24.1–2, pp. 119–149.

Gourier, Elise, Walter Farkas and Donato Abbate (Fall 2009). "Operational risk quantification using extreme value theory and copulas: from theory to practice". *Journal of Operational Risk* 4.3, pp. 1–24.

Hanson, Samuel G., Anil K. Kashyap and Jeremy C. Stein (Winter 2011). "A Macroprudential Approach to Financial Regulation". *Journal of Economic Perspectives* 25.1, pp. 3–28.

Härdle, Wolfgang Karl and Ostap Okhrin (Mar. 2010). "De copulis non est disputandum: Colulae: an overview". *AStA Advances in Statistical Analysis* 94.1, pp. 1–31.

References VI

- Hennessy, David A. and Harvey E. Lapan (2006). "On the nature of certainty equivalent functionals". *Journal of Mathematical Economics* 43.1.
- Hull, John (2009). *Options, futures and other derivatives*. 7th. Pearson Prentice Hall.
- Hurst, Simon (Sept. 1995). *The characteristic function of the Student t distribution*. research report FMRR 006-95. Centre for Financial Mathematics, Australian National University.
- Keynes, John Maynard (1921). *A treatise on probability*. London: MacMillan and Co., Ltd.
- Kolanovic, Marko and Rajesh T. Krishnamachari (18 May 2017). *Big Data and AI Strategies*. research report. JP Morgan.
- Kritzman, Mark (July 2011). "Long live quantitative models!" *CFA Institute Magazine*, pp. 8–10.

References VII

Langetieg, Terence C., Martin L. Leibowitz and Stanley Kogelman (Sept. 1990). "Duration Targeting and the Management of Multiperiod Returns". *Financial Analysts Journal* 46.5, pp. 35–45.

Levy, Haim (Apr. 1992). "Stochastic dominance and expected utility: survey and analysis". *Management Science* 38.4, pp. 555–593.

Malkiel, Burton G. (Winter 2003). "The Efficient Market Hypothesis and its Critics". *Journal of Economic Perspectives* 17.1, pp. 59–82.

– (2016). *A random walk down Wall Street: the time-tested strategy for successful investing*. 11th. W.W. Norton & Co.

Markowitz, H. (1952). "Portfolio selection". *Journal of Finance* 7.1, pp. 77–91.

McNeil, Alexander J., Rüdiger Frey and Paul Embrechts (2015). *Quantitative Risk Management: Concepts, Techniques, and Tools*. revised. Princeton Series in Finance. Princeton University Press.

References VIII

Meinhold, Richard J. and Nozer D. Singpurwalla (May 1983). "Understanding the Kalman filter". *The American Statistician* 37.2, pp. 123–127.

Meucci, Attilio (2005). *Risk and Asset Allocation*. Springer Finance. Springer.

– (July 2009). "Review of Discrete and Continuous Processes in Finance: Theory and Applications". mimeo.

– (Mar. 2010). "The Black-Litterman Approach: Original Model and Extensions". mimeo.

– (2017). *ARPM Lab*. URL: https://www.arpm.co/lab (visited on 14/09/2017).

Morini, Massimo (2011). *Understanding and Managing Model Risk: A Practical Guide for Quants, Traders and Validators*. Finance. Wiley. ISBN: 978-0-470-97761-3.

Nelsen, Roger B. (2006). *An Introduction to Copulas*. 2nd. Springer.

References IX

- Rebonato, Riccardo (2007). *Plight of the Fortune Tellers: Why We Need to Manage Financial Risk Differently*. Princeton University Press.
- Resti, Andrea and Andrea Sironi (2007). *Risk Management and Shareholders' Value in Banking*. Wiley Finance. John Wiley & Sons.
- Schroeder, Alice (2009). *The snowball: Warren Buffett and the business of life*. Bloomsbury Publishing PLC.
- Sharpe, William F. (Jan. 1991). "The Arithmetic of Active Management". *Financial Analysts' Journal* 47.1, pp. 7–9.
- Sherif, Nazneen (June 2016). "US model risk rules put lions back in their cages". *Risk*.
- Shin, Hyun Song (2010). *Risk and liquidity*. Clarendon Lectures in Finance. Oxford: Oxford University Press.
- Stefanica, Dan (2011). *A Primer for the Mathematics of Financial Engineering*. 2nd. Financial Engineering Advanced Background Series. New York: FE Press.

