STAD70 Course Work

Richard Ye

Contents

Chapter 1

Introduction

Work from the STAD70 course taken Winter 2023.

Chapter 2

Financial Data and Returns

https://richardye101.github.io/STAD70/

- Prices of financial instruments (stocks, bonds, futures, options, etc)
-
- LIBOR: London Inter-Bank Offered Rate
  - Avg interest rate that major London Banks would charge when borrowing from each other
- FX Rates: Decentralized market that sets currency prices

There are two types of *raw (tick)* data:

- **Quote data**
  - Record of bid/ask prices from the order book, usually just the top N lines of the order book on the bid and ask sides
- **Trade data**
  - Trade records (filled orders)

Data that is identified using time:

- **Intraday Data**
  - Data of the most current bids/asks, last price, volume etc
  - Describes the data of a stock within a given day
- **Daily data**
  - For longer term analysis

2.1 Trading

There are two types:

- **OTC (Over the counter)**

 - Negotiated and traded directly between parties, not for the public to bid on

- **Through an Exchange**
  - The TSX, NYSE, LSE (London)
  - Uses an **auction** system where there are many potential buyers and sellers
  - Uses a **continuous double auction (CDA)**, where the bids and asks are matched in real-time to determine which trades to execute.
    * A matching order (bid $\geq$ ask) is executed right away
    * Outstanding orders are maintained in an **order book**

2.2 Order Types

- Limit order: buy/sell are no more/les than a specified price
  - It is maintained in the order book if not filled immediately
- Market order: buy/sell at the current market price immediately
  - No control over the price at which the order will execute
- Iceberg order: Only a portion of the total (giant) order is displayed in the order book, and when it is filled, a new portion is shown
  - Used to maintain anonymity

2.3 Returns

- The ratio of money gained/lost on an investment relative to the invested amount
- It is defined relative tot he holding period (daily, monthly, annual)

2.3.1 Net Returns

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} = \frac{P_t}{P_{t-1}} - 1$$

Where P_t is the price at time T.

Gross return is just $\frac{P_t}{P_{t-1}} = 1 + R_t$

2.3.2 Log Returns

$$\begin{aligned} r_t &= \log(1 + R_t) \\ &= \log\left(\frac{P_t}{P_{t-1}}\right) \\ &= \log(P_t) - \log(P_{t-1}) \end{aligned}$$

You can obtain the log return from the net return (for small returns, $< 1\%$) using the Taylor Approximation:

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2}f''(x_0)(x - x_0)^2 + ...$$

$$\begin{aligned}
r_t &= \log(1 + R_t) \\
&\approx \log(x_0) + \log'(x)|_{x=x_0} \cdot (\underbrace{1 + R_t}_{x} - x_0) + ... \\
&\text{Expand around } x_0 = 1 \\
&= 0 + \frac{1}{1} \cdot R_t \\
&\approx R_t
\end{aligned}$$

We like to work with log returns because they are also easy to aggregate!

$$\begin{aligned}
R_{1-22} &= (1 + R_1) \cdot (1 + R_2) \cdot \dots \cdot (1 + R_{22}) - 1 \\
r_{1-22} &= r_1 + r_2 + \dots + r_{22} \\
&= \log\left(\frac{P_1}{P_0} \cdot \frac{P_2}{P_1} \cdot \dots \cdot \frac{P_{22}}{P_{21}}\right) \\
&= \log\left(\frac{P_{22}}{P_0}\right)
\end{aligned}$$

2.3.3 Returns accounting for dividends

If a dividend was paid just before time t, and after $t - 1$, then we would have:

$$R_t = \frac{P_t - D_t}{P_{t-1}} - 1$$

$$r_t = \log(P_t + D_t) - \log(P_{t-1})$$

2.3.4 Returns accounting for splits

$$\begin{aligned}
R_t &= \frac{P_t}{P_{t-1}/s} - 1 \\
r_t &= \log(P_t) - \log(P_{t-1}/s) \\
&\text{Where } s = \text{number of shares recieved per 1 share owned}
\end{aligned}$$

2.3.5 Adjusted returns

These returns have already accounted for dividends and splits, so returns should always be calculated on **adjusted returns. They should not be used as prices (ie the adjusted close)!**

2.3.6 Random Walk Model

We say the additive log returns

$$\log\left(\frac{P_t}{P_0}\right) = r_1 + r_2 + \cdots + r_t$$

Follow a random walk model if:

$$r_t \sim (\mu, \sigma^2)$$

$$\begin{aligned} \text{Where} \quad \mathbb{E}[r_i] &= \mu \text{ is drift} \\ \sqrt{Var[r_i]} &= \sigma \text{ is volatility} \end{aligned}$$

However they follow a Normal random walk if:

$$r_t \sim N(\mu, \sigma^2)$$

Which implies that log-returns are Normally distributed.

The random walk from $1-n$ has:

$$r_{1-n} = r_1 + r_2 + \cdots + r_n$$

$$\begin{aligned} \text{Mean: } & \mathbb{E}(r_{1-n}) = n\mu \\ \text{Variance: } & \sqrt{Var[r_{1-n}]} = \sqrt{n}\sigma \end{aligned}$$

Which can be converted to the asset price simply using

$$P_t = P_0 \cdot e^{r_1 + r_2 + \cdots + r_t}$$

Which is referred to as the Exponential/Geometric Random walk

Random walks are not a good description of reality, but they are useful for modelling.

Chapter 3

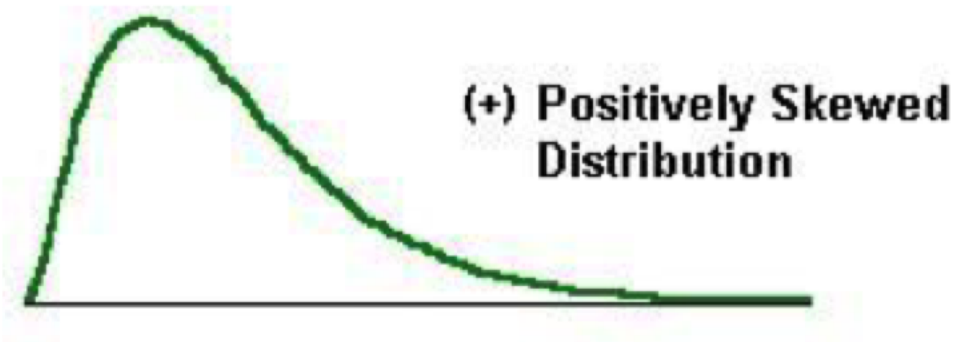
Return Distributions

3.1 Skewness

Skewness measures symmetry.

- Positive skew $\rightarrow$ right skewed
- Negative skew $\rightarrow$ left skewed

$$Sk = \mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^3\right]$$

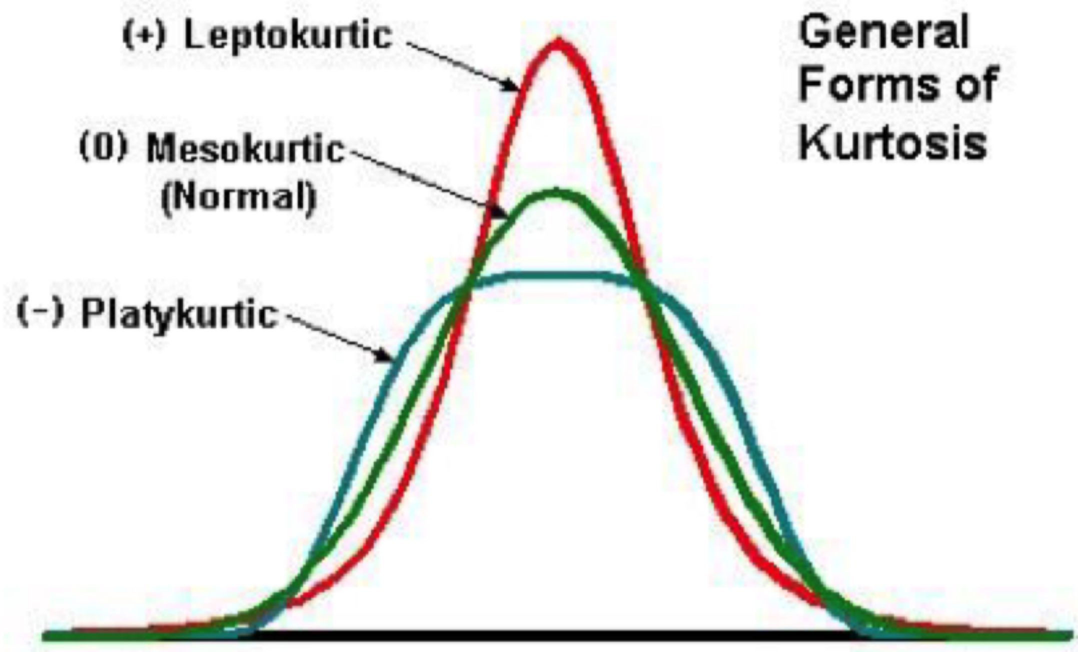


3.2 Kurtosis

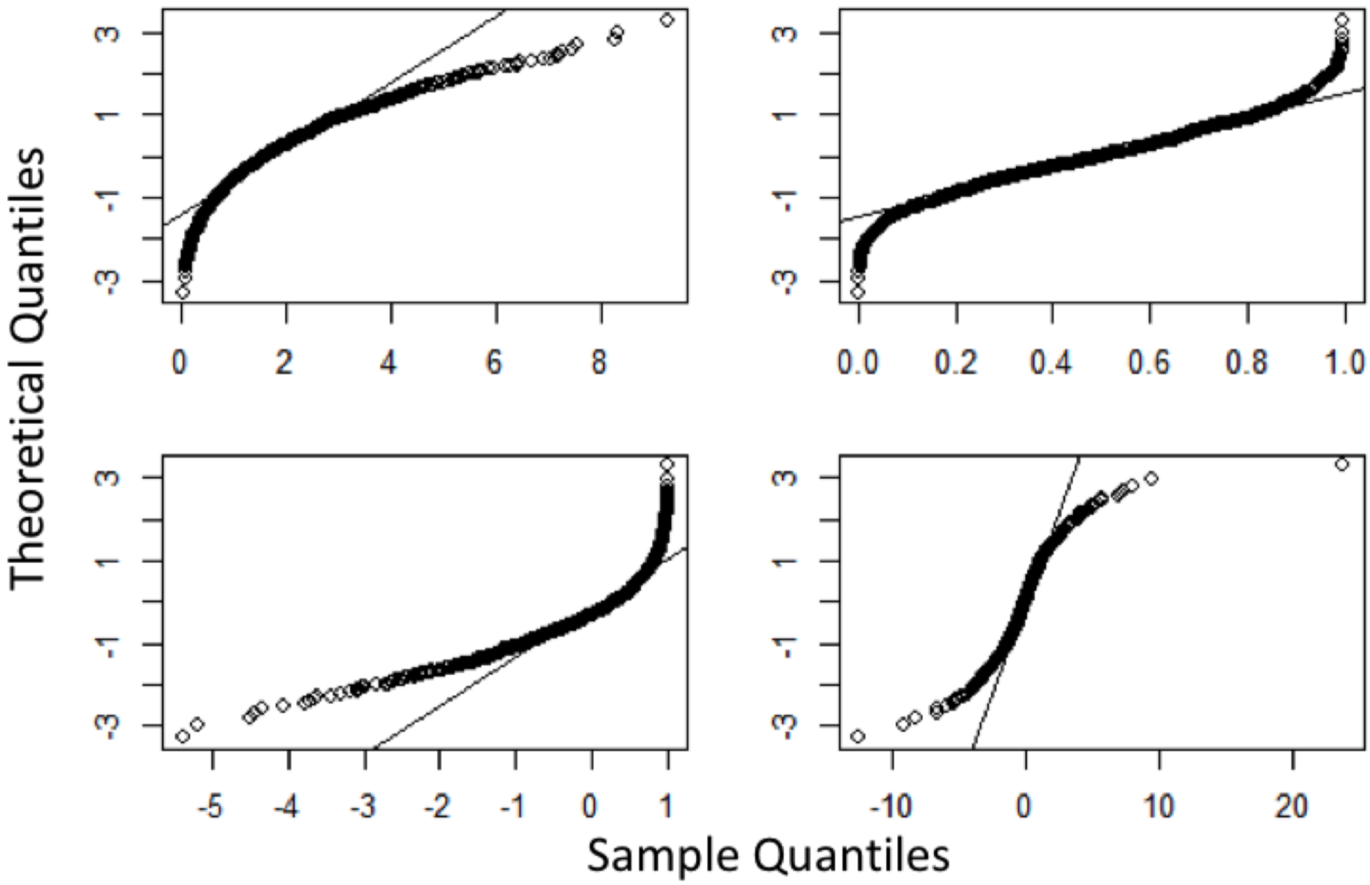
Kurtosis measures how concentrated the data is around the mean, or how heavy the tails are. You can only measure kurtosis (meaningfully) if the distribution has 0 skew.

Kurtosis of a distribution is measured against the kurtosis of the Normal distribution, which is 3.

$$Kur = \mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] - 3$$


(+) Leptokurtic
General
Forms of
Kurtosis
(0) Mesokurtic
(Normal)
(-) Platykurtic

3.3 QQ Plot



- Top left: The sample quantiles spread further on the right than the theoretical quantiles, which implies a **right/pos skewness**
- Top right: The sample quantiles for the tails are very concentrated compared to the theoretical quantiles, which implies the data has short/finite tails implying it is **platykurtic/neg kurtosis**
- Bottom left: The sample quantiles spread further on the left than the theoretical quantiles, which implies a **left/neg skewness**
- Bottom right: The sample quantiles for the tails are very spread compared to the theoretical quantiles, which implies the data has long/heavy tails implying it is **leptokurtic/pos kurtosis**

3.4 Heavy Tail Distributions

A distribution $f(x)$ is said to have:

- Exponential tails (short/finite tails) if

$$f(x) \propto e^{-x/\lambda}$$

- Polynomial tails (long/heavy tails) if

$$f(x) \propto x^{-(1+\alpha)}$$

Where α is the tail index

The smaller the tail index, the heavier the tail

Heavy tailed distributions can also have infinite moments (including the mean!)

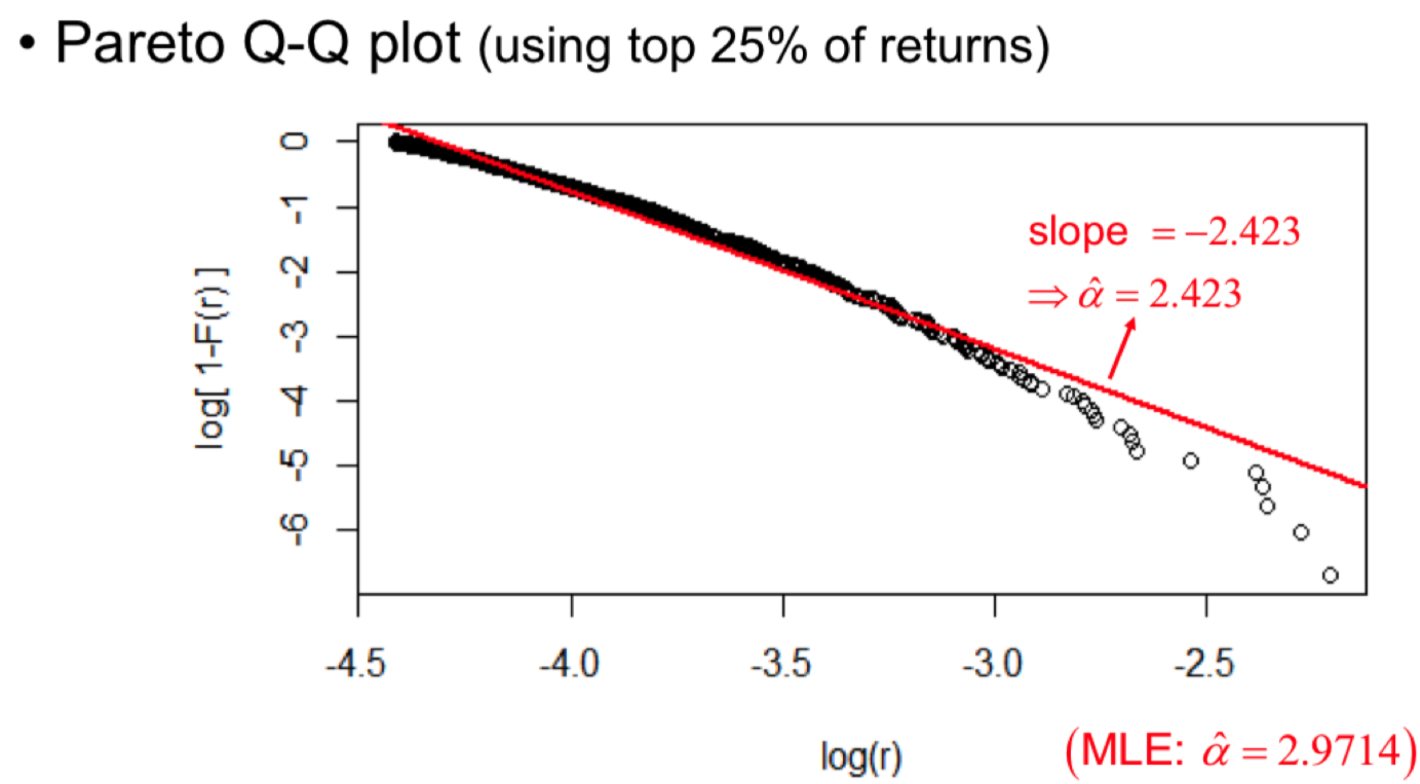
$$\mathbb{E}(X^k) = \infty \text{ for } k \geq \alpha$$

So if the tail index is high, the tail is lighter, and the first α moments exist.

The tail index can be estimated using: - MLE approximation, which takes the derivative log of the product of n the distributions when it equals 0, and solving for α - $\hat{\alpha} = \frac{n}{\sum_{i=1}^{n} \ln(r_i/r_{min})}$

$$L(\alpha) = \prod_{i=1}^{n}[f(x)]$$
$$\log(L(\alpha)) = l(\alpha) = \sum_{i=1}^{n} f(x)$$
$$l'(\alpha) = \sum_{i=1}^{n} f'(x) = 0$$

- Pareto Q-Q plot of the empirical CDF and the returns in log x log scale - α is the estimated slope of the line of best fit



3.4.1 Common Heavy Tail Distributions

3.4.1.1 Pareto

$$f(x) = \frac{a\ x^{-(1+\alpha)}}{l^{-\alpha}},\ x \geq l$$

Can be used to model **absolute** returns above a cutoff r_{min}

$$\bar{F}(r) = \left(\frac{r_{min}}{r}\right)^{\alpha} \quad \forall\ r > r_{min}$$

3.4.1.2 Standard Cauchy

$$\begin{aligned} f(x) &= \frac{1}{\pi(1+x^2)} \\ &= \frac{1 + x^{-(\alpha+1)}}{\pi} \implies \alpha = 1 \end{aligned}$$

Which is the t-distribution when df = 1.

$$f(x) \sim t(df = 1)$$

3.4.1.3 Students t

$$f(x) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi}\Gamma(v/2)} \left(1 + \frac{x^2}{v}\right)^{-\frac{v+1}{2}}$$

$$\text{Where } \alpha = v = \text{degrees of freedom}$$

The Students t distributions offers a tractable heavy-tail model of the entire return distribution (not just the tail). It is typically adjusted for the location and scale:

$$Y = \mu + \sigma X \quad \text{where } X \sim t(df = v)$$

3.4.2 Stable Distributions

If we let log returns $r_i \sim$ heavy tail distribution with $0 < \alpha < 2$ (the first and second moment - mean/variance exist)

The aggregate return $r_{1 \to n} = r_1 = r_2 + ... r_n \sim$ Stable Distribution

3.4.2.1 Generalized Central Limit Theorem

Stable distributions have no closed form expression, although the share heavy tails and the same tail index α as it's individual distributions.

As you sum independent, **STABLE** random variables, the sum will follow a Stable Distribution.

3.4.2.2 Measuring tail behaviour

We use the complementary CDF

$$\bar{F}(x) = 1 - F(x) = P(X > x) \sim x^{-\alpha}$$

3.5 Mixture Distributions

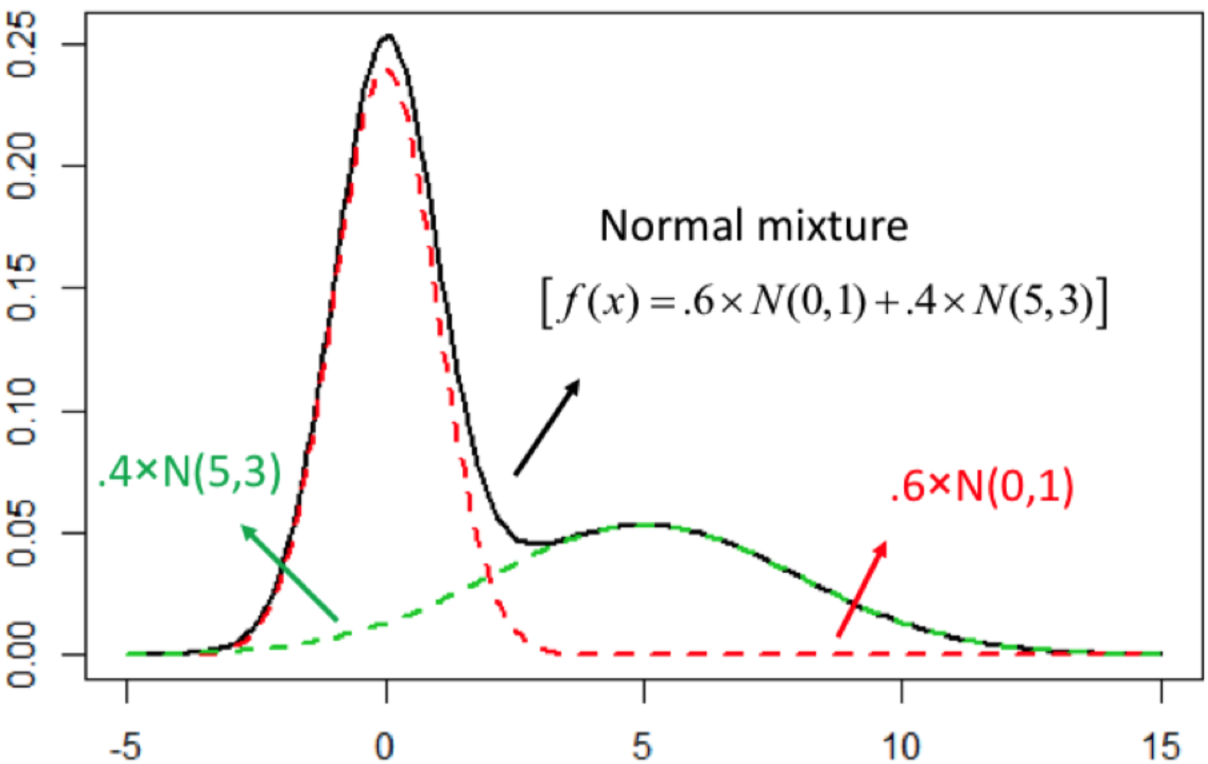
We can generate a random variable using one of out a selection of a family of distributions, choosing the distribution using another distribution.

They are easy to generate, but harder to work with analytically.

We can select from a discrete and finite set of distributions, or of a continuous family of distributions (possibly countable) known as compound distributions.

Example:

$$\text{Generate RV from: } \begin{cases} N(0,1) & p = 60\% \\ N(5,3) & p = 40\% \end{cases}$$



3.5.0.1 Normal scale mixture

$Y = \mu + \sqrt{V} \cdot Z$ Where V is a RV with non-negative mixing distribution and represents a random sd of Y.

Example (probably don't need to memorize...) which is used when simulating returns, as seem in PS2 Q4c part ii and iii.

$$\text{t-dist } t = Z\sqrt{v/W} \quad \text{where } W \sim \chi^2(df = v)$$

$$\text{GARCH model } r_t = \mu + \sigma_t Z_t \quad \text{where } \sigma_t^2 = \omega + \sum_{i=1}^{p} a_i r_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2$$

Chapter 4

Modelling Extreme Events

4.1 Stylized Facts

Characteristics of typical empirical asset returns:

1. Absence of simple autocorrelations
    - Returns are not correlated with past time-steps of itself
2. Volatility clustering
    - Large amounts of volatility often occur in clusters through time
3. Heavy tails
    - Returns often have very abnormal (large) values suggesting heavy tail distributions
4. Intermittency
5. Aggregation changes distribution
6. Gain/loss asymmetry
    - The market tends to go up over time

4.2 Extreme Value Theory

This theory helps with modelling extreme events that have small probabilities of occurring.

4.2.1 Two main results

- Maxima of i.i.d. sequences
- Values exceeding threshold

4.2.2 1st Theorem: Fisher-Tippet-Gnedenko

Theres no need to prove any results

If $X_1, X_2, ...$ are i.i.d. RVs, we can create a RV that simply takes the maximum of n RVs called $M_n = max(X_1, X_2, ... X_n)$.

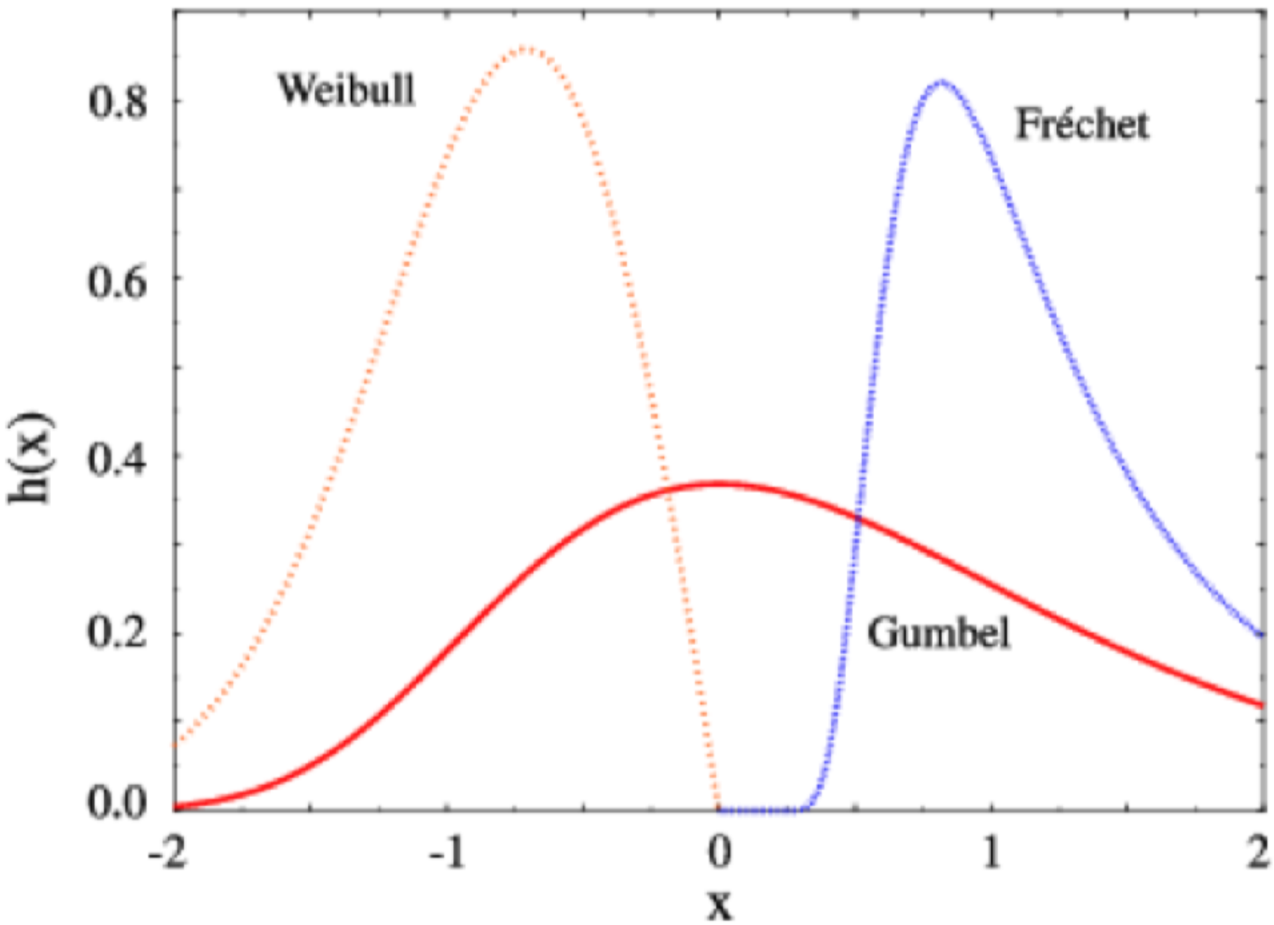
In certain cases, we can find normalizing constants $a_n > 0, b_n$ such that they can be transformed into one of the three distributions (identified by $H(x)$) which can be much easier to work with.

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) = [F(a_n x + b_n)]^n \to H(x)$$
$$F(a_n x + b_n) = \text{Single RV}$$

This is just saying that the normalized max RV is less than some value x is just the probability of each individual RV being less than the transformed value of x.

$$H(x) = \begin{cases} Gumbel & \exp\{-e^{-x}\} \quad\quad x \in \mathbb{R} \\ Frechet & \begin{cases} 0 & x < 0 \\ \exp\{-x^{-\alpha}\} & x > 0 \end{cases} \\ Weibull & \begin{cases} \exp\{-|x|^{\alpha}\} & x < 0 \\ 1 & x > 0 \end{cases} \end{cases}$$

Where $\alpha > 0$ for the Frechet and Weibull distributions.



4.2.3 Generalized Extreme Value (GEV) Distribution

The three types of distributions can be represented using:

$$H(x) = \exp\left\{-\left(1+\xi\frac{x-\mu}{\sigma}\right)_+^{-1/\xi}\right\}$$

$$\text{Where } \mu = \text{location}$$
$$\sigma = \text{scale}$$
$$\xi = \text{shape parameters}\begin{cases}\xi > 0 & \text{heavy tails (Frechet)}\\ \xi = 0 & \text{exponential tails (Gumbel)}\\ \xi < 0 & \text{short/light tails (Weibull)}\end{cases}$$

Where the ξ value describes the tail behaviour

The log-transformation of the Frechet($\alpha = 1$) is the Gumbel distribution!

$$\begin{aligned}
\ln(H_F(x)) &= \ln\left[\exp\{-x^{-1}\}\right] \\
&= -\Big(1 + \frac{x-\mu}{\sigma}\Big)^{-1} \\
&= H_G = \exp\left\{-\Big(1 + 0 \cdot \frac{x-\mu}{\sigma}\Big)^{-1/0}\right\}
\end{aligned}$$

As an example, we'll show that the normalized maximum of i.i.d. Uniform(0,1) aka $F_n(x) = x$ with $a_n = 1/n, b_n = 1$ converges to the Weibull distribution.

$$\begin{aligned}
P\Big(\frac{M_n - b_n}{a_n} \le x\Big) &= P\Big(\frac{M_n - 1}{1/n} \le x\Big) \\
&b_n \text{ shifts the Uniform from (0,1) to (-1,0)} \\
&= P\Big(M_n \le \frac{x}{n} + 1\Big) \\
&= P\Big(\max(U_1, U_2, \ldots U_n) \le 1 + \frac{x}{n}\Big) \\
&= \prod_{i=1}^{n} \underbrace{P\Big(U_i \le 1 + \frac{x}{n}\Big)}_{(1+\frac{x}{n})} \\
&= \Big(1 + \frac{x}{n}\Big)^n \\
&\text{because } x < 0, \quad 1 + (-.4) = 1 - |-.4| \\
&= \Big(1 - \frac{|x|}{n}\Big)^n
\end{aligned}$$

$$\begin{aligned}
\lim_{n\to\infty} \Big(1 - \frac{|x|}{n}\Big)^n &\to \exp\{((-1)|x|)^{-1/-1}\} \\
&= \exp\{-|x|^1\} \implies Weibull(\alpha = 1)
\end{aligned}$$

4.2.4 2nd Theorem: Pickands-Balkema-De Haan

For any RV X with CDF $F(\cdot)$, it's conditional distribution when exceeding a certain threshold u is:

$$F_u(y) = \frac{F(u+y) - F(u)}{1 - F(u)}, \quad 0 \le y \le x_F - u$$

Where $x_F = sup\{x \in \mathbb{R} : F(x) < 1\}$ is the right endpoint of F

x_F could be finite, or ∞. As $u \to x_F$, the conditional distribution converges to something belonging to the **Generalized Pareto Distribution (GPD).**

$$F_u(y) \to G_{\xi,\sigma}(y)$$

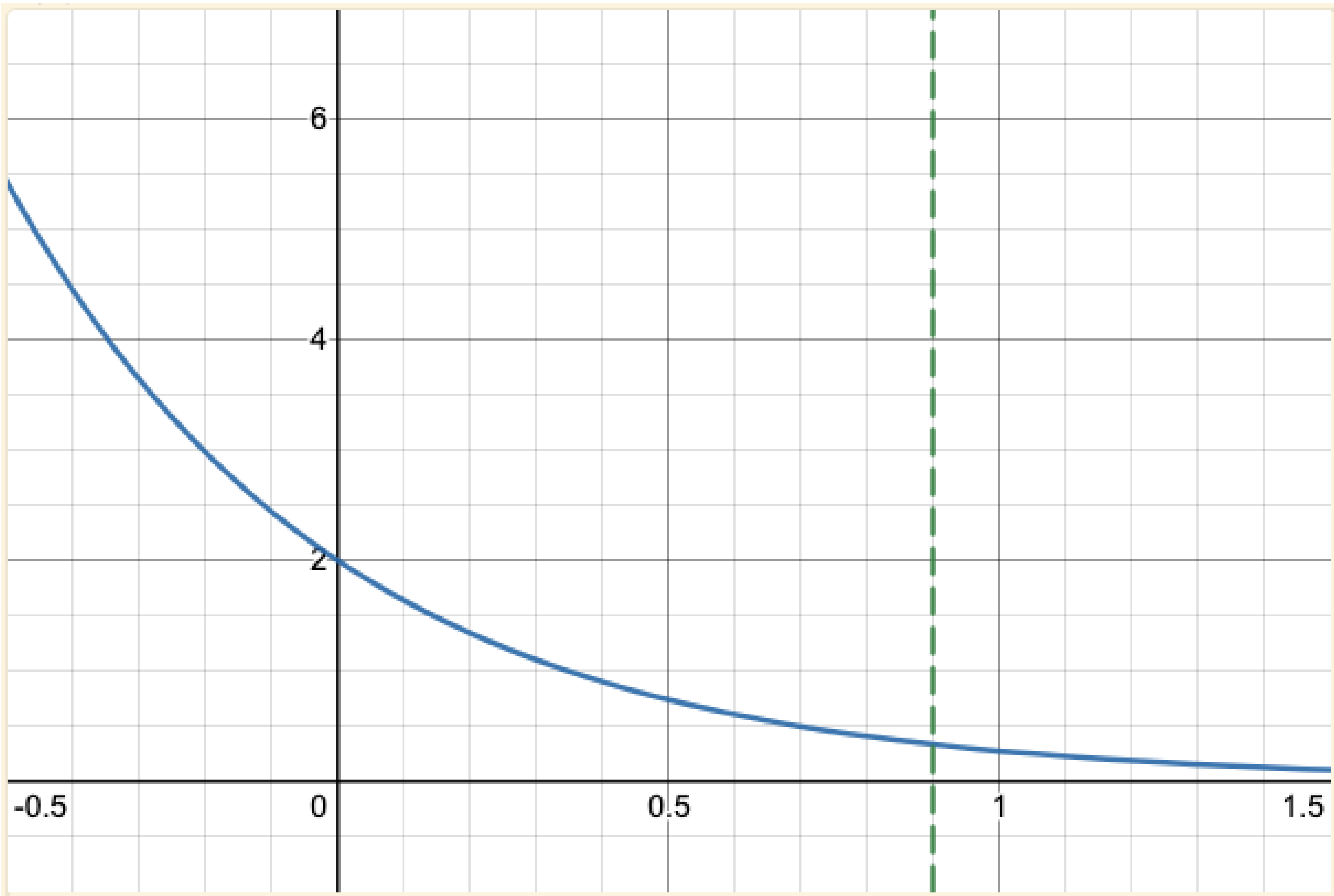
The **Generalized Pareto Distribution (GPD)** is given by:

$$G_{\xi,\sigma}(y) = 1 - \Big(1 + \xi\frac{y}{\sigma}\Big)_+^{-1/\xi} = \begin{cases} G_{\xi,\sigma}(y) = 1 - (1 + \xi\frac{y}{\sigma})^{-1/\xi} & \xi \ne 0 \\ G_{\xi,\sigma}(y) = 1 - \exp\{-\frac{y}{\sigma}\} & \xi = 0 \end{cases}$$

Where $\sigma > 0, y \ge 0, \quad y \le -\dfrac{\sigma}{\xi}$ when $\xi < 0$

$f(x) = \lambda e^{-\lambda x}, \lambda > 0$

```
bottom=-1; left=-0.5; right=1.5;
---
y=2 \exp(-2x)
x=.9|dashed
```



The exponential distribution is used to model waiting times, no matter how much time has passed. The conditional distribution

Chapter 5

Multivariate Return Modelling

When modelling returns of multiple assets, it's common to assume the correlation between them is constant.

Many investment strategies combine multiple assets together, but how those assets are related is an important piece of information to understand. It's also difficult to model. Say we start with a simple multivariate Normal, where dependence $\leftrightarrow$ covariance, so we can model their relations with the covariance matrix.

5.1 Covariance & Correlation Matrix

For a linear combination of assets, we can attempt to use the following covariance matrix to describe their linear dependence:

$$Cov[A^T R] = A^T Cov[R] A, \quad \text{where} \begin{cases} A = \text{Constant matrix} \\ R = \text{random vector} \end{cases}$$

Unfortunately, sample covariance estimation is very sensitive to extreme values, which happen frequently with return (See heavy tails)

Returns are often treated as independent samples, but that is not realistic.

5.2 Multivariate Student's t Distribution

We should **never remove outliers** in finance. We must model the heavy tails (extreme returns) somehow, and Student's t distribution is one way to do that.

A multivariate Normal scale mixture model can be used to describe a multivariate Student's t distribution:

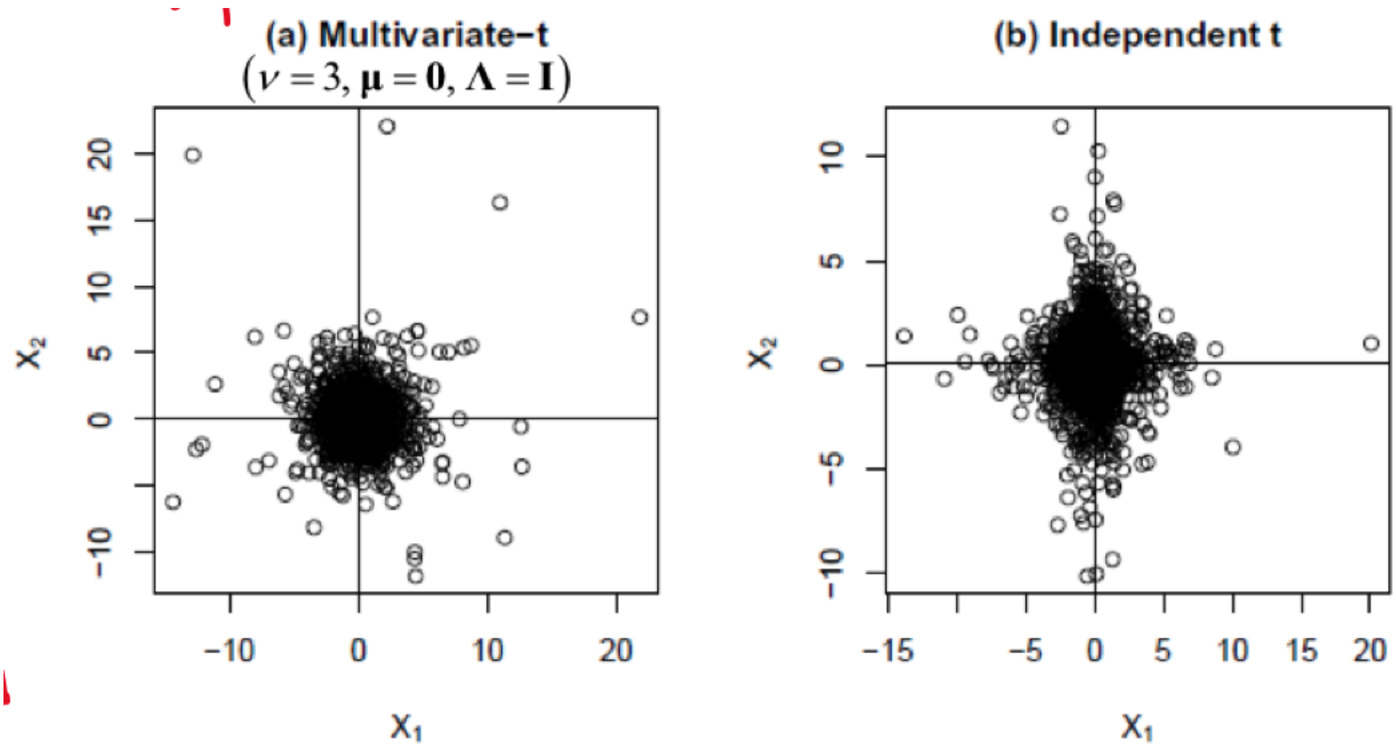
$$\mathbf{R} = \mu + \mathbf{Z}\sqrt{v/W} \sim t_v(\mu, \Lambda)$$

$$\begin{aligned} \text{Where } W &\sim \chi^2(df = v) \\ \mathbf{Z} &\sim N(\mathbf{0}, \Lambda) \quad \text{where } \Lambda = \mathit{Indentity\ Mat} \end{aligned}$$

It's mean and variance are:

$$\begin{aligned}
\mathbb{E}(\mathbf{R}) &= \mathbb{E}\left(\mu + \mathbf{Z}\sqrt{\frac{v}{W}}\right) \\
&= \mu + \underbrace{\mathbb{E}(\mathbf{Z})}_{=0} \cdot \mathbb{E}\left(\sqrt{\frac{v}{W}}\right) \\
&= \mu
\end{aligned}$$

$$\begin{aligned}
Cov(\mathbf{R}) &= Cov\left(\mu + \mathbf{Z}\sqrt{\frac{v}{W}}\right) \\
&= Cov(\mathbf{Z}) \cdot Cov\left(\sqrt{\frac{v}{W}}\right) \\
&= \Lambda \cdot \frac{v}{v-2} \quad \text{where } v > 2
\end{aligned}$$



From these plots, we see that with the Multivariate t distribution, the extreme values between the two RVs are more evenly spread (the round shape of the points) and imply extreme values for both assets happen together. This is because they both share the same Λ.

This shows that the multivariate t distribution is more practical and realistic than using the Normal to model multiple assets. We can linearly combine multivariate t-distributions with the same degrees of freedom:

$$\mathbf{Y} \sim t_v(\mu, \Lambda) \implies \mathbf{w}^T\mathbf{Y} \sim t_v(\mathbf{w}^T\mu, \mathbf{w}^T\Lambda\mathbf{w})$$

$$\begin{aligned}
\mathbb{E}[\mathbf{w}^T\mathbf{Y}] &= \mathbf{w}^T\mathbb{E}(\mathbf{Y}) \\
Var[\mathbf{w}^T\mathbf{Y}] &= \mathbf{w}^T Var\left(\mu + \mathbf{Z}\sqrt{\frac{v}{W}}\right)\mathbf{w} \\
&= \mathbf{w}^T Var(\mathbf{Z}) Var\left(\sqrt{\frac{v}{W}}\right)\mathbf{w} \\
&= \frac{v}{v-2}(\mathbf{w}^T\Lambda\mathbf{w})
\end{aligned}$$

However, as seen above, the multivariate t distribution is **restrictive** because all the marginal distributions share the same degrees of freedom, as they all depend on the variance of $W \sim \chi^2$.

Chapter 6

Copulas

A more flexible way of modelling dependencies of RVs is using copulas.