References X

Supervisory Guidance on model risk management (Apr. 2011). Bulletin OCC 2011-12. Board of Governors of the Federal Reserve System; and Office of the Comptroller of the Currency.

Taylor, Stephen J., Pradeep K. Yadav and Yuanyuan Zhang (2010). "The information content of implied volatilies and model-free volatility expectations: evidence from options written on individual stocks". *Journal of Banking and Finance* 34, pp. 871–881.

Normally distributed percentage changes

- suppose that $Y \sim N\left(\mu, \sigma^2\right)$ and $X \equiv e^Y \sim LogN\left(\mu, \sigma^2\right)$
- the percentage change in X is then $\frac{X_t - X_{t-1}}{X_{t-1}} \times 100$
- define $Z \equiv \frac{X_t - X_{t-1}}{X_{t-1}}$
- for small Z, Taylor expansion yields

$$\ln(1+Z) = Z - \frac{Z^2}{2} + \cdots \approx Z$$

- thus, for small Z, $e^Z \approx 1 + Z = \frac{X_t}{X_{t-1}}$ so $Z \approx \ln X_t - \ln X_{t-1}$
- are we done?

◂ LogN

Multivariate Student's t is dependent

- let $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with $\boldsymbol{\Sigma}$ diagonal
  - the individual components of $\boldsymbol{X}$ are statistically independent
  - the diagonal $\boldsymbol{\Sigma}$ ensures that the major axes of the distribution align with the coordinate axes
  - each diagonal entry assigns a variance to its X_i, with no effect on the X_j
- but, if $\boldsymbol{Y} \sim St(\nu, \boldsymbol{\mu}, \boldsymbol{\Sigma})$, then

$$\boldsymbol{Y} - \boldsymbol{\mu} \stackrel{d}{=} \boldsymbol{X} / \sqrt{Z/\nu};$$

  where $Z \sim \chi^2_\nu$ and is independently distributed from X
  - why does the same intuition for a diagonal $\boldsymbol{\Sigma}$ not hold?
  - now each component is now divided by a common stochastic factor
  - were Z deterministic, learning X_1 would not inform about X_2
  - as Z is stochastic, learning X_1 does inform about X_2
- why divide by a common, stochastic Z? A: physical motivation — corrects for sample mean's variance

◂ $\sim St$

Probability integral transform

Proof.

Let $U \equiv F_X(X)$ where F_X is an invertible CDF; thus, U is a rv. By definition

$$
\begin{aligned}
F_U(u) &\equiv \mathbb{P}\{U \leq u\} \\
&= \mathbb{P}\{F_X(X) \leq u\} \\
&= \mathbb{P}\{X \leq Q_X(u)\} \, \forall u \in [0,1] \\
&= F_X(Q_X(u)) \, \forall u \in [0,1] \\
&= u \forall u \in [0,1];
\end{aligned}
$$

where the restriction to $u \in [0,1]$ arises from the domain of Q_X. As a CDF is non-decreasing, and $F_U(0) = 0$ and $F_U(1) = 1$, it follows that $U \sim U([0,1])$. □

◂ prob. trans. How would you correct the naïve statement that $F_X(X) = \mathbb{P}\{X \leq X\}$? ▸ hint

The following proof is taken from Meucci (2005, www.3.2):

Proof.