Definition: A copula (C) is a multivariate CDF with Uniform(0,1) marginals

$$C(u_1, u_2, \dots u_d) = P(u_1, u_2, \dots u_d) \in [0,1], \quad \forall\ u_1, \dots u_d \in [0,1]$$

$$\text{Where } \begin{cases} C(0,0,\dots,0) = 0 \\ C(1,1,\dots,1) = 1 \\ C(\dots, u_{i-1}, 0, u_{i+1}, \dots) = 0 \\ C(1,\dots,1,u_i,1,\dots,1) = u_i \end{cases}$$

Third case: The cumulative probability of one RV being less than or equal to 0 regardless of what the other RVs are in a copula is 0. Fourth case: The cumulative probability that the other RVs have values less than 1 is 1 but the i^{th} RV is u_i is simply u_i

6.1 The independence copula

Definition: $C_{indep}(u_1, \dots u_d) = u_1 \times \cdots \times u_d$

By the **Frechet-Hoeffding** theorem, any/every copula is bounded by

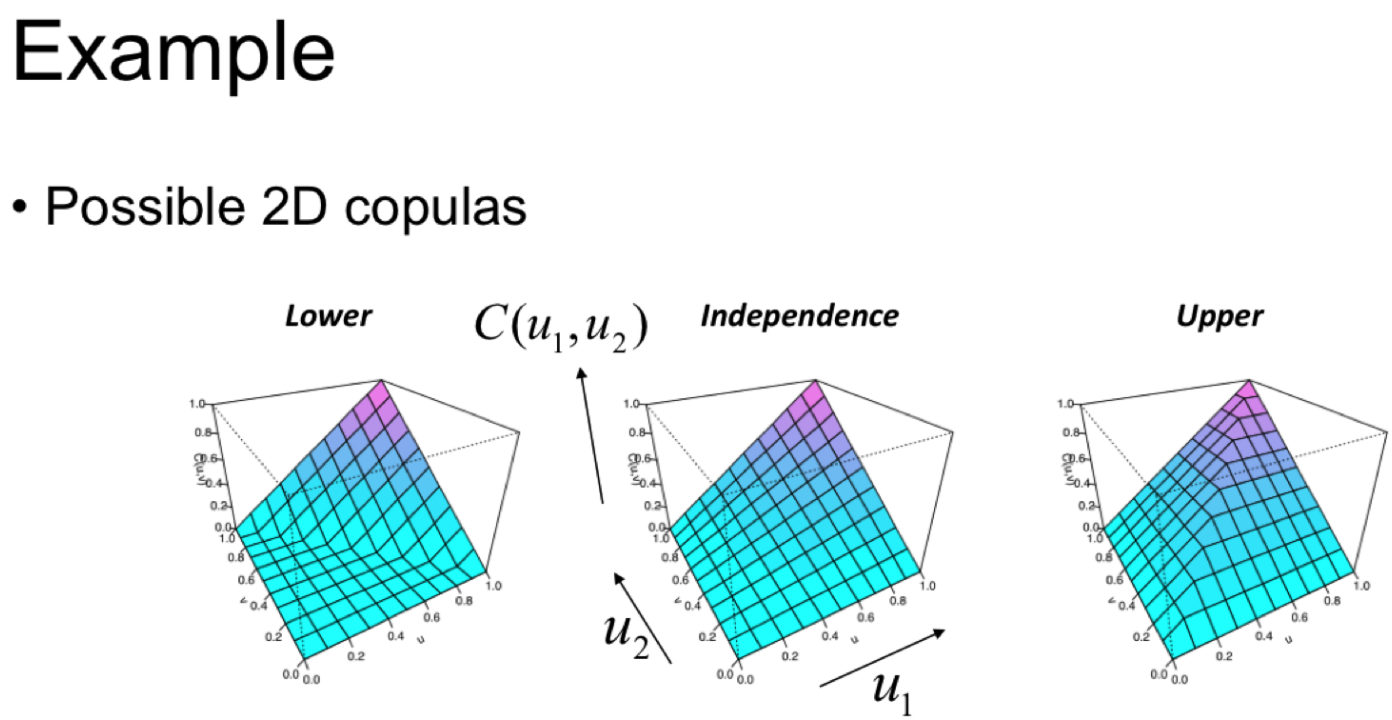
$$\underline{C}(u_1, \dots, u_d) \le C(u_1, \dots, u_d) \le \bar{C}(u_1, \dots, u_d)$$
$$\text{where } \begin{cases} \underline{C}(u_1, \dots, u_d) = \max\left\{1 - d + \sum_{i=1}^d u_i, 0\right\} = \max\left\{1 - \left(\sum_{i=1}^d 1 - u_i\right), 0\right\} \\ \bar{C}(u_1, \dots, u_d) = \min\{u_1, \dots, u_d\} \end{cases}$$

$\max\left\{1 - d + \sum_{i=1}^d u_i, 0\right\}$ is 1 minus number of uniforms plus the values of the uniforms

- (if $\mathbf{d} = \mathbf{5}, u_1 = .5, u_2 = .4, u_3 = .7, u_4 = .2, u_5 = .6$
  - then $\max(1 - 5 + 2.4 = \mathbf{-1.6}, 0)$)

- This implies the maximum of the Copula is bounded by 0, or higher if the average value of the uniforms are $\geq \frac{d-1}{d}$.

$\min\{u_1, \ldots, u_d\}$ implies the max of the Copula is bounded by the smallest marginal probability, which makes sense as that is the only one limiting the **cumulative probability**.



6.2 Sklar's Theorem

Any continuous multivariate CDF $F(x_1, \ldots x_d)$ with marginal (1D) CDF's $F_i(x_i)$ $\forall$ $i = 1, \ldots, d$ can be expressed in terms of a copula C, as

$$F(x_1, \ldots x_d) = C(F_1(x_1), \ldots, F_d(x_d))$$

Inverse is also true, where any copula combined with marginal CDF's can give a multivariate CDF.

- Copula's model dependence separately from the marginal distributions of the RVs

If $X \sim F \implies F(x) \sim Unif(0,1) \implies F^{-1}(Unif) \sim F$ *If an RV follows some CDF F, then ...*

> If you take a copula of a bunch of marginal CDF's, you can obtain the multivariate CDF of all the marginals. The inverse is true, where you can take a multivariate CDF and come up with a copula to represent the dependency between the marginal distributions, and the marginal distributions themselves. (I think is what this is saying.)

6.2.1 Example

For a continuous CDF $F(x_1, \dots, x_d)$ with marginals $F_i(x_i)$, the copula is given by:

$$C(F_1(x_1), \dots, F_d(F_d)) = F(x_1, \dots, x_d) \text{ by Sklar's Thm}$$

$$\text{Let } u_i = F_i(x_i) \implies x_i = F_i^{-1}(u_i)$$

$$\therefore \mathbf{C}(\mathbf{u_1}, \dots, \mathbf{u_d}) = \mathbf{F}(\mathbf{F_1^{-1}}(\mathbf{u_1}), \dots, \mathbf{F_d^{-1}}(\mathbf{u_d}))$$

6.3 Gaussian Copula

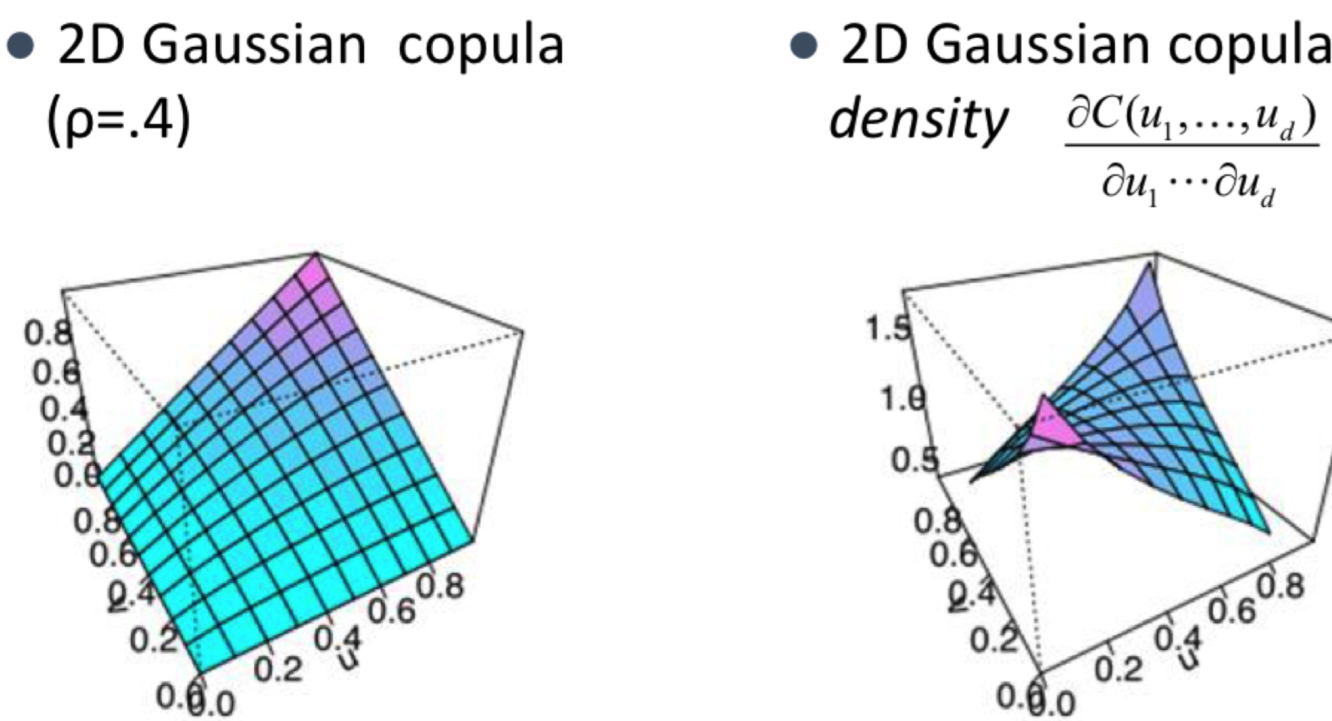
We can also construct copula's of non uniform distributions, such as this multivariate Normal CDF:

$$\text{Let } \mathbf{X} \sim N_d(\mu, \Sigma) \quad \text{with Correlation Mat } \rho$$

We can find the copula C_p of $\mathbf{X}$ using:

$$C_p(u_1, \dots u_d) = \Phi_{\mu,\Sigma}(\Phi^{-1}_{\mu_1,\sigma_1^2}(u_1), \dots, \Phi^{-1}_{u_d,\sigma_d^2}(u_d))$$

For multivariate distributions that have a Gaussian Copula are called **meta-Gaussian** distributions. These distributions themselves do not need to be Gaussian.



The second plot would be flat/uniform if the two independent distributions were not correlated ($\rho = 0$). This plot shows that the probability of both being 1 or 0 is very high, but the probability that one is 1 and the other is 0 is virtually 0. This supports our goal of modelling distributions where extreme values occur together.

6.4 Creating Copula's from Multivariate Distributions

We can create Copula's from known multivariate Distributions such as the Normal or t distributions.

We copy the dependence structure of known distributions (allowing us to use different marginals for modelling)

Copula from multivariate Normal CDF with correlation matrix ρ

$$C_\rho(u_1, \dots, u_d) = \Phi_\rho(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d))$$

$$\text{Where } \begin{cases} \Phi_\rho \text{ is multivariate Normal CDF with correlation } \rho \\ \\ \Phi \text{ is Standard univariate Normal CDF} \end{cases}$$

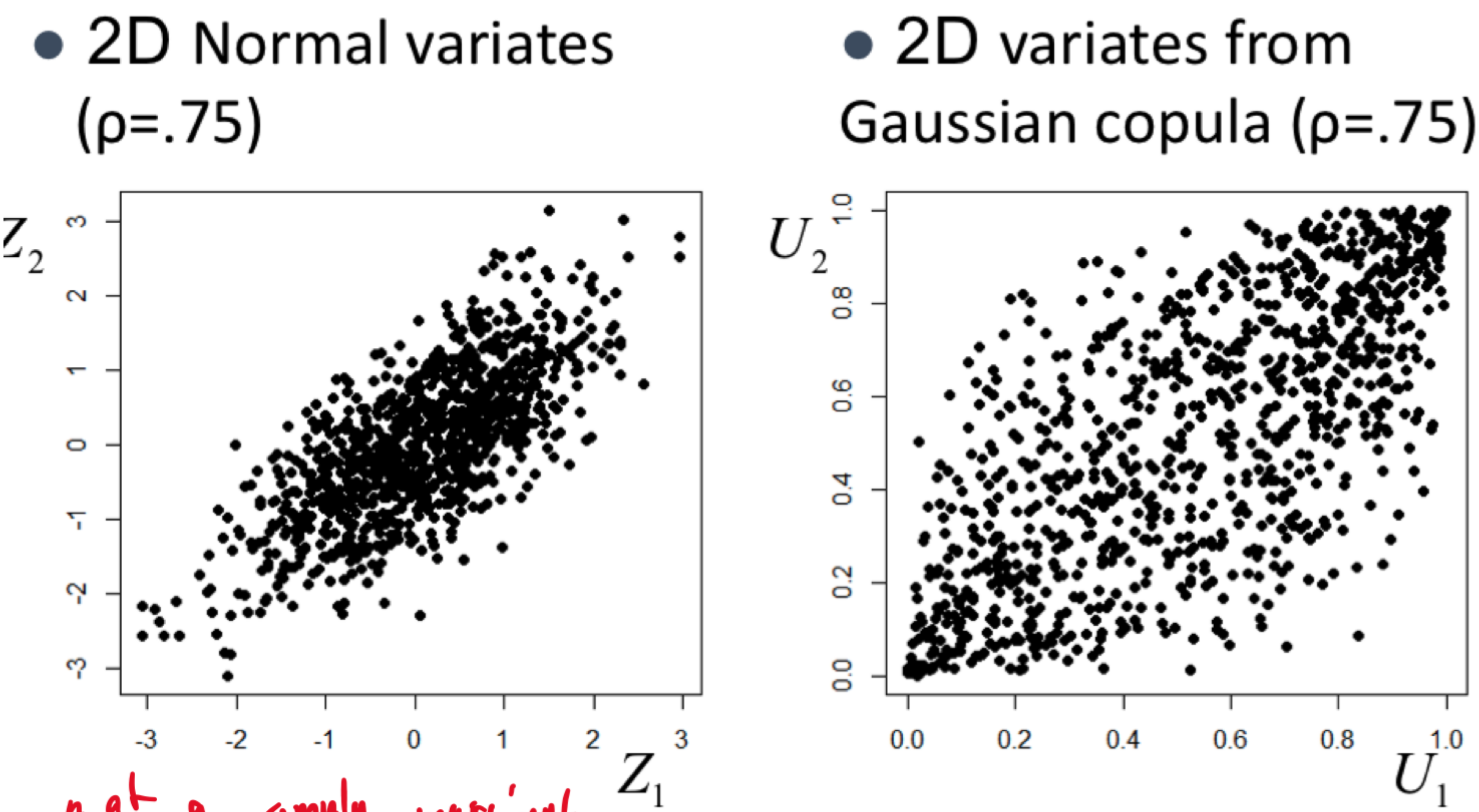
Simulating from a Copula Simulating from a distribution with a copula (dependence between marginals) and marginals themselves can be done with the following steps:

1. Generate (dependent) uniforms from the copula:
   - $$(U_1, \dots U_d) \sim C$$
   - This is done by generating from a multivariate normal with correlation ρ
     $$\mathbf{Z} = \begin{bmatrix} Z_1 \\ \vdots \\ Z_d \end{bmatrix} \sim N_d(0, \rho)$$
   - Calculate uniforms as their marginal CDF's (value of uniform var is a probability $P(Z \leq Z_i)$)
   - $$U_i = \Phi(Z_i),\ i = 1, \dots, d$$
   - Then we can use the uniforms with any other marginal
   - This is a very involved process, that we don't need to go into
2. Generate target variates from marginals, using the inverse CDF method
   - $$X_i = F_i^{-1}(U_i)\ \forall\ i$$

It's difficult for simulating uniforms from multivariate copula.

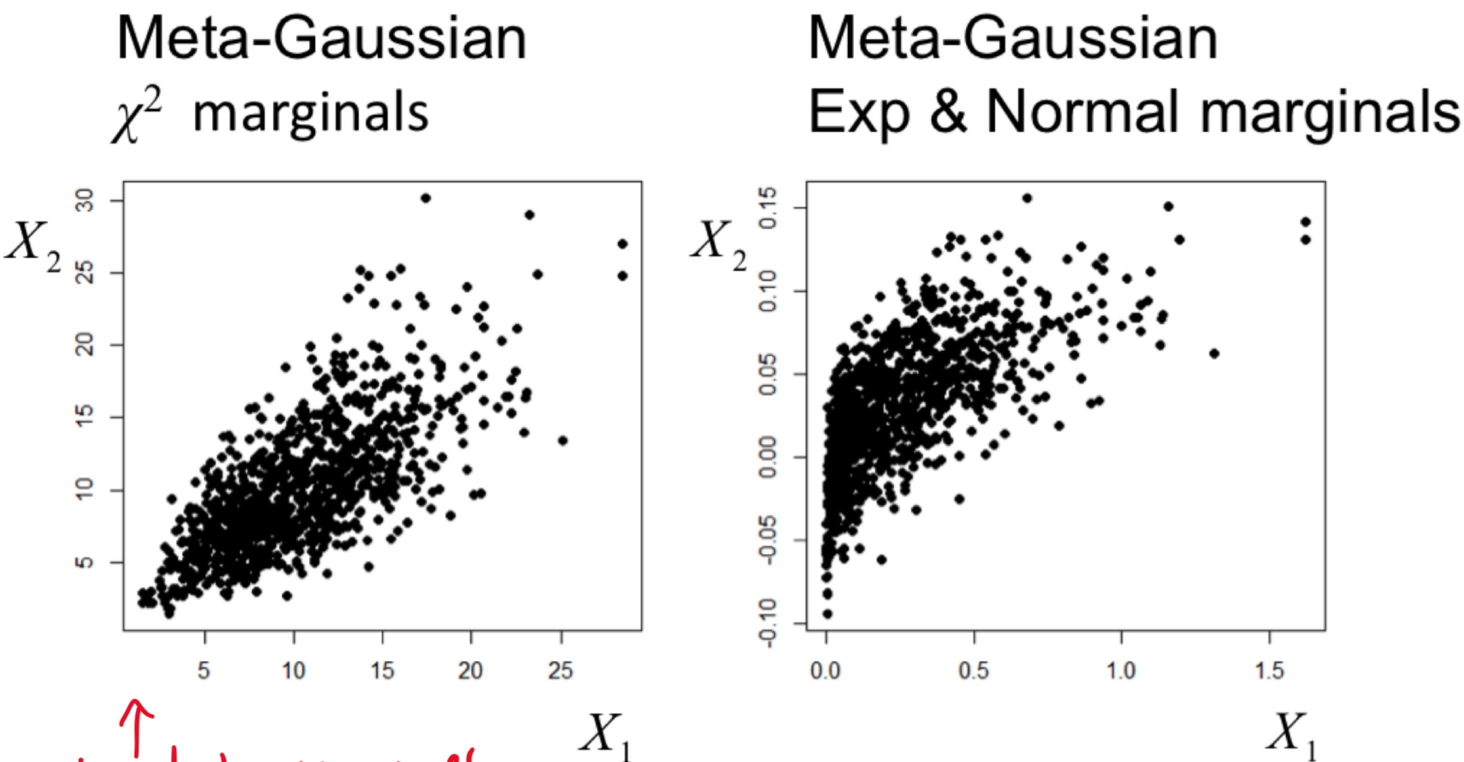
Simulating from just the 2D Multivariate normal with a correlation of 75% compared to simulating from the Gaussian Copula with the same correlation between variables.

6.4.1 Copula example plots



not a copula, marginals

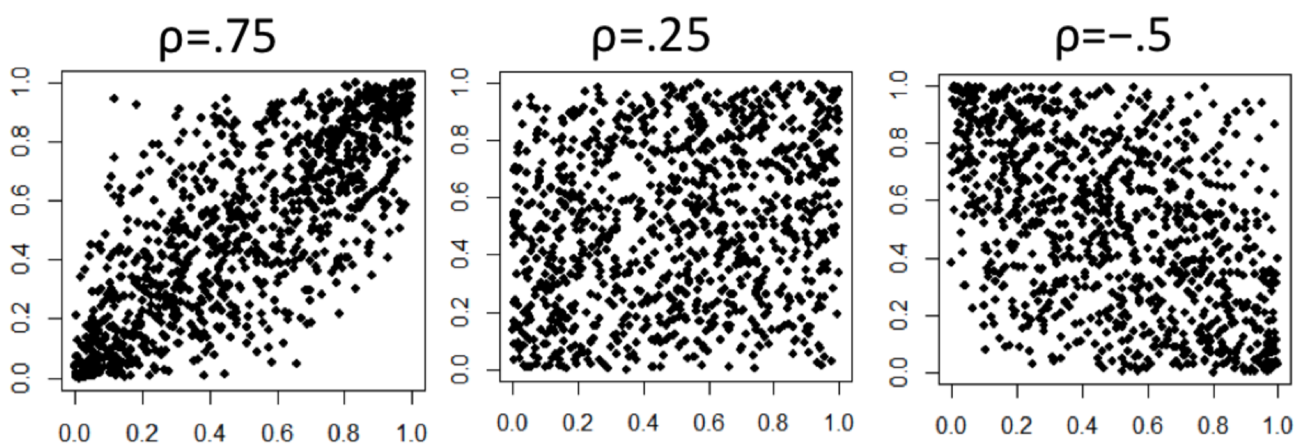
We can use two different marginals (χ^2 in this case), which are meta-Gaussian as they have a gaussian copula.



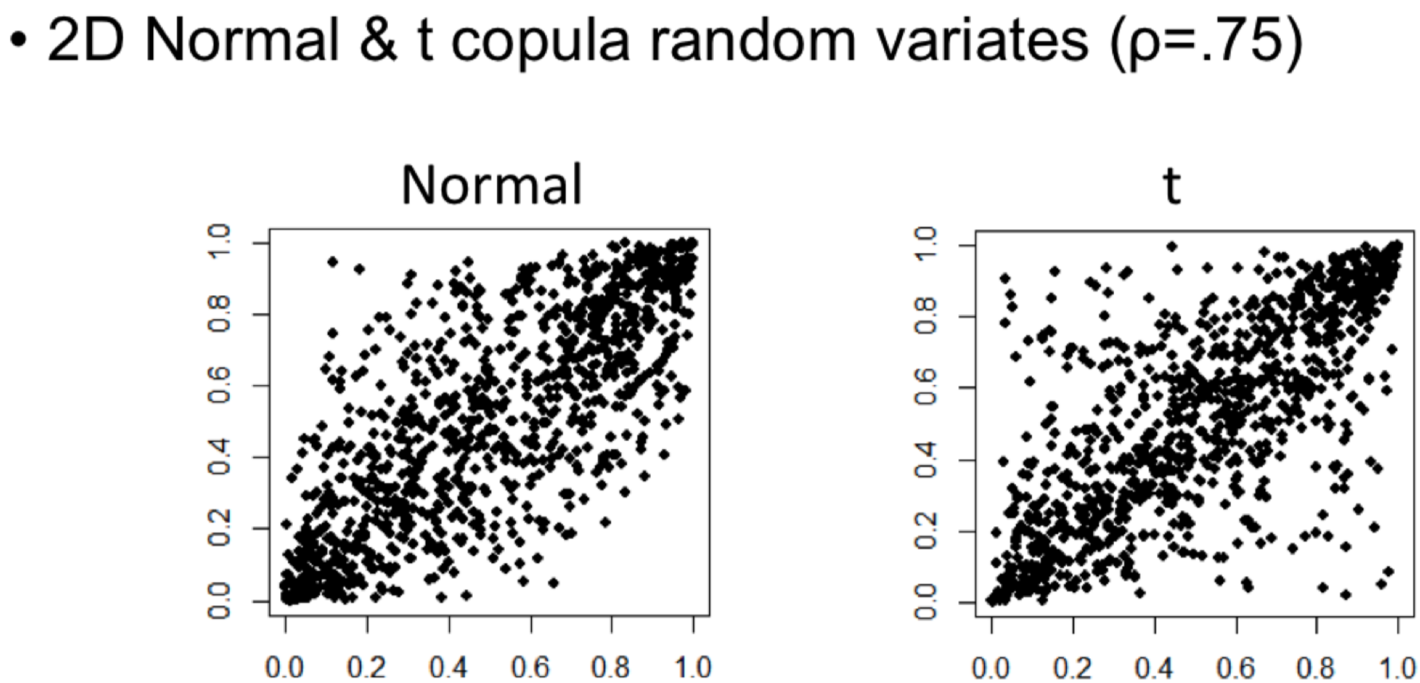
dependent chi squares

Here we see the simulated copula values for different correlation values.

- 2D Normal copula random variates



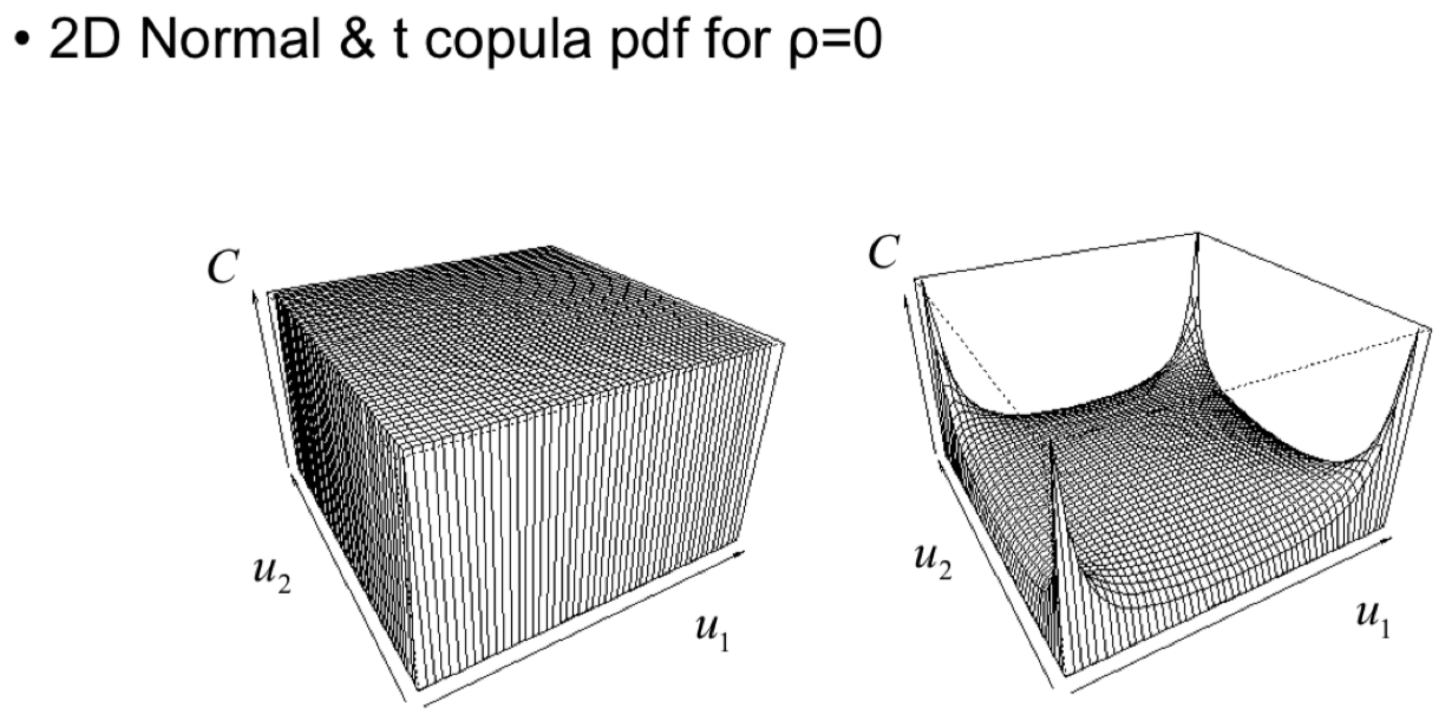
We can also see the differences between a Copula created using Normal and t distributions



The Normal copula shows u_1, u_2 are independent with 0 correlation as there is equal probability any value is sampled from u_1 and any value of u_2.

On the right, the t copula shows that values on the borders ($u_1 = 0 \mid\mid u_2 = 0$) drop to probability 0, and at the extreme values, jump very high.

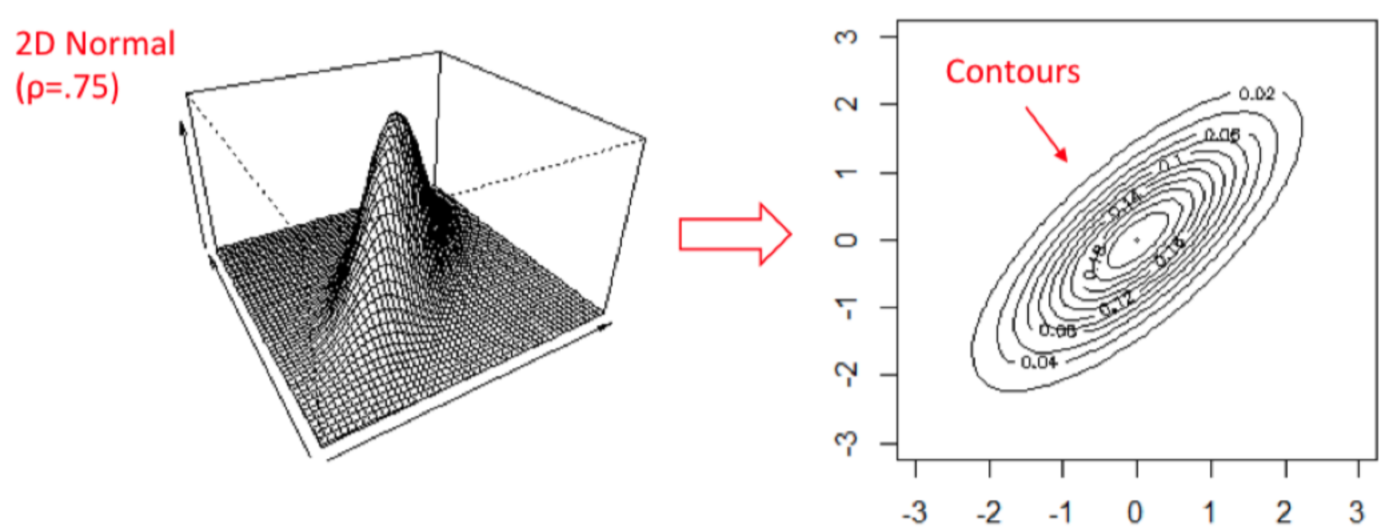
Peaks at extreme combinations $\rightarrow$ manifestation of tail dependence.

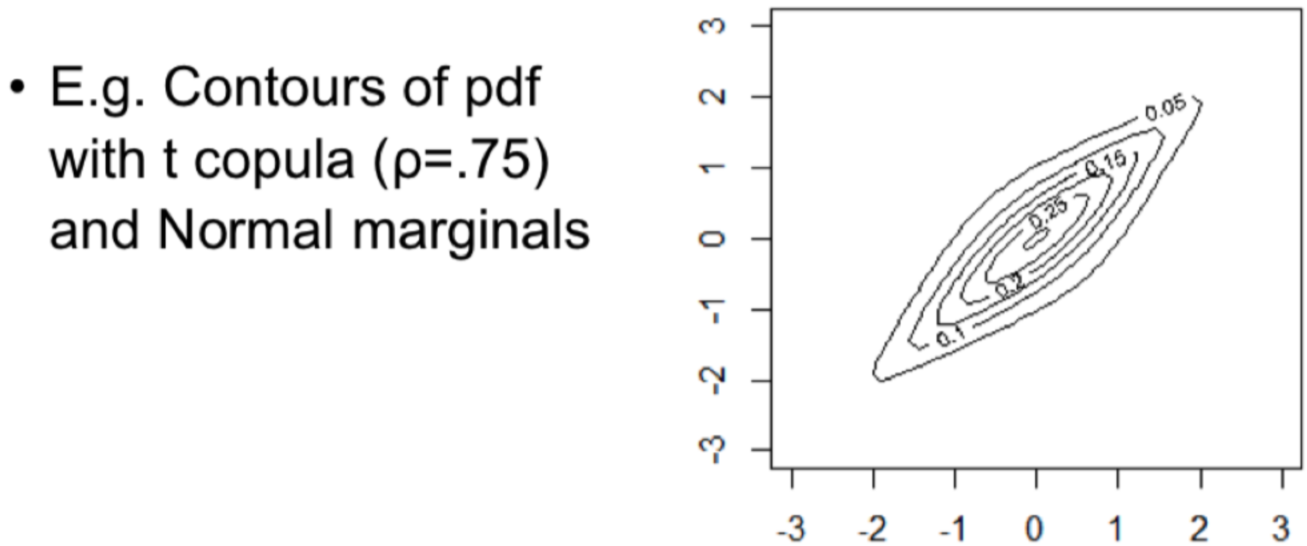


6.5 Elliptical Copulas

The Normal and t distributions have a specific type of dependence: **elliptical dependence**.

The ellipses describe the contours of the multivariate Normal and t distributions:





The ellipses are determined by the covariance matrix of the distributions. This implies a symmetry in the dependency structure, where **the strength** is the same for positively and negatively correlated variables.

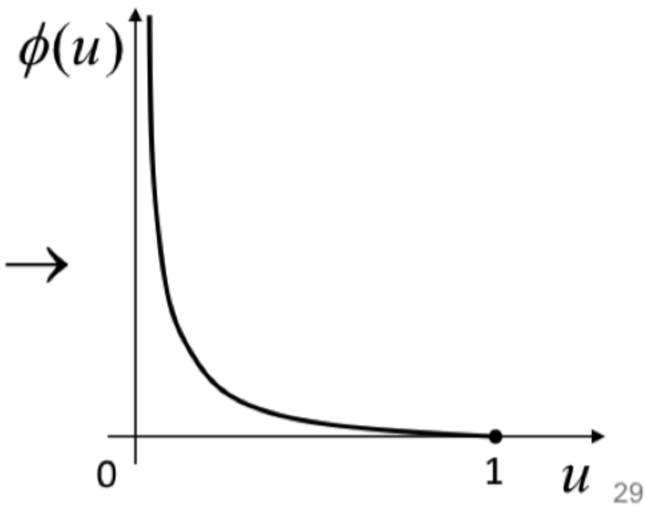
6.6 Archimedean Copulas

A Family of copula's whose form is given by

$$C(u_1, \dots, u_d) = \phi^{-1}[\phi(u_1) + \dots + \phi(u_d)]$$

Where ϕ:

- is a **continuous convex** generator function
- maps from $[0, 1] \to [0, \infty]$
- $\phi(0) = \infty, \phi(1) = 0$



This is just an example of a possible ϕ

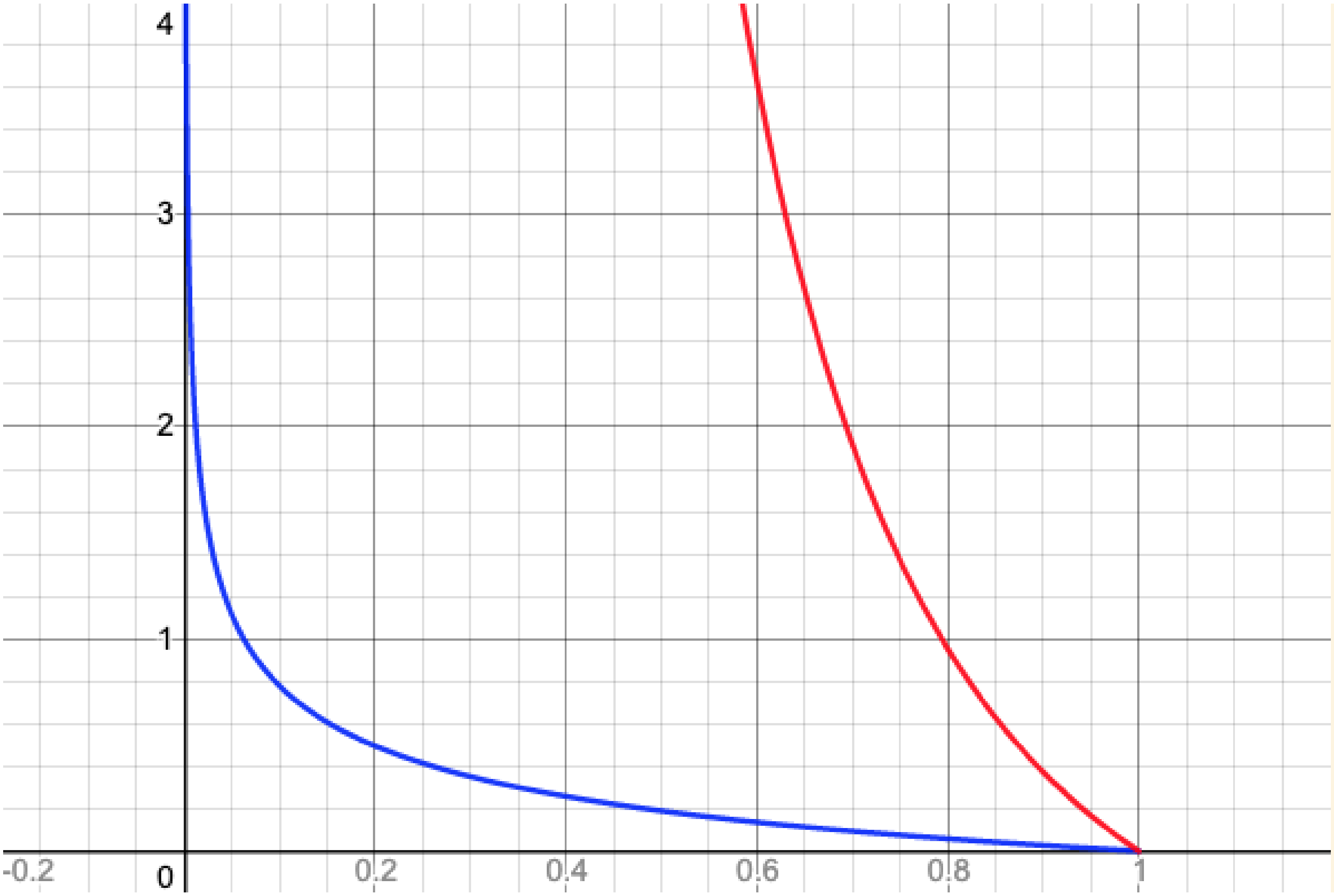
There are infinitely many ϕ but some popular choices for ϕ are:

Name	Generator $\phi(t)$	Generator Inverse $\phi^{-1}(t)$	Parameter
Clayton	$t^{-\theta} - 1$	$(1+s)^{-1/\theta}$	$\theta \geq 0$
Frank	$-\ln \frac{e^{-\theta t}-1}{\theta^{-\theta}-1}$	$-\frac{1}{\alpha}\ln(1+e^{-s}(e^{-\theta}-1))$	$\theta \geq 0$
Gumbel	$(-\ln t)^{\theta}$	$\exp\{-s^{-1/\theta}\}$	$\theta \geq 1$

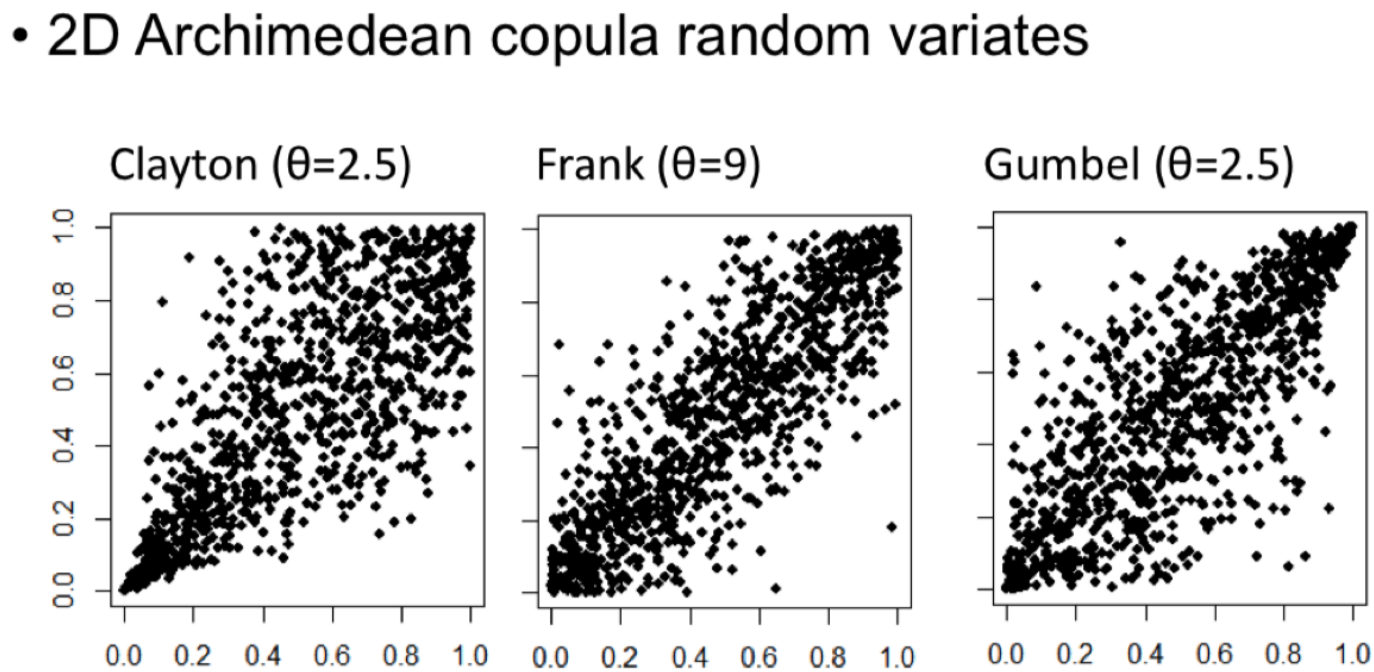
θ seems to represent how extreme the function varies between its domain, with larger θ's implying larger values throughout the domain.

For example, the Clayton generator with $\theta = 0.25$ in blue, and $\theta = 3$ in red.

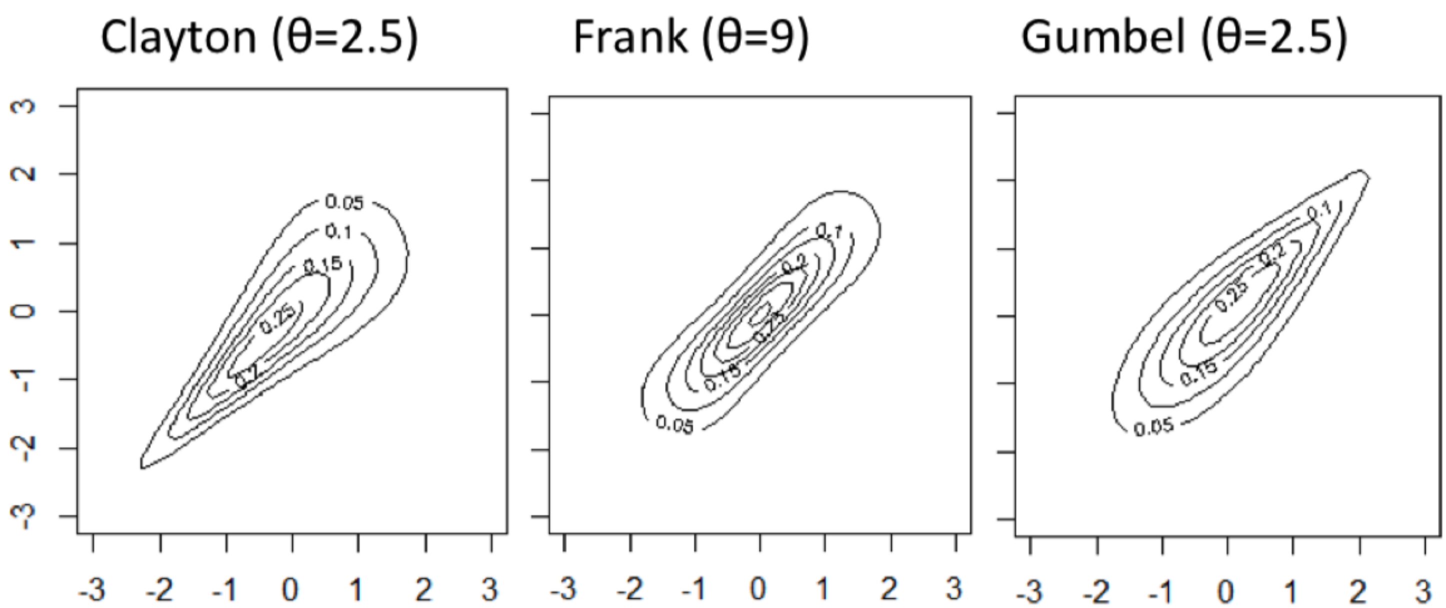
```
bottom=-0.2;top=4;
right=1.2;left=-0.2;
---
y=1/x^{0.25}-1|0<=x<=1|blue
y=1/x^{3}-1|0<=x<=1|red
```



Samples from the Archimedean copula's:



The dependency contours of each type of archimedean copula's base distribution covariance



This plot shows that we can use Archimedean copula's to model **asymmetric dependencies**, but suffer limitation in ≥ 3 dimensions.

The Archimedean Copula's value is constant for any permutation of coordinates $u_1, \dots, u_d$

$$\begin{aligned} C(u_1, u_2, \dots, u_d) &= \phi^{-1}(\phi(u_1) + \phi(u_{\mathbf{2}}) + \dots + \phi(u_{\mathbf{d}})) \\ &= \phi^{-1}(\phi(u_1) + \phi(u_{\mathbf{d}}) + \dots + \phi(u_{\mathbf{2}})) \end{aligned}$$

All pairs of coordinates(variables) have the same dependence, **which is not the case for elliptical copulas**.

> There exist copulas that can both model asymmetric dependencies, **and** differences in pairwise dependence called **vine copula's**.

6.7 Fitting Copula's

Given a copula and marginal distributions, the MLE method can be applied to fit multivariate distribution parameters to sample data. This could however lead to a very high number of parameters.

Instead, *pseudo-MLE* could be used to break down the problem into the marginals and copula.

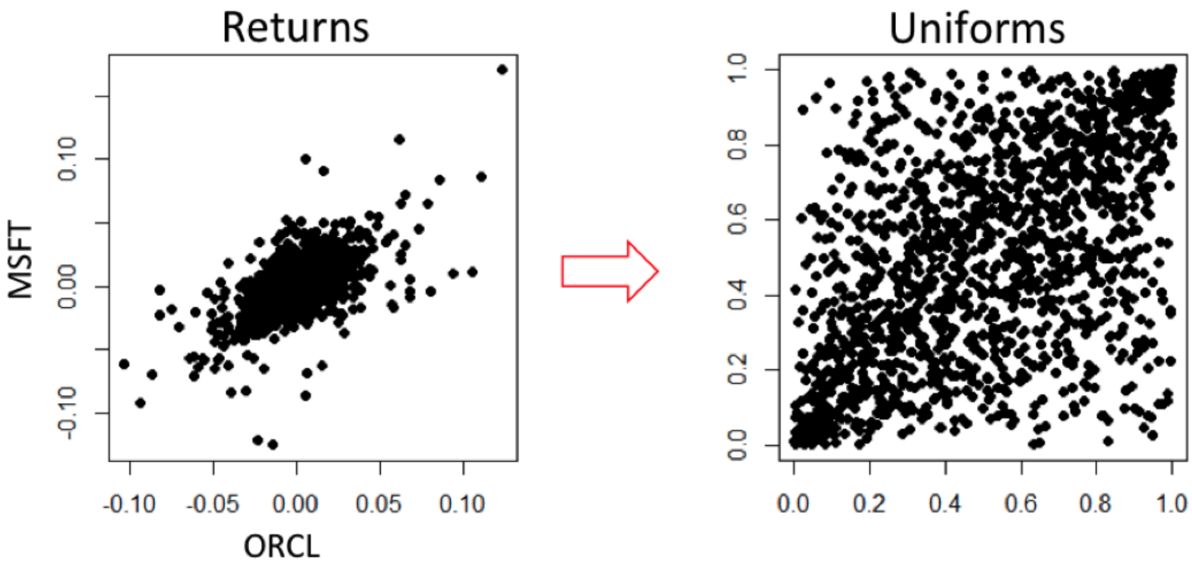
$$U_i^{(j)} = \hat{F}_j(X_i^{(j)}) \quad \forall \; i, \dots, n \; j = 1, \dots, d$$

> Each n uniforms in d dimensions of the copula can be created from the actual multivariate distribution. Alternatively, the empirical CDF could be used to obtain the uniforms.

We can then estimate the copula using MLE on the uniforms.

Example

- ORCL & MSFT daily returns (2005-10)
  - Marginal empirical cdf *“converts”* data to [0,1]



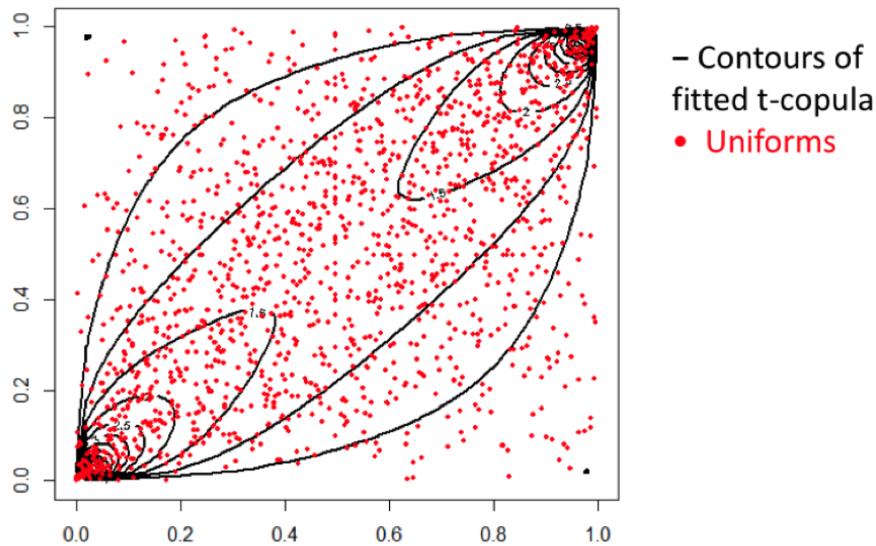
Guess: The uniforms are created by plugging in each asset return into their own marginal CDF.

Example

how to empirically model dependence

- Pseudo-MLE of t-copula, based on marginal empirical quantiles

$$\rightarrow \hat{\rho} = .5767 \ \& \ \hat{v} = 2.8889$$



36

The *pseudo-MLE* can give the uniform marginals used to construct/fit a t-copula.

Chapter 7

Portfolio Theory

Portfolio theory deals with how an asset manager can form a portfolio that optimizes their goals, whether that be lowest risk, highest return, or some other measure of performance.

How to pick stocks is a complicated science and there are many ways to go about it.

7.1 Assumptions

We first make some assumptions when dealing with the theory. - Static multivariate return distribution determined by assets' mean and covariance. - This implies a normal or elliptical distribution for returns - The investors have the same views on mean and variance - Investors also want minimum risk for maximum return - Investors measure risk by portfolio's variance - No borrowing or short-selling restrictions - No transaction costs

7.2 Dealing with Two Assets

We'll start with a simple example with two risky assets, S_1, S_2

We assume that net returns (as time $0 \to t$) satisfy:

$$\left[\begin{array}{c} R_1 \\ R_2 \end{array}\right] \sim N\left(\left[\begin{array}{c} \mu_1 \\ \mu_2 \end{array}\right], \left[\begin{array}{cc} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{array}\right]\right) \quad \text{where } R_i = \frac{S_i(t) - S_i(0)}{S_i(0)}, i = 1, 2$$

We can form a portfolio with x_i units of asset S_i, which gives us the following equation:

$$V(t=0) = x_1 \cdot S_1(0) + x_2 \cdot S_2(0)$$

We can derive the initial *weights* of each asset that we've invested into using:

$$w_i = \frac{x_i S_i(0)}{V(0)}$$

7.2.1 Portfolio Return

We can calculate the return of this two asset portfolio and show that: $R_p = w_1 R_1 + w_2 R_2$

$$\begin{aligned}
R_p &= \frac{V(t)-V(0)}{V(0)} = \frac{[x_1 S_1(t) + x_2 S_2(t)] - [x_1 S_1(0) + x_2 S_2(0)]}{V(0)} \\
&= \frac{x_1[S_1(t)-S_1(0)] + x_2[S_2(t)-S_2(0)]}{V(0)} \\
&= x_1 \underbrace{\frac{S_1(t)-S_1(0)}{S_1(0)}\frac{S_1(0)}{V(0)}}_{\text{Introduce } S_1(0) \text{ to get} R_1} + x_2 \underbrace{\frac{S_2(t)-S_2(0)}{S_2(0)}\frac{S_2(0)}{V(0)}}_{\text{Introduce } S_2(0) \text{ to get} R_2} \\
&= R_1 \underbrace{x_1 \frac{S_1(0)}{V(0)}}_{w_1} + R_2 \underbrace{x_2 \frac{S_2(0)}{V(0)}}_{w_2} \\
&= R_1 w_1 + R_2 w_2
\end{aligned}$$

Which shows that the net returns of a portfolio is exactly the weight combination of the net returns of the assets within.

We can also find the distribution of the portfolio returns (which may not surprise you, is just the linear combination of the individual assets returns)

$$\begin{aligned}
&R_p = \underline{w}^T \underline{R} = [w_1 \ w_2] \begin{bmatrix} R_1 \\ R_2 \end{bmatrix} \quad \text{where } \underline{R} \sim N_{2D}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix} \right) \\
&\implies R_p \sim N_{1D}(\mu_p, \sigma_p) \ \text{ where}
\end{aligned}$$

$$\begin{aligned}
\mu_p &= \mathbb{E}(R_p) = \mathbf{E}[\underline{w}^T \underline{R}] \\
&= \underline{w}^T \mathbb{E}[\underline{R}] = \underline{w}^T \underline{\mu} \\
&= \sum_i w_i \mu_i
\end{aligned}$$

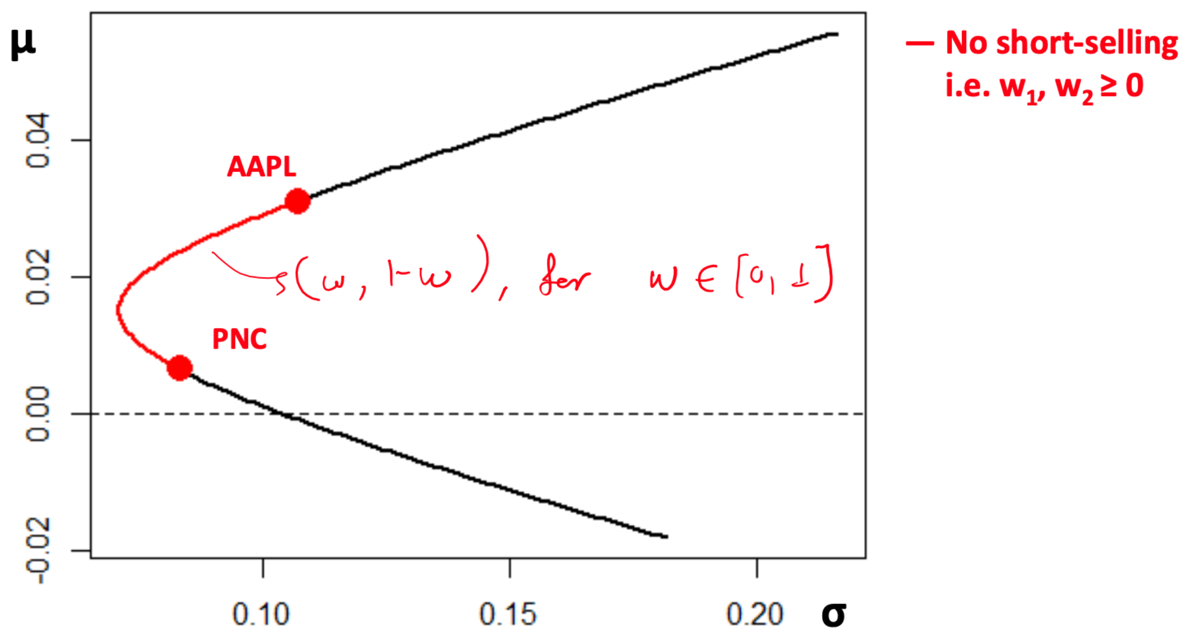
$$\begin{aligned}
\sigma_p^2 &= \mathbb{V}[R_p] = \mathbb{V}[\underline{w}^T \underline{R}] \\
&= \underline{w}^T \mathbb{V}(\underline{R}) \underline{w} = \underline{w}^T \Sigma \underline{w} \\
&= [w_1 \ w_2] \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \\
&= w_1^2 \sigma_1^2 + 2 w_1 w_2 \sigma_{12} + w_2^2 \sigma_2^2
\end{aligned}$$

$$\therefore R_p \sim N\left(\sum_i w_i p_i, \quad w_1^2\sigma_1^2 + 2w_1 w_2 \sigma_{12} + w_2^2 \sigma_2^2 \right)$$

7.3 Efficiency frontier

A plot of mean μ by standard deviation σ , where you can load up entirely on `AAPL` with $w_1 = 1, w_2 = 0$ or entirely on `PNC` with the opposite.

• 2-asset portfolio mean return vs st.dev. (risk)



But how do we know the optimal amount of weight to put in each asset to **minimize variance**? We can simply find the w that minimizes the derivative of the variance w.r.t. w

$$
\begin{aligned}
&\text{Let } w_1 = w \quad \& \; w_2 = 1 - w \\
&\implies \sigma_p^2 = w^2\sigma_1^2 + 2w(1-w)\sigma_{12} + (1-w)^2\sigma_2^2 \\
&\text{Let } \frac{\delta\sigma_p^2}{\delta w} = 0 \\
&\quad 0 = 2w\sigma_1^2 + 2(1-w)\sigma_{12} - 2w\sigma_{12} - 2(1-w)\sigma_2^2 \\
&\quad 0 = w(\sigma_1^2 - \sigma_{12}) + (1-w)(\sigma_{12} - \sigma_2^2) \\
&\quad 0 = w(\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}) + (\sigma_{12} - \sigma_2^2) \\
&\quad w = \frac{-(\sigma_{12} - \sigma_2^2)}{(\sigma_1^2 + \sigma_2^2 - 2\sigma_{12})}
\end{aligned}
$$

7.4 Multiple asset portfolio

Consider n risky assets with returns $R_1, \ldots, R_n$ so that:

$$
\mathbf{R} = \begin{bmatrix} R_1 \\ \vdots \\ R_n \end{bmatrix} \sim N\left(\mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix}, \Sigma = \begin{bmatrix} \sigma_1^2 & \cdots & \sigma_{1n} \\ \vdots & \ddots & \vdots \\ \sigma_{n1} & \cdots & \sigma_n^2 \end{bmatrix}\right)
$$

We want to create portfolio using weights $\mathbf{w} = \begin{bmatrix} w_1 & \cdots & w_n \end{bmatrix}^\top$ such that: $\sum_{i=1}^n w_i = \mathbf{w}^\top \mathbf{1} = 1$

The distribution of returns for this portfolio would simply be:

$$\begin{aligned}
&\mathbf{R_p} \sim \mathbf{N_{1D}}(\mathbf{_p}, \mathbf{^2_p}) \text{ where} \\
&\mathbf{_p} = \underline{\mathbf{w}}^{\mathbf{T}} \cdot \\
&\mathbf{^2_p} = \underline{\mathbf{w}}^{\mathbf{T}} \cdot \cdot \underline{\mathbf{w}}
\end{aligned}$$

When fixing an expected return μ_p, we can derive weights that minimize variance:

$$min_w\{\mathbf{w^T w}\} \ s.t. \ \mathbf{w^T} = \mu_p, \mathbf{w^T 1} = 1$$

We can use Lagrange multipliers to define the objective function:

> Lagrange formula: $\mathcal{L}(x, \lambda) = f(x) + \lambda g(x)$ If we want to find the max/min of the function $f(x)$ subject to the constraint $g(x) = 0$, we can form this Lagrangian and find where it's derivative is 0. In this case, $x = \underline{w}$ and our constraint $\mathbf{w^T} \cdot \mathbf{1} = 1$ can be represented as $g(w) = \mathbf{w^T} \cdot \mathbf{1} - 1$

$$\begin{aligned}
&\mathcal{L}(\underline{w}, \lambda) = \underline{w}^T \underline{\Sigma w} - \lambda(\underline{w}^T \cdot \underline{1} - 1) \\
&\text{Let } \frac{\delta \mathcal{L}}{\delta \underline{w}} = 0 \\
&\quad 0 = 2 \cdot \underline{\Sigma} \cdot \underline{w} - \lambda \cdot \underline{1} \\
&\implies \underline{w} = \frac{\lambda}{2} \underline{\Sigma}^{-1} \cdot \underline{1} \\
&\text{Given } \underline{w}^T \cdot \underline{1} = 1 \implies \underline{1}^T \underline{w} = 1
\end{aligned}$$

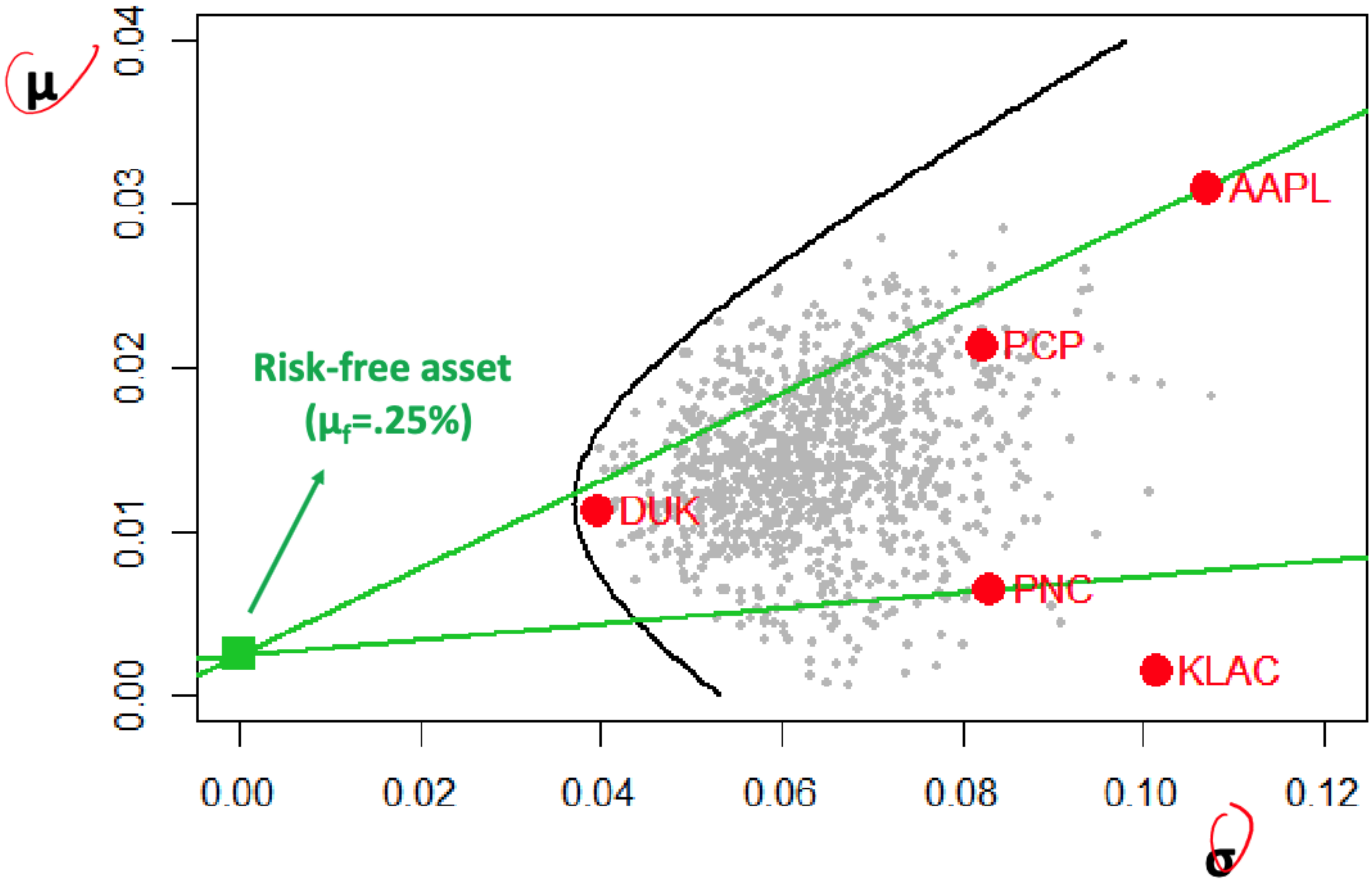
$$\begin{aligned}
&\implies \frac{\lambda}{2} \cdot \underline{1}^T \cdot \underline{\Sigma}^{-1} \cdot \underline{1} = \underline{1}^T \underline{w} = 1 \\
&\implies \lambda = \frac{2}{\underline{1}^T \cdot \underline{\Sigma}^{-1} \cdot \underline{1}} \\
&\implies \underline{w}^* = \frac{\frac{2}{\underline{1}^T \cdot \underline{\Sigma}^{-1} \cdot \underline{1}}}{2} \underline{\Sigma}^{-1} \cdot \underline{1} \\
&\qquad = \frac{\underline{\Sigma}^{-1} \cdot \underline{1}}{\underline{1}^T \cdot \underline{\Sigma}^{-1} \cdot \underline{1}}
\end{aligned}$$

> $\underline{\Sigma}^{-1}$ represents the precision matrix, where if there is a 0 at (i, j) then i and j are conditionally independent (given all other assets are held constant, there is no dependence/correlation between these two assets). If it exists, then it is positive definite.
>
> The numerator is a vector containing the L1 norms of each row's (asset) precision. The denominator represents the total precision of the entire Precision matrix. The final value is the scaled weights given to each asset **depending** on the total precision.

Instead of only having n risky assets, we can introduce a risk-free asset with constant return $R_f = \mu_f > 0$, with variance $\sigma_f = 0$. A real life example is a government bond.

With this new asset, we can determine how much to invest in the risky portfolio and the risk-free asset with weights w_p and $(1 - w_p)$.

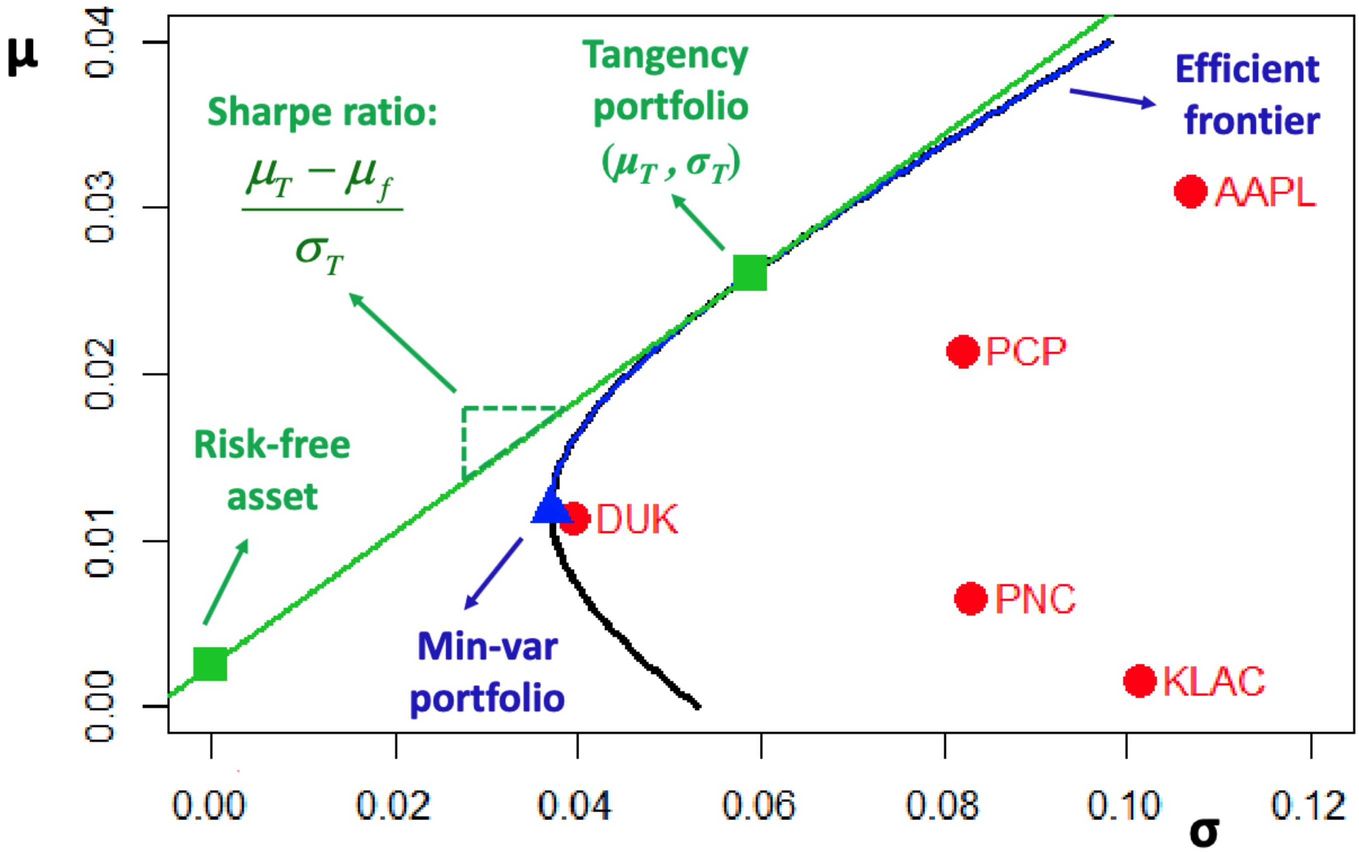
- Return: $R = w_p R_p + (1 - w_p) R_f$
- Mean Return: $\mathbb{E}[R] = w_p \mathbb{E}[R_p] + (1 - w_p)\mathbb{E}[R_f] = w_p \mu_p + (1 - w_p)\mu_f$
- Variance of returns: $\mathbb{V} = \mathbb{V}(w_p R_p + (1 - w_p R_f)) = w_p^2 \mathbb{V}(R_p) = w_p^2 \cdot \sigma_p^2$



The best portfolio (in terms of risk and return) lie on this green line, where this line is tangent to the efficient frontier.

The slope of this line is called the **Sharpe** ratio, which measure the excess return per unit risk (**risk premium**).

The **tangency portfolio** is exactly the portfolio on the efficient frontier that also belongs to the tangent line. (It is a risky asset only portfolio with μ_T, σ_T)



When you have a risk-free asset, the optimal investment strategy is to have a combination of the tangency portfolio and the risk-free asset. To find this tangency portfolio, we *maximize* the Sharpe ratio.

$$max \left\{ \frac{\mu_p - \mu_f}{\sigma_p} \right\} = max_w \left\{ \frac{\mathbf{w^T} - \mu_f}{\sqrt{\mathbf{w^T w}}} \right\} \quad \text{subject to } \mathbf{w^T 1} = 1$$

The tangency portfolio weights are given by:

$$\mathbf{w_t} = \frac{^{\mathbf{-1}}(- \mu_f \cdot \underline{\mathbf{1}})}{\mathbf{1^{T-1}}(- \mu_f \cdot \underline{\mathbf{1}})}$$

7.5 CAPM - Capital Asset Pricing Model

The CAPM theory suggests that all investors hold some form of the tangency/market portfolio. This model assumes the mean-variance analysis *stretched to it's logical consequences* can explain **all financial market behaviour**.(Widely known to be false)

The slope of the tangency portfolio is also known as the Sharpe Ratio, or **Market Price of Risk**.

If every investor followed this mean-variance analysis, and the market is in *equilibrium*: >idk what it means to be in equilibrium

- Every investor holds some portion of the same tangency portfolio
- The entire financial market will be made up of the **same** mix of risky assets
- **Making the tangency portfolio equal to the market value-weighted index**

7.5.1 Market Portfolio

Using the above assumptions, the **composition of the tangency portfolio is equivalent to the market** and we can find the weights for each asset as:

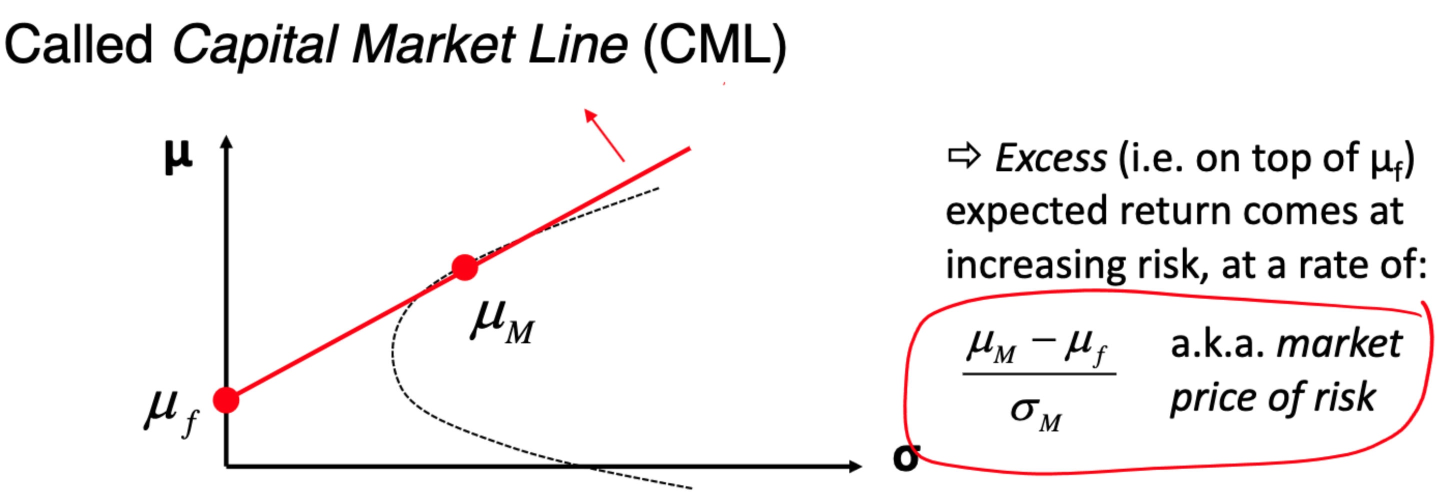
$$w_i = \underbrace{\frac{S_i \times O_i}{\sum_{i=1}^{N} S_i \times O_i}}_{\text{Market Capitalization \%}} \qquad \text{where} \begin{cases} S_i = \text{price of asset i} \\ O_i = \text{shares outstanding} \end{cases}$$

Therefore the tangency portfolio is exactly a market cap weighted index, such as the S&P500 or FTSE100.

7.5.2 Capital Market Line

Every mean-variance efficient portfolio (μ_p, σ_p) lies on a straight line, which is determined by

$$\mu_p = \mu_f + \frac{\mu_M - \mu_f}{\sigma_M}, \quad \text{where} \begin{cases} (\mu_M, \sigma_M) = \text{market portfolio} \\ \mu_f = \text{risk-free rate} \end{cases}$$



7.5.3 Security Market Line

For **each individual asset**, CAPM implies the following risk/reward relationship (Security Market Line)

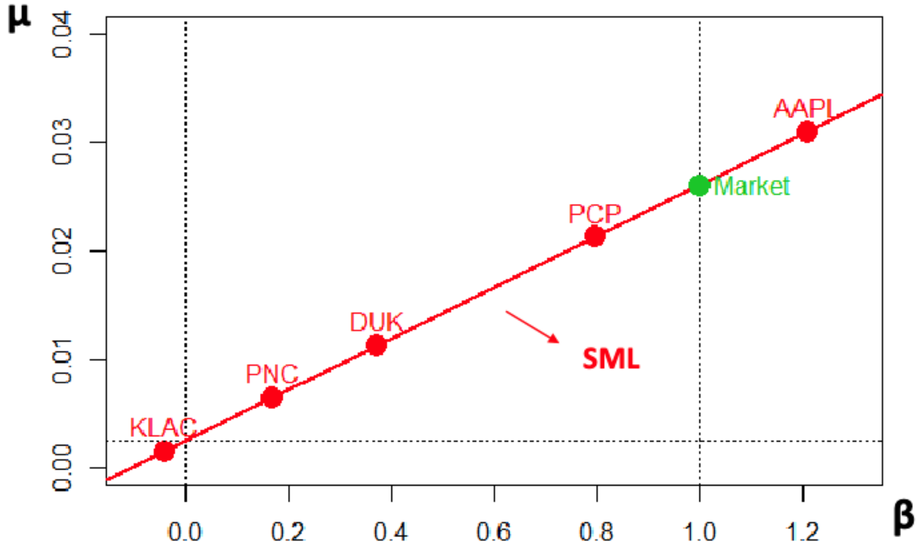
$$\mu_i = \mu_f + \beta_i \underbrace{(\mu_M - \mu_f)}_{Slope} \quad \text{Where } \begin{cases} \beta_i = \frac{\sigma_{iM}}{\sigma_M^2} \\ \sigma_{iM} = Cov(R_{i,R_M}) \end{cases}$$
$$= \mu_f + \left(\frac{\mu_M - \mu_f}{\sigma_M}\right) \cdot \sigma_p$$

- β_i is different for each asset, it captures how related the return on the given asset is with the Market return.
- σ_M is the market risk AKA undiversifiable risk
- Imagine the line is in (μ, β)-space, with slope $(\mu_M - \mu_f)$.

To derive this line, we need to find a market portfolio that maximizes the Sharpe ratio (using first order derivative)

$$f(\underline{w}) = \frac{\underline{w}^T\underline{\mu} - \mu_f}{\sqrt{\underline{w}^T\underline{\Sigma}\underline{w}}} = \frac{\underline{w}^T \cdot (\underline{\mu} - \underline{1}\mu_f)}{\sqrt{\underline{w}^T\underline{\Sigma}\underline{w}}} \quad \text{maximized when } \frac{\delta f}{d\underline{w}} = 0$$
$$\implies 0 = \frac{(\underline{\mu} - \underline{1}\mu_f) \cdot \sqrt{\underline{w}^T\underline{\Sigma}\underline{w}} - \frac{1}{2}(\overbrace{\underline{w}^T\underline{\Sigma}\underline{w}}^{\sigma_m^2})^{-1/2} \cdot 2\underline{\Sigma}\underline{w} \cdot (\overbrace{\underline{w}^T \cdot (\underline{\mu}}^{\mu_M} - \underline{1}\mu_f)}{\underline{w}^T\underline{\Sigma}\underline{w}}$$
$$= \frac{(\underline{\mu} - \underline{1}\mu_f)\sigma_m - \frac{1}{\sigma_M}\underline{\Sigma}\underline{w}(\mu_M - \mu_f)}{\sigma_M^2}$$
$$\implies (\underline{\mu} - \underline{1}\mu_f) = \frac{1}{\sigma_M^2 \underline{\Sigma}\underline{w}(\mu_M - \mu_f)}$$
$$\implies \begin{bmatrix} \mu_1 - \mu_f \\ \vdots \\ \mu_n - \mu_f \end{bmatrix} = \frac{\mu_M - \mu_f}{\sigma_M^2} \cdot \begin{bmatrix} Cov(R_1, R_M) \\ \vdots \\ Cov(R_n, R_M) \end{bmatrix}$$
$$= \frac{(\mu_M - \mu_f)}{\sigma_M^2} \cdot \begin{bmatrix} \sigma_{1,M} \\ \vdots \\ \sigma_{n,M} \end{bmatrix}$$

This is because $Cov(\underline{R}, R_M) = Cov(\underline{R}, \underline{w}^T \cdot \underline{R}) = \underline{w}^T \underbrace{Cov(\underline{R}, \underline{R})}_{\underline{\Sigma}} = \underline{w}^T\underline{\Sigma}$



7.5.3.1 Ex: Find market portfolio weights and SML with N i.i.d assets

Each asset $\sim N(\mu, \sigma^2)$ returns and the risk-free return $\mu_f < \mu$.

$$\begin{bmatrix} R_1 \\ \vdots \\ R_N \end{bmatrix} \sim N_{N\ Dim}(\underline{\mu} = \mu \cdot \underline{1}, \underline{\Sigma} = \sigma^2 \cdot \underline{I})$$

So we know:

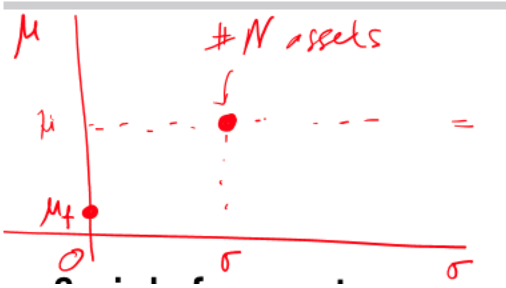
$$\forall\ \underline{w} \quad s.t.\ \underline{w}^T \cdot \underline{1} = 1 \implies \underline{w}^T \cdot \mu \cdot 1 = \mu$$

Which implies that the market portfolio is equal to the minimum variance portfolio.

$$\underline{w}^\star = \frac{\underline{\Sigma}^{-1} \cdot \underline{1}}{\underline{1}^T \underline{\Sigma}^{-1} \cdot \underline{1}} = \frac{\frac{1}{\sigma^2} \underline{I} \cdot \underline{1}}{\frac{1}{\sigma^2} \underline{1}^T \underline{I} \cdot \underline{1}} = \frac{\underline{1}}{N}$$

$$\implies \underline{w} = \frac{\underline{1}}{N} = \begin{bmatrix} \frac{1}{N} \\ \vdots \\ \frac{1}{N} \end{bmatrix} \implies w_i = \frac{1}{N}\ \forall i = 1, \dots, N$$

$$\begin{aligned} \implies \text{min variance } &= \underline{w}^T \underline{\Sigma} \underline{w} \\ &= \left(\frac{1}{N}\right)^2 \cdot \underline{1}^T (\sigma^2 \cdot \underline{I}) \cdot \underline{1} \\ &= \frac{\sigma^2}{N} \end{aligned}$$



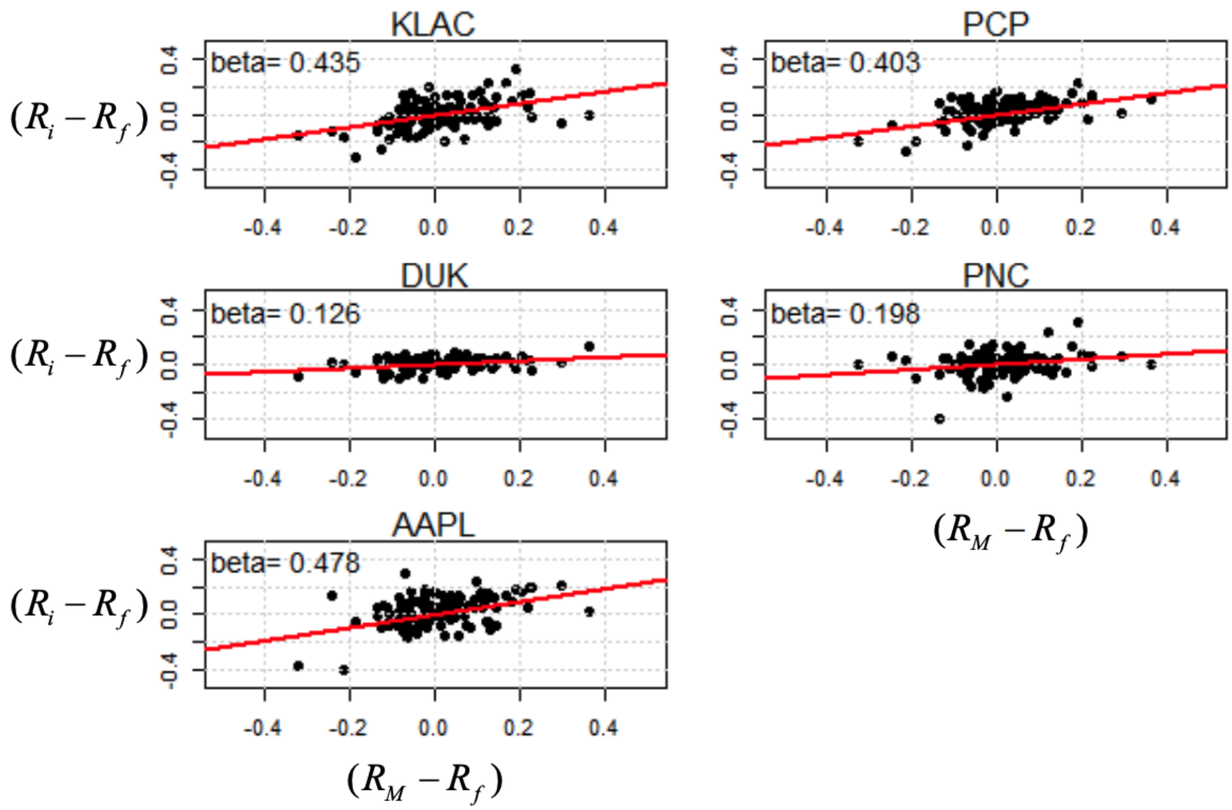
7.5.4 Security Characteristic Line

The β_i's are found empirically, by regressing $(R_i - R_f)$ on $(R_M - R_f)$ - Where R_M is the market return (Proxy by large market index, ex S&P500) - R_f is the risk free rate (proxy by T-bill)

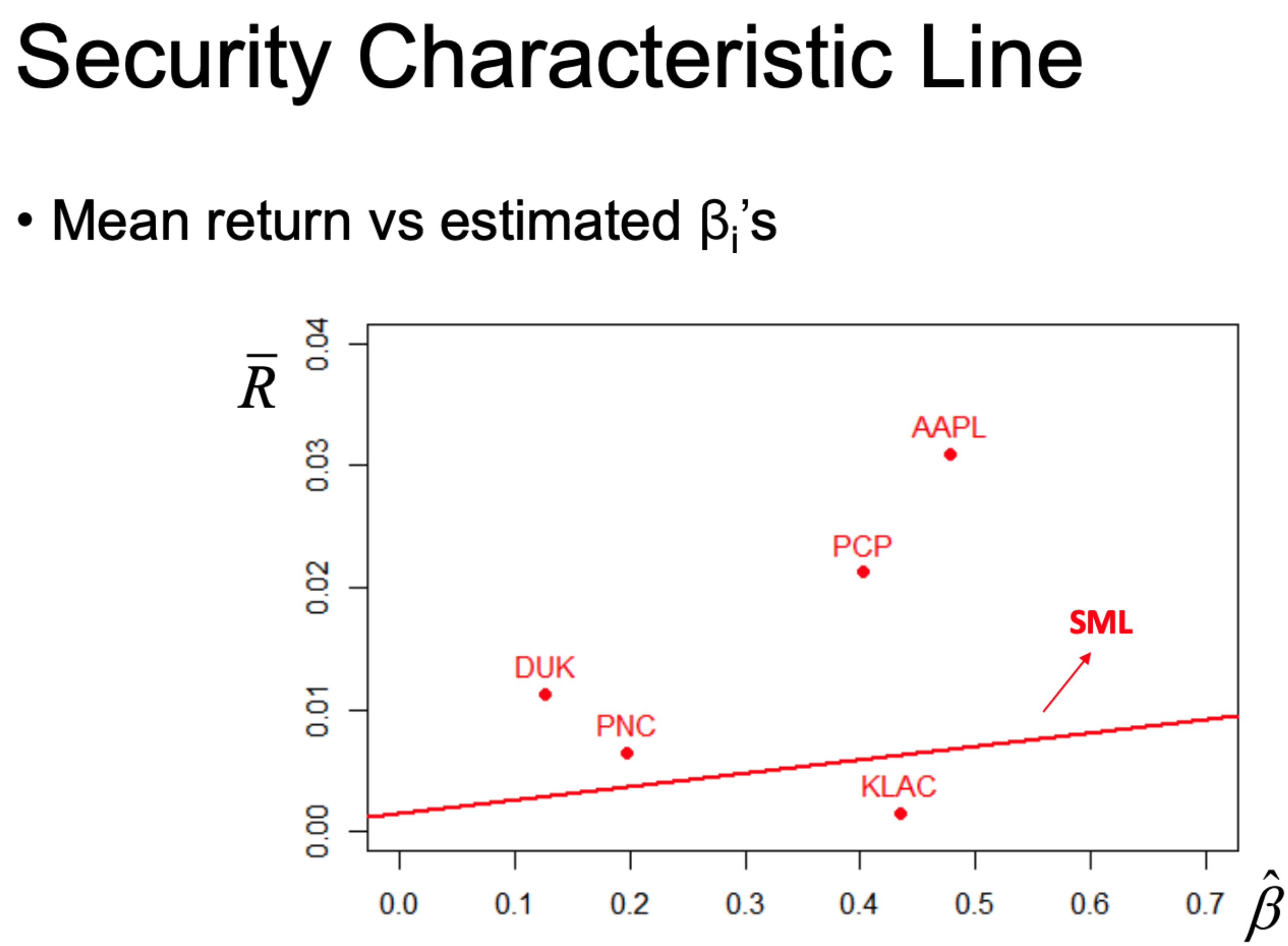
$$(R_{i,t} - R_{f,t}) = \beta_i(R_{M,t} - R_{f,t}) + \epsilon_t, \quad \text{where } \epsilon_t \sim N(0, \sigma^2_{\epsilon,i})$$

By fitting this, we can then extract the β_i's!

Security Characteristic Line



However the mean expected return is much different than the actual return. Therefore the CAPM model doesn't work very well, or hold.



An iteration of this model includes an intercept, α

$$(R_{i,t} - R_{f,t}) = \alpha_i + \beta_i(R_{M,t} - R_{f,t}) + \epsilon_t$$

Finding the mean and variance:

$$\begin{aligned}
\mu_i = \mathbb{E}[R_i] &= \mathbb{E}[R_{f,t} + \alpha_i + \beta_i(R_{M,t} - R_{f,t}) + \epsilon_t] \\
&= \underbrace{\mathbb{E}[R_f]}_{R_f} + \alpha_i + \beta_i \underbrace{\mathbb{E}[R_m - R_f]}_{\mu_{m-R_f}} + \underbrace{\mathbb{E}[\epsilon_t]}_{=0} \\
&= R_f + \alpha_i + \beta_i(\mu_m - R_f)
\end{aligned}$$

$$\begin{aligned}
\sigma_i^2 = \mathbb{V}[R_i] &= \mathbb{V}[\underbrace{R_f + \alpha_i}_{Var=0} + \beta_i(R_m - R_f) + \epsilon_i)] \\
&= \mathbb{V}[\beta_i(R_m - \underbrace{R_f}_{Var=0}] + \mathbb{V}[\epsilon_i] \\
&= \beta_i^2 \cdot \mathbb{V}[R_M] + \sigma_{\epsilon,i}^2 \\
&= \beta_i^2 \cdot \sigma_m^2 + \sigma_{\epsilon,i}^2
\end{aligned}$$

Time doesn't matter for the risk free or market returns

7.5.4.1 Alpha & Beta

An asset's beta β_i can be seen as a measure of both risk & reward

$$CAPM \to \mu_f = \mu_f + \beta_i(\mu_M - \mu_f) \implies \mu_i = \mu_f + \beta_i \underbrace{\left(\frac{\mu_M - \mu_f}{\sigma_M}\right)}_{\text{Sharpe Ratio}} \sigma_M$$

β_i measures extent of the return of a given asset is related to the market

> *If you believe in CAPM, then the bigger beta is, the more risk you will have but also the more reward you will have.*

α_i measure how much the asset outperforms the market consistently, the amount returned on top of the β_i

7.5.5 Legacy of CAPM

CAPM is wrong. A lot of it's assumptions do not hold in reality.

However, it had immense practical impact on investing, specifically in terms of - Diversification: concept of decreasing risk by spreading portfolio over different assets - Index investing: justification for common investing strategy of tracking some broad index with mutual funds or ETF's - Benchmarking: Measuring performance of investment relative to market / index

7.5.6 Performance Evaluation

CAPM says the best portfolio you can create is the tangency/market portfolio. This implies the best you can do is get the broadest index and combine it with a T-bill.

> Aside: In the past, you needed to use a mutual fund because exchange traded funds didn't exist. Thats how most people invested, even though they cost high fees as there was a lot of human involvement.