By definition of the cf, the cf of $\boldsymbol{X}_{T+\tau,\tilde{\tau}} + \boldsymbol{X}_{T+\tau-\tilde{\tau},\tilde{\tau}} + \cdots + \boldsymbol{X}_{T+\tilde{\tau},\tilde{\tau}}$ is

$$\begin{aligned}
\phi_{\boldsymbol{X}_{T+\tau,\tilde{\tau}}+\boldsymbol{X}_{T+\tau-\tilde{\tau},\tilde{\tau}}+\cdots+\boldsymbol{X}_{T+\tilde{\tau},\tilde{\tau}}}(\boldsymbol{\omega}) &= E\left\{e^{i\boldsymbol{\omega}'\left(\boldsymbol{X}_{T+\tau,\tilde{\tau}}+\boldsymbol{X}_{T+\tau-\tilde{\tau},\tilde{\tau}}+\cdots+\boldsymbol{X}_{T+\tilde{\tau},\tilde{\tau}}\right)}\right\} \\
&= E\left\{e^{i\boldsymbol{\omega}'\boldsymbol{X}_{T+\tau,\tilde{\tau}}} \times \cdots \times e^{i\boldsymbol{\omega}'\boldsymbol{X}_{T+\tilde{\tau},\tilde{\tau}}}\right\} \\
&= E\left\{e^{i\boldsymbol{\omega}'\boldsymbol{X}_{T+\tau,\tilde{\tau}}}\right\} \times \cdots \times E\left\{e^{i\boldsymbol{\omega}'\boldsymbol{X}_{T+\tilde{\tau},\tilde{\tau}}}\right\} \\
&= \phi_{\boldsymbol{X}_{T+\tau,\tilde{\tau}}}(\boldsymbol{\omega}) \times \cdots \times \phi_{\boldsymbol{X}_{T+\tilde{\tau},\tilde{\tau}}}(\boldsymbol{\omega}) \\
&= \left(\phi_{\boldsymbol{X}_{t,\tilde{\tau}}}(\boldsymbol{\omega})\right)^{\frac{\tau}{\tilde{\tau}}}
\end{aligned}$$

where the antepenultimate equality comes from independence, and the ultimate from identicality. □

ATMF liquidity

- futures
  - standardised, exchange traded contracts
  - settled by delivery
- forwards
  - customisable, OTC contracts
  - quoted on Pink Quote, OTCBB
  - can be settled in cash (Stefanica, 2011, §1.10)
- Bank for International Settlements: OTC market much larger than exchange-traded (Hull, 2009, p.3)
- why are ATM options most liquid?
  - most data are ATM; anything away from that relies on assumptions
  - focal point

imp vol

Why set $E\left\{\boldsymbol{U}\right\} = \mathbf{0}$?

Lemma

Choice of $\boldsymbol{m}_p$ in the R^2 maximisation problem leads to $E\left\{\boldsymbol{U}_p\right\} = \mathbf{0}$.

Proof.

By definition,

$$E\left\{\boldsymbol{U}_p\right\} = \left(\boldsymbol{I} - \boldsymbol{B}_p\boldsymbol{A}_p'\right) E\left\{\boldsymbol{X}_{t,\tilde{\tau}}\right\} - E\left\{\boldsymbol{m}_p\right\}.$$

Thus, $\boldsymbol{m}_p$ can be freely chosen to produce any desired value for $E\left\{\boldsymbol{U}_p\right\}$. For simplicity, work with the univariate U, and decompose it into $U = V + k$ for some constant k and rv V such that $E\left\{V\right\} = 0$. We seek to minimise

$$E\left\{U^2\right\} = \int_{-\infty}^{\infty} u^2 f_U\left(u\right) du = \int_{-\infty}^{\infty} v^2 f_U\left(v+k\right) dv + 2kE\left\{V\right\} + k^2$$

which, as k is irrelevant to the integral's value, is achieved by $k = 0$. □

Lemma

$E\left\{\Psi_{\alpha}\right\}$ can be written as the unweighted average of the quantiles

$$E\left\{\Psi_{\alpha}\right\} = \int_{-\infty}^{\infty} \psi f_{\Psi_{\alpha}}\left(\psi\right) d\psi = \int_{0}^{1} Q_{\Psi_{\alpha}}\left(p\right) dp$$

Proof.