There are several ways to measure an asset's performance, based on CAPM

Sharpe ratio: (excess return per unit risk)

$$S_i = \frac{\mu_i - \mu_f}{\sigma_i}$$

- If you combine different assets, you can reduce the σ_i of the portfolio, so it's not objective because two assets may be really risky, but negatively correlated reducing σ_i.
- It doesn't however measure how correlated a portfolio is to the market.

Treynor index: (excess return per unit non-diversifiable risk)

$$T_i = \frac{\mu_i - \mu_f}{\beta_i}$$

- You would look at this to find the best return

Jensen's alpha: (excess return on top of the return explained by the market)

$$\alpha_i = \hat{\alpha}_i$$

- Usually the most important measure a portfolio manager tries to use to convince people to invest in them.

Chapter 8

Factor Models

8.1 Three Factor models

The CAPM model assumes the market is a single factor that drives asset returns. - In practice, CAPM does not adequately describe real-world returns

We can improve this model by including more factors in the regression.

There are three types of factor models we will look at:

- Macroeconomic: Factors are *observable* economic and financial time series data (eg return of the S&P 500)
- Fundamental: Created explicitly or implicitly from observable asset characteristics
- Statistical (Latent variable model): Factors are unobservable and extracted from asset returns

All three types follow some form of

$$R_i(t) = \beta_{i,0} + \beta_{i,1} F_1(t) + \cdots + \beta_{i,p} F_p(t) + \epsilon_i(t), \quad \forall \begin{cases} i = 1, \ldots, N \\ t \in \mathbb{R} \end{cases}$$

where:

- $R_i(t)$ is return on the i^{th} asset at time t
- $F_j(t)$ is the j^{th} common factor at time t
- $\beta_{i,j}$ is the factor loading/beta of i^{th} asset on the j^{th} factor
- $\epsilon_i(t)$ is the idiosyncratic/unique return of asset i^{th}

We want the errors to be independent from the factors,

$$Cov[\epsilon(t), F(t)] = 0$$

In matrix form,

$$
\begin{aligned}
&\Leftrightarrow\\
&\begin{bmatrix} R_1(t) \\ \vdots \\ R_N(t) \end{bmatrix} = \begin{bmatrix} \beta_{1,0} \\ \vdots \\ \beta_{N,0} \end{bmatrix} + \begin{bmatrix} \beta_{1,1} & \cdots & \beta_{1,p} \\ \vdots & \ddots & \vdots \\ \beta_{N,1} & \cdots & \beta_{N,p} \end{bmatrix} \begin{bmatrix} F_1(t) \\ \vdots \\ F_p(t) \end{bmatrix} + \begin{bmatrix} \varepsilon_1(t) \\ \vdots \\ \varepsilon_p(t) \end{bmatrix}\\
&\Leftrightarrow\\
&\mathbf{R}(t) = \beta_0 + \beta^\top \mathbf{F}(t) + \varepsilon(t)
\end{aligned}
$$

The factors $F_j(t)$ are stationary with moments:

$$
R(t) = \beta_0 + \beta^T F(t) + \epsilon(t)
$$

$$
\begin{aligned}
\mu_r &= \mathbb{E}(R(t)) = \mathbb{E}[\beta_0 + \beta^T F(t) + \epsilon(t)]\\
&= \beta_0 + \beta^T \underbrace{\mathbb{E}[F(t)]}_{\mu_F} + \underbrace{\mathbb{E}[\epsilon(t)]}_{0}\\
&= \beta_0 + \beta^T \mu_F
\end{aligned}
$$

$$
\begin{aligned}
\Sigma_R &= \mathbb{V}[\beta_0 + \beta^T F(t) + \epsilon(t)]\\
&\underbrace{= \mathbb{V}[\beta^T F(t)] + \mathbb{V}[\epsilon(t)]}_{\text{As } F(t) \text{ indep of } \epsilon_(t)}\\
&= \beta^T \underbrace{\mathbb{V}[F(t)]}_{\Sigma_F} \beta + \Sigma_\epsilon\\
&= \beta^T \Sigma_F \beta + \underbrace{\Sigma_\epsilon}_{\text{Diagonal}}
\end{aligned}
$$

We can also find the moments of the portfolio with weights $w = [w_1, \dots w_N]^T$

$$
R_{port} = w^T R \implies \begin{cases} \mu_{port} = \mathbb{E}[R_{port}] = w^T \mathbb{E}[R] = w^T(\beta_0 + \beta_{\mu_F}) \\ \sigma^2_{port} = \mathbb{V}[R_{port} = \mathbb{V}[w^T R] = w^T \mathbb{V}[R] w = w^T(\beta^T \Sigma_F \beta + \Sigma_\epsilon) w \end{cases}
$$

8.2 Factor Model Assumptions

- Factors $F_j(t)$ are stationary, with moments:

$$
E[\mathbf{F}(t)] = \mu_F \quad \& \quad \mathrm{Var}[\mathbf{F}(t)] = \Sigma_F
$$

- Asset-specific errors $\varepsilon_i(t)$ are uncorrelated with common factors:

$$
\mathrm{Cov}[(t), \mathbf{F}(t)] = \mathbf{0}
$$

- Errors are serially & contemporaneously uncorrelated across assets

$$
\mathrm{Var}[\varepsilon(t)] = \mathrm{diag}\left[\left\{\sigma^2_{\varepsilon_i}\right\}_{i=1,\dots,N}\right] = \Sigma_\varepsilon \quad \& \, \mathrm{Cov}[\varepsilon(t), \varepsilon(s)] = \mathbf{0}
$$

8.3 Time Series Regression Models

Consider model for which factor values are known (e.g. macro/fundamental model)

Estimate betas & risks (variances) one asset at a time, using time series regression

- For each fixed $i = 1, \dots N$, fit regression model:

$$R_i(t) == \beta_{i,0} + \beta_{i,1} F_1(t) + \cdots + \beta_{i,p} F_p(t) + \epsilon_i(t)$$

Most models will always include some proxy for the overall economy (eg the market)

Instead of Fama-French, there is BARRA which uses the cross section of returns? Still regression tho

8.3.1 Fama-French 3 Factor Model

Additionally to the market, they looked at two other factors that are consistent over many datasets and time periods.

Small Minus Big: (SMB) One factor tries to capture the size(market cap) of the company/stock, which empirically showed that size played a role in its return

- Regressed returns on how companies of a certain size/group did
  - They took the smallest and biggest companies, and looked at their avg return over some period, creating a difference to use to separate the groups of companies

High Minus Low (HML): The second factor looks at whether a company is a **value** one or not. Measured using book-to-market ratio.

These two factors were found to have statistically significant coefficients in multiple regression.

People frequently use the factor model to estimate the return covariance matrix. They can use the sample covariance matrix of the factors and apply the beta vector in order to obtain the return covariance matrix

$$\mathrm{Var}(\mathbf{R}) = \hat{\Sigma}_R = \hat{\beta}^\top \hat{\Sigma}_F \hat{\beta} + \hat{\Sigma}_\varepsilon$$

where:

$$\begin{aligned} \hat{\beta} &= \text{ beta coefficient matrix (from regressions)} \\ \hat{\Sigma}_F &= \text{ factor sample covariance matrix} \\ \hat{\Sigma}_\varepsilon &= \text{ diagonal error variance matrix (from residuals)} \end{aligned}$$

This gives more stable estimates than sample covariance.

8.4 Statistical Factor Models

In this model, the factors are **unknown (latent) and unobserved**. This implies that we need to estimate both β and F.

Unusual constraint: Because the factors are uncorrelated (not necessarily i.i.d.)

$$\Sigma_F = Cov(F) = I\sigma_F \quad \& \quad \mu_F = \mathbb{E}(F) = 0$$

This gives us the resulting moments of the returns:

$$\mu_R = \mathbb{E}[R] = \mathbb{E}[\beta_0 + \beta^T F + \epsilon] = \beta_0 + \beta^T \underbrace{\mathbb{E}[F]}_{0} + \underbrace{\mathbb{E}[\epsilon]}_{0} = \beta_0$$

$$\Sigma_R = \mathbb{V}[R] = \mathbb{V}[\beta_0 + \beta^T F + \epsilon] = \beta^T \mathbb{V}[F]\beta + \Sigma_\epsilon = \beta^T(I \cdot \sigma_F)\beta + \Sigma_\epsilon$$

8.5 Principle Component Analysis (PCA)

This technique helps reduce the dimensionality of a problem. Consider a factor model without errors:

$$\mathbf{R}(t) = \beta_0 + \beta^T \mathbf{F}(t) \implies \Sigma_R = \beta^T \Sigma_F \beta$$

Given a set of N assets, we can construct a set of variables (components) capturing most of the variability.

- PCA is a linear transformation
- The goal is to capture as much variation as possible.

When we apply PCA to the statistical model, we find n factors set as the PCA components. These components are **uncorrelated**, and have **maximum variance.**

$$F_1, \dots, F_n \text{ are factors which are the Principle Components}$$
$$F_1 = \gamma_1^T \mathbf{R} = \gamma_{11} R_1 + \cdots + \gamma_{1n} R_n$$
$$\vdots$$
$$F_n = \gamma_n^T \mathbf{R} = \gamma_{n1} R_1 + \cdots + \gamma_{nn} R_n$$

Where each factor:

$$F_i = \gamma_i^T \mathbf{X} \quad \text{Maximizes} \quad Var(F_i) = \gamma_i^T \Sigma_R \gamma_i \quad s.t. \gamma_i^T \gamma_i = 1$$

$$Cov(F_i, F_j) = \gamma_j^T \Sigma_R \gamma_i = 0$$

$$\text{Where } i, j \in 1, \dots, n, \quad i > j$$

Supposedly does not imply that $\gamma_j^T \gamma_i = 0$ which may mean they are not orthogonal? But PCA components should be orthogonal?

Principal Component Analysis

$$\begin{pmatrix} \mathbf{P}^{-1} = \mathbf{P}^{\mathsf{T}} \\ \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N \geq 0 \end{pmatrix}$$

- Solution given by eigen-decomposition of $\mathbf{\Sigma}_\mathrm{R}$

$$\mathbf{\Sigma}_R = \lambda_1 \mathbf{e}_1 \mathbf{e}_1^{\mathsf{T}} + \cdots + \lambda_N \mathbf{e}_N \mathbf{e}_N^{\mathsf{T}} = \mathbf{P\Lambda P}^{\mathsf{T}}$$

$$\left(\mathbf{P} = \begin{bmatrix} | & & | \\ \mathbf{e}_1 & \cdots & \mathbf{e}_N \\ | & & | \end{bmatrix} \text{ and } \mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_N \end{bmatrix} \right)$$

- Principal Components: $\mathbf{F} = \mathbf{P}^{\mathsf{T}}\mathbf{R} = \begin{bmatrix} - & \mathbf{e}_1^{\mathsf{T}} & - \\ & \vdots & \\ - & \mathbf{e}_N^{\mathsf{T}} & - \end{bmatrix} \mathbf{R} \Rightarrow$

PC 1: $F_1 = \mathbf{e}_1^{\mathsf{T}}\mathbf{R} = e_{11}R_1 + \cdots + e_{1N}R_N$

PC 2: $F_2 = \mathbf{e}_2^{\mathsf{T}}\mathbf{R} = e_{21}R_1 + \cdots + e_{2N}R_N$

$\vdots$

PC j: $F_j = \mathbf{e}_j^{\mathsf{T}}\mathbf{R} = e_{j1}R_1 + \cdots + e_{jN}R_N \quad (\forall j = 1, \ldots, N)$

31

We can find the $Cov(F)$ and beta (loading factor) of R_i on F_j:

$$\begin{aligned} \mathbb{V}(\mathbf{F}) &= \mathbb{V}(\mathbf{P}^T\mathbf{R}) \\ &= \mathbf{P}^T\mathbb{V}(\mathbf{R})\mathbf{P} \\ &= \mathbf{P}^T\Sigma_R\mathbf{P} \\ &\text{By Eigen-decomposition} \\ &= \mathbf{P^T}(\mathbf{PP^T})\mathbf{P} \\ &\text{Where } \mathbf{P^TP} = 1 \\ &= \Lambda \end{aligned}$$

It's kind of circular logic. Also, Λ is sorted with the eigenvalue associated with the eigenvector that captures the most variance is first.

Fact: Total variance of all PC's is equal to original variable variance.

$$\begin{aligned} \text{PC Total Variance} = \text{Population Total Variance} &\iff \\ tr(\Lambda) = tr(\Sigma) &\iff \\ \lambda_1 + \cdots + \lambda_P = \sigma_1^2 + \cdots + \sigma_N^2 & \end{aligned}$$

The proportion of total variance that is explained by each PC is:

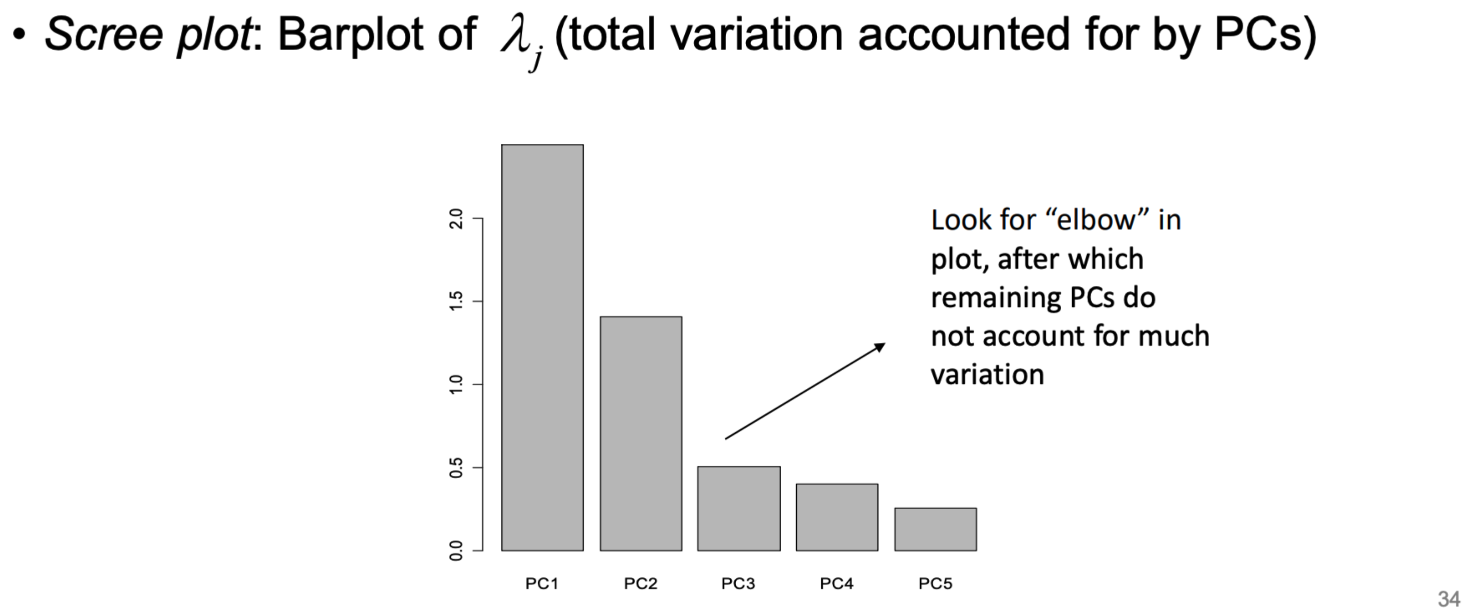
$$Var(PC_j) = \frac{\lambda_j}{\lambda_1 + \lambda_2 + \cdots + \lambda_N} \quad \text{Where } j = 1, \ldots, N$$

8.5.1 Selecting Components

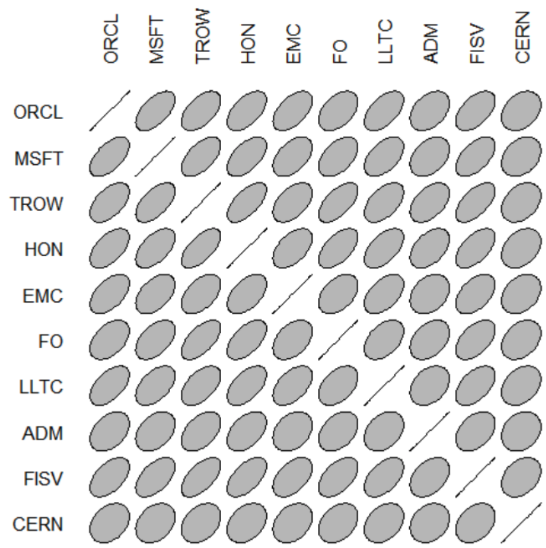
The important part of PCA is how many components to select to capture a suitable amount of variability.

You don't need many components if the data is very correlated, and a few components may represent most of the variance in the data.

We can also use a scree plot and pick the number of components before some "elbow" point.



- Correlation matrix of 10 stocks' returns



35

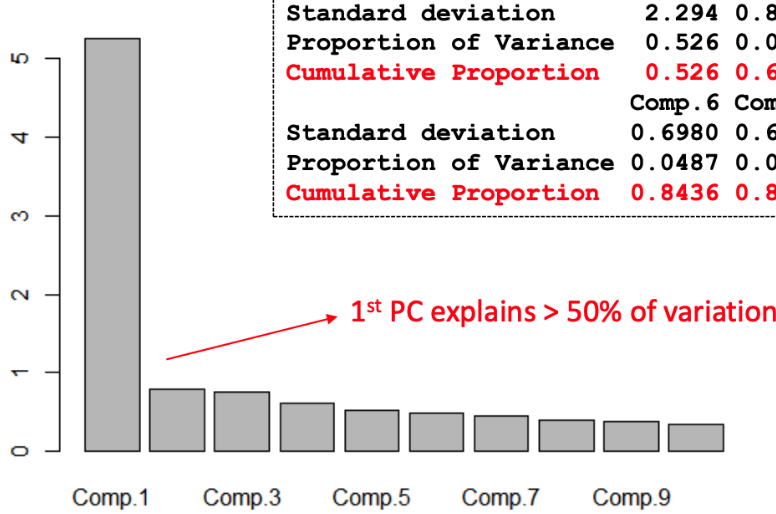
PCA run on the correlation matrix (which is the standardized covariance matrix, which is preferred when there is very large range of variances between variables)

- PCA suggests using 1PC:

Importance of components:

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
Standard deviation	2.294	0.8932	0.8682	0.7838	0.7195
Proportion of Variance	0.526	0.0798	0.0754	0.0615	0.0518
Cumulative Proportion	0.526	0.6062	0.6816	0.7430	0.7948

	Comp.6	Comp.7	Comp.8	Comp.9	Comp.10
Standard deviation	0.6980	0.6723	0.6264	0.6179	0.5810
Proportion of Variance	0.0487	0.0452	0.0393	0.0382	0.0338
Cumulative Proportion	0.8436	0.8888	0.9280	0.9662	1.0000

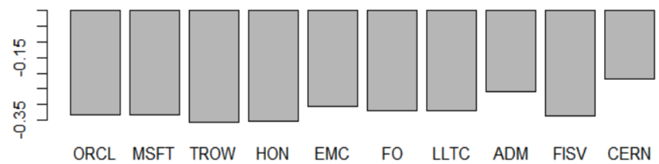


36

When we look at the first component, we can see it is almost an equal weighting of all the assets.

PCA Example

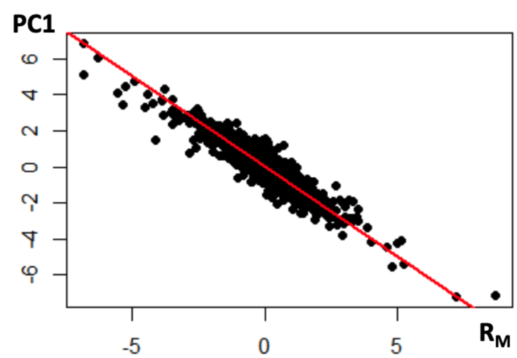
- PC1 loadings:



  - Almost equally weighted average (ignore sign)

- Plot of PC1 vs market returns:

$$\rho(R_M, PC1) = -0.932$$



37

We can use PCA to identify components that explain overall variation of the data, although it is not guaranteed to give meaning components. For a proper data-generating model, we must use Factor Analysis.

Principal Component Analysis

- PCA can be used to identify components that explain overall variation (behavior) of data
- PCA does not always give meaningful PC's
  - PCs are just transformations trying to capture the most variability
  - PCs are not meant to explain how data was generated
- For proper data-generating model, use Factor Analysis

39

8.6 Factor Analysis

He didn't go into much detail, just went over how the factor analysis works at a high level and how you can't really try to interpret the factors.

Factor Analysis

- Assume $\mathbf{\Sigma}_F = \mathbf{I} \rightarrow$ return variance becomes

$$\mathbf{\Sigma}_R = \boldsymbol{\beta}^\mathsf{T}\boldsymbol{\beta} + \mathbf{\Sigma}_\varepsilon$$

- Factor analysis splits return variance into:
  - *Communality*: $\boldsymbol{\beta}^\mathsf{T}\boldsymbol{\beta}$
  - *Uniqueness*: $\mathbf{\Sigma}_\varepsilon = \operatorname{diag}\{\sigma^2_{\varepsilon_i}\}_{i=1,\ldots,N}$
- Need to estimate **β** & unique variances $\sigma^2_{\varepsilon_i}$
  - Fit using Maximum Likelihood
  - Model can also be used to generate data

40

If the variance of the factors is simply the identity matrix, the variance of the returns is just the cross products of the factor variables plus some error variance.

Factor Analysis

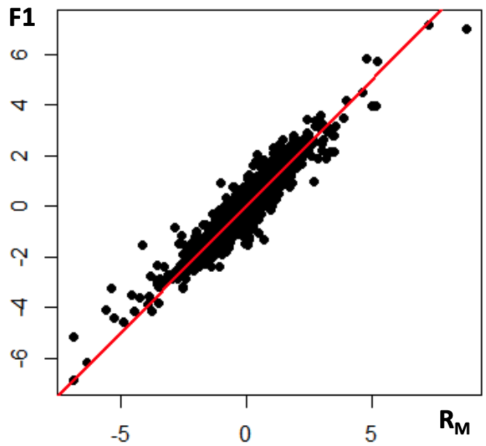
- Even with constraints, any *rotation* of **β** (i.e. multiple of *orthogonal* matrix $\mathbf{P}$ with $\mathbf{P}^{-1} = \mathbf{P}^{\mathsf{T}}$) results in the same model:

$$\mathbf{\Sigma}_R = (\boldsymbol{\beta}^{\mathsf{T}}\mathbf{P})(\mathbf{P}^{\mathsf{T}}\boldsymbol{\beta}) + \mathbf{\Sigma}_\varepsilon = \boldsymbol{\beta}^{\mathsf{T}}(\mathbf{P}\mathbf{P}^{\mathsf{T}})\boldsymbol{\beta} + \mathbf{\Sigma}_\varepsilon = \boldsymbol{\beta}^{\mathsf{T}}\boldsymbol{\beta} + \mathbf{\Sigma}_\varepsilon$$

- For estimation, need further constraints on **β**
  - Common constraint is to rank factors by "explained" variability, i.e. similar to PCA
  - *Rotation* constraints favor certain interpretations
    - E.g. *Varimax* rotation maximizes squared loadings → leads to few non-0 loadings for each factor

Factor Analysis Example

- F1 vs market returns

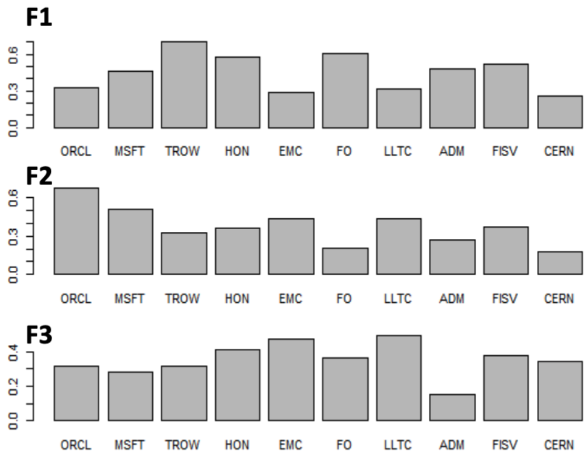


Factor Correlations
with Fama-French

	Factor1	Factor2	Factor3
XMT	0.939	-0.1032	-0.0215
SMB	0.180	0.0142	0.2013
HML	0.282	-0.3894	-0.0531

Factor Analysis Example

• Using Varimax rotation



Factor Correlations
with Fama-French

	Factor1	Factor2	Factor3
XMT	0.5442	0.2803	0.1995
SMB	-0.0116	-0.0532	0.2398
HML	0.4547	-0.1315	-0.0232

Chapter 9

Risk Measures

We will primarily focus on market risk.

Financial risk describes the possibility of an investment losing money. It can have several sources, the main ones being:

- Market risk: due to changes in market prices
- Credit risk: counterparty doesn't honor obligations
- Liquidity risk: lack of asset tradability
- Operational risk: from organization's internal activities
  - E.g., legal, fraud, or human error risk
- Risk management is the process of identifying, measuring, and controlling risk

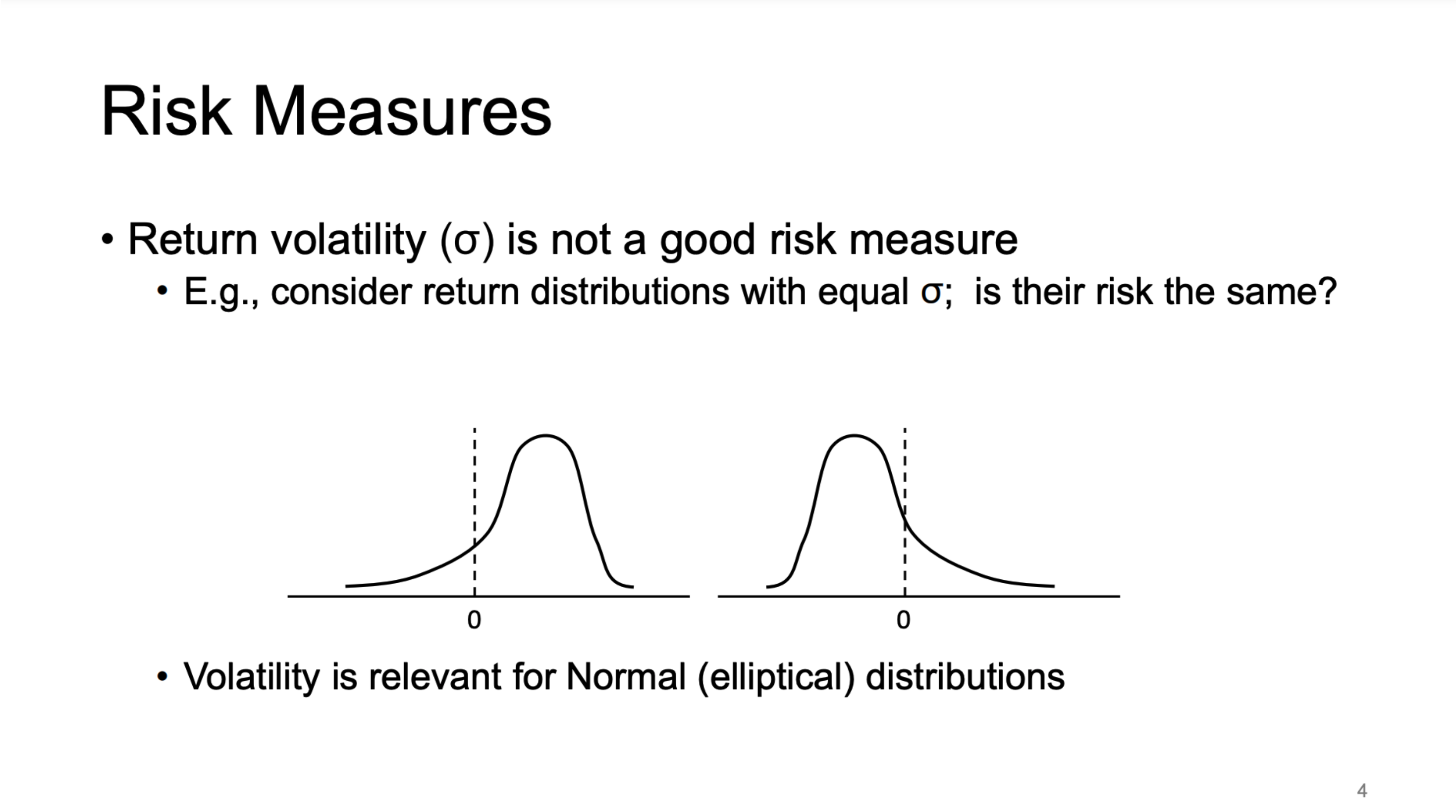
The first thing we want to do is quantify the risk somehow.

To operationalize risk management, we use risk measures to define some capital requirement that the organization should hold in a liquid asset/cash.

The first thing that comes to mind is measuring risk by the variance of returns. **But it is not a good risk measure**. In the following example, the variances are the same, but clearly the risk is different.

WIth the left distribution, the long tail goes to the left which means your returns could become large negative values. (Reasonable to assume it is "more risky")

For the right distribution, the average return is negative although there is no chance of very large negative returns. (Implies that it is "less risky")



The point of risk mgmt: We want to know the amount of money we should set aside to cover expected losses.

9.1 Formulas

Assuming $Port \sim N(\mu, \sigma^2)$

$$VaR_\alpha = \mu + z_\alpha \cdot \sigma$$
$$ES_\alpha = \mu + \frac{\phi(z_\alpha)}{\alpha} \cdot \sigma$$
$$EVaR_\alpha = \mu + \sigma \cdot \sqrt{-2\ln(\alpha)}$$

9.2 Value At Risk (VAR)

Was really common, although recently it has decreased in use (discredited) although there are variations still in use.

Let RV L be the loss of an investment over some time period T

- Loss is negative of gain/revenue $(R) : L = -R$

  We look at losses because we only capture the negative losses from this asset

Value at Risk at confidence level $(1-\alpha)$ & time horizon T is defined as the:

$(1-\alpha)$**-quantile of** L **for some** $\alpha \in (0,1)$

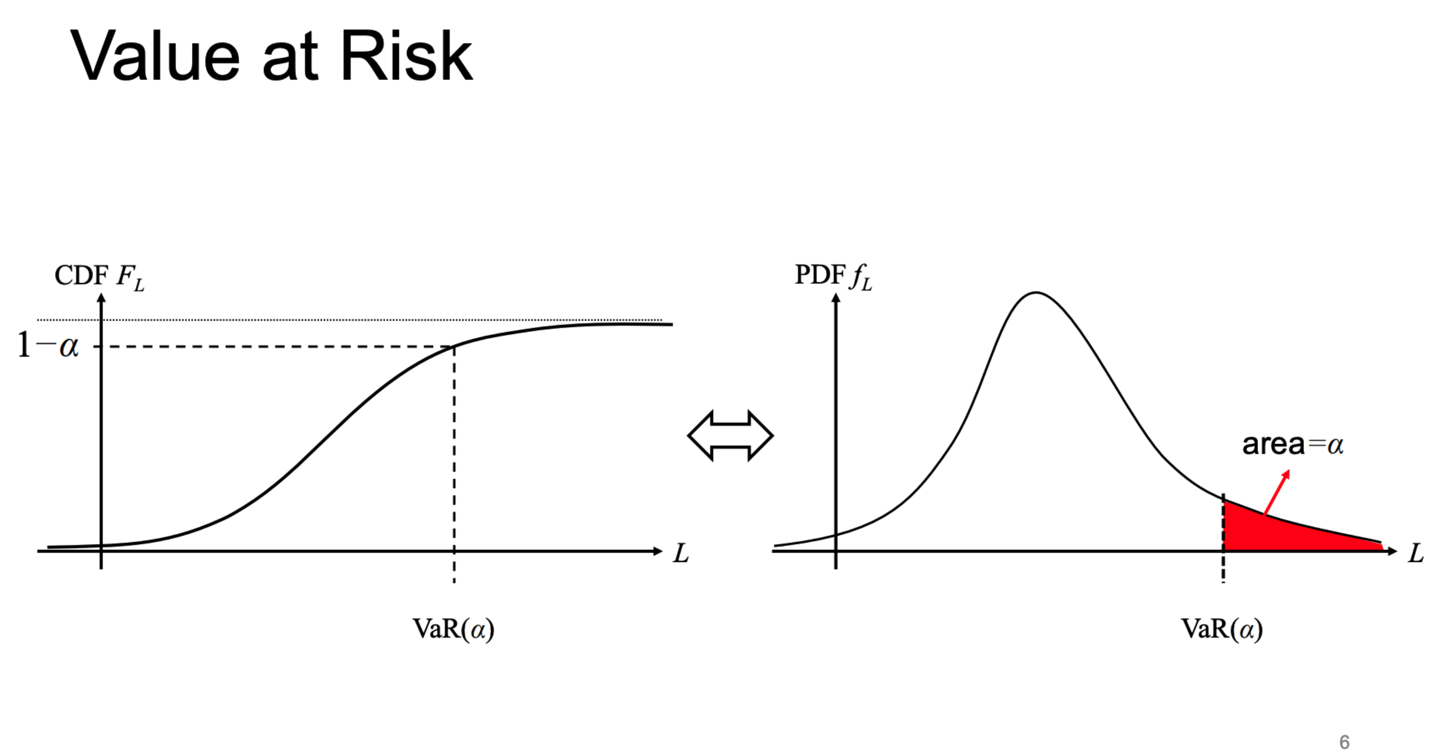
$$\text{VaR}(\alpha) = \inf\{x : P(L \leq x) \geq 1-\alpha\} = \inf\{x : P(L > x) \leq \alpha\}$$

Infimum (smallest value of x) is only used for discrete distributions such that the the probability of L will not exceed some value x and we want this to happen $(1-\alpha)$ of the time. (Usually want it to be less than 5%)

- For continuous RV with CDF F_L , we don't need the infimum anymore and can just do: $\Rightarrow \text{VaR}(\alpha) = F_L^{-1}(1-\alpha)$

  Basically the quantile of the distribution

- VaR represents amount that covers losses with probability $(1-\alpha)$



We want the quantile at which the probability is at $(1-\alpha)$

The VaR(α) value indicates what value of x we need to hold to cover our losses $(1-\alpha) = 95\%$ of the time.

9.2.1 Example

Consider asset with $N(\mu = .03, \sigma^2 = .04)$ annual log-returns. Find the 95% confidence level annual VaR for a S_0 =$1000 investment in this asset. Let $R = S_T - S_0$ represent revenue (-losses)

We want to find the VaR(α) s.t

$$\begin{aligned}
P(L > VaR) &= 5\% \\
&= P(-R > VaR) \\
&= P(R < -VaR)
\end{aligned}$$

$$\begin{aligned}
X &\sim N(0.03, \sigma^2 = 0.04) \\
\implies S_T &= S_0 \cdot e^X
\end{aligned}$$

$$\begin{aligned}
P(S_T - S_0 < -VaR) &= P(S_T < S_0 - VaR) \\
&= P(S_0 e^X < S_0 \cdot VaR) \\
&= P\left(X < \log\left(\frac{S_0 - VaR}{S_0}\right)\right) \\
&= P\left(\underbrace{\frac{X - 0.03}{.2}}_{Z \sim N(0,1)} < \frac{\log\left(1 - \frac{VaR}{S_0}\right) - 0.03}{.2}\right) = 0.05 \\
&= P\left(Z < \underbrace{\frac{\log\left(1 - \frac{VaR}{S_0}\right) - 0.03}{.2}}_{z}\right) = 0.05 \\
\implies z &= -1.645 \\
\frac{\log\left(1 - \frac{VaR}{S_0}\right) - 0.03}{.2} &= -1.645
\end{aligned}$$

$$\begin{aligned}
\mathbf{VaR} &= (1 - \exp\{-1.645 \cdot 0.20 + 0.03\}) \cdot S_0 \\
\mathbf{VaR} &= (1 - \exp\{-0.299\}) \cdot 1000 \\
&= \mathbf{258.44}
\end{aligned}$$

VaR is basic and widely used as a risk measure to simply indicate

> *"What is the most I can lose on this investment?"*

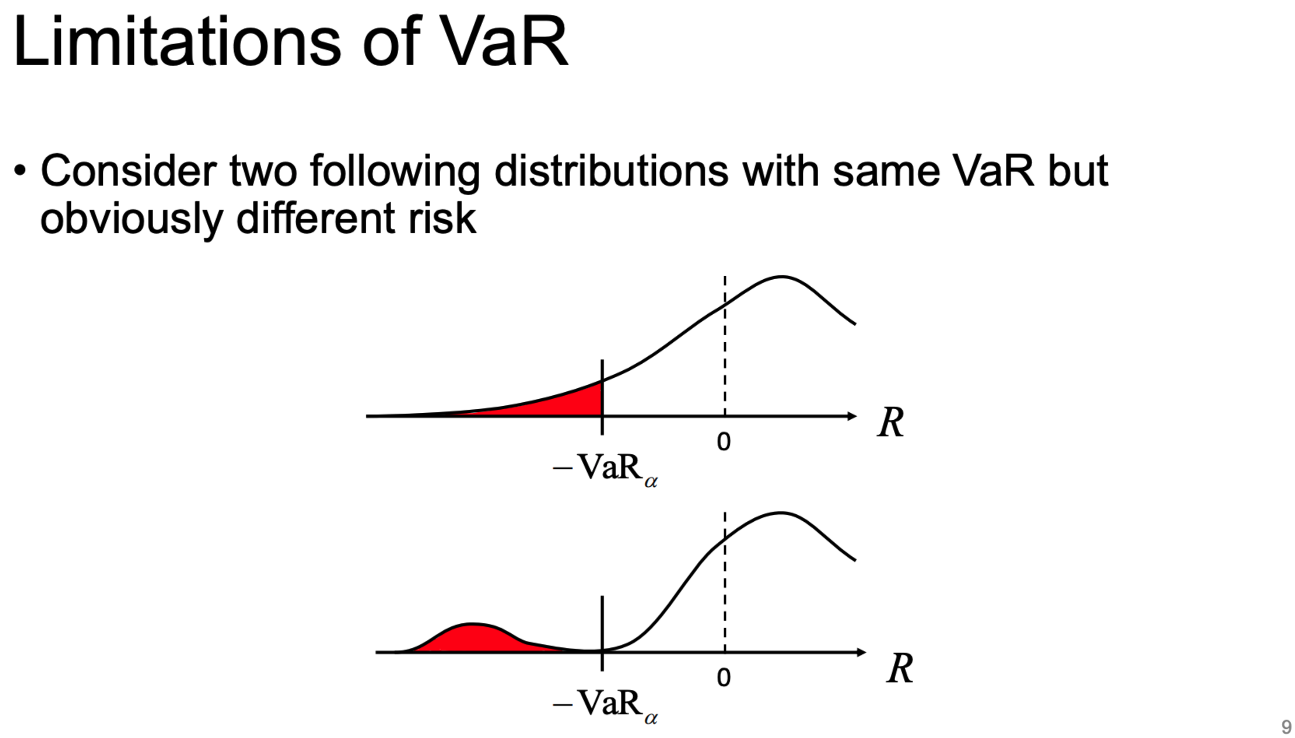
It has been widely used partly because of the Basel framework which is some risk management standard for banks and banks implemented Basel to comply.

However it has some glaring issues. No one trusts it as a proper risk management tool.

Flaws: - VaR hides tail risk (hides risk of very extreme events) although there is a way to remedy it - It discourages diversification

The additional resources talk about why VaR is not a good measure of risk: [[../../Lectures/Risk Mismanagement.pdf]] [[../../Lectures/Against VaR.pdf]]

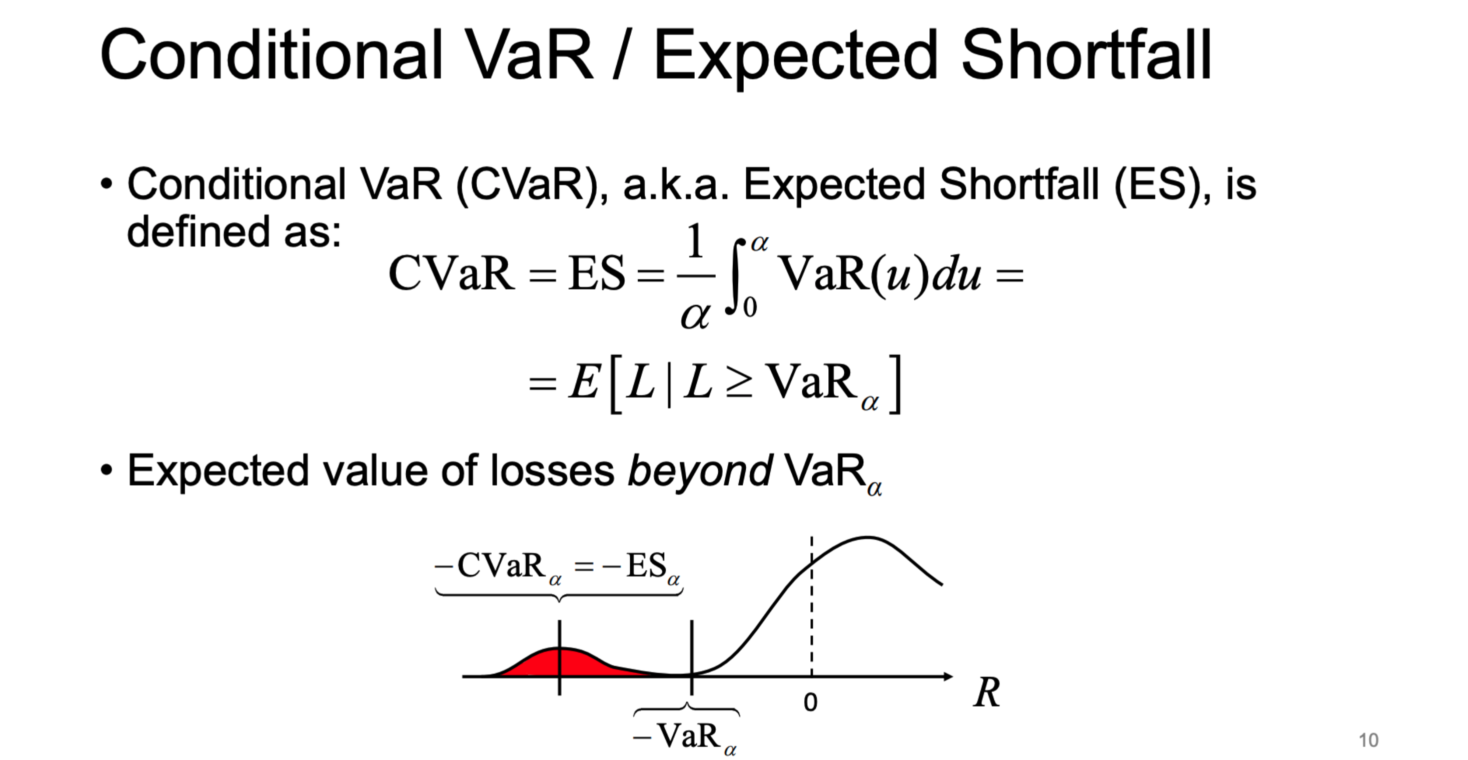
9.3 VaR hides tail risk



Var doesn't care what happens beyond the cut off, just the actual z-score (or whatever score from some distribution). If we push some density of losses very far (hiding a small probability of really bad events), we can actually concentrate the density of profits higher.

9.3.0.1 Mitigation: Cond VaR or Expected Shortfall

AKA CVar or ES



This gives us the expected losses that can occur given that we are beyond VaR_α

> Imagine we cut off the loss distribution beyond the VaR(α)= x point, and we have another little distribution of losses which are way larger/further from the bulk of the distribution. We take the mean/average/expected value of this distribution which tells us "hey in the 5% scenario our losses are so huge, what can we **expect** to lose?"

9.3.0.2 Example

If $R \sim N(0, 1)$, find ES at confidence level α

$R \sim N(0,1) \implies L \sim N(0,1)$

Let $Z_\alpha$denote the α-quantile of the std normal

$$
\begin{aligned}
ES_\alpha = \mathbb{E}[L|L > Z_\alpha] &= \int_{Z_\alpha}^{\infty} x\ \phi(x|L > z_\alpha)\, dx \\
&= \int_{Z_\alpha}^{\infty} x \underbrace{\frac{\phi(x)}{P(L > Z_\alpha)}}_{\alpha} dx \\
&= \frac{1}{\alpha}\int_{Z_\alpha}^{\infty} x \frac{1}{2\sqrt{\pi}} e^{-x^2/2}\, dx \\
&= \frac{1}{\alpha}\int_{Z_\alpha}^{\infty} \frac{1}{2\pi} \lfloor -e^{-x^2/2} \rfloor\, dx \\
&= \frac{1}{\alpha}\frac{1}{2\pi}\left[-e^{-x^2/2}\right]_{x=z_\alpha}^{\infty} \\
\implies ES_\alpha = CVaR_\alpha &= \frac{1}{\alpha}\frac{1}{\sqrt{2\pi}} e^{-z_\alpha^2/2} \\
&= \frac{1}{\alpha}\phi(z_\alpha)
\end{aligned}
$$

This means that generally for $L \sim N(\mu, \sigma^2)$,

$$
\begin{aligned}
ES_\alpha &= \mu + \frac{\phi(z_\alpha)}{\alpha} \cdot \sigma \\
VaR_\alpha &= \mu + z_\alpha \cdot \sigma
\end{aligned}
$$

and for small α, $\frac{\phi(z_\alpha)}{\alpha} > z_a$ which implies $ES_\alpha \geq VaR_\alpha$

9.4 Risk Measure Properties

We now look at formal set of requirements for risk measures (5 properties)

We let $\rho(L)$ represent the risk measure, which takes in a distribution of a some returns and outputs a number indicating the capital requirements (money to be kept aside) should be. We let L is a random variable and let ρ is a functional distribution of L.

For this measure ρ to reasonably quantify risk, it must:

1. be normalized $\rho(0) = 0$ (risk of holding no assets is 0)
2. Translation invariance: $\rho(L + c) = \rho(L) + c\ \forall c \in \mathbb{R}$
   - adding a loss c to the portfolio increases risk by exactly c
3. Positive Homogeneity: $\rho(bL) = b\rho(l)$
   - scaling portfolio returns also will scale risk
4. Monotonicity: $L_1 \geq L_2 \implies \rho(L_1) \geq \rho(L_2)$
   - The ordering of the random variables is almost surely $P(L_1 \geq L_2) = 1$

5. Sub-additivity: $\rho(L_1 + L_2) \leq \rho(L_1 + L_2)$
   - Only when the two losses are perfectly correlated, then equal
   - The risk of two combined portfolios cannot exceed the sum of the two portfolio risks

If a risk measure has all 5 properties, then it is called a **coherent risk measure**.

9.4.1 Example that VAR and CVar satisfy prop 2 & 3

We first show that VAR and CVaR satisfy the second and third property of translation invariance and positive homogeneity.

$$VaR_\alpha(L) = inf\{x : P(L > x) \leq \alpha\}$$

$$\begin{aligned}
\text{Let } \ L' &= bL + c \\
\implies VaR(L') &= inf\{x' : P(L' > x') \leq \alpha\} \\
&= inf\{x' : \mathbb{P}(bL + c > x') \leq \alpha\} \\
&= inf\left\{x' : \mathbb{P}\left(L > \frac{x' - c}{b} \leq \alpha\right)\right\} \\
&= inf\{bx + c : \mathbb{P}(L > x) \leq \alpha\} \\
&= b \cdot inf\{x : \mathbb{P}(L > x) \leq \alpha\} + c \\
&= bVaR(L) + c
\end{aligned}$$

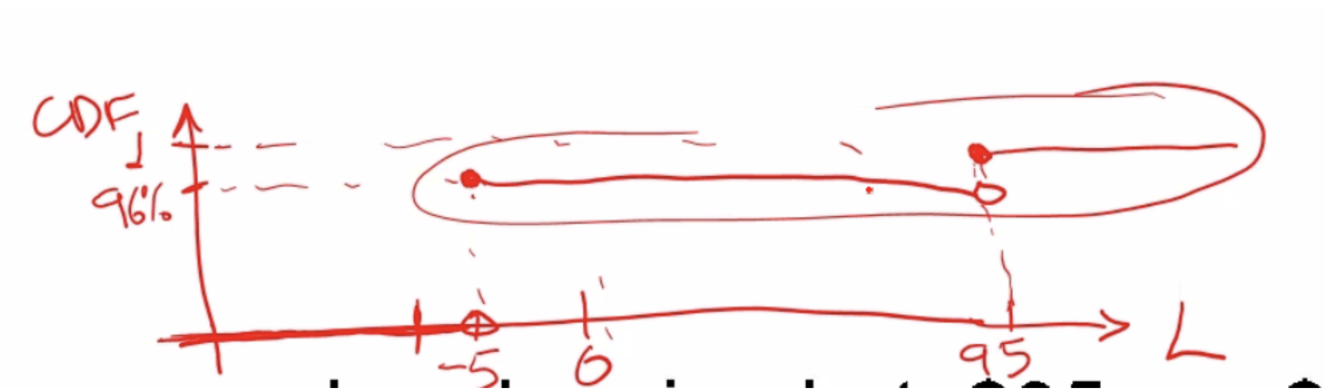
$$\begin{aligned}
CVaR_\alpha(L) &= \frac{1}{\alpha}\int_0^\alpha VaR_u(L)\, du \\
\implies CVaR_\alpha(L') &= \frac{1}{\alpha}\int_0^\alpha (bVaR_u(L) + c)\, du \\
&= b\left(\frac{1}{\alpha}\int_0^\alpha VaR_u(L)\, du\right) + c \\
&= b \cdot CVaR_\alpha(L) + c
\end{aligned}$$

We will show that VaR is not sub-additive

Consider two risky zero-coupon bonds priced at \$95 per \$100 face value. If each one has 4% independent default probability, show that $VaR_{5\%}$ is not sub-additive

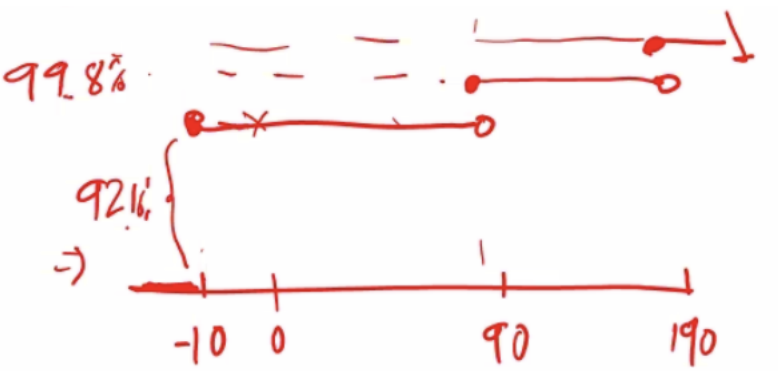
$$\begin{aligned}
\text{Distribution of } L_{1/2} &= \begin{cases} -5 & w.p.96\% \\ +95 & w.p.4\% \end{cases} \\
\text{By definitioin, VaR} &= inf\{x : \mathbb{P}(L > x) < 5\%\} \\
&= inf(x : \mathbb{P}(L \leq x) \geq 95\%) \\
\implies VaR_{5\%}(L_{1/2}) &= -5
\end{aligned}$$

As -5 is the smallest value that satisfies $\mathbb{P}(L \leq x) \geq 95\%$ as $\mathbb{P}(L \leq -5) = 96\%$



$$\text{Distribution of } L_1 + L_2 = \begin{cases} -5-5=-10 & w.p. 96\%^2 = 92.16\% \quad \text{when both don't default} \\ -5+95=90 & w.p. 96\% \cdot 4\% = 7.68\% \quad \text{when only one defaults} \\ 95+95=190 & w.p. 4\%^2 = .16\% \quad \text{both default} \end{cases}$$

$$VaR_{5\%}(L_1+L_2) = inf\{x : \mathbb{P}(L_1+L_2 < x) \geq 95\%\} = 90$$



$$\implies VaR_{5\%}(L_1+L_2) = 90 > VaR_{5\%}(L_1) + VaR_{5\%}(L_2) = -5-5 = 10$$

Which suggests holding both bonds is more risky than holding them separately.

We can show that CVaR is sub-additive.

$$\begin{aligned} CVaR_{5\%}(L_1) &= \frac{1}{\alpha}\int_0^{\alpha} VaR_u(L_1)\,du \\ &= \frac{1}{5\%}(95 \cdot 4\% + (-5)\cdot 1\%) = 73\% \\ CVaR_{5\%}(L_1+L_2) &= \frac{1}{5\%}\int_0^{5\%} VaR_u(L_1+L_2)\,du \\ &= \frac{1}{5\%}[190 \cdot .16\% + 90 \cdot 4.84\%] = 93.2 \end{aligned}$$

$$\begin{aligned} \implies CVaR_{5\%}(L_1+L_2) &= 93.2 \leq CVaR_{5\%}(L_1) + CVaR_{5\%}(L_2) \\ &= 2 \cdot 73 = 146 \end{aligned}$$

9.5 Entropic VaR

A coherent alternative to VaR based on the Chernoff bound for tail probability

For positive RV X, the Chernoff inequality gives:

$$P(X \geq c) \leq \frac{\mathbb{E}(x)}{c} \quad \forall c > 0$$

For loss RV L with mgf $M_L(z) = E\left[e^{zL}\right] < \infty, \forall z > 0$ we have $P(L \geq c) \Leftrightarrow P(e^{zL} \geq e^{zc}) \leq M_L(z)e^{-zc}$ Limiting bound to $M_L(z)e^{-zc} \leq \alpha$ and solving for c we get $c = z^{-1} \ln\left(M_L(z)/\alpha\right)$

EVaR defined as: $\text{EVaR}_\alpha = \inf_{z>0}\left\{z^{-1}\ln\left(M_L(z)/\alpha\right)\right\}$

> Can define a risk measure we can use the MGF instead of the PDF or CDF which in some cases is convenient. We can just minimize the MGF

9.5.1 EVaR for a Normal Distribution

$$\begin{aligned}
\text{Let } L &\sim N(\mu, \sigma^2) \\
M_L(z) &= e^{\mu z + 1/2\sigma^2 z^2} = \mathbb{E}[e^{z \cdot L}] \quad \forall z
\end{aligned}$$

$$\begin{aligned}
\text{Know: } EVaR_\alpha &= inf_{z>0}\left\{z^{-1}\ln(M_L(z)/\alpha)\right\} \\
\text{Let } f(z) &= z^{-1}\ln(M_L(z)/\alpha) \\
&= z^{-1}\ln(e^{\mu z + 1/2\sigma^2 z^2}/\alpha) \\
&= \frac{1}{z}\left[\mu z + \frac{\sigma^2 z^2}{2} - \ln\alpha\right] \\
&= \mu + \frac{z\sigma^2}{2} - \frac{\ln(\alpha)}{z}
\end{aligned}$$

So to find the smallest z, we take the derivative:

$$\begin{aligned}
\implies f'(z) &= \frac{\sigma^2}{2} + \frac{\ln(\alpha)}{z^2} \\
\implies z &= \frac{\sqrt{-2\ln(\alpha)}}{\sigma}
\end{aligned}$$

$$\begin{aligned}
\implies EVaR_\alpha(L) &= \mu + \frac{\sigma^2}{2} \cdot \frac{\sqrt{-2\ln(\alpha)}}{\sigma} - \frac{\ln(\alpha)}{\frac{\sqrt{-2\ln(\alpha)}}{\sigma}} \\
&= \mu + \sigma \cdot \sqrt{-\frac{\ln(\alpha)}{2}} + \sigma \frac{\sqrt{(-\ln(\alpha))^2}}{\sqrt{-2\ln(\alpha)}} \\
&= \mu + \sigma \cdot \sqrt{-\frac{\ln(\alpha)}{2}} + \sigma \underbrace{\frac{\sqrt{(\ln(\alpha^{-1}))^2}}{\sqrt{2\ln(\alpha^{-1})}}}_{=\sqrt{\frac{\ln(\alpha^{-1})}{2}} = \sqrt{-\frac{\ln(\alpha)}{2}}} \\
&= \mu + 2\sigma \cdot \sqrt{-\frac{\ln(\alpha)}{2}} \\
&= \mu + \sigma \cdot \sqrt{-2\ln(\alpha)}
\end{aligned}$$

9.6 Calculating Risk Measures

1. Parametric Modelling
   - Not used because theres usually no closed form distribution for the losses
2. Historical Simulation
   - Not favoured because past performance does not predict future performance
3. Monte Carlo Simulation
   - 85% of large banks use historical simulation, while the remaining use MC simulation
   - Usually done by sampling from some estimated parametric model for individual assets

Other ways to test are stress-testing (the worst case scenario) and use the Extreme Value Theorem

> Rest of the slides were not gone over in detail, just showed how the CVaR performed better than VaR (gave lower z-score -> allocation more money for risk purposes), how parametric modelling can't really be done, how time series models are independent over time but can correlate between assets and show volatility clustering, how RiskMetrics is a time series model but didnt get into it, and the fact GARCH models are used.

Chapter 10

Betting Strategies

There are betting/gambling games where a player has an edge, which just says your expected payoff > 0.

Consider roulette, where the "house" has an edge: - P(win) = 1/37 - odds: 35:1 >In this scenario, the house has a slight edge as the odds don't directly reflect the actual probability.

Consider blackjack, in the 60's, Ed Thorp, a mathematician figured out how to beat the house. Unfortunately, casino's have strategies against card counting, like using multiple decks, burning cards, dealing only some of a deck of cards, etc.

Given a sequence of bets, we can use a **kelly criterion** to determine how to bet optimally.

10.1 Problem Setup

Consider a sequence of independent and identical gambles. $P(win) = \frac{1}{2} < p < 1 \implies P(lose) = q = 1 - p$

For each \$1 bet: - Get \$1 if you win - Lose \$1 if you lose.

By starting with an initial wealth V_0 what is the best strategy for placing bets?

10.1.1 Scenario 1: Bet fixed amount

Bet a fixed amount X at every step. *We ignore ruin ($V_t \leq 0$ for some $t > 0$) for now*

The expected wealth V_n after n steps:

$$
\begin{aligned}
\text{Let } I_i &= \begin{cases} 1 & \text{w.p. } p \\ 0 & o/w \end{cases} \\
V_1 &= V_0 + (2I_1 - 1)X \\
V_2 &= V_1 + (2I_2 - 1)X \\
&= V_0 + (2I_1 - 1)X + (2I_2 - 1)X \\
&\vdots \\
V_n &= V_0 + X\left(2 \underbrace{\sum_{i=1}^{n} I_i}_{\sim\text{Binomial(n,p)}} - n\right) \\
\implies \mathbb{E}(V_n) &= \mathbb{E}\left(V_0 + X\left(2\sum_{i=1}^{n} I_i - n\right)\right) \\
&= V_0 + X\left(2\mathbb{E}\left(\sum_{i=1}^{n} I_i\right) - n\right) \\
&= V_0 + X(2np - n) \\
&= V_0 + nX(2p - 1) \\
\text{Since } \frac{1}{2} &\leq p < 1 \implies 2p - 1 > 0
\end{aligned}
$$

With this probability and bet X per step, then we can expect to have some positive wealth at the n^{th} step which increases linearly, and the variance increases quadratically, but the probability of ruin is very small as $p > \frac{1}{2}$

We didn't consider the probability of ruin, as if V_0 is very large and X is very small, the probability of ruin is very small.

10.1.1.1 Hidden probability of ruin

By betting \$1 at each step with V_0 starting wealth, the probability of eventual ruin at some time step n is:

$$
\begin{aligned}
&\text{Let } \pi_i = \mathbb{P}(\text{eventual ruin for } V_0 = i) \quad \forall\ i \geq 1 \\
&\text{Let } \pi_0 = 1 \quad \text{if you start with no money, you are already ruined} \\
\\
&\implies \pi_i = \pi_{i+1} \cdot p + \pi_{i-1} q \quad \forall i \geq 1
\end{aligned}
$$

Which is a second order linear recurrence relation. This can solved like a differential equation.

We will assume a solution of the form: $\pi_i = y^i$

$$
\begin{aligned}
\implies y^i &= py^{i+1} + qy^{i-1} \\
y &= py^2 + q \\
=> py^2 - y + q &= 0
\end{aligned}
$$

Which leaves a quadratic equation:

$$\begin{aligned}
y &= \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \\
&= \frac{-(-1) \pm \sqrt{(-1)^2 - 4pq}}{2p} \\
&= \frac{1 \pm \sqrt{1 - 4p(1-p)}}{2p} \\
&= \frac{1 \pm \sqrt{(2p-1)^2}}{2p} \\
\implies y &= \begin{cases} \frac{1+2p-1}{2p} = \frac{2p}{2p} = 1 & \text{reject} \\ \frac{1-2p+1}{2p} = \frac{1-p}{p} = \frac{q}{p} \end{cases} \\
\implies \pi_i &= y^i = \left(\frac{q}{p}\right)^i \quad \forall i \geq 0
\end{aligned}$$

Intuitively, if $q < p$ then the probability of ruin decreases geometrically fast. If they're equal, then the chances of ruin are also practically 0.

10.1.2 Scenario 2: Bet entire wealth

What is $\mathbb{E}(V_n)$ after n steps?

$$V_n = \begin{cases} V_0 \cdot 2^n & w.p. p^n \\ 0 & o/w \end{cases}$$

$$\begin{aligned}
\mathbb{E}(V_n) &= 2^n V_0 p^n + 0(1 - p^n) \\
&= V_0 \cdot (2p)^n \to \infty \\
&\text{Because } p > \frac{1}{2} \implies 2p > 1
\end{aligned}$$

We've achieved exponential growth rate by betting everything every step, but should only be used if $p = 1$ or you are very very desperate. We are combining very extreme events which means the expected value is kind of meaningless.

10.1.3 Scenario 3: Exp growth w/o ruin

Betting a fixed fraction of wealth f at each step.

$$\begin{aligned}
V_1 &= \begin{cases} V_0 \cdot (1+f) & \text{win} \\ V_0 \cdot (1-f) & \text{lose} \end{cases} \\
\implies V_i &= \begin{cases} V_{i-1}(1+f) & \text{win with } p \\ V_{i-1}(1-f) & \text{lose with } q \end{cases} \\
V_n &= V_{n-1} \cdot (1+f)^{I_n} \cdot (1-f)^{1-I_n} \\
&= V_0 \cdot (1+f)^{\sum_{i=1}^n I_i} \cdot (1-f)^{n - \sum_{i=1}^n I_i} \\
&\text{Let } w = \sum_{i=1}^n I_i \\
&= V_0 (1+f)^w (1-f)^{n-w} \quad \text{where } W \sim Bin(n, p)
\end{aligned}$$

So the expected value becomes:

$$\begin{aligned}
\mathbb{E}(V_n) &= \mathbb{E}(V_0(1+f)^w(1-f)^{n-w}) \\
&= V_0 \cdot (1-f)^n \mathbb{E}\left(\frac{1+f^w}{1-f}\right)
\end{aligned}$$

From Prob Gen Fnc (Laplace Trans) of Bin(n,p)
$G_w(z) = \mathbb{E}(z^w) = (q+pz)^n$

$$\begin{aligned}
\implies \mathbb{E}(V_n) &= V_0 \cdot (1-f)^n \underbrace{\mathbb{E}\left(\frac{1+f^w}{1-f}\right)}_{G_w\left(\frac{1+f}{1-f}\right)} \\
&= V_0 \cdot (1-f)^n \left(q + p\left(\frac{1+f}{1-f}\right)\right)^n \\
&= V_0 \cdot (q(1-f) + p(1+f))^n \\
&= V_0 \cdot (1 - qf + pf)^n \\
&= V_0(1 + f \cdot (p-q))^n \\
&= V_0(1 + f\underbrace{(2p-1)}_{>0})^n
\end{aligned}$$

Which gives geometric growth and has very low bankruptcy probability.

10.2 Kelly Criterion

We can use this criterion to determine the optimal fraction f.

Definition: The fraction f is the amount that maximizes the expected log returns.

The log can be interpreted as because you are compounding your wealth, you don't just want to maximize the next step but the entire sequence of bets.

By maximizing the *expected log-return* $\leftrightarrow$ maximizing expected log of V_n $\leftrightarrow$ maximizing geometric avg of returns.

By jensens inequality,

$E\left[\log\left(V_n\right)\right] \neq \log\left(E\left[V_n\right]\right)$

10.2.1 Finding the Optimal Value

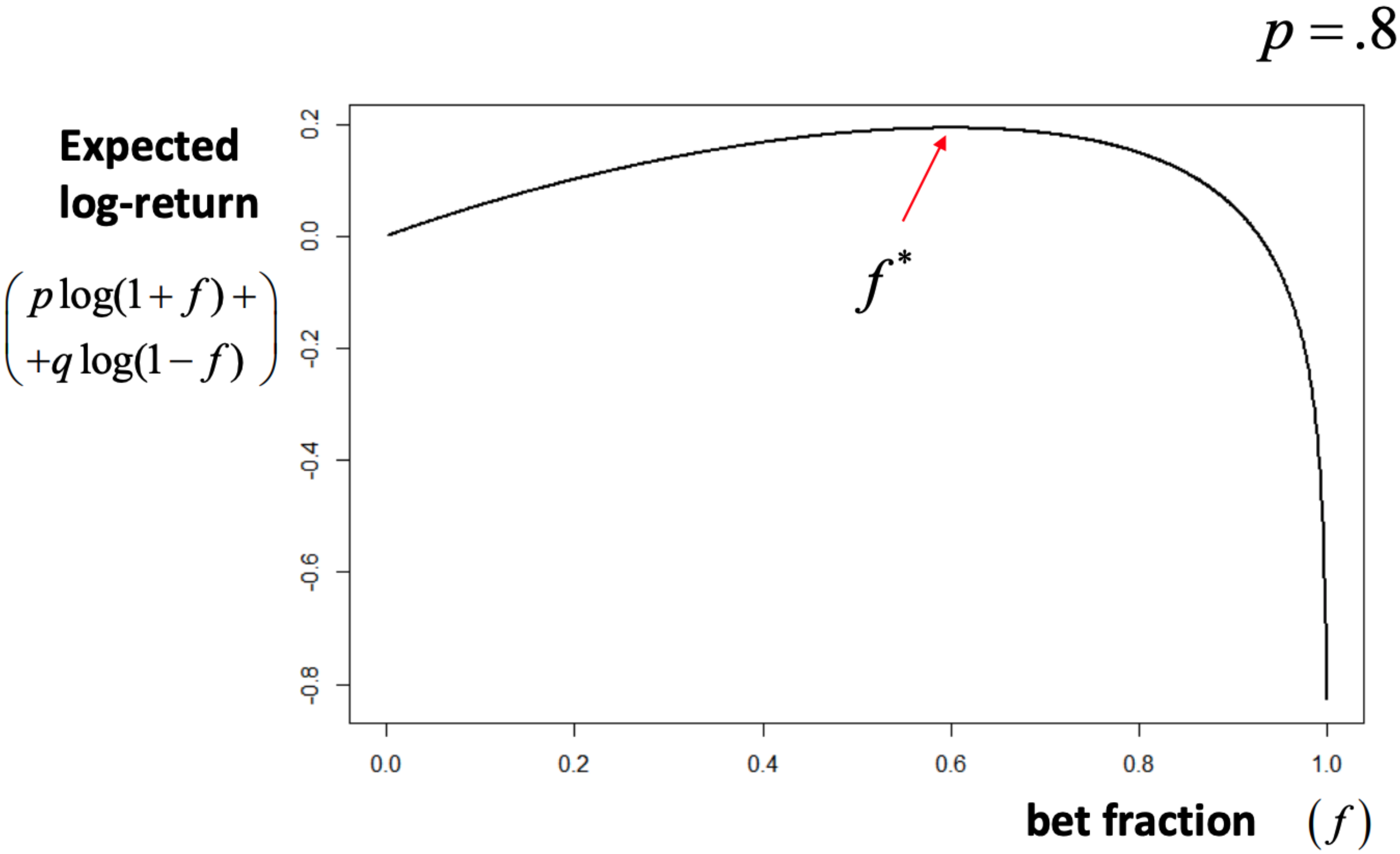
Finding the optimal value of fraction f^* according to the Kelly criterion

$$
\begin{aligned}
\mathbb{E}\left(\log\left(\frac{V_n}{V_0}\right)\right) &= \mathbb{E}\left(\log\left(\frac{V_0\cdot(1+f)^w\cdot(1-f)^{n-w}}{V_0}\right)\right)\\
&= \mathbb{E}[w\log(1+f)) + (n-w)\log(1-f)]\\
&= \log(1+f)\mathbb{E}(w) + \log(1-f)(n-\mathbb{E}(w))\\
&= \log(1+f)np + \log(1-f)(\underbrace{n-np}_{n(1-p)=nq})\\
&= \underbrace{\log(1+f)np + \log(1-f)nq}_{G(f)}
\end{aligned}
$$

We want to maximize $G(f)$ with respect to the fraction f. We can do this by taking the derivative:

$$
\begin{aligned}
\frac{\delta G(f)}{\delta f} &= \frac{\delta}{\delta f}(\log(1+f)np + \log(1-f)nq)\\
\text{Set } 0 &= n\left(\frac{1}{1+f}p + \frac{1}{1-f}(-1)q\right)\\
\implies \frac{p}{1+f} &= \frac{p}{1-f}\\
p(1-f) &= q(1+f)\\
p-q &= f\cdot(p+q)\\
&\implies \mathbf{f^* = p - q = 2p - 1}
\end{aligned}
$$

So the fraction you bet depends on *the edge you have on this bet*.



10.2.2 Geometric average

Assuming a fixed fraction f of wealth is bet at each time step. What is the geometric avg of the returns as $n \to \infty$?

$$
\begin{aligned}
&\text{Let growth rate } r_i \text{ be } : \\
&\text{For } i-1 \to i : r_i = \frac{V_i}{V_{i-1}} \\
\text{Geometric Average: } &: (r_1 \times r_2 \cdots \times r_n)^{1/n} = \left[\prod_{i=1}^{n} (1+f)^{I_i} (1-f)^{1-I_i}\right]^{1/n} \\
&= (1+f)^{\overbrace{\sum_{i=1}^{n} I_i/n}^{W_n}} \times (1-f)^{1/n \sum_{i=1}^{n}(1-I_i)} \\
&= (1+f)^{W_n/n} \times (1-f)^{1-W_n/n} \quad \text{where } W_n \sim Binom(n,p)
\end{aligned}
$$

So as $n \to \infty$, $\frac{W_n}{n} \to p$ from LLN, which implies the geometric average converges to:

$$(1+f)^p(1-f)^{1-p}$$

from the continuous mapping theorem. This is also the expression where if you take the log, you get $\underbrace{\log(1+f)np + \log(1-f)nq}_{G(f)}$ from the [[#Finding the Optimal Value]] section.

10.2.3 Simple example

Consider general sequence of bets, where

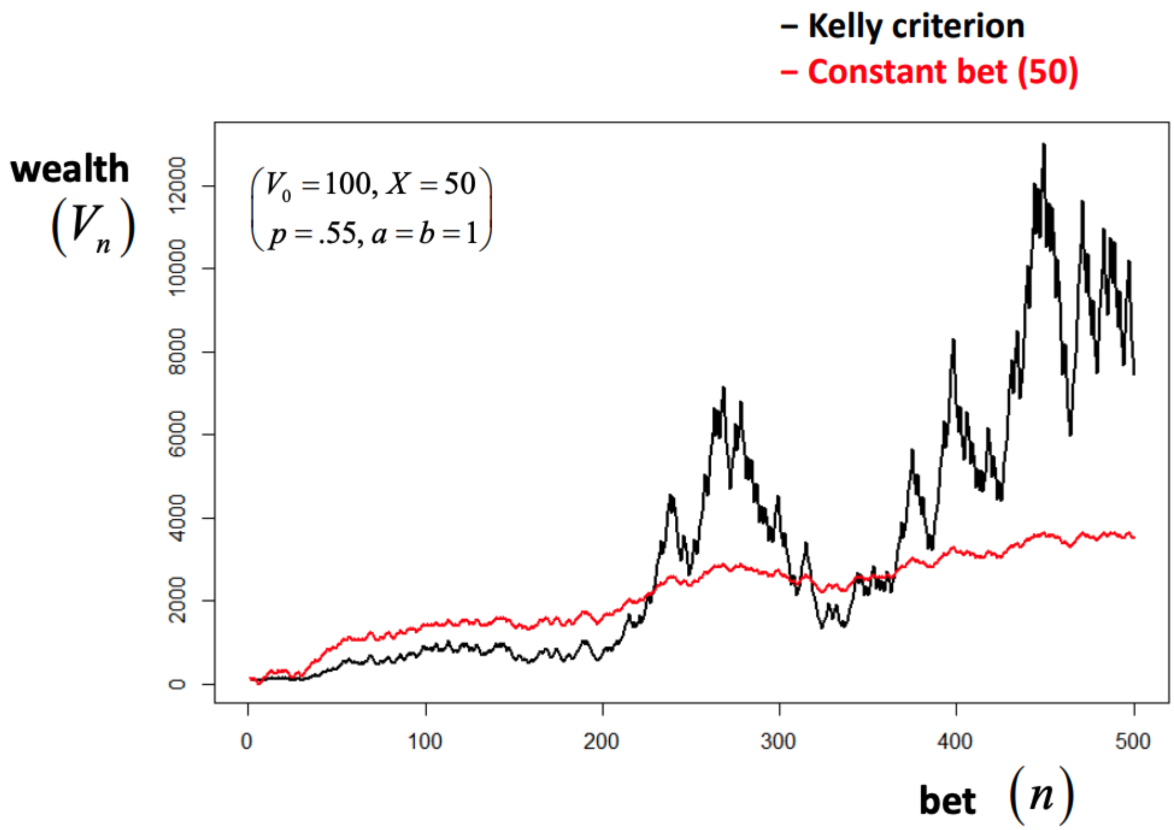
- $1bet \implies +\$a$ if win and $-\$b$ if lose
- For previous examples $a = b = 1$
- $P(win) = p$ and $P(lose) = q$
- Bet is favorable, i.e. $p \cdot a > q \cdot b$
- Kelly criterion optimal fraction to bet is
$$f^* = \frac{pa - qb}{ab}$$

So

$$\text{Given } a = 1, b = 1$$

$$f^* = \frac{1p - 1q}{1 \cdot 1} = p - q = .55 - .45 = .10$$

Example



10.2.4 Kelly Criterion in Investing

Consider a situation where you have: - risk free asset with return r_i - One risky asset with return R_i, with $\mathbb{E}(R) = \mu, \mathbb{V}(R) = \sigma^2$

Assume we invest fraction f in the risky asset, and $1 - f$ in the risk free asset.

We first need to show that:

$$\mathbb{E}\left[\log\left(\frac{V_n}{V_0}\right)\right] \approx \log(1 + r_f) + f\left(\frac{\mu - r_f}{1 + r_f}\right) - f^2\left(\frac{\sigma^2 + (\mu - r_f)^2}{2(1 + r_f)^2}\right)$$

$$\begin{aligned}
\text{Know: } V_1 &= V_0 \cdot f \cdot (1+R) + V_0(1-f)(1+r_f) \\
&= V_0[1 + fR + (1-f)r_f] \\
&= V_0[1 + r_f + f\underbrace{(R - r_f)}_{\text{Excess ret}}] \\
\implies \frac{V_1}{V_0} &= \log(1 + r_f + f(R - r_f))
\end{aligned}$$

$$\begin{aligned}
\implies \log\left(\frac{V_n}{V_0}\right) &= \log\left(\prod_{t=1}^{n} \frac{V_t}{V_{t-1}}\right) \\
&= \sum_{t=1}^{n} \log(1 + r_f + f(R_t - r_f))
\end{aligned}$$

$$\begin{aligned}
\implies \mathbb{E}\left[\log\left(\frac{V_n}{V_0}\right)\right] &= \sum_{t=1}^{n} \mathbb{E}[\log(1 + r_f + f(R_t - r_f))] \\
&\text{Assuming i.i.d. returns} \\
&= n\mathbb{E}[\log(1 + r_f + f(R - r_f))]
\end{aligned}$$

We can simplify this using a taylor approx of the function $\log(1 + r_f + f(R - r_f))$ around $1 + r_f$:

$$\underbrace{\log}_{g}\left(\underbrace{1 + r_f}_{x_0} + \underbrace{f(R_t - r_f)}_{\delta}\right) \approx \underbrace{\log(1 + r_f)}_{g(x_0)} + \underbrace{\frac{1}{1 + r_f} f(R_t - r_f)}_{g'(x_0)} + \underbrace{\frac{1}{2}\left(\frac{f^2}{(1 + r_f)^2}\right)\delta^2}_{g''(x_0)}$$

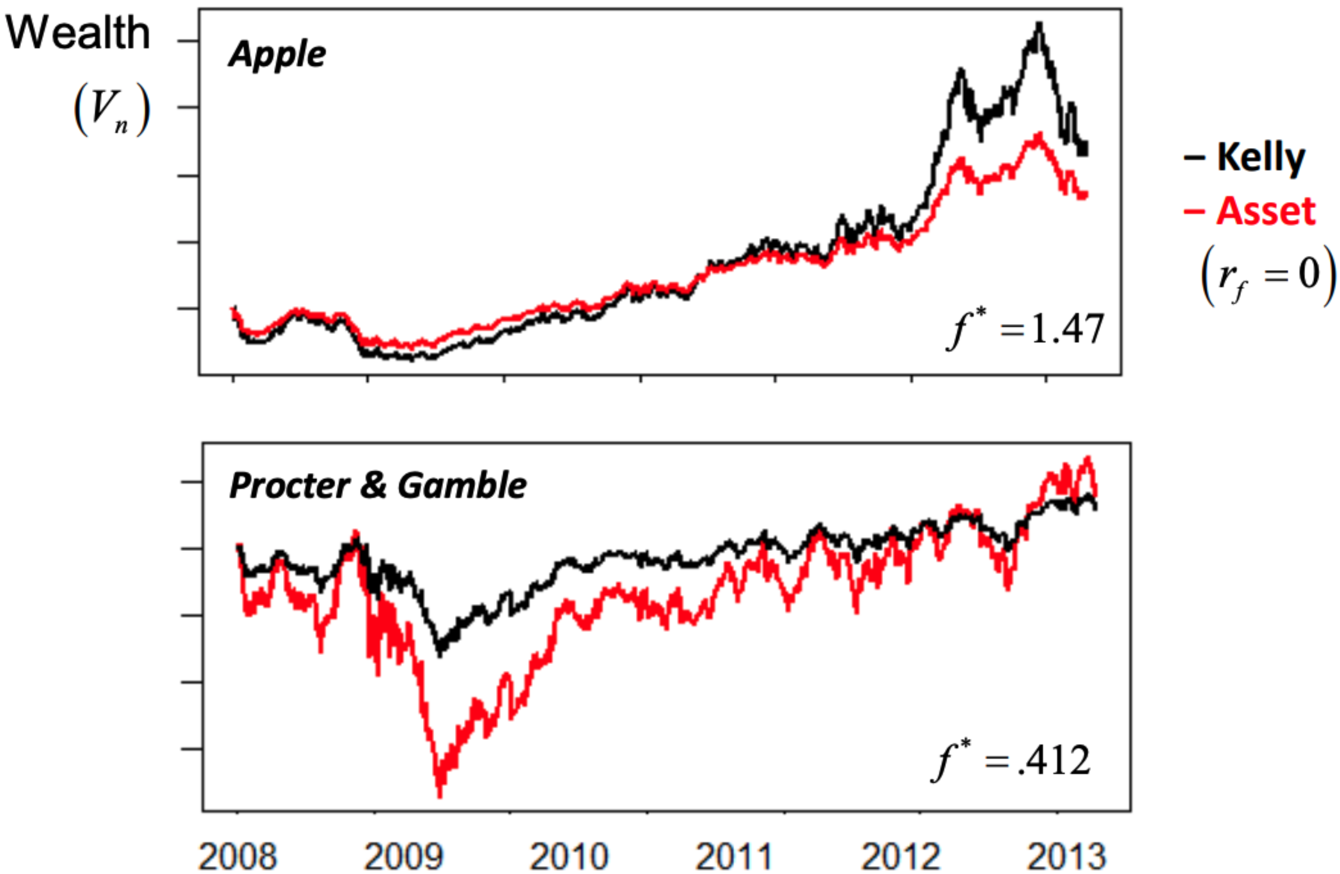
$$\begin{aligned}
\mathbb{E}[g(x_0 - \delta)] &\approx \log(1 + r_f) + \frac{1}{1 + r_f} f \underbrace{\mathbb{E}(R_t - r_f)}_{\mu - r_f} + \frac{1}{2}\left(\frac{f^2}{(1 + r_f)^2}\right)\mathbb{E}[(R_t - r_f)^2] \\
&\text{Where } \mathbb{E}[(R_t - r_f)^2] = (\sigma^2 + \mu^2 + r_f^2 - 2\mu r_f)
\end{aligned}$$

We can now show that $f^* \approx (1 + r_f)\frac{(\mu - r_f)}{\sigma^2}$

$$\text{Let } G(f) = \log(1 + r_f) + \frac{f}{1 + r_f}(\mu - r_f) - \frac{f^2}{2(1 + r_f)^2}(\sigma^2 + \mu^2 + r_f^2 - 2\mu r_f)$$

$$\implies \mathbf{f^* \approx (1 + r_f)\frac{- r_f}{2}}$$

10.2.4.1 Example in investing



10.2.5 Theoretical properties

In the long term $n \to \infty$, with probability 1, a strategy based on a Kelly criterion will: - Maximize the limiting exponential growth rate of wealth - Maximizes median of final wealth - Half of distribution is above median & half below it - Minimizes the expected time required to reach a specified goal for the wealth

10.2.6 Criticisms

It tends to be risky

It works well if you have clear information (what exactly your edge is) and that you have an infinite sequence of gambles

In practice, people will take half of the kelly criterion, as a precaution against high volatility

Most people retire at some point and therefore do not have infinite gambles

Chapter 11

Statistical Arbitrage

Mathematical definition of a "free lunch". In the real world, there is never 0 risk. Statistical arbitrage uses "statistical mispricing" of assets to make money. This involves long and short positions simultaneously and are typically short-term and market-neutral.

Examples:

- Pairs Trading
- Index Arbitrage
- Volatility Arbitrage
- Algorithmic and High Frequency Trading

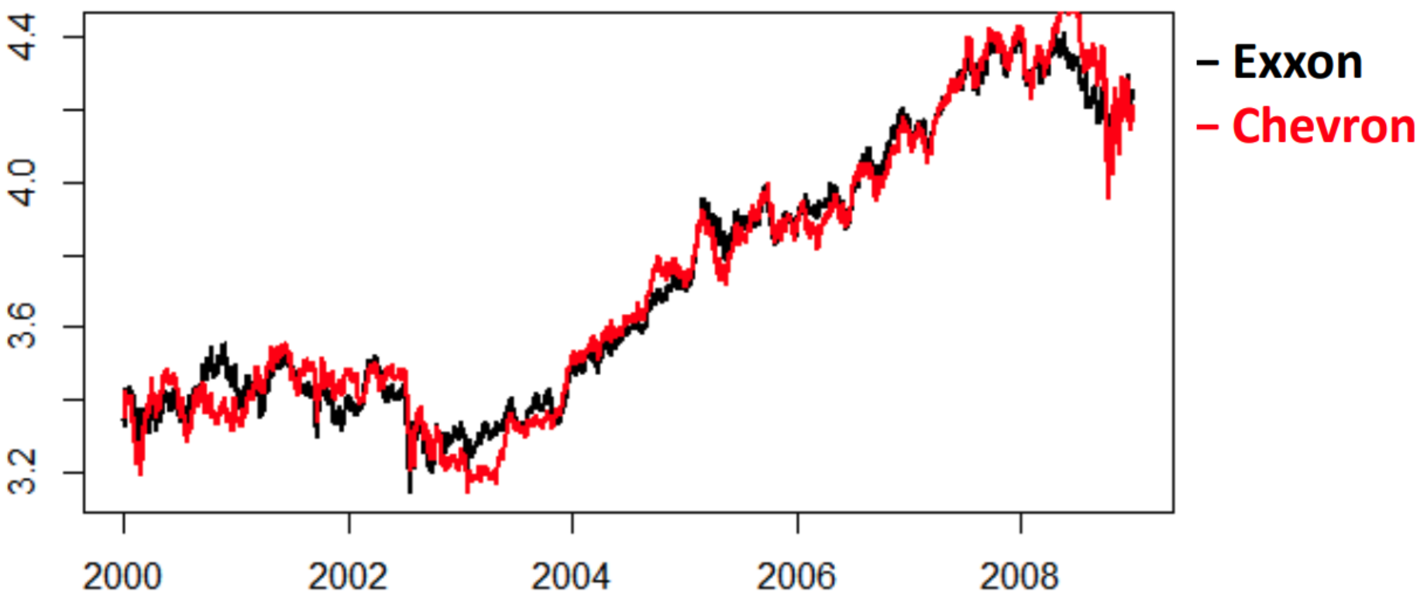
11.1 Pairs Trading

Developed in the 80's by Morgan Stanley quants, who made ~50 Million in profits in 1987.

The idea: Find pairs of stocks that tend to move together, and once they diverge, you expect them to come back together. If they cross some threshold of divergence, buy the low priced one and sell the higher priced one until they converge.

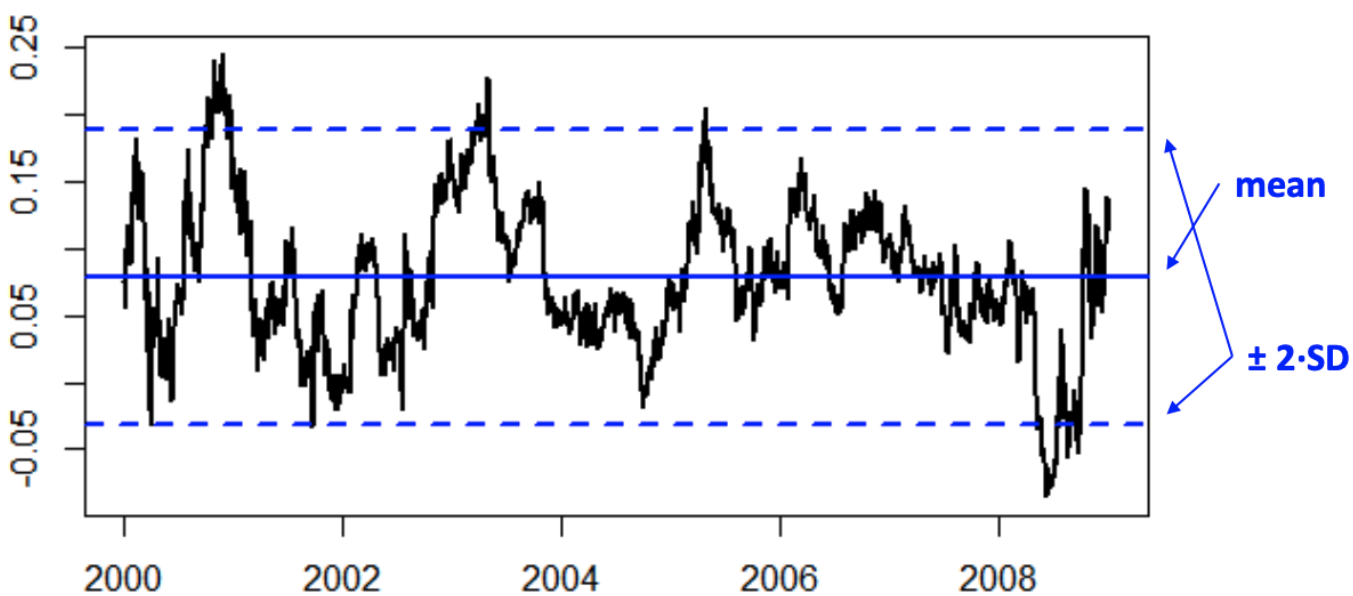
Problem: If the prices don't converge.

Example of two pairs:



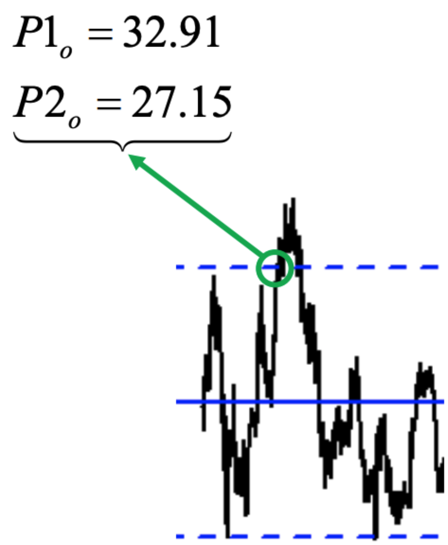
We use log asset prices because it's easier to model with Brownian motion.

$$\log\left(\text{Exxon}/\text{ Chevron}\right) = \log\left(P1/P2\right) = \log(P1) - \log(P2)$$



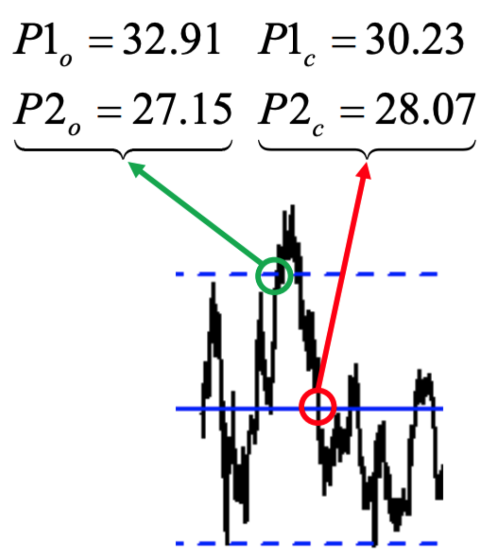
Using this chart, we can decide the threshold at which to trade the two stocks:

- When stock prices get "out of balance", take opposite positions in stocks
  - Buy \$1 worth of low asset
  - Sell \$1 worth of high asset
  - ⇨ Cost of position is 0 (assuming no transaction/short selling costs)
- For our example:
  - Buy $1/P2_o$ units of Chevron
  - Short-sell $1/P1_o$ units of Exxon



- Key: Buy and sell the **same dollar amount**.
- Buy $\frac{1}{P2_0}$ where $P2_o$ is the opening price of stock 2

- Unwind position when assets get "back in balance" and collect profit
  - Sell long & buy short asset
- For our example:
  - Sell $1/P2_c$ units of Chevron @ $P2_c$
  - Buy $1/P1_c$ units of Exxon @ $P1_c$
  - Profit = ?