For any continuous g, g' and h, integration by substitution allows

$$\int_{a}^{b} h\left(g\left(\psi\right)\right) g'\left(\psi\right) d\psi = \int_{g(a)}^{g(b)} h\left(p\right) dp$$

Thus, if $a = -\infty, b = \infty, g\left(\cdot\right) \equiv F_{\Psi_{\alpha}}\left(\cdot\right)$, and $h\left(\cdot\right) \equiv Q_{\Psi_{\alpha}}\left(\cdot\right)$:

$$\int_{-\infty}^{\infty} Q_{\Psi_{\alpha}}\left(F_{\Psi_{\alpha}}\left(\psi\right)\right) f_{\Psi_{\alpha}}\left(\psi\right) d\psi = \int_{0}^{1} Q_{\Psi_{\alpha}}\left(p\right) dp$$

which, as $Q_{\Psi_{\alpha}}$ and $F_{\Psi_{\alpha}}$ are mutual inverses, establishes the result. □

McNeil, Frey and Embrechts (2015, Example 2.25)

1. fully concentrated portfolio
   - portfolio consists of a single debt instrument with 1% chance of default
   - distribution of Ψ_α is discrete: .99 weight on zero; .01 weight on full loss
   - $VaR_{95} = 0$ as default occurs within the tail
2. diversified portfolio
   - portfolio consists of 100 independent debt instruments, each with 1% chance of default
   - binomial distribution: probability of k non-defaults from n trials, each with success probability p

$$\mathbb{P}\{X = k\} = \binom{n}{k} p^k (1-p)^{n-k}$$

   - thus, the probability of no defaults is only

$$1 - \mathbb{P}\{X = n\} = 1 - \binom{n}{n} p^n (1-p) = 1 - .99^{100} \approx .634$$

   - thus, defaults occur well before the VaR threshold, so that $VaR_{95} > 0$

The Lorentz cone: $\mathbb{K}^M$

Theorem

$\mathbb{K}^M$ *is a cone.*

incomplete.

1. closed under positive multiplication:
2. closed under addition: by the triangular inequality for Euclidean norms

$$\|\boldsymbol{x}_{-M} + \boldsymbol{y}_{-M}\| \leq \|\boldsymbol{x}_{-M}\| + \|\boldsymbol{y}_{-M}\|;$$

   where $-M$ denotes all dimensions except for the M^{th}.
   By definition of the cone, $\|\boldsymbol{x}_{-M}\| \leq x_M$ and $\|\boldsymbol{y}_{-M}\| \leq y_M$, so that their sum is less than or equal to $x_M + y_M$.
3. pointed:

□

◂ Lorentz cone

The semidefinite cone: $\mathbb{S}_{+}^{M} \equiv \{\boldsymbol{S} \succcurlyeq \mathbf{0}\}$

Theorem

$\mathbb{S}_{+}^{M}$ *is a cone.*

Proof.

Recall that, for PSD matrices, $\boldsymbol{S}$: $\operatorname{tr}(\boldsymbol{S}) \geq 0$, $\boldsymbol{z}'\boldsymbol{S}\boldsymbol{z} \geq 0$ for all $\boldsymbol{z} \neq \mathbf{0}$, and all principal minors of $\boldsymbol{S}$ are non-negative.

1. if $|\boldsymbol{S}_m|$ is any $m \times m$ principal minor of $\boldsymbol{S}$, then the corresponding principal minor of $\lambda\boldsymbol{S}$ is $\lambda^m |\boldsymbol{S}_m|$. As $\lambda \geq 0$, these share signs: $\boldsymbol{S}$ and $\lambda\boldsymbol{S}$ thus share sign definiteness.
2. given PSD matrices, $\boldsymbol{S}$ and $\tilde{\boldsymbol{S}}$: $\boldsymbol{z}'\left(\boldsymbol{S} + \tilde{\boldsymbol{S}}\right)\boldsymbol{z} = \boldsymbol{z}'\boldsymbol{S}\boldsymbol{z} + \boldsymbol{z}'\tilde{\boldsymbol{S}}\boldsymbol{z} \geq 0$.
3. $\operatorname{tr}(-\boldsymbol{S}) = -\operatorname{tr}(\boldsymbol{S}) \leq 0$.

□

◂ SD cone