Per \$1 invested, the profit will be:

$$\frac{1}{P2_o} \times P2_c - \frac{1}{P1_o} \times P1_c = \frac{28.07}{27.15} - \frac{30.23}{32.91} = .11532$$

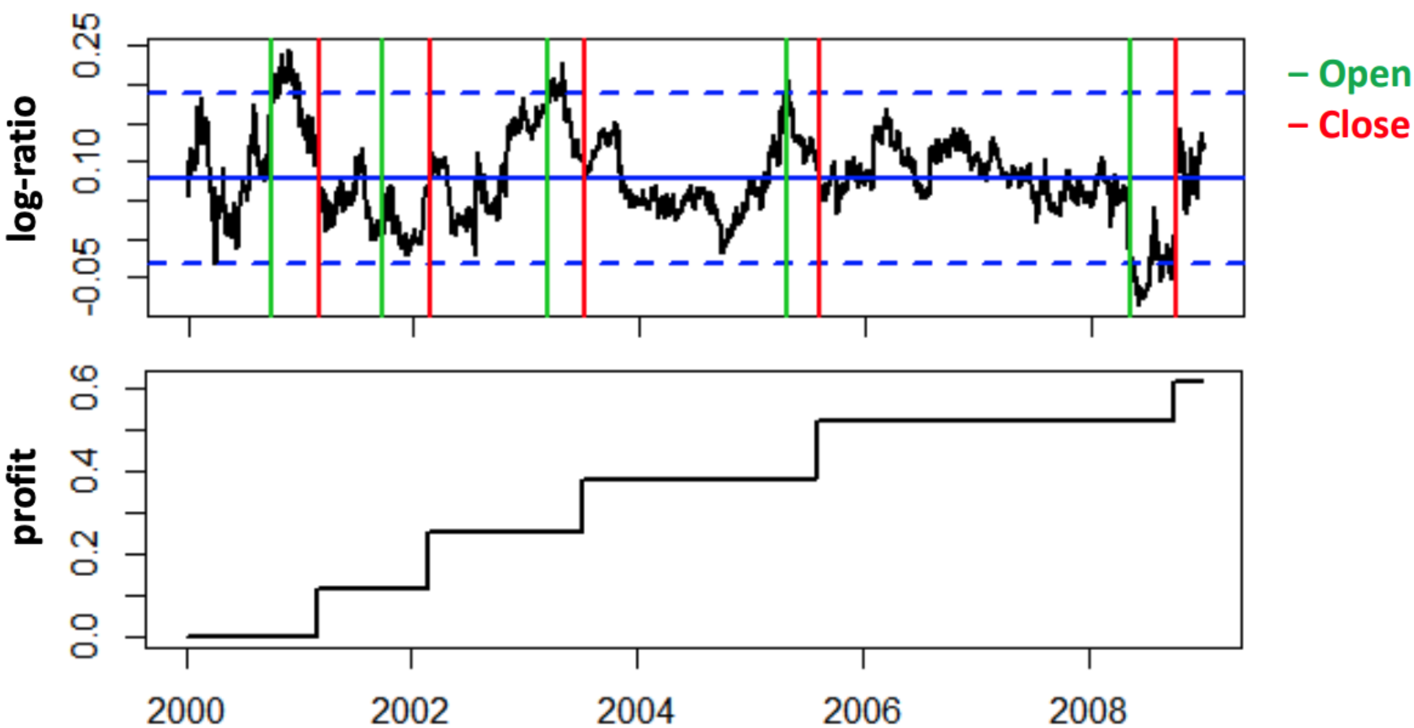
The profitability is determined by the behaviour of the log asset-price ratio.

$$\text{Profitable if: } \frac{P1_c}{P1_o} - \frac{P2_c}{P2_o} \neq 0$$
$$\implies \frac{P1_c}{P1_o} < \text{ or } > \frac{P2_c}{P2_o}$$

which is equivalent to asking

$$\frac{P2_o}{P1_o} < \text{or} > \frac{P2_c}{P1_c}$$
$$\iff$$
$$\log\left(\frac{P2_o}{P1_o}\right) < \text{or} > \log\left(\frac{P2_c}{P1_c}\right)$$

>Profitability is determined by log-ratio of prices. We say $<$ or $>$ just because we only care that there exists a difference, but not which one exactly is the larger of the two.



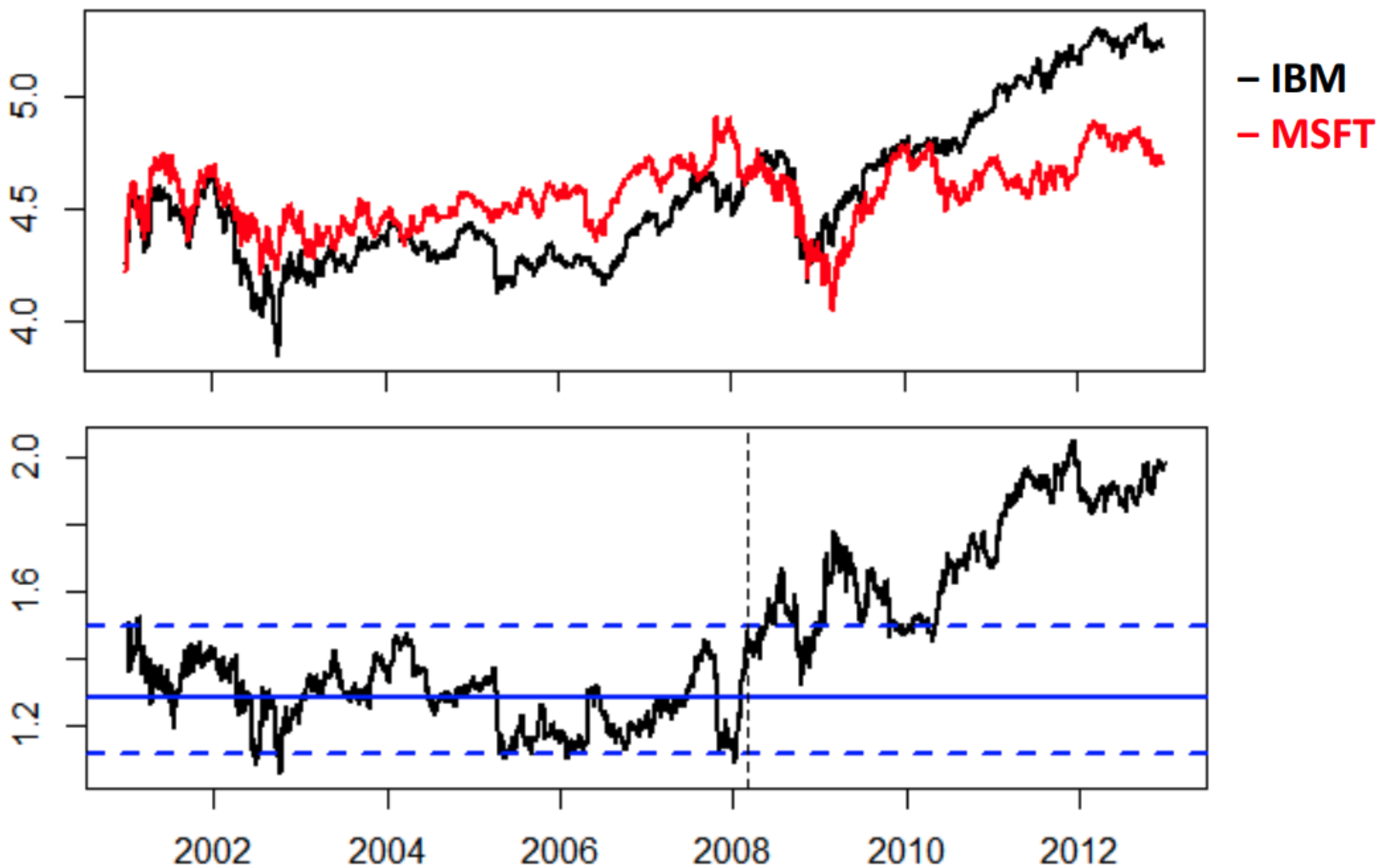
We can also show market-neutrality:

$$CAPM \implies \begin{cases} R1 = \beta \cdot R_M + \epsilon_1 \\ R2 = \beta \cdot R_M + \epsilon_2 \end{cases} \implies \begin{cases} \frac{P1_c}{P1_o} = 1 + R1 = 1 + \beta \cdot R_M + \epsilon_1 \\ \frac{P2_c}{P2_o} = 1 + R2 = 1 + \beta \cdot R_M + \epsilon_2 \end{cases}$$
$$\implies \frac{P1_c}{P1_o} - \frac{P2_c}{P1_o} = \frac{P2_c}{P2_o} = (1 + \beta \cdot R_M + \epsilon_1) - (1 + \beta \cdot R_M + \epsilon_2) = \epsilon_1 - \epsilon_2$$

Where $\epsilon_1 - \epsilon_2$ is independent of market return R_M.

11.1.1 What could go wrong

When the prices don't converge, and the log ratio of prices don't go back to the mean. (Mean-reverting)



11.1.2 Other things to consider

1. Determine which pairs to trade
2. When to open a position (the threshold)
3. What amounts to buy/sell
4. When to close trade
5. When to bail out of trade

Need to create a statistical/mathematical model to help make decisions on these items to consider.

11.1.2.1 1. Determining which pairs to trade

For a market with N assets, there are max ${}_nC_2$ pairs that could be made which is in the order of n^2.

We only want pairs who's log-ratio has **strong mean reversion**. This is not the same as simply having a constant mean. It depends on the dynamics of log-ratio process. For example, ARMA models have mean reversion properties but random walk processes/brownian motion do not have the mean reversion property.

11.1.2.2 Example 1: Log

Let $X_t = \log(\frac{P1_t}{P2_t}) \sim^{iid} N(0, \sigma^2),\ \forall t = 1, 2, ...$

If $X_0 = 2\sigma$, what's the expected time until $X_T \leq 0$?

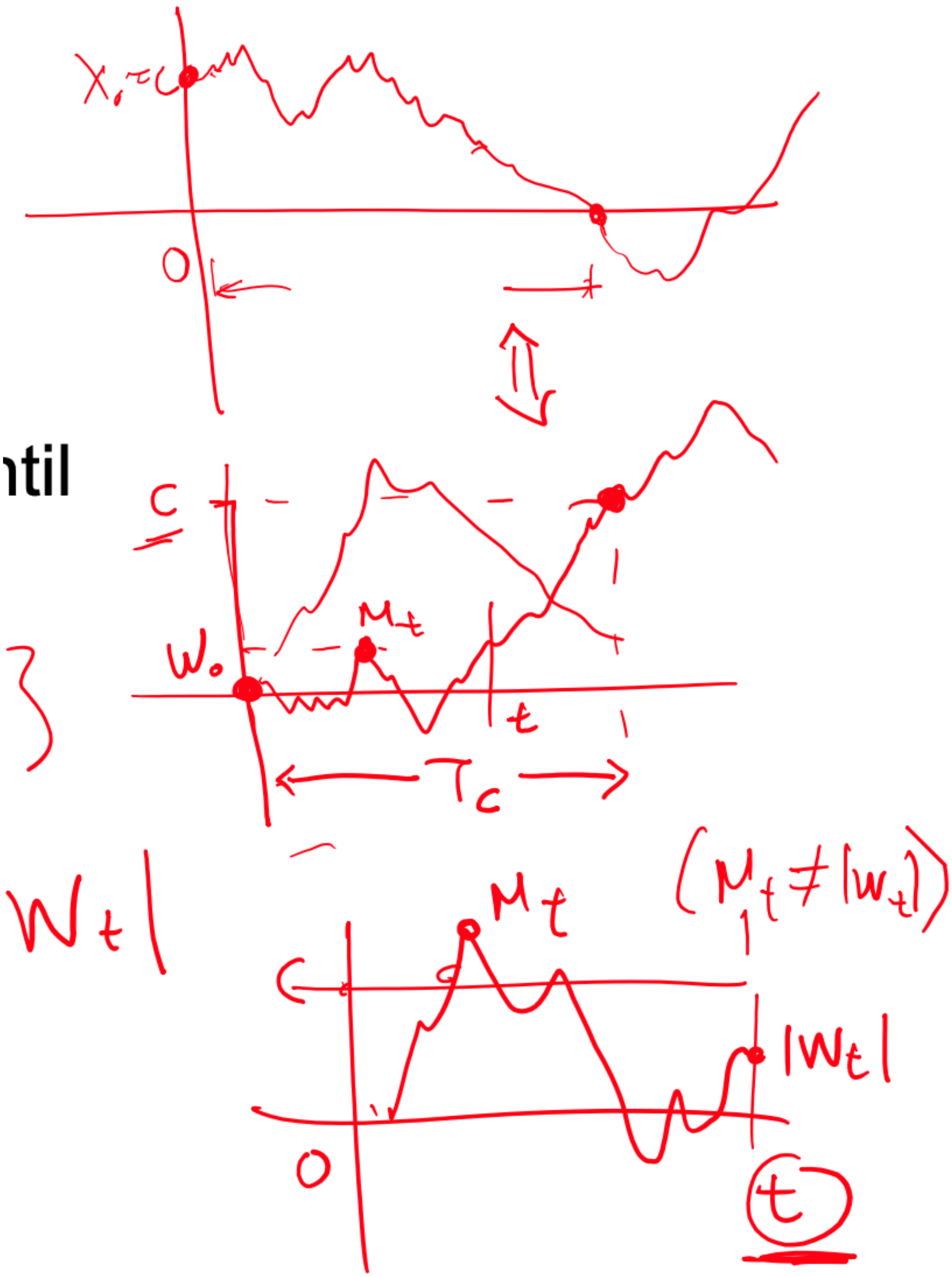
On any day t, $\mathbb{P}(X_t \leq 0) = 0.5$ by symmetry

$$\begin{aligned}
\text{Let } T &= \#\text{ days until } X_t \leq 0 \text{ for the first time}\\
&= (\text{hitting time})\\
&= \#\text{ trials until 1st success } x_t \leq 0\\
\implies T &\sim \text{Geometric}\left(p = \frac{1}{2}\right)\\
PMF\ \ p_T(t) &= \left(\frac{1}{2}\right)^t \quad \forall t \geq 1\\
\mathbb{E}(T) &= \frac{1}{p} = 2
\end{aligned}$$

11.1.2.3 Example 2: BM

Let $X_t = \log(\frac{P1_t}{P2_t}) \sim$ Brownian Motion (BM) (continuous time Random Walk) For any $X_0 = c > 0$, show that the expected time until $X_T \leq 0$ is infinite.

$$\begin{aligned}
\text{Let } T_c &= \{\text{first time std BM w}/W_0 = 0 \text{ hits level c}\}\\
\text{Let } M_t &= \max\{W_u; 0 \leq u \leq t\} \implies M_t \sim |W_t|
\end{aligned}$$



Which tells you that the distribution of the M_t follows the same distribution of the absolute value of the Brownian Motion.

This means:

$$\mathbb{P}(T_c \leq t) = \mathbb{P}(M_t \geq c) = \mathbb{P}(|\underbrace{W_t}_{\sim N_(0,t)}| \geq c) = 2 \cdot \Phi\left(-\frac{c}{\sqrt{t}}\right)$$

Where Φ is the CDF of the std normal.

As we look at the two tails of the normal after standardizing, but we take the area under the lower tail (-c).

$$\implies \text{PDF of } T_c \text{ is given by } f(t)$$

$$\begin{aligned}
f(t) &= \frac{\delta}{\delta t}\underbrace{\mathbb{P}(T_c \leq t)}_{F(t)} = \frac{\delta}{\delta t}\left[2 \cdot \Phi\left(-\frac{c}{\sqrt{t}}\right)\right] \\
&= 2\frac{\phi\left(-\frac{c}{\sqrt{t}}\right)\delta}{\delta t}\left(-\frac{c}{\sqrt{t}}\right) \\
&= 2\phi\left(-\frac{c}{\sqrt{t}}\right) \cdot \left(-\left(-\frac{1}{2}\frac{c}{\sqrt{t^3}}\right)\right) \\
&= \frac{1}{\sqrt{2\pi}}e^{-1/2 \cdot c^2/t} \cdot \frac{c}{\sqrt{t^3}}
\end{aligned}$$

So the expected value of hitting time is:

$$\begin{aligned}
\mathbb{E}(T_c) &= \int_0^\infty t \cdot f(t)\, dt \\
&= \int_0^\infty t\frac{c}{\sqrt{2\pi t^3}}e^{-1/2 \cdot c^2/2}\, dt \\
&= \frac{c}{\sqrt{2\pi}}\int_0^\infty \frac{1}{\sqrt{t}}e^{-c^2/2t}\, dt \\
&\geq \frac{c'}{\sqrt{2\pi}}\int_0^\alpha \frac{1}{\sqrt{t}}\, dt + \int_\alpha^\infty \frac{1}{\sqrt{t}}e^{-c^2/2t}\, dt \to \infty
\end{aligned}$$

$$\implies \mathbb{E}[T_c] = \infty$$

A stationary process ensures mean reversion. Thats why we use stationary processes to model the log-ratio.

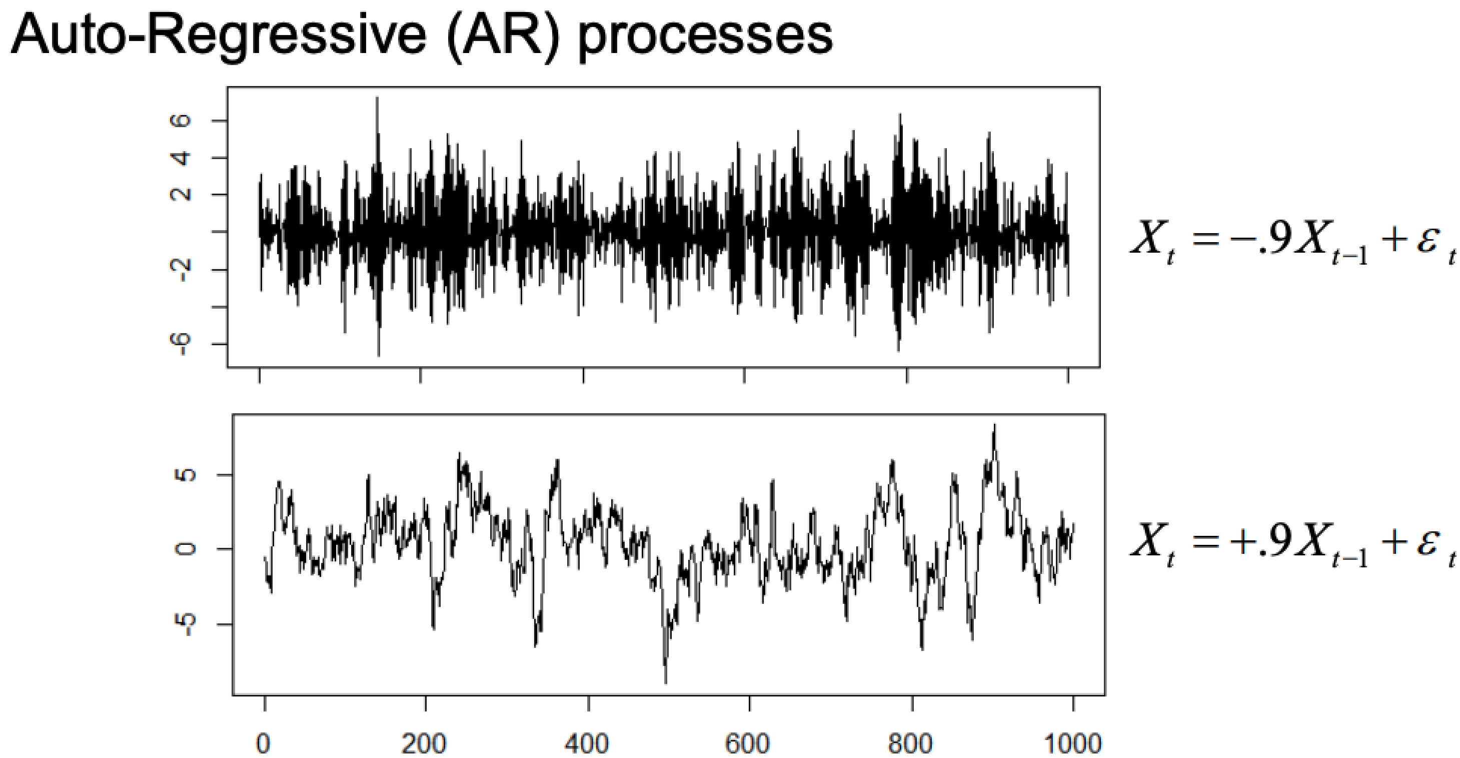
11.1.3 Mean Reversion

Mean reversion suggests log-ratio process are stationary.

Stationary processes are guaranteed to converge back to their mean within a reasonable time.

- Marginal distribution has constant mean, variance, and covariance between two stationary variables only depends on the distance between times
- Auto-Correlation Function (ACF) $\rho(h), \forall h = 0, 1, ...$ describes (linear) dependence at lag $h = |t - s|$

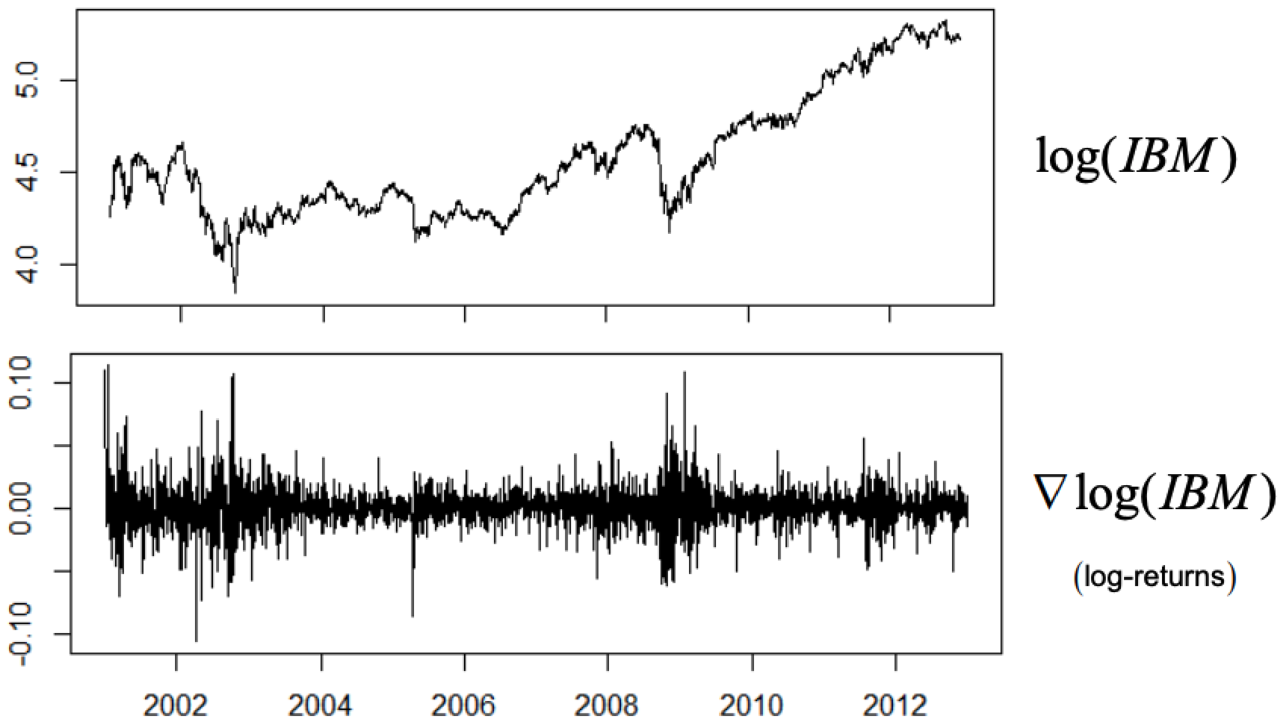
$$\begin{aligned}
\mathbb{E}[X_t] &= \mu \quad \forall t \\
\mathbb{V}[X_t] &= \sigma < \infty \quad \forall t \\
Cov(X_t, X_s) &= Cov(X_{t+r}, X_{s+r}) \quad \forall r, s, t
\end{aligned}$$



11.1.4 Integrated Series

In general, asset log-prices which are random walks, are not stationary. They are unpredictable. However, the **log returns** $r_t = \log(\frac{S_t}{S_{t-1}})$ follow a stationary process, whereas the asset log-prices $\log(S_t) = \log(S_0) + \sum_{i=1}^{t} r_i$ are random walks.

A variable that is not stationary but it's differences are, is called an integrated series. $\{X_t\}$ is not stationary but $\{\nabla X_t = X_t - X_{t-1}\}$ is.

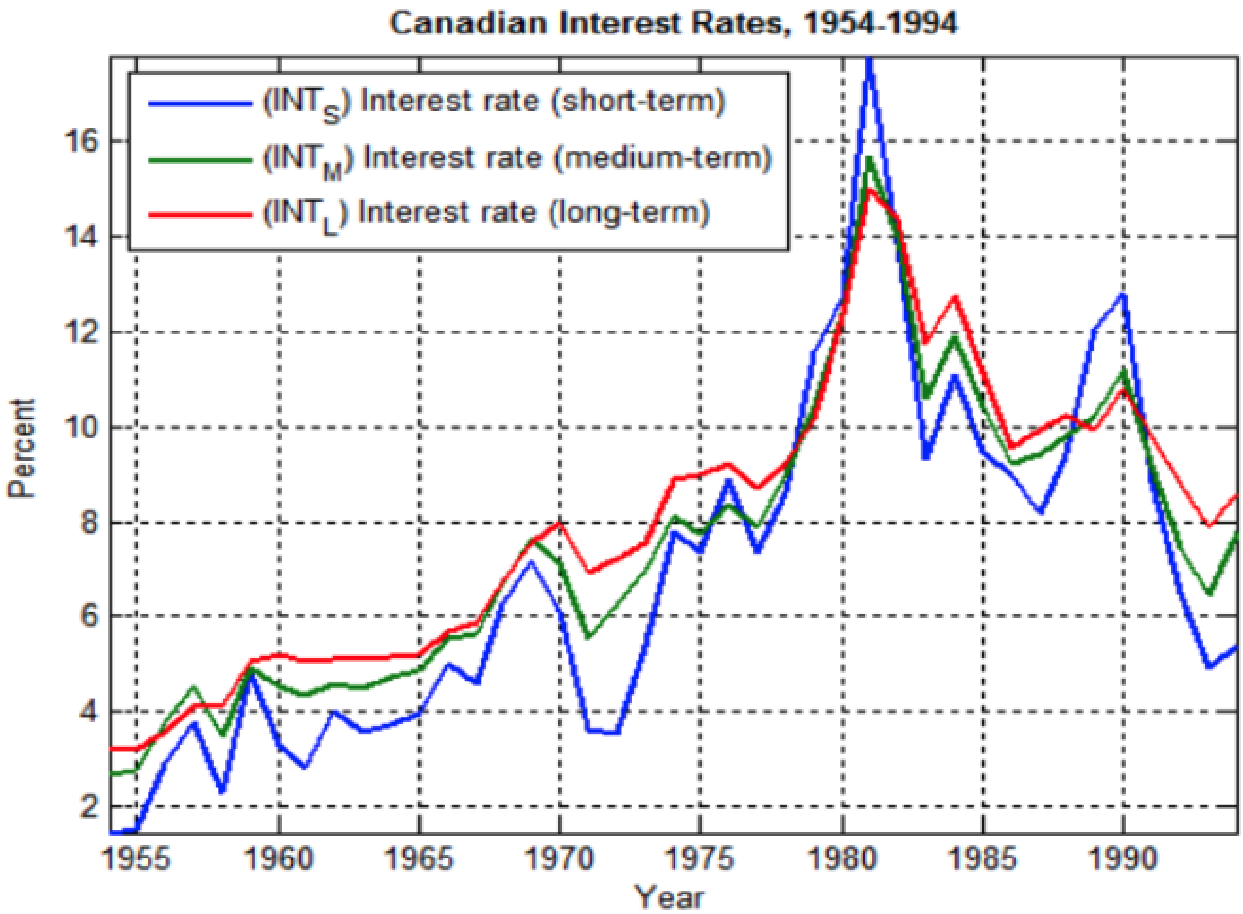


11.1.5 Cointegration

Consider two integrated series $\{X_t, Y_t\}$ which behave as random walks, but if they seem to have some constant (stationary) relationship when linearly combined then they are called cointegrated.

$$\exists\ \alpha\ s.t.\ X_t + \alpha Y_t \sim Stationary$$

11.1.5.1 Ex: Yield Rate



The short, medium, and long term rates are considered cointegrated as they individually follow a random walk, but they do it together.

11.1.5.2 Ex: Math

Let $\{W_t\}$ be a random walk, and

$$\begin{cases} X_t = W_t + \epsilon_t \\ Y_t = W_t + \nu_t \end{cases}$$

where $\epsilon, \nu \sim^{iid} N(0, \sigma^2)$

We can show $\{X_t, Y_t\}$ are cointegrated. (They are not stationary, but their 1st order different is stationary)

$$\mathbb{V}(X_t) = \mathbb{V}[W_t + \epsilon_t] = \overbrace{\mathbb{V}(W_t)}^{t\sigma_w^2} + \overbrace{\mathbb{V}(\epsilon_t)}^{\sigma^2} = t\sigma_w^2 + \sigma^2 \implies \text{not stationary}$$

Which is the same case for Y_t. Looking at the first order differences:

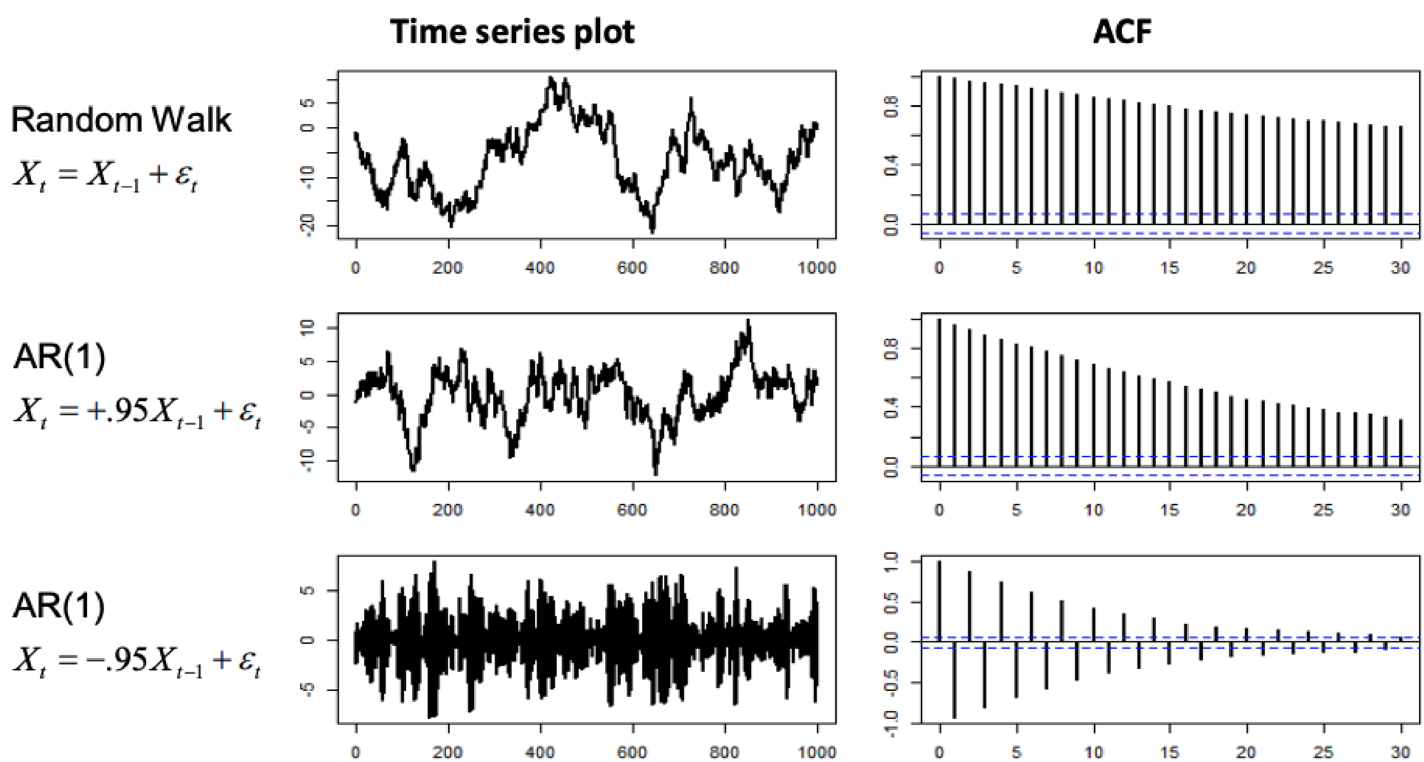
$$\nabla X_t = X_t - X_{t-1} = W_t + \epsilon_t - W_{t-1} + \epsilon_{t-1} = (W_t - W_{t-1}) + \epsilon_t + \epsilon_{t-1}$$
$$\implies \mathbb{V}(\nabla X_t) = \sigma^2 \text{stationary}$$

To show cointegration, we much show $X_t - Y_t$ is stationary.

$$X_t - Y_t = (W_t + \epsilon_t) - (W_t + \nu_t) = \epsilon_t + \nu_t \text{sum of iid sequences}$$
$$\implies \begin{cases} \mathbb{E}(X_t - Y_t) = 0 \\ \mathbb{V}(X_t - Y_t) = 2\sigma^2 \\ Cov(X_t - Y_t, X_{t+h} - Y_{t+h}) = \rho|h|?? \end{cases}$$

For pairs trading, we want assets which are cointegrated (their log difference is stationary)

11.1.6 Stationarity



Time series is not always a good indicator of a random walk. We can use a ACF to determine it. Linear decline -> random walk

11.1.6.1 Tests

There also exists stationarity tests:

Hypothesis test for $\begin{cases} H_0: \text{series is integrated} \\ H_1: \text{series is stationary} \end{cases}$ - Idea: fit $X_t = \beta X_{t-1} + \varepsilon_t$ to data and test $\begin{cases} H_0: \beta = 1 \\ H_1: \beta < 1 \end{cases}$ E.g. For $n = 1000$

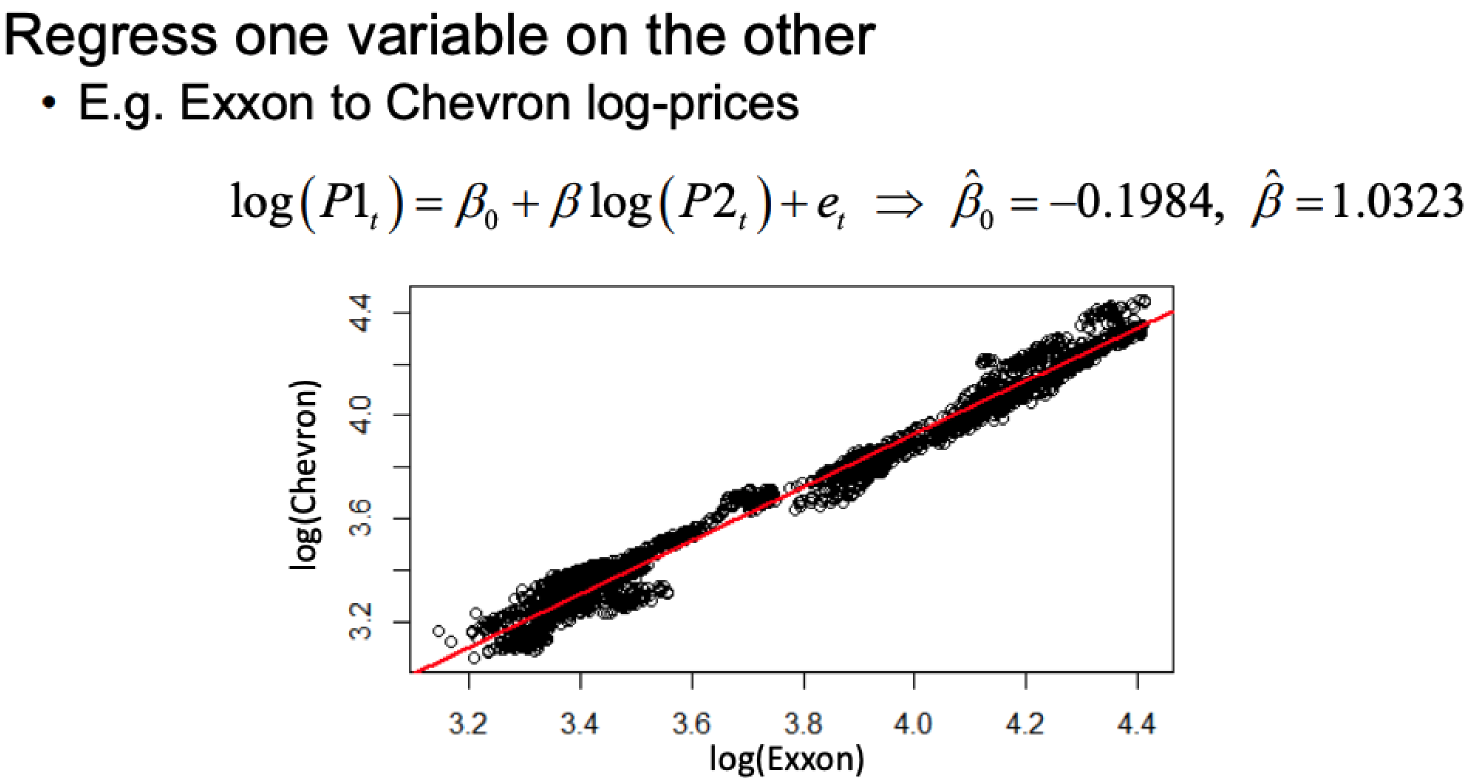
Model	Test statistic	P-value
$X_t = X_{t-1} + \varepsilon_t$	-1.9027	0.6195
$X_t = +.95X_{t-1} + \varepsilon_t$	-5.6161	<< .01
$X_t = -.95X_{t-1} + \varepsilon_t$	-232.4851	< .01

But if we're trying to check if a linear combination of assets is stationary, its difficult because we don't know the linear combination to look for. There are two tests for cointegration:

- **Engle-Granger two-step method**: Run linear regression and look for potential cointegrated relationship, finding the combination that minimizes squared errors. Then the errors are tested for stationarity (converge to 0)
- **Vector Error Correction models (VECM)**: Covered in time series, a part of VAR models

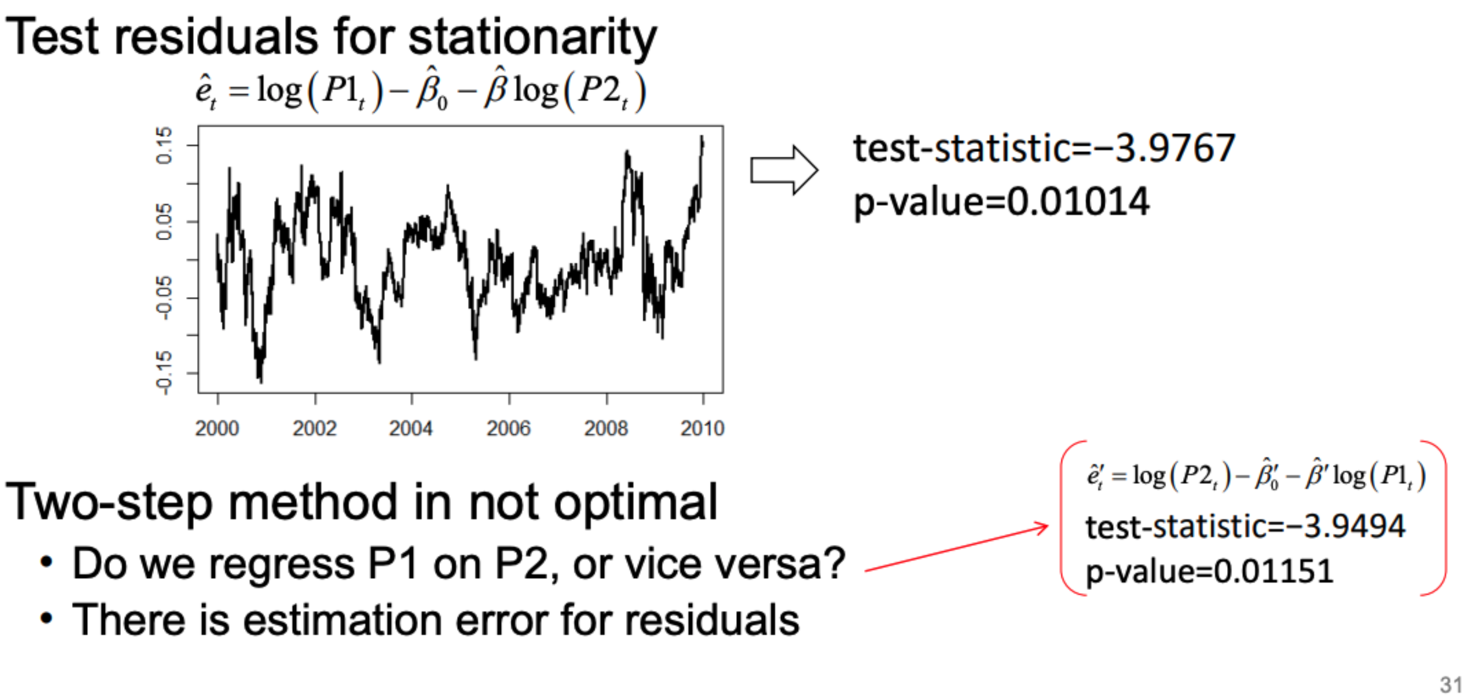
11.1.6.2 Two step method example

First we fit a linear regression to obtain the slope (not useful), and the intercept (need to subtract from errors).



Now we can look at the errors (minus the intercept so its 0 centered) and run stationarity tests on them. We would trade stocks with the highest metrics, although this method is not optimal since our results could be arbitrary.

If we did the regression swapping the variables, the p-value could differ a lot



11.1.7 Spurious regression

Two totally independent stocks may show some false linear relationship.

Consider 2 independent random walks $\{W_t, V_t\}$ - When you regress $W_t = \beta_0 + \beta V_t + e_t, t = 1, \ldots, n$ you are NOT guaranteed that $\hat{\beta} \to 0$ as the sample size $n \to \infty$ (i.e. not consistent)!!! Effect called spurious (fake) regression - Results of random walk (integrated series) regressions are NOT reliable - To address this problem & estimation errors, use Phillips-Ouliaris stationarity test in 2 step method

> Bottom line is using VAR models is superior than the two step test/using linear regression. A multivariate time series model is best.

11.2 Financial Indices

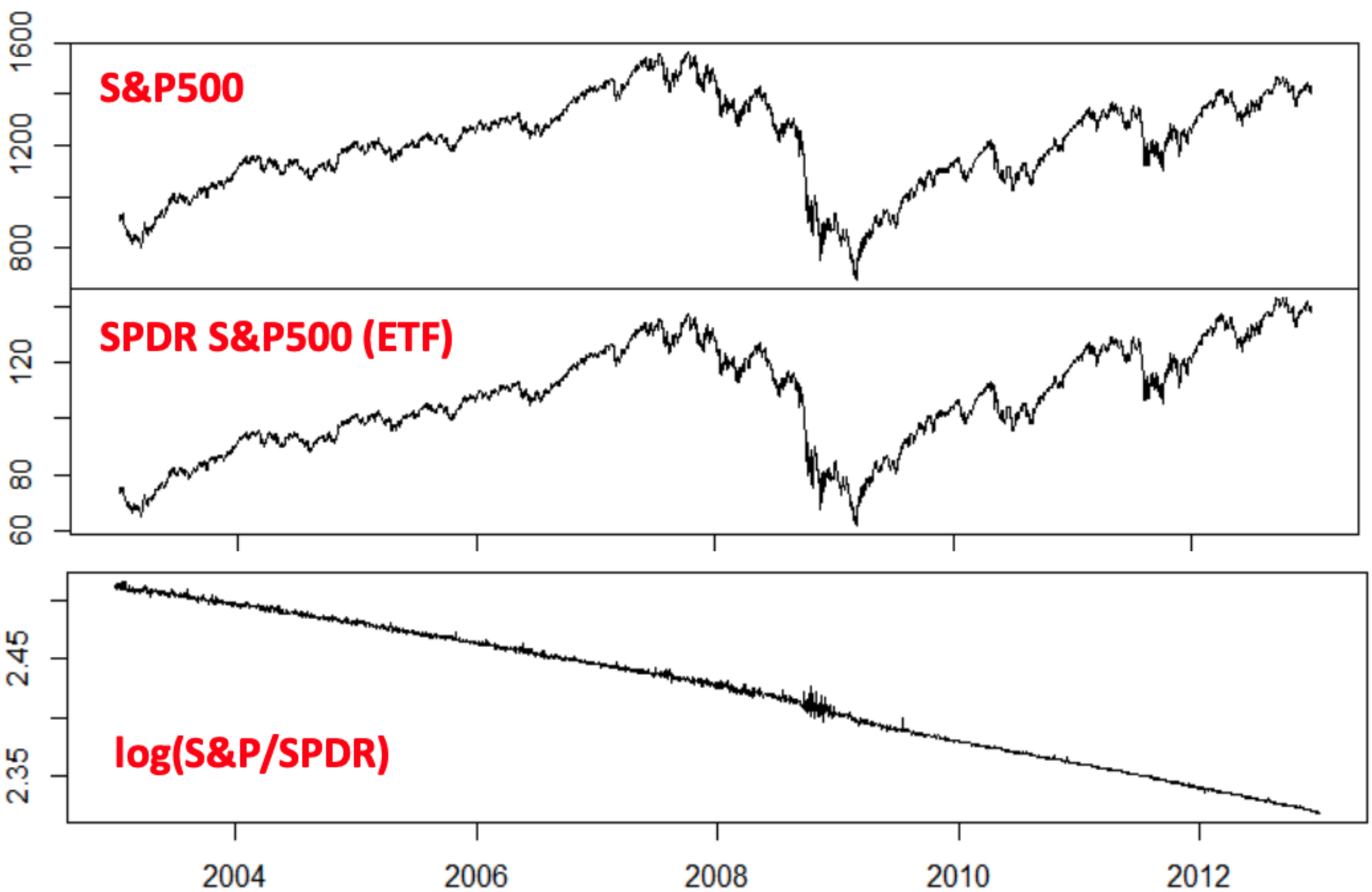
Indices measure value/performance of financial markets. E.g., - Dow-Jones Industrial Average (DJIA): Simple average of 30 major US stock prices (since 1896) - Standard & Poor (S&P) 500: Weighted (cap-base) average of 500 large NYSE & NASDAQ listed companies Financial indices are NOT traded instruments However, there are many financial products whose value is directly related to indices: - Mutual funds: e.g., Vanguard 500 Index Fund - Exchange-Traded-Funds (ETF's): e.g., SPDR or iShares S&P500 Index - Futures: e.g., E-Mini S&P futures

11.3 Index Arbitrage

- Financial products based on indices essentially offer a sophisticated version of multivariate cointegration

- Index of #N assets $\{S_i\}_{i=1}^N$ w/ weights $\{w_i\}_{i=1}^N \Rightarrow$ Index level: $I(t) = \sum_{i=1}^N w_i \times S_i(t)$
- Instrument tracking index $F(t)$ (e.g. futures)
- Known cointegration relationship:

$$F(t) - I(t) = F(t) - \sum_{i=1}^{N} w_i \times S_i(t) \sim \text{ stationary}$$



11.4 Volatility Arbitrage

- VolArb is implemented with derivatives, primarily options

To fix ideas, consider European options: - For Black-Scholes formula, only unobserved input is volatility σ, which has to be estimated - How does volatility affect Call/Put prices? - Implied volatility σ_i is input which makes Black-Scholes price equal to observed market price - σ_i is not estimated from underlying asset dynamics - Imagine you know volatility will increase in the future, beyond what current options prices warrant (implied vol i). - How can you take advantage of this? - Eliminate effects of asset movement by delta-neutral strategy - Profit by large moves, irrespective of direction

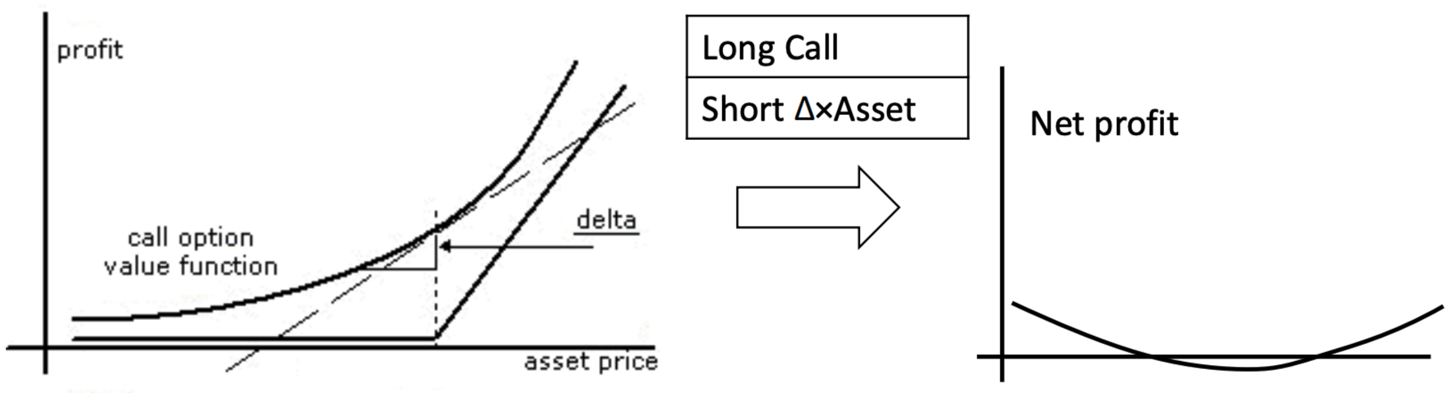
- VolArb relies on predicting (implied) volatility of underlying asset

- Common approach is to describe the evolution of volatility with GARCH (Generalized AutoRegressive Conditional Heteroskedasticity) models

$$y_t = \sigma_t \cdot \varepsilon_t, \varepsilon_t \sim^{iid} N(0,1)$$
$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^{p} \alpha_j y_{t-j}^2 + \sum_{k=1}^{q} \beta_k \sigma_{t-k}^2$$

- Essentially, an autoregressive model for conditional variance

11.4.1 E.g. Delta-hedged long call



11.5 High Frequency Trading

- HFT uses algorithmic trading over very short holding periods, profiting from very small price discrepancies by trading frequently and at large volumes
- HFT employs predictive algorithms for machine learning and data mining
  - Essentially, tries to discover patterns of trading activity & profit by preempting them
  - This includes traditional method like index arbitrage, but also others which might not have any intuitive interpretation

Chapter 12

Monte-Carlo Simulation

Primarily dealing with option pricing. There are three basic numerical option pricing methods:

- Binomial Trees (BT)
  - Don't need sophisticated mathematical or stochastic analysis, just backward induction
  - Useful in many cases
- Finite Difference (FD)
  - Uses the Black-Scholes PDE
  - Not discussed much in the stats program
- Monte-Carlo Simulation
  - Based on the stochastic differential equation (determines evolution of asset-price)
  - Easy to program and risk-neutral
  - Really shines with multiple dimensions (many factors, path/multi-asset)

	European options	Early exercise	Path dependence	Multi-asset dependence
BT	✓	✓	✗	✗
FD	✓	✓	✗	✗
MC	✓	✗	✓	✓

12.1 Brownian Motion

Any martingale can be represented as an integral of brownian motion.

Brownian Motion (BM) forms the building block of continuous stochastic models Standard BM $\{W_t\}$ is such that

$$W_0 = 0 \quad \& \quad (W_t - W_s) \mid W_s \sim N(0, t-s)$$

Arithmetic BM (ABM) $\{X_t\}$ with drift μ & volatility σ is

$$X_0 = 0 \quad \& \quad (X_t - X_s) \mid X_s \sim N\left(\mu(t-s), \sigma^2(t-s)\right)$$

- In form of Stochastic Differential Equation (SDE)

$$dX_t = \mu dt + \sigma dW_t \Leftrightarrow X_t - X_0 = \mu t + \sigma\left(W_t - W_0\right)$$

12.2 Properties of the Multivariate Normal

If

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} \sim \mathbf{N}\left(\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}\right)$$

$$\mathbf{X} = \left[\begin{array}{c} \mathbf{X}_1 \\ \hdashline \mathbf{X}_2 \end{array}\right] \sim \mathbf{N}\left(\boldsymbol{\mu} = \left[\begin{array}{c} \boldsymbol{\mu}_1 \\ \hdashline \boldsymbol{\mu}_2 \end{array}\right], \boldsymbol{\Sigma} = \left[\begin{array}{c:c} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \hdashline \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{array}\right]\right)$$

, then:

Property	Formula
Marginals:	$\mathbf{X}_1 \sim \mathbf{N}\left(\mu_1, \Sigma_{11}\right)$
Linear combinations:	$\mathbf{a} + \mathbf{B}^\top \mathbf{X} \sim N\left(\mathbf{a} + \mathbf{B}^\top \mu, \mathbf{B}^\top \Sigma \mathbf{B}\right)$
Conditionals:	$\mathbf{X}_1 \mid (\mathbf{X}_2 = \mathbf{x}) \sim \mathrm{N}\left(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}\left(\mathbf{x}\mu_2\right), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)$

If $\Sigma_{12} = 0$ then $\mathbf{X_1}|\mathbf{X_2} = x \sim N(\mu_1, \Sigma_{11})$

Using these properties, we can do calculations for brownian motion

12.2.1 Ex: Distribution given X__s

Finding the distribution of $X_t|X_s = x, t > s$ for $dX_t = \mu dt + \sigma dW$

$$\begin{bmatrix} X_s \\ X_t \end{bmatrix} \sim N(\begin{bmatrix} \mu_s \\ \mu_t \end{bmatrix}, \begin{bmatrix} \sigma^2 s & \sigma s \\ \sigma s & \sigma^2 t \end{bmatrix}) \iff \begin{bmatrix} X_t \\ X_s \end{bmatrix} \sim N(\mu \begin{bmatrix} t \\ \hline s \end{bmatrix}, \sigma^2 \begin{bmatrix} t & s \\ s & s \end{bmatrix})$$

$$\begin{bmatrix} X_s \\ X_t \end{bmatrix} \sim N(\begin{bmatrix} \mu_s \\ \mu_t \end{bmatrix}, \begin{bmatrix} \sigma^2 s & \sigma s \\ \sigma s & \sigma^2 t \end{bmatrix}) \iff \begin{bmatrix} X_t \\ \hdashline X_s \end{bmatrix} \sim N(\mu \begin{bmatrix} t \\ \hdashline s \end{bmatrix}, \sigma^2 \begin{bmatrix} t & s \\ s & s \end{bmatrix})$$

Because

$$\begin{aligned} Cov(W_s, W_t) &= Cov(W_s, W_s + (W_t - W_s)) \\ &= Cov(W_s, W_s) + \underbrace{Cov(W_s, (W_t - W_s))}_{=0} \\ &= Var(W_s) = s \end{aligned}$$

Which gives:

$$\begin{aligned} X_t | X_s = x &\sim N\left(\mu t + \sigma^2 s \frac{1}{\sigma^2 s}(x - \mu s), \sigma^2 \left(t - s\frac{1}{8}8\right)\right) \\ &\sim N(x + \mu(t-s), \sigma^2(t-s)) \end{aligned}$$

12.2.2 Ex: Brownian Bridge (Distribution given X_t)

Find the distribution of $X_s | X_t = x, s \in (0, t)$ for $dX_t = \mu dt + \sigma dW_t$

$$\begin{bmatrix} X_s \\ X_t \end{bmatrix} \sim N(\mu \begin{bmatrix} s \\ t \end{bmatrix}, \sigma^2 \begin{bmatrix} s & s \\ s & t \end{bmatrix})$$

$$\begin{bmatrix} X_s \\ X_t \end{bmatrix} \sim N(\mu \begin{bmatrix} s \\ \hdashline t \end{bmatrix}, \sigma^2 \left[\begin{array}{c:c} s & s \\ \hdashline s & t \end{array}\right])$$

$$\begin{aligned} \implies X_s | X_t = x &\sim N\left(\underbrace{\mu s + \frac{s}{t}(x - \mu t)}_{\mu s + \frac{s}{t}x - \mu s}, \sigma^2\left(s - \frac{s}{t}s\right)\right) \\ &\sim N\left(\frac{s}{t}x, \sigma^2 \frac{s(t-s)}{t}\right) \end{aligned}$$

May be useful for generating brownian motion paths where I know that at time T, I have value of x.

12.3 Geometric Brownian Motion

A transformation of the arithmetic brownian motion. We use this to avoid negative values.

Process $\{S_t\}$ whose logarithm follows ABM

$$\begin{aligned} &\log(S_t) - \log(S_0) = \log\left(\frac{S_t}{S_0}\right) = X_t \sim N(\mu t, \sigma^2 t) \Leftrightarrow \\ &\Leftrightarrow S_t = S_0 \exp\{X_t\} \sim S_0 \times \log\text{Normal}(\mu t, \sigma^2 t) \end{aligned}$$

As the log of $\exp\{X_t\} = X_t$ and $X_t \sim Normal(\mu t, \sigma^2 t)$

Expressed in terms of SDE:

$$dX_t = d\log(S_t) = \mu dt + \sigma dW_t \Leftrightarrow dS_t = \left(\mu + \frac{\sigma^2}{2}\right) S_t dt + \sigma S_t dW_t$$

12.4 Risk-Neutral Pricing

- Assuming Geometric BM for asset $\{S_t\}$ and risk-free interest rate r, there exists a probability measure such that
  - $$dS_t = rS_t dt + \sigma S_t dW_t \Rightarrow S_t \sim S_0 \times \log N\left(\left(r - \frac{\sigma^2}{2}\right)t, \sigma^2 t\right)$$
  - Under this measure, discounted asset prices are martingales
  - Called risk-neutral (RN) or equivalent martingale measure
- The arbitrage-free price of any European derivative with payoff $G_T = f(S_T)$ is given by discounted expectation w.r.t. RN measure
  $$G_0 = \mathbb{E}\left[e^{-rT}G_T\right] = \mathbb{E}\left[e^{-rT}f(S_T)\right]$$

  We just need to figure out that expected value to obtain the price at time 0.

12.4.1 Ex: RN Pricing

We'll show that under risk-neutral pricing measure,

$$\mathbb{E}(S) = S_0 e^{rt}$$

(a martingale) then we would have

$$\mathbb{E}\left(\frac{S_t}{e^{rt}}\right) = S_0 \quad \text{or more generally } \mathbb{E}\left(\frac{S_t}{e^{rt}}|S_s\right) = \frac{S_s}{e^{rt}}$$

Proof:

$$\begin{aligned}\mathbb{E}(S_t) &= \mathbb{E}(S_0 e^{\log(S_t/S_0)}) \\ &= S_0\mathbb{E}(e^Y) \text{ where } Y = \log\left(\frac{S_t}{S_0}\right) \sim N\left(\left(r - \frac{\sigma^2}{2}\right)t, \sigma^2 t\right)\end{aligned}$$

We can use the Normal MGF $\mathbb{E}(e^X)$ if $X \sim N(\mu, \sigma^2)$:

$$m_X(z) = e^{\mu z + (1/2)\cdot\sigma^2 z^2}$$

In this case, we have

$$m_Y(1) = \exp\left\{\overbrace{\left(r - \frac{\sigma^2}{2}\right)t}^{\mathbb{E}(Y)} + \overbrace{\frac{1}{2}\sigma^2 t}^{\mathbb{V}(Y)}\right\} = \exp\left\{rt - \frac{\sigma^2}{2}t + \frac{\sigma^2}{2}t\right\} = S_0 e^{rt}$$

12.4.2 Ex: Find price of forward contract (no dividends)

$$G_T = f(S_T) = (S_T - F_{0,T}) \quad \text{where } F_{0,T} = S_0 \cdot e^{rT}$$

We know $G_0 = 0$ (forward contracts involve no cashflow at $t = 0$)

By the RN pricing: $G_0 = \mathbb{E}[e^{-rT} G_T]$

$$\begin{aligned}
\implies 0 &= \mathbb{E}\left[\underbrace{e^{-rT}}_{\text{const, can remove}} (S_T - F_{0,T})\right] \\
\implies 0 &= \mathbb{E}(S_T) - F_{0,T} \\
\implies F_{0,T} &= \mathbb{E}(S_T) = e^{rT}\mathbb{E}(\underbrace{e^{-rT} S_T}_{S_0 e^{rT} = S_T \sim mtgl}) = e^{rT} S_0
\end{aligned}$$

12.4.3 Estimating Expectations

If $\mathbb{E}(e^{-rT} f(S_t))$ cannot be calculated exactly, it can be estimated/approximated by simulation.

- Generate # N independent random variates $S_i(T), i = 1, \dots, N$ based on RN measure
- By Strong LLN
  -
  $$\hat{G}_0 = \frac{1}{N}\sum_{i=1}^{N} e^{-rT} f(S_i(T)) \to \mathbb{E}(e^{-rT} f(S_T))$$

    with probability 1
- By the CLT
  -
  $$\frac{\hat{G}_0 - G_0}{s_G/\sqrt{n}} \sim^{appr} .N(0,1), \text{ where } s_G^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left[e^{-rT} f(S_i(T)) - \hat{G}_0\right]^2$$

  If you can express the quantity you're looking for as an expected value, you can run a simulation and it will approximate the value you're looking for

  How do you run a simulation experiment to approximate a probability? $\mathbb{P}(A) = \mathbb{E}(I_A)$ the indicator variable of a given event A.

12.4.4 Ex: Show estimator is consistent

Estimator: $\mathbb{E}[e^{-rT} f(S_T)]$

Build a 95% confidence interval for G_0 as well

$$\begin{aligned}
\mathbb{E}(\hat{G}_0) &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} e^{-rT} f(S_i(T))\right] \\
&= \frac{1}{n}\sum_{i=1}^{n} \underbrace{\mathbb{E}[e^{-rT} f(S_i(T))]}_{G_1?} = \frac{1}{n} n G_0 = G_0
\end{aligned}$$

Confidence interval:

$$\hat{G}_0 \pm 1.96 \times \frac{SG}{\sqrt{n}}$$

Need 4 times the samples to double the accuracy (scaled by $\sqrt{n}$)

12.5 European Call

Estimating European call price w/ simulation

- Asset price dynamics: $dS_t = rS_t dt + \sigma S_t dW_t$
- Payoff function for strike K and maturity T: $f(S_t) = (S_T - K_T)$

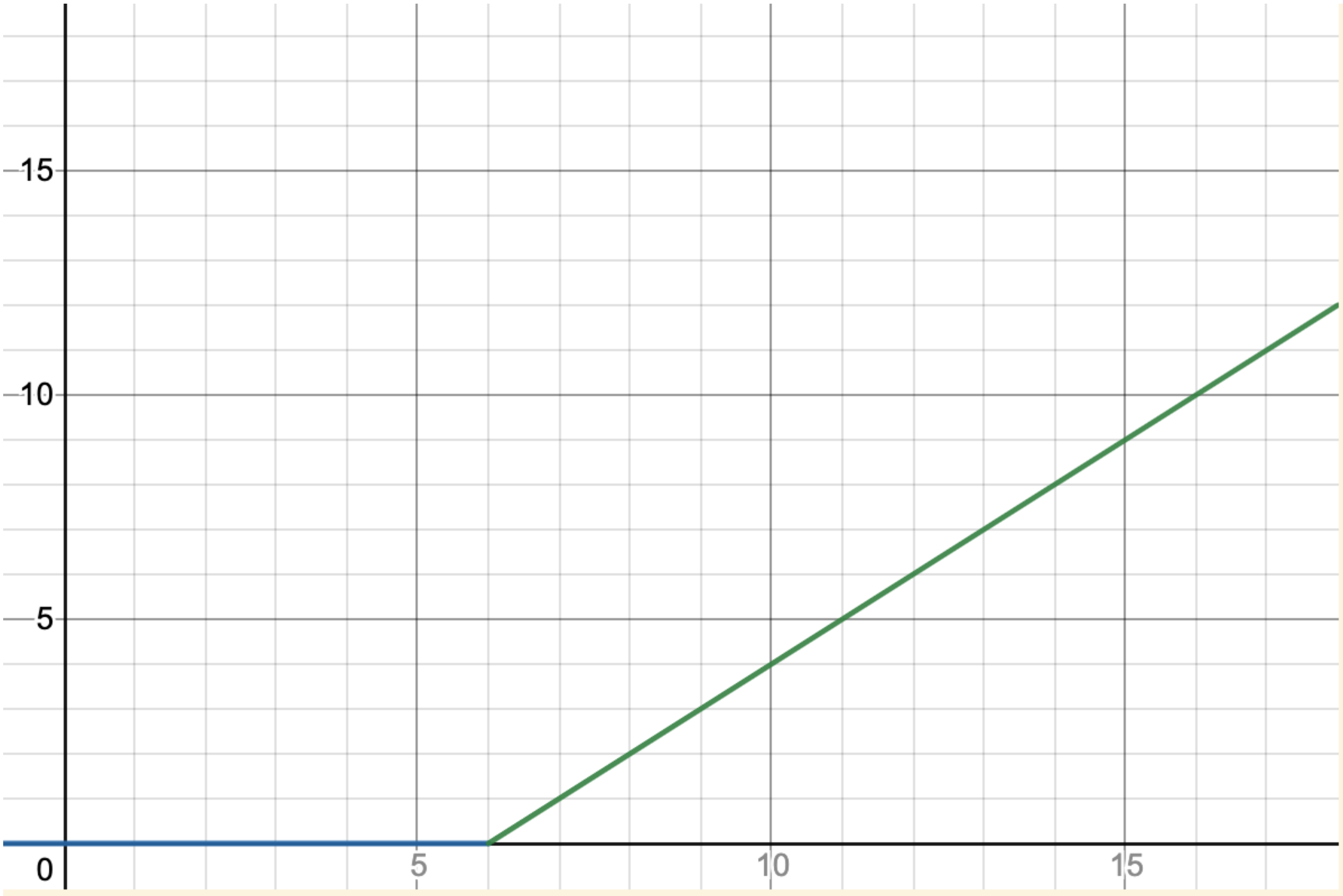
Generate random asset price variates as:

$$S_i(T) = S(0) \times \exp\left\{\left(r - \frac{\sigma^2}{2}\right)T + \sigma\sqrt{T} \times Z_i\right\}$$

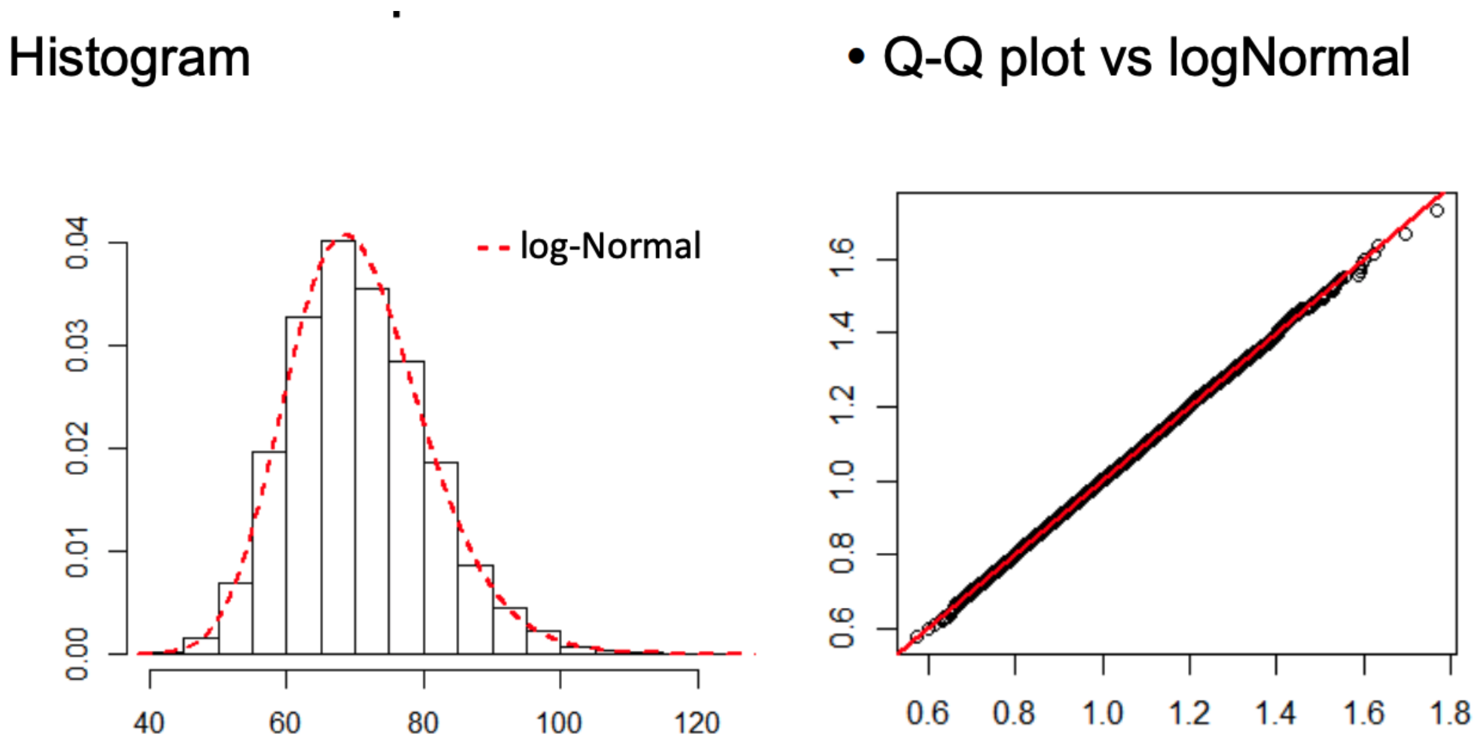
where Z_i is standard Normal variate

Where the x-axis is S_t, and y-axis is the payoff G_t

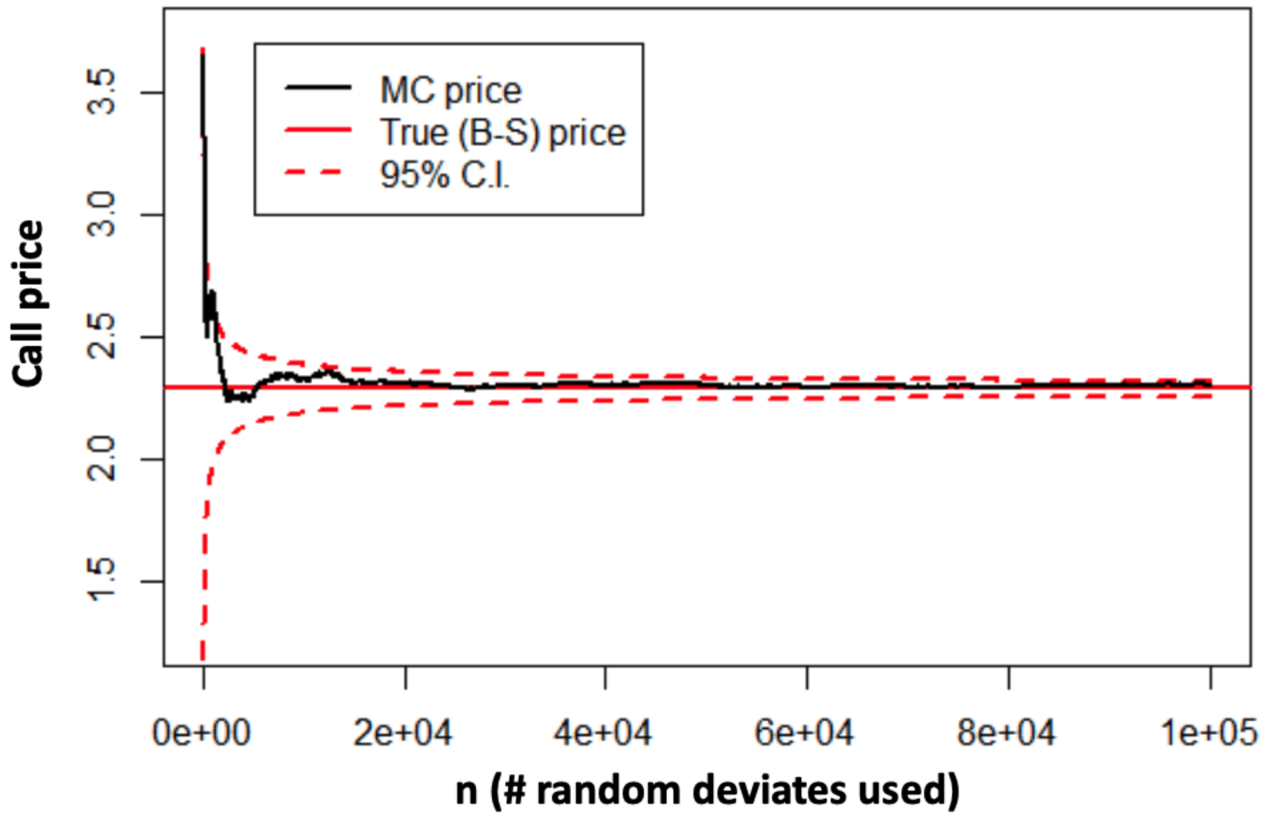
```
top=19;bottom=-1;left=-1;right=18
---
y=0|x<k
y=x-k|x>=k
k=6
```



After simulation, we get asset-prices that follow:



And we can see that the Monte Carlo simulated price converges to the true price:



When model options with depend on multiple assets, we need to consider those assets in the model.

12.6 Multiple Assets

The payoff of some options depends on the prices of multiple assets.

E.g. Exchange (outperformance) option w/ payoff

$$max\left\{S_1(T) - S_2(T), 0\right\} = (S_1(T) - S_2(T))$$

MC option pricing requires simulating & averaging multiple asset prices/paths - Cannot simply simulate each asset separately - Need to consider cross-asset dependence

Assets may have cross-asset dependence, so we cannot simply just simulate each asset independently until time T.

12.6.1 Multivariate Brownian Motion

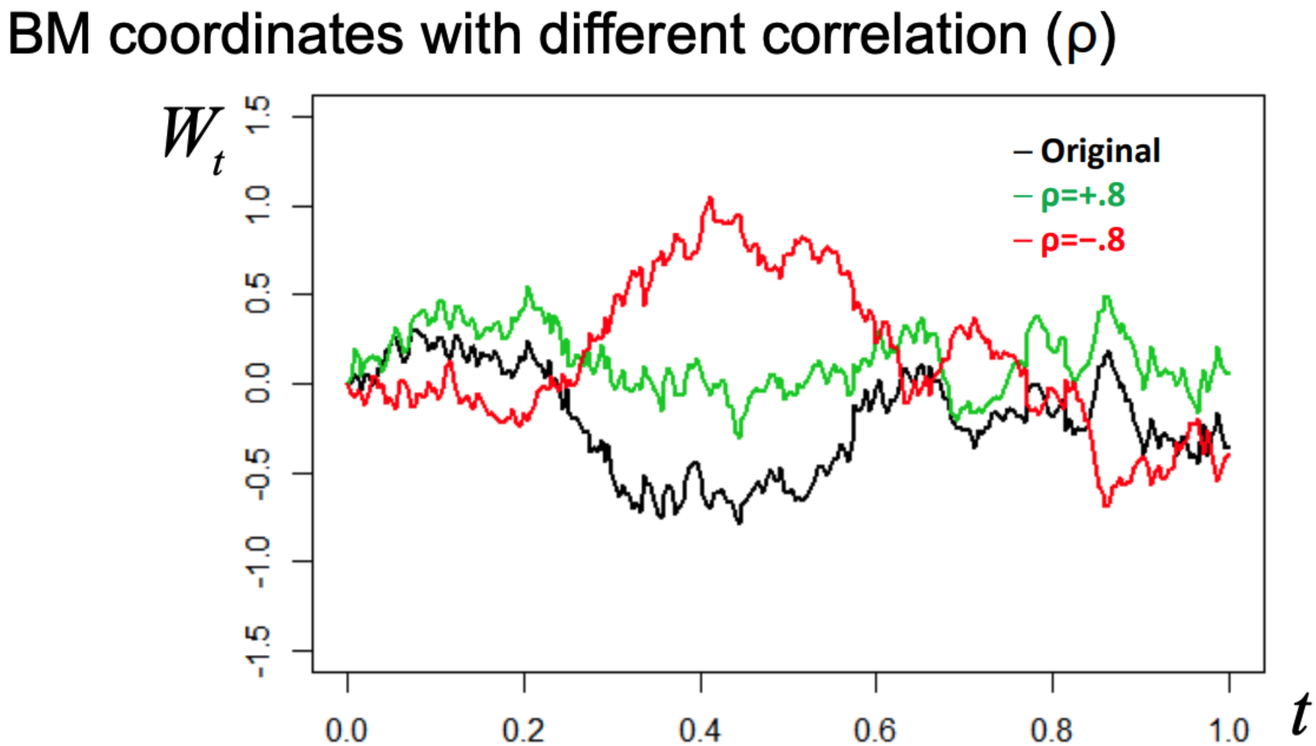
Define d-dimensional standard BM $\mathbf{W}(t) = \begin{bmatrix} W_1(t) \\ \vdots \\ W_d(t) \end{bmatrix}$ with correlation matrix $\rho = \begin{bmatrix} 1 & \cdots & \rho_{1d} \\ \vdots & \ddots & \vdots \\ \rho_{d1} & \cdots & 1 \end{bmatrix}$ to have independent Normal increments

$$\mathbf{W}(t) - \mathbf{W}(s) \mid \mathbf{W}(s) \sim N_d(\mathbf{0}, (t-s)\rho)$$

- Note: increments are independent over time, but can be dependent across dimensions!

Increments are independent of the past, and follow a d dimensional normal with mean 0, with the corr mat being the cov matrix multiplied by time (proportional to time).

3-dimensional standard Brownian Motion



12.6.2 Multivariate Arithmetic BM

- $\{\mathbf{X}(t)\}$ $w/$ SDE $d\mathbf{X}(t) = \mu dt + \sigma d\mathbf{W}(t)$, where

$$\mu = \begin{bmatrix} \mu_1 & \cdots & \mu_d \end{bmatrix}^\top, \sigma = \begin{bmatrix} \sigma_1 & \cdots & \sigma_d \end{bmatrix}^\top$$

  $\{\mathbf{W}(t)\} \sim d$-dim. standard BM W/ correlations ρ
- $\mathbf{X}(t) - \mathbf{X}(s) \mid \mathbf{X}(s) \sim N_d((t-s)\mu, (t-s)\Sigma)$, where

$$\Sigma = \left[\left\{\sigma_i \sigma_j \rho_{ij}\right\}_{i,j=1}^{d}\right] = \begin{bmatrix} \sigma_1^2 & \cdots & \sigma_1\sigma_d\rho_{1,d} \\ \vdots & \ddots & \vdots \\ \sigma_1\sigma_d\rho_{1,d} & \cdots & \sigma_d^2 \end{bmatrix} = (\sigma^\top) \circ \rho$$

How to go from a univariate uniform and invert to a multivariate distribution? We show for multivariate normal below:

12.7 Cholesky Factorization

Analog for square root for positive definite matrices

Simple way to generate correlated Normal variates from independent ones:

If $\mathbf{Z} \sim N_d(\mathbf{0}, \mathbf{I})$ and $\Sigma = \mathbf{L}\mathbf{L}^\top$ is the Cholesky factorization of the covariance matrix Σ, then $\mathbf{X} = \mathbf{L}\mathbf{Z} \sim N_d(\mathbf{0}, \Sigma)$ - Note: $\mathbf{L}$ is lower diagonal $\mathbf{L} = \begin{bmatrix} l_{11} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ l_{d1} & \cdots & l_{dd} \end{bmatrix}$ - $\mathbb{V}(L \cdot Z) = \underline{L}\underbrace{\mathbb{V}(\underline{Z})}_{I}\underline{L}^T = \underline{L} \cdot \underline{L}^T = \underline{\Sigma}$

So the process goes: - Get d uniforms - Apply a linear transformation

12.7.1 Example

$$\begin{bmatrix} W_1 \\ W_2 \\ W_3 \end{bmatrix} \sim N\left(\underline{0}, \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix}\right)$$

$$\text{Let } L = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \implies \Sigma = L \times L^T = \begin{bmatrix} z_1 \\ z_1 + z_2 \\ z_1 + z_2 + z_3 \end{bmatrix}$$

$$\implies Z_i \sim N(0, \Sigma)$$

12.7.2 Multivariate Geometric BM

$\{\mathbf{S}(t)\} w/\ \mathrm{SDE}\, d\mathbf{S}(t) = \mu \circ \mathbf{S}(t)dt + \sigma \circ \mathbf{S}(t)d\mathbf{W}(t)$ - Solution given by $\mathbf{S}(t) = \exp\{\mathbf{X}(t)\}$ where

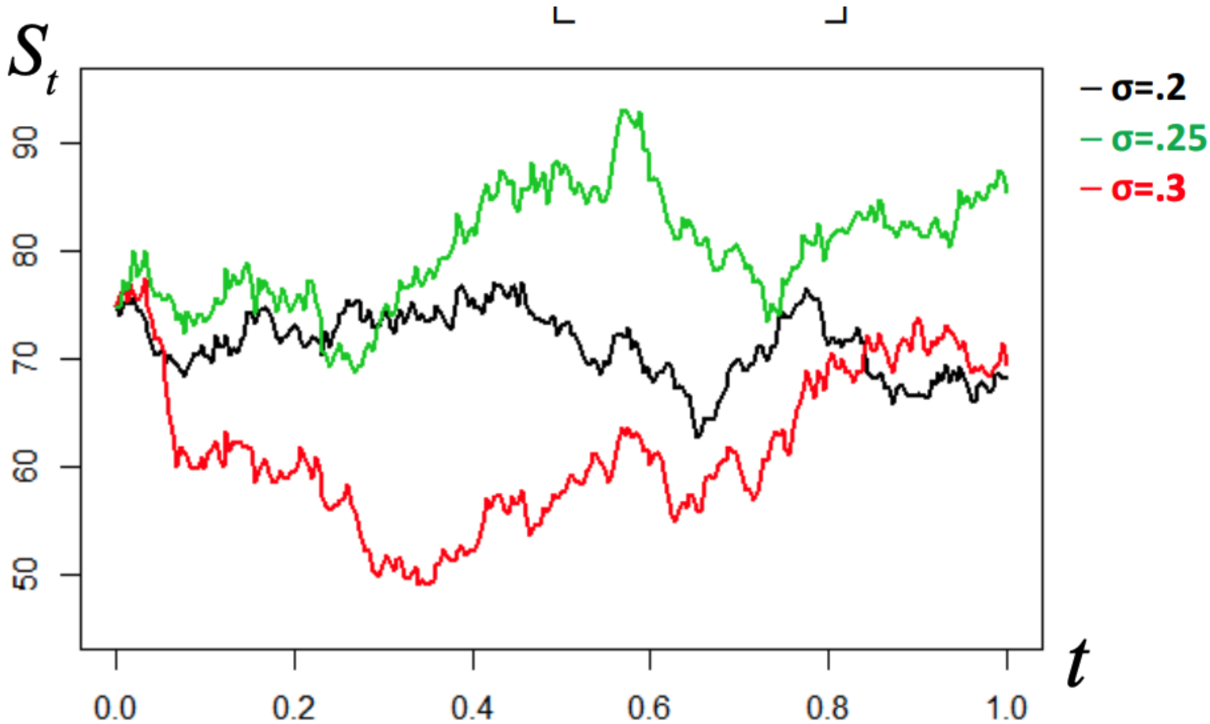
$$d\mathbf{X}(t) = \left(\mu - \sigma^2/2\right) dt + \sigma d\mathbf{W}(t)$$

Generate geometric BM variates as

$$\mathbf{S}\left(t_i\right) = \mathbf{S}\left(t_{i-1}\right) \circ \exp\left\{\left(\mu - \frac{\sigma^2}{2}\right)\Delta t + \mathbf{L}\mathbf{Z}_i\sqrt{\Delta t}\right\}, i = 1, \ldots, m$$
$$\text{where } \mathbf{Z}_i \sim^{iid} N_d(\mathbf{0}, \mathbf{I}), \mathbf{L}\mathbf{L}^\top = \; = \left(\sigma\sigma^\top\right) \circ$$

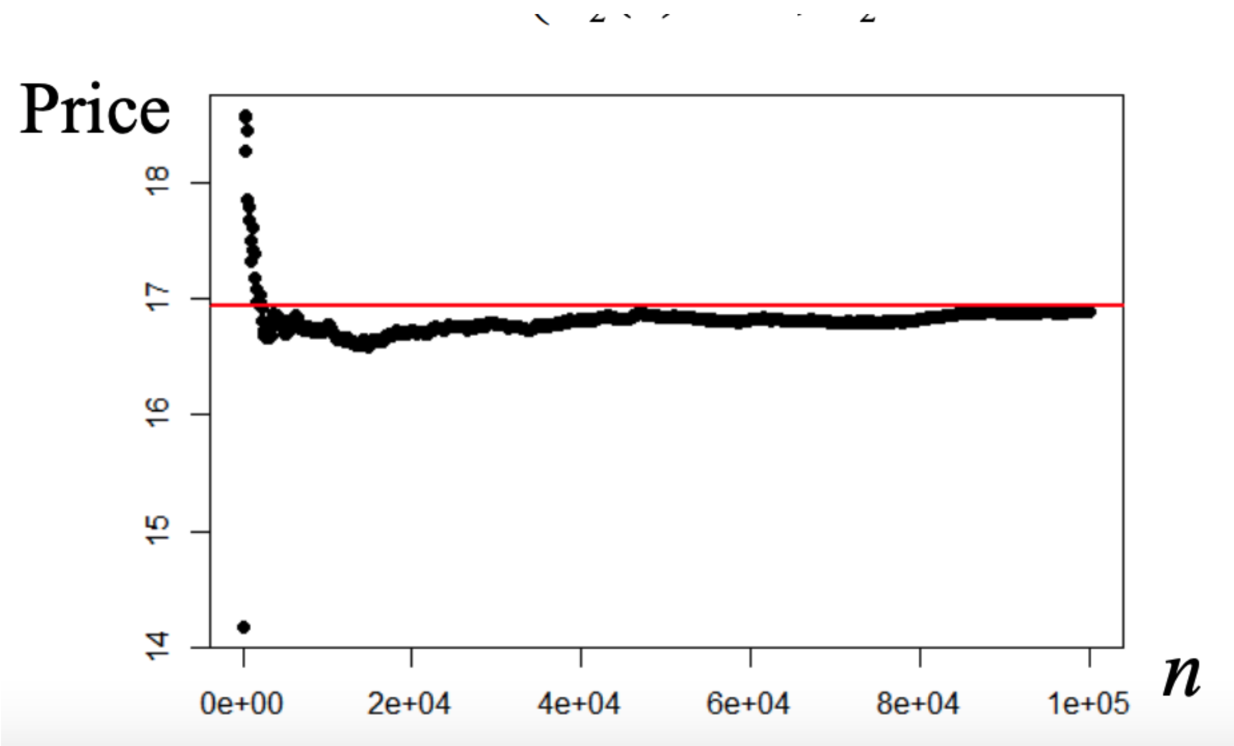
Example

$$\text{Geometric BM paths w/ } \rho = \begin{bmatrix} 1 & .4 & .4 \\ .4 & 1 & .4 \\ .4 & .4 & 1 \end{bmatrix}$$



Exchange option MC price

$$\begin{pmatrix} S_1(0) = 50, \sigma_1 = .4 \\ S_2(0) = 35, \sigma_2 = .3 \end{pmatrix}, r = 3\%, \rho = .3$$


Price
14
15
16
17
18
0e+00
2e+04
4e+04
6e+04
8e+04
1e+05
n

Chapter 13

Pricing Derivatives

13.1 Path Dependent Options

First we'll look at options where the price depends on (aspects of) the entire asset price path. Unlike a European option, whose payoff depends only on the asset price at expiry.

13.1.1 Barrier Options

An example is a **barrier option**, where the payoff depends on whether the asset price crosses some barrier prior to expiry. At expiry the payoff will be similar to a call/put, but differs between whether the barrier is hit (activated/knocked out)

4 types of Barrier options:

1. Up-and-out (U&O): price starts below barrier & has to move up for option to be knocked out
2. Down-and-out (D&O): price starts above barrier & has to move down for option to be knocked out
3. Up-and-in (U&I): price starts below barrier & has to move up for option to become activated
4. Down-and-in (D&I): price starts above barrier & has to move down for option to become activated

13.1.1.1 Example

Let C/P be price of plain Euro call/put option, CD&O be that of Euro down-&-out call, etc

- Find $C_{U\&O}$ when barrier $B < K$ strike

  The option needs to reach K in order to have a positive pay off, but in order to reach that price, it needs to pass the barrier B. However once it passes B, then the call option is worthless. $\therefore C_{U\&O} = 0$

- Find $P_{U\&I} + P_{U\&O}$, where options have same B, K, T, etc

  We can create Put-Call Parity, the sum of both of them should be equivalent to the price of a normal European put $P = P_{U\&I} + P_{U\&O}$. If under barrier, only $P_{U\&O}$ has value, while if over the barrier, only $P_{U\&I}$ has value.

Similarly,

- $C = C_{U\&I} + C_{U\&O}$
- $P = P_{D\&I} + P_{D\&O}$
- $C = C_{D\&I} + C_{D\&O}$

This is because having both of these options, it is equivalent to having an active option at all times.

13.1.1.2 Barrier options only depend on min/max

$$M_T = max\left\{S_t\right\}_{0\leq t\leq T} \quad \& \quad m_T = min\left\{S_t\right\}_{0\leq t\leq T}$$

so where $\mathbb{I}_{\{M_T<B\}}$ is an indicator variable that the max of S_T over time T crosses B or not,

- $C_{U\&O} = e^{-rT}\mathbb{E}\left[(S_T - K)_+\mathbb{I}_{\{M_T<B\}}\right]$
  - We can simulate S_T and M_T dependently, as S_T is some random walk
  - Use a Brownian Bridge to do it
- $C_{D\&O} = e^{-rT}\mathbb{E}\left[(S_T - K)_+\mathbb{I}_{\{m_T>B\}}\right]$
  - Want the minimum to be greater than the barrier for this option to keep it's value.
- $C_{U\&I} = e^{-rT}\mathbb{E}\left[(S_T - K)_+\mathbb{I}_{\{M_T>B\}}\right]$
- $C_{U\&O} = e^{-rT}\mathbb{E}\left[(S_T - K)_+\mathbb{I}_{\{m_T<B\}}\right]$

13.2 Simulating Geometric Brownian Motion (GBM) Paths

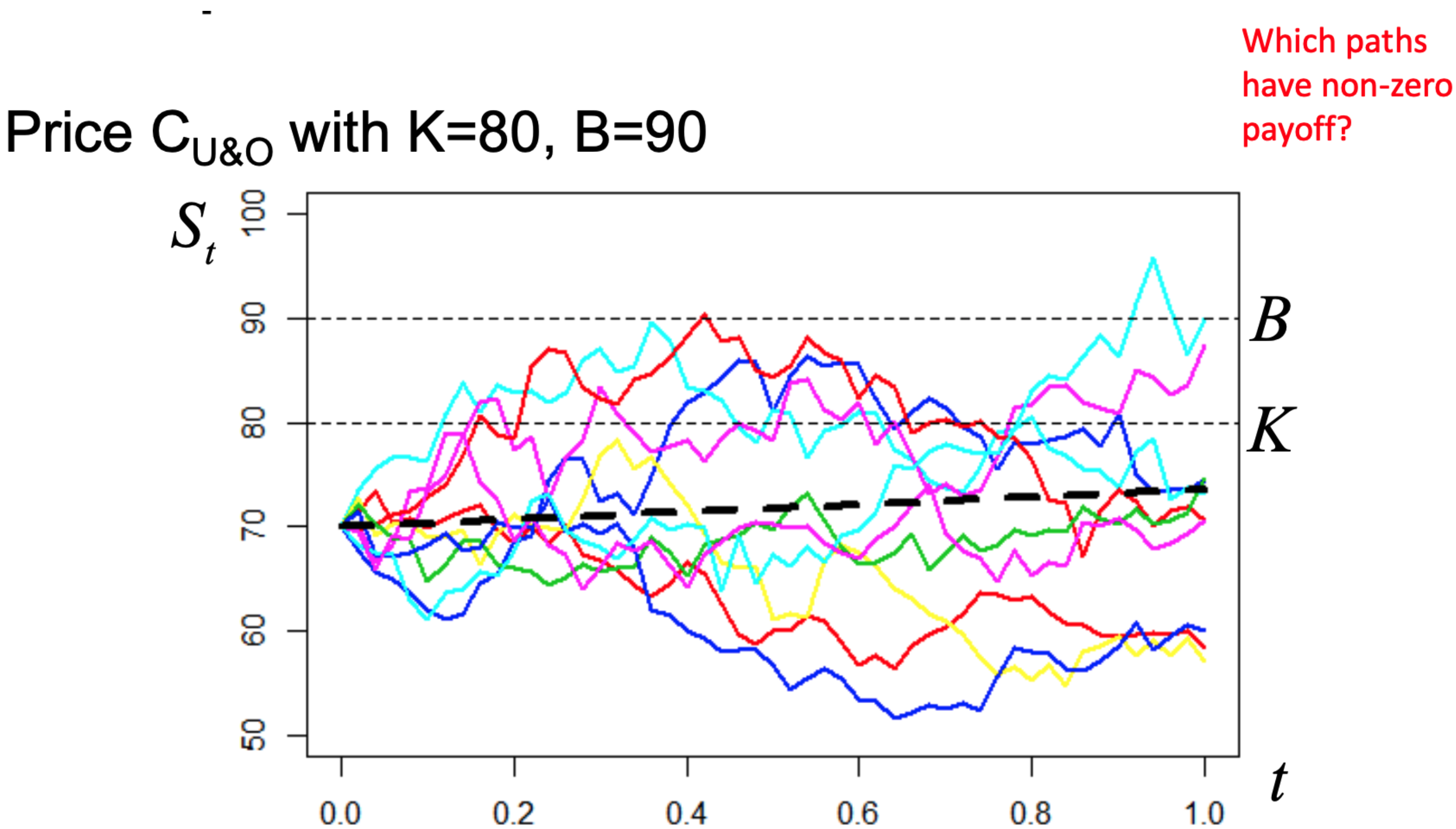
To price general path -dependent options, we need to simulate asset price paths $\left\{S_t\right\}_{0\leq t\leq T}$

In practice, we must discretize time: simulate asset price at #m points

$$\left\{S(t_i)\right\}_{i=0}^m \text{ where } t_i = i\frac{T}{m} = i\cdot\Delta t, \forall i = 1,\ldots,m$$

Which allows us to create GBM $dS_t = rS_tdt + \sigma S_tdW_t$ (Risk-Neutral Measure)

$$S(t_i) = S(t_{i-1})\exp\left\{\left(r - \frac{\sigma^2}{2}\right)\Delta t + \sigma\sqrt{\Delta t}\times Z_i\right\} \text{where } \begin{cases}\Sigma t = \frac{T}{m}\\ Z_i \sim^{iid} N(0,1)\\ i = 1,\ldots,m\end{cases}$$



Discounted prices are martingales(?) Only the purple path above K at time T will have a non-zero payoff, the cyan path crossed the barrier above so it is out. The rest are under the strike and worthless.

13.3 Monte Carlo for Barrier Options

The above will be biased, even if the number of iterations is large. This is because there may still be some time point that we did not simulate, the barrier could have been crossed and the value of the an option, ex $C_{U\&O}$ would have been worthless. Hence the simulation will always overestimate the value of an option, because it underestimates the maximum.

To address this bias, you could increase m, and let it go to ∞, we could be more accurate but the computation would become increasingly expensive.

There is a tradeoff between # paths(n) and # steps (m),

$$n \uparrow \iff Var \downarrow, m \uparrow \iff bias \downarrow \quad \text{which represents the Bias/Variance Trade-off}$$

13.3.1 Example: Find maximum by time T

For a standard BM $\{W_t\}$ find the distribution of maximum by time T:

-Assume std BM always starts at 0

$$M_t = max\left\{W_t\right\}_{0\le t\le T} \quad \text{WTF CDF } P(M_T \le y)$$

$$\text{First look at: for } x \le y, P(W_T \le x, M_T \ge y) = \underbrace{P(W_T \ge 2y - x,\ \overbrace{M_T \ge y}^{\text{always true}})}_{\text{By relfection principle}}$$

All paths that start at 0, cross above y at some point, and end at some value x which is below y.

By the reflection principle, the path is just as likely to have the reflected path. We first consider the first time the path hits y, and we take the symmetric path to the existing one (equally likely as the true path).

So the probability that the path ends below x and hits y at some point, we can just look at when it hits y and find when we cross $(y - x)$ above y, which is equivalent to hitting x from y eventually. This quantity becomes $y + (y - x) = 2y - x$.

$$\begin{aligned}\text{For } x \le y,\ \underbrace{P(W_T \le x, M_T \le y)}_{P(A\cap B)} &= \underbrace{P(W_T \le x)}_{P(A)} - \underbrace{P(W_T \le x, M_T \ge y)}_{P(A\cap B^C)} \\ &= P(W_T \le x) - P(W_T \ge 2y - x)\end{aligned}$$

$$\begin{aligned}\implies \text{For } x = y, P(W_T \le y,\ \overbrace{M_T \le y}^{\implies W_T \le y}) &= P(M_T \le y) = P(W_T \le y) - P(W_T \ge 2y - y) \\ &= P(W_T \le y) - P(W_T \ge y)\end{aligned}$$

We know the distribution of W_T is normal with $\mu = 0, \sigma^2 = T$ (for Brownian motion), and because of symmetry of the Normal distribution,

$$\begin{aligned}&= P(W_T \le y) - P(W_T \le -y) \\ &= P(-y \le W_T \le y) \\ &= P(|W_T| \le y) \quad \text{As } y \text{ will always be positive}\end{aligned}$$

$$\implies M_T \sim |W_T| \quad \text{Which is the folded normal, only pos } x > 0$$

The maximum is distributed just like the absolute value of the Std Brownian motion at time T

13.3.2 Example: Find Prob of W_T hitting barrier B by time T

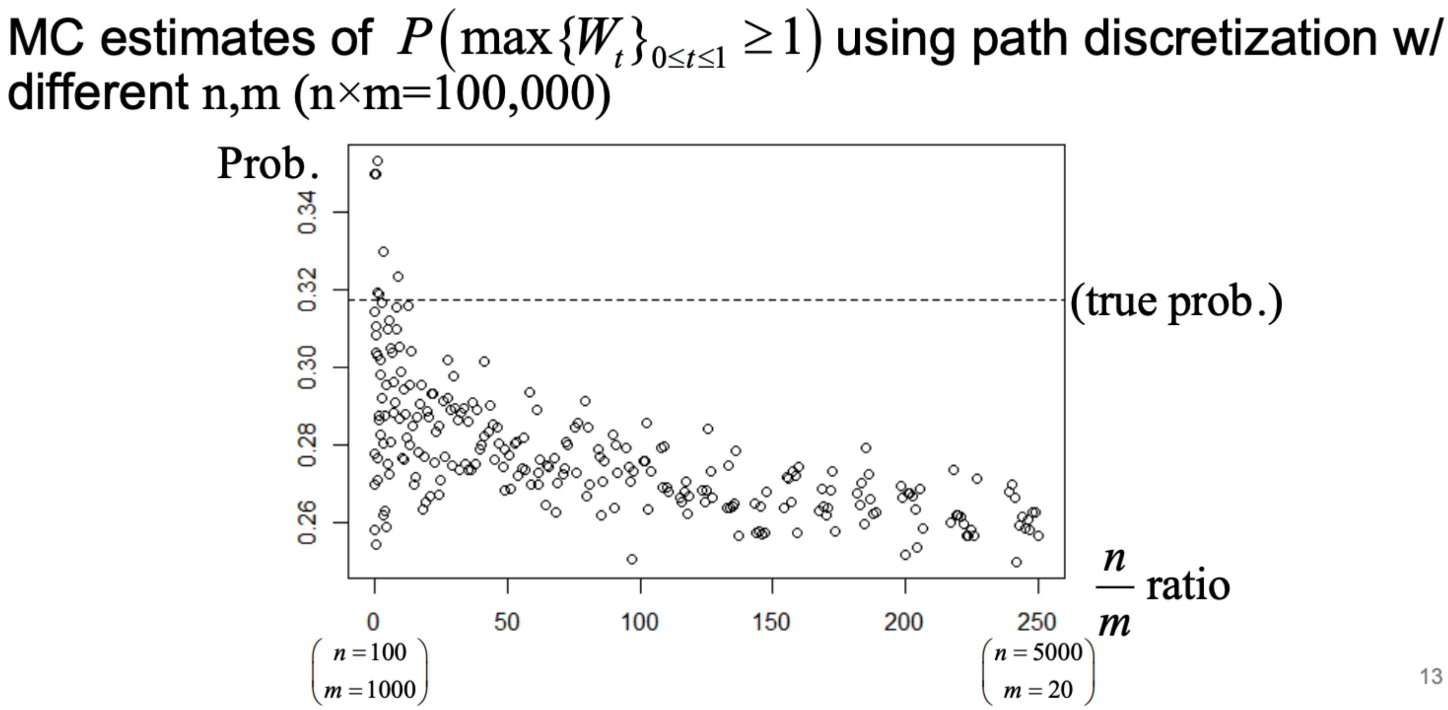
The probability of standard BM $\{W_t\}$ hits barrier $B = 1$ before time $T = 1$ is

$$\begin{aligned}M_T \sim |W_T| \implies &P(\text{BM hits 1 before } T = 1) \\ &= P(M_1 \ge 1) = P(\overbrace{|W_T|}^{\sim N(0,1)} \ge 1) \\ &= 2P(Z \ge 1) = 2 \cdot \overbrace{\Phi(-1)}^{\sim \text{CDF of std Normal}} \\ &= 0.317862\end{aligned}$$

13.3.2.1 By simulation

If we don't know the reflection principle, and we wanted to simulate it instead:

- Simulate m random Normals $\sim N\left(0, \frac{1}{m}\right)$ and take their cumulative sum to create a path which ends at time 1
- Do this for n paths
- Calculate the proportion of paths which had a max M_T above the barrier B



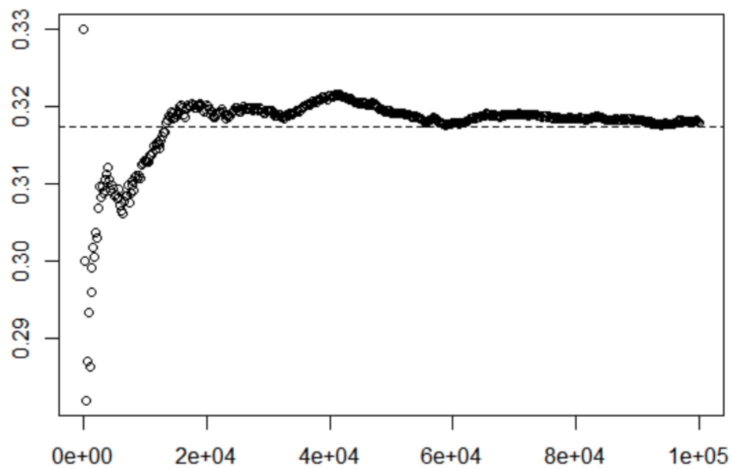
This plot shows the trade off between granularity (m) and number of paths (n) when estimating the probability When m is low the bias is high but the variance is low, whereas on the left side the variance is really high but the bias is low (close to true prob)

13.3.3 Example: MC Simulation w/o bias

Estimate probability that standard BM hits 1 before time 1, with MV but without bias? We can look at the distribution of maximum.

We can generate values of M_T directly by generating a standard brownian motion W_T and setting $M_T = |W_T|$. We then estimate the probability that $|W_T|$ crosses 1, and as we generate more Normal RVs (W_T), the probability will converge.

MC estimates of $P\left(\max\{W_t\}_{0\leq t\leq 1} \geq 1\right)$ using direct simulation of $\max\{W_t\}$ w/ n=100,000



15

13.4 Extrema of Brownian Motion

- However, one can easily simulate random deviates of maximum using Brownian bridge
  - Construction allows for general treatment of extrema of various processes

The reflection principle doesn't work for arithmetic BM, due to drift. (The paths that go in the direction of the drift will have higher probability than the paths that oppose it.)

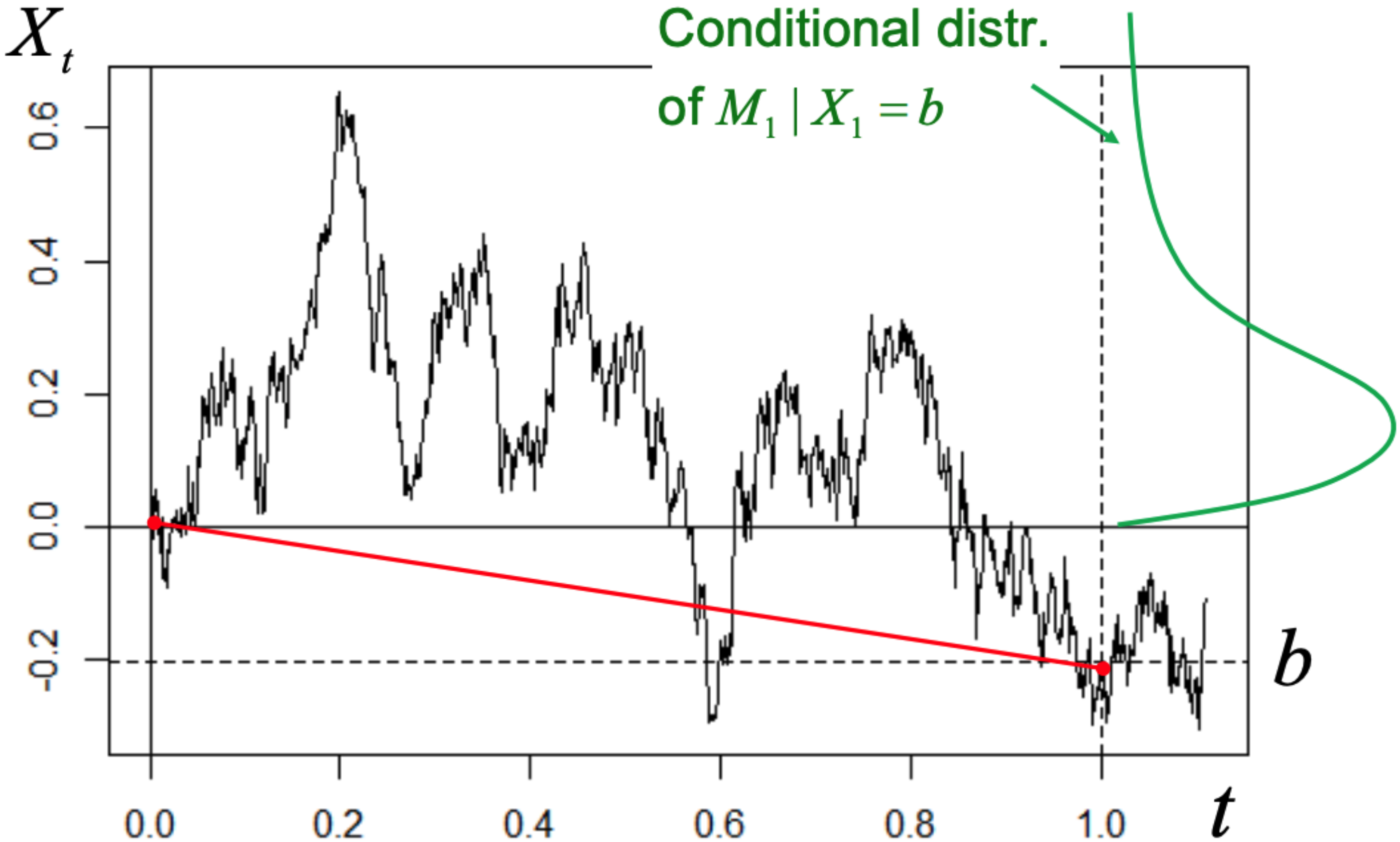
If we fix the starting and ending point of our arithmetic BM however, we can still kind of use the reflection principle.

Conditional on a final point $X_T = h$ then the maximum $(M_T|X_T) = max_t(X_t|X_T)$ of the Brownian bridge process has a **Rayleigh** distribution.

$$P(M_T \leq m|X_T = b) = 1 - \exp\left\{-2\frac{m(m-b)}{\sigma^2 T}\right\} \quad \forall m \geq (0 \cup B)$$

Note that distribution of conditional maximum is independent of the drift, given $X_T = b$

The lower the ending point, the closer the max value is to 0, the starting point.

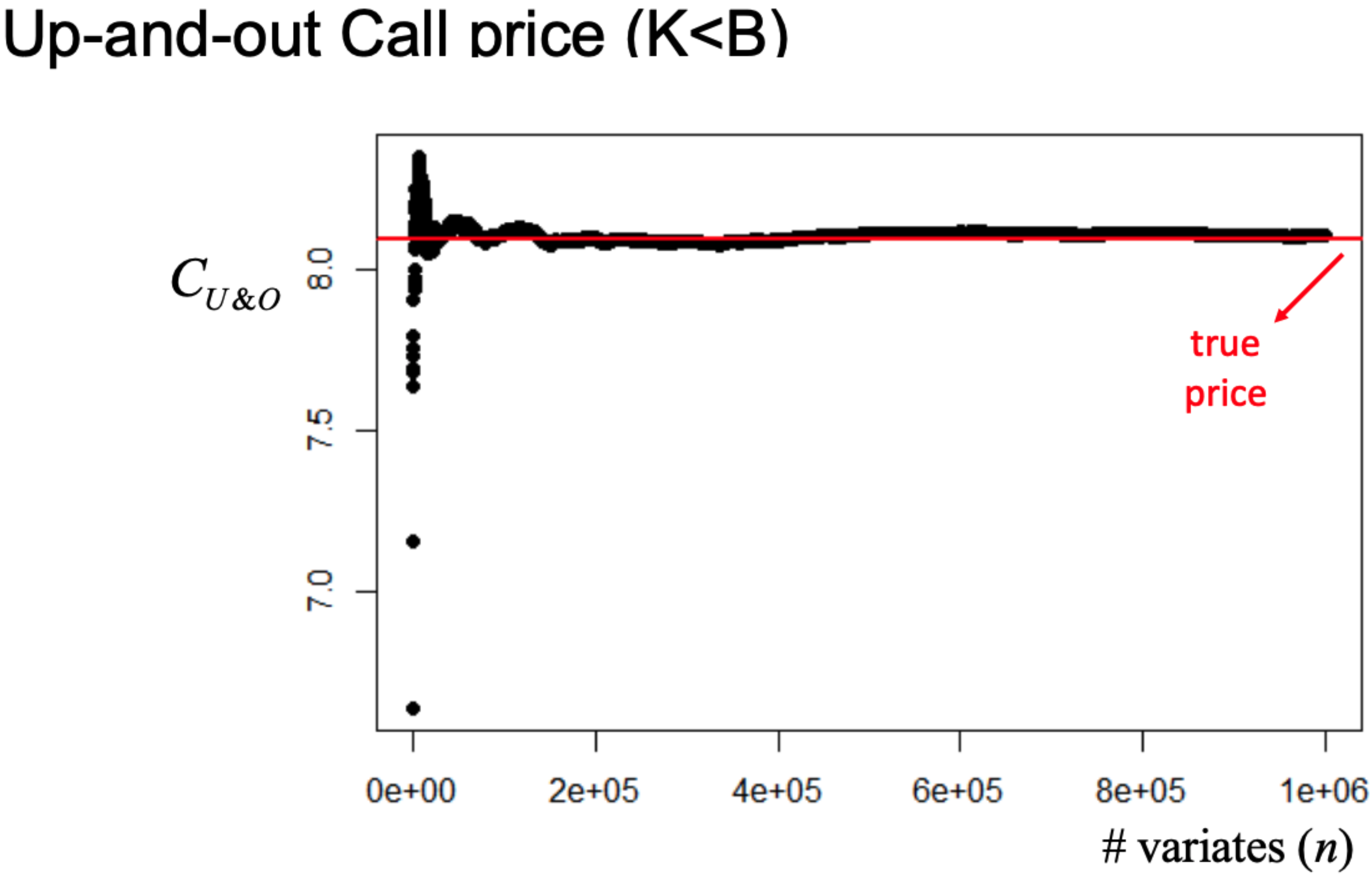


13.4.1 Procedure to generate maxima

Procedure for simulating maxima of arithmetic BM: 1. Generate $X_T \sim N(\mu T, \sigma^2 T)$ 2. Generate $U \sim \text{Uniform}(0, 1)$ 3. Calculate $M_T \mid X_T = \frac{X_T + \sqrt{X_T^2 - 2\sigma^2 T \log(U)}}{2}$

For maxima of geometric BM, exponentiate arithmetic BM result

13.4.2 Example: U&O Call



13.4.3 Example: Simulate minimum of arithmetic BM based on max

If we have an arithmetic BM with some drift μ and volatility σ, we can simulate from the exact distribution the max of the process and the ending price.

We can also use this to simulate the minimum by using symmetry. You'd simulate arithmetic BM paths with negative drift $-\mu$ and find it's maximum. That would become minimum of the normal arithmetic BM with positive μ.

13.5 Time Discretization

What happens when a stochastic process which is not straight forward? (Not GBM or arithmetic?), for example with stochastic drift and stochastic volatility.

The process is no longer log Normal, so we must use discretization.

$$dS_T = \mu_t dt + \sigma_t dW_t$$

- Path-dependent options generally require simulation of entire discretized path
  - Exceptions are options depending on maximum (e.g. barrier, lookback)
- If prices do not follow GBM, it is not generally possible to simulate from exact distribution of asset prices
  - Need to approximate sample path distribution over discrete times

13.5.1 Euler Discretization

Consider general SDE where drift/volatility can depend on time (t) and/or process (St)

$$dS_t = \mu\left(t, S_t\right) dt + \sigma\left(t, S_t\right) dW_t$$

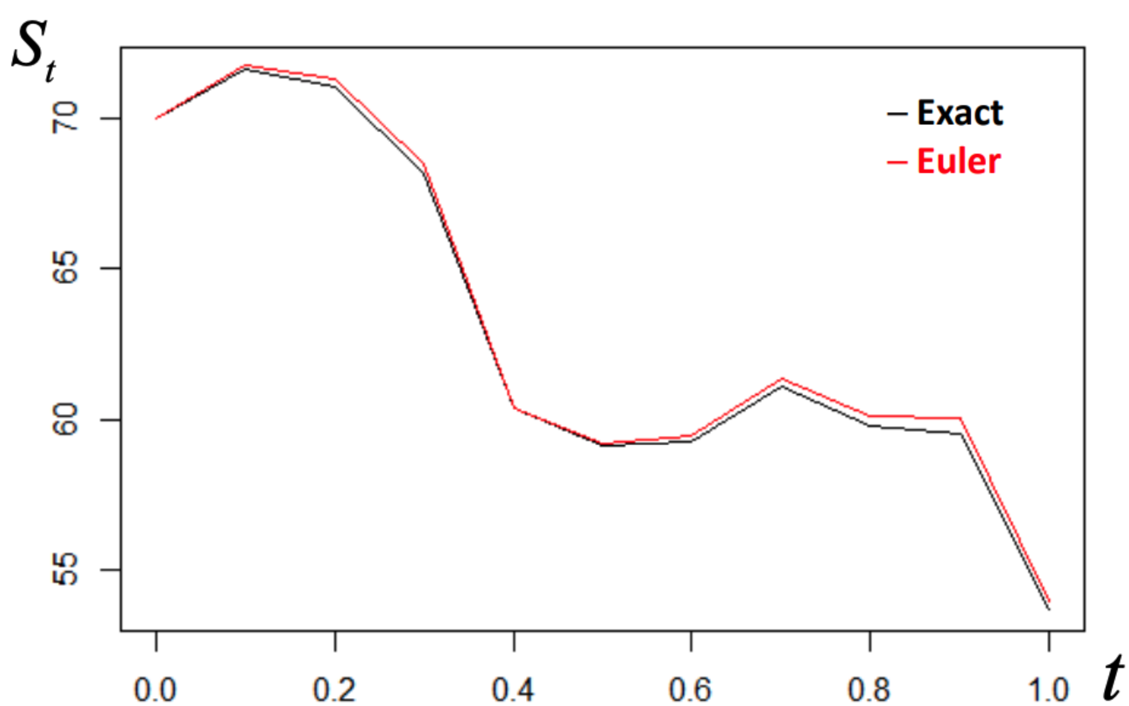
- There is no general explicit solution for S_t - Distribution of St is unknown (in closed form) - Notable exceptions are Arithmetic/Geometric BM - Behavior of S_t can be approximated using discretization scheme

SDE: $dS_t = \mu\left(t, S_t\right) dt + \sigma\left(t, S_t\right) dW_t$ Discretize time $t_i = i(T/m) = i\Delta t, i = 0, \ldots, m$ Simulate (approx.) path recursively, using

$$S\left(t_i\right) = S\left(t_{i-1}\right) + \mu\left(S\left(t_{i-1}\right), t_{i-1}\right) \Delta t + \sigma\left(S\left(t_{i-1}\right), t_{i-1}\right) \sqrt{\Delta t} Z_i$$
$$\text{for } i = 1, \ldots, m, \text{ where } Z_i \sim^{iid} N(0,1)$$

- To approximate distribution of $S(T)$, generate multiple (#n) discretized paths - Method called Euler (or 1st order) discretization

- **Exact vs Euler discretization of GBM (m=10)**



Chapter 14

Variance Reduction

Running simulation basically invokes SLLN. To sample the expectation of some RV Z, we can sample a fixed n variables and the sample means will converge to $\mathbb{E}(Z)$ at the rate $\frac{1}{\sqrt{n}}\sigma$ (scaled by the variance).

We can converge faster simply by increase the number of samples, **or we can decrease the variance**. We'll tackle the second method.

14.1 Techniques

14.1.1 Antithetic Variables

Idea: For any Normal variate Z_i consider it's negative $-Z_i$. (It will follow the same distribution, if it's centered at 0. If not, you can shift it)

- They come from the same distribution, but are *dependent*
- Generally, for Uniform (0,1), use U_i and $1-U_i$

We can calculate the discounted payoff, denoted by Y under both:

$$Y_i = f(Z_i), \tilde{Y}_i = \tilde{f}(-Z_i)$$

And then estimate price as

$$\bar{Y}_{AV} = \frac{1}{2n}\left(\sum_{i=1}^{n} Y_i + \sum_{i=1}^{n} \tilde{Y}_i\right) = \frac{1}{n}\sum_{i=1}^{n}\frac{Y_i + \tilde{Y}_i}{2}$$

- Balance payoffs of paths with "opposite" returns.

  Instead of generating many pseudo random numbers, just take the antithetic (negative) variable (faster)

We can find the asymptotic distribution of antithetic variable estimator in terms of moments of $\frac{Y_i+\tilde{Y}_i}{2}$

$$\bar{Y}_{AN} \text{ is sample mean of iid } RV_i : \frac{Y_i + \tilde{Y}_i}{2}$$

$$\text{by CLT } \bar{Y_{AN}} \sim^{approx} N\left(\mathbb{E}\left[\frac{Y_i + \tilde{Y}_i}{2}\right], \frac{1}{n}\mathbb{V}\left[\frac{Y_i + \tilde{Y}_i}{2}\right]\right)$$

$$\begin{aligned}
\text{where } \mathbb{E}\left[\frac{Y_i + \tilde{Y}_i}{2}\right] &= \frac{1}{2}\left[\mathbb{E}(Y_i) + \mathbb{E}(\tilde{Y}_i)\right] \\
&= \frac{1}{2}2\mathbb{E}(Y) = \mathbb{E}\left(f(Z)\right) \\
\mathbb{V}\left[\frac{Y_i + \tilde{Y}_i}{2}\right] &= \frac{1}{n}\frac{1}{4}\left(\mathbb{V}[Y_i] + \underbrace{\mathbb{V}[\tilde{Y}_i]}_{=\mathbb{V}[Y_i]} + 2Cov(Y_i, \tilde{Y}_i)\right) \\
&= \frac{1}{n}\frac{1}{4}\left(2\mathbb{V}(Y_i) + 2Cov(Y_i, \tilde{Y}_i)\right) \\
&= \frac{1}{n}\frac{1}{2}\left(\mathbb{V}(Y_i) + Cov(Y_i, \tilde{Y}_i)\right)
\end{aligned}$$

Antithetic variables may be simple and easy to generate but they won't always help, although they will **if the original and antithetic variables are negatively related. If they are unrelated/positively related, then you get worse results.**

Using the antithetic variable vs Generating twice the number of samples

Proof that $\mathbb{V}[\bar{Y}_{AV}] < \mathbb{V}\left[\frac{1}{2n}\sum_{i=1}^{2n} Y_i\right]$

$$\begin{aligned}
\mathbb{V}\left[\frac{1}{2n}\sum_{i=1}^{2n} Y_i\right] &= \frac{1}{2n}\mathbb{V}(Y) \\
\mathbb{V}[\bar{Y}_{AV}] &= \frac{1}{2n}\cdot\left(\mathbb{V}(Y) + Cov(Y_i, \tilde{Y}_i)\right) \\
\left(\mathbb{V}(Y) + Cov(Y_i, \tilde{Y}_i)\right) &\leq \mathbb{V}(Y) \text{ iff } Cov(Y_i, \tilde{Y}_i) \leq 0
\end{aligned}$$

This shows that the payoff is only worthwhile if $f(Z)$ is not an even function, as if it was, then $\frac{Y_i + \tilde{Y}_i}{2}$ would simply be Y_i and that wouldn't help us decrease varaince.

Antithetic Variable pricing of European call

n	$\bar{Y}$	$\bar{Y}_{AV}$	$s.e.(\bar{Y})$	$s.e.(\bar{Y}_{AV})$
50	5.8626	4.7646	0.7617	0.4623
250	4.7019	4.7439	0.3018	0.2362
500	4.3211	4.7834	0.2013	0.1722
2500	4.7537	4.6539	0.1017	0.0734
5000	4.6634	4.6923	0.0704	0.0531
25000	4.7503	4.7024	0.0317	0.0238
50000	4.7046	4.6941	0.0224	0.0166

true Black-Scholes price = 4.7067

For a European call, using the antithetic variables will give us improved results because the payoff of the European call is not an even function. There is an improvement in the accuracy of prediction using antithetic variables.

Antithetic variables are a very crude way of estimation, but it's easy to program. It ensures that you have as many numbers on one side of the distribution as the other.

14.1.2 Stratification

This is inspired by statistical sampling. Idea: Split RV domain into equi-probable strata and draw equal number of variates from within each one.

- E.g. (2 strata) draw equal number of independent positive and negative Z_i

Stratification ensures equal representation of each stratum in the RV domain

- Always reduces variance, but could be marginal improvements.
- It's worth doing stratification when **the target RV (payoff) changes over its domain.** Like a payoff function that may be exponential.

  Why don't we always use it? It's expensive, as you may need to calculate conditional distribution, and many details that require work to figure out. However, for the Normal, a conditional Normal is still Normal, which makes things easy in that particular case. Otherwise, it may require numerical computation and may not be worth it.

Consider $\#m$ equi-probable Normal strata $\{A_j\}$

$$P\big(Z \in A_j\big) = 1/m \text{ for } Z \sim N(0,1), j = 1, \dots, m$$

Stratified estimator of $Y = f(Z)$ (payoff, but in general could be any function)

$$\begin{aligned}
&\bar{Y}_{Str} = \frac{1}{m}\sum_{j=1}^{m} \bar{Y}^{(j)}, \text{ where } \bar{Y}^{(j)} = \frac{1}{n}\sum_{i=1}^{n} f\left(Z_i^{(j)}\right) \\
&Z_i^{(j)} \sim^{\text{iid}} N\left(0, 1 \mid Z_i^{(j)} \in A_j\right), j = 1, \dots, m
\end{aligned}$$

- $\bar{Y}^{(j)}$ is estimator within each stratum j

We can verify that $\bar{Y}_{Str}$ is an unbiased estimator of $\mathbb{E}(f(Z))$.

$$\begin{aligned}
\mathbb{E}(\bar{Y}_{Str}) &= \mathbb{E}\left[\frac{1}{m}\sum_{j=1}^{m}\bar{Y}^{(j)}\right] = \frac{1}{m}\sum_{j=1}^{m}\mathbb{E}[\bar{Y}^{j}] \\
&= \frac{1}{m}\sum_{j=1}^{m}\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} Y_j^{(i)}\right) \\
&= \frac{1}{m}\sum_{j=1}^{m}\frac{1}{n}\sum_{i=1}^{n}\underbrace{\mathbb{E}\left[Y_i^{(j)}\right]}_{\mathbb{E}[Y|Z\in A_j]=\mathbb{E}\left[f(Z)|Z\in A_j\right]} \\
&= \frac{1}{m}\sum_{j=1}^{m}\frac{1}{n}n\mathbb{E}[f(Z)|Z \in A_j] \\
&= \frac{1}{m}\sum_{j=1}^{m}\mathbb{E}(f(Z)|Z \in A_j) \\
&= \sum_{j=1}^{m}\mathbb{E}\left[f(Z)|Z \in A_j\right] \cdot P(A \in A_j) \\
\text{By Law of Tot. Prob } &= \mathbb{E}(f(Z)) \cdot \mathbb{E}(Y)
\end{aligned}$$

We want to show that stratified sampling cannot do worse (higher variance) than simple random sampling ie $\mathbb{V}(\bar{Y}_{Str}) \leq \mathbb{V}(\bar{Y})$ where $\bar{Y} = \frac{1}{nm}\sum_{i=1}^{nm} f(Z_i)$, with n numbers per m strata

Stratification can't always be used because it's computationally expensive

$$\begin{aligned}
\mathbb{V}(\bar{Y}) = \mathbb{V}\left(\frac{1}{nm}\sum_{i=1}^{nm} Y_i\right) &= \frac{1}{nm}\mathbb{V}(Y_i) \\
&= \frac{1}{nm}\left(\mathbb{E}[Y_i^2] \cdot [\mathbb{E}(Y_i)]^2\right) \\
&= \frac{1}{nm}\left[\mathbb{E}(f^2(z))\right] - \mu^2
\end{aligned}$$

$$
\begin{aligned}
\mathbb{V}(\bar{Y}_{Str}) &= \mathbb{V}\left(\frac{1}{m}\sum_{j=1}^{m}\bar{Y}^{(j)}\right) \\
&= \frac{1}{m^2}\sum_{j=1}^{m}\mathbb{V}(\bar{Y}^{(j)}) \\
&= \frac{1}{m^2}\sum_{j=1}^{m}\mathbb{V}\left[\frac{1}{n}\sum_{i=1}^{n}\underbrace{Y^{(j)}}_{\text{iid in each sample}}\right] \\
&= \frac{1}{m^2n^2}\sum_{j=1}^{m}\sum_{i=1}^{n}\mathbb{V}\left(Y_i^{(j)}\right) \\
&= \frac{1}{m^2n^2}\sum_{j=1}^{m}n\mathbb{V}(Y^{(j)}) \\
&= \frac{1}{m^2n}\sum_{j=1}^{m}\mathbb{V}(\underbrace{Y}_{f(z)}|z \in A_j) \\
&= \frac{1}{m^2n}\sum_{j=1}^{m}\left[\mathbb{E}\left(f^2(z)|z \in A_j\right) - \left(\mathbb{E}(f(z))|z \in A_j\right)^2\right] \\
&= \frac{1}{mn}\left\{\overbrace{\sum_{j=1}^{m}\mathbb{E}\left[f^2(z)|z \in A_j\right]\cdot\underbrace{\frac{1}{m}}_{P(z\in A_j)}}^{\text{By LTP, } \mathbb{E}[f^2(z)]} - \frac{1}{m}\sum_{j=1}^{m}\underbrace{\left(\mathbb{E}(f(z)|z \in A_j)^2\right)}_{\mu_j}\right\} \\
&= \frac{1}{mn}\left\{\mathbb{E}\left[f^2(z) - \frac{1}{m}\sum_{j=1}^{m}\mu_j^2\right]\right\} \\
&\le \mathbb{V}\left[\bar{Y}\right] = \frac{1}{mn}\left\{\mathbb{E}\left[f^2(z) - \mu^2\right]\right\} \\
&\iff \frac{1}{m}\sum_{j=1}^{m}\mu_j^2 \ge \mu^2
\end{aligned}
$$

Which is true by Jensens Inequality.

Note $\mu = \frac{1}{m}\sum_{j=1}^{m}\mu_j = \sum_{j=1}^{m}\underbrace{\mathbb{E}\left[f(z)|z \in A_j\right]}_{\mu_j}\underbrace{P(A_j)}_{1/m}$

For convex functions:

$$
g(x) = x^2, \text{then } \mathbb{E}(g(X)) = \underbrace{\mathbb{E}(X^2)}_{\frac{1}{m}\sum_{j=1}^{m}\mu_j^2} \ge \underbrace{\left(\mathbb{E}(X)\right)^2}_{\left(\frac{1}{m}\sum_{j=1}^{m}\mu_j\right)^2=\mu^2}
$$

Because we're subtracting a larger value in the variance, then we get a smaller variance. If we have a payoff that has lots of jumps/large values, then this would give us lots of benefit.

Stratified pricing of European call

m	n	$\bar{Y}_{Str}$	$s.e.(\bar{Y}_{Str})$
1	10000	4.6908	0.0712
10	1000	4.7174	0.0207
20	500	4.7303	0.0136
50	200	4.7065	0.0088
100	100	4.7062	0.0054
200	50	4.7124	0.0046
500	20	4.7047	0.0024
1000	10	4.7068	0.0021

true Black-Scholes price = 4.7067

The more strata used, the bigger the benefit. This may be known as quasi Monte Carlo, but in the multivariate case it becomes a large pain.

14.1.3 Control Variates

Comes from regression

To estimate $\mathbb{E}(Y) = \mathbb{E}[f(z)]$ using MC, we generate iid Z_i and use

$$\bar{Y} = \sum_{i=1}^{n} \frac{Y_i}{n} = \sum_{i=1}^{n} \left(\frac{f(Z_i)}{n} \right)$$

For our purposes, $f(\cdot)$ is the option's discounted payoff

We assume there is another option with payoff $g(\cdot)$ whose price we already know $\mathbb{E}(X) = \mathbb{E}(g(Z))$

Idea: Use MC with same variates to estimate both $\mathbb{E}(Y)$ and $\mathbb{E}(X)$, but adjust estimate $\bar{Y}$ to take into account the error of estimate $\bar{X}$

- E.g., if $\bar{X}$ underestimates $\mathbb{E}[X]$ then we can adjust $\bar{Y}$ upward because it most likely also underestimates $\mathbb{E}(Y)$.

For our purposes, we'll focus on linear adjustments.

14.1.3.1 Example

Adjust $\bar{Y}$ for estimation error $\bar{X} - \mathbb{E}(X)$ linearly, as

$$\bar{Y}(b) = \bar{Y} - b(\bar{X} - \mathbb{E}(X))$$

No matter what value of b we choose, the estimator will be unbiased (even though it may inefficient)

We will show that this is an unbiased estimator.

$$\begin{aligned}
\mathbb{E}(\bar{Y}(b)) &= \mathbb{E}[\bar{Y} - b(\bar{X} - \mathbb{E}(X))] \\
&= \underbrace{\mathbb{E}(\bar{Y})}_{\mathbb{E}(Y)} - b(\underbrace{\mathbb{E}(\bar{X})}_{\mathbb{E}(X)} - \mathbb{E}(X)) \\
&= \mathbb{E}(Y) - b \cdot (\mathbb{E}(X) - \mathbb{E}(X))
\end{aligned}$$

On average the adjustment will always be 0, but for certain choices of b, we can make the estimator much more efficient. We find b using regression on the random variables of Y, X that we've simulated.

We can also find the variance of $\bar{Y}(b)$

$$\begin{aligned}
\mathbb{V}[\bar{Y}(b)] &= \mathbb{V}[\bar{Y} - b(\bar{X} - \mathbb{E}(X))] = \mathbb{V}[\bar{Y} - b\bar{X}] \\
&= \mathbb{V}(\bar{Y}) + b^2\mathbb{V}(\bar{X}) - 2bCov(\bar{Y}, \bar{X}) \\
&= \frac{1}{n}\mathbb{V}(Y) + b^2\frac{1}{n}\mathbb{V}(X) - 2bCov\left(\frac{1}{n}\sum_{i=1}^{n}\underbrace{f(z_i)}_{Y_i}, \quad \frac{1}{n}\sum_{i=1}^{n}\underbrace{g(z_i)}_{X_i}\right) \\
&= \frac{1}{n}\underbrace{\mathbb{V}(Y)}_{\sigma_y^2} + b^2\frac{1}{n}\underbrace{\mathbb{V}(X)}_{\sigma_x^2} - 2b\frac{1}{n^2}nCov(f(Z)), g(Z)) \\
&= \frac{1}{n}\left[\sigma_y^2 + b^2\sigma_x^2 - 2b\sigma_{xy}\right]
\end{aligned}$$

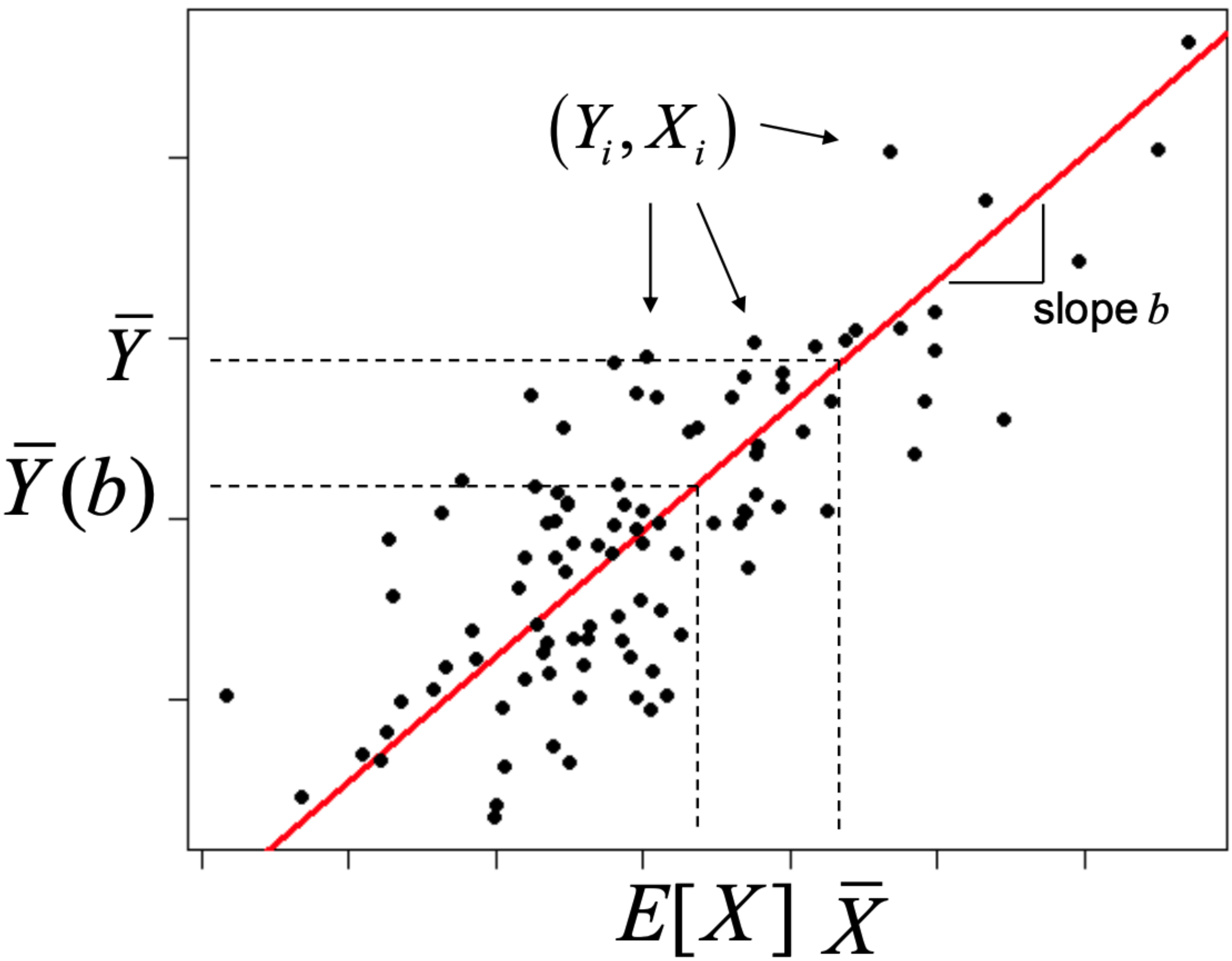
We can also show that the optimal value of b is $b^\star = \frac{Cov(X,Y)}{Var[X]}$ which is the regression slope coefficient

$$\begin{aligned}
\frac{\partial}{\partial b}\mathbb{V}(\bar{Y}(b)) = 0 &\implies \frac{\partial}{\partial b}\left(\frac{1}{n}[\sigma_y^2 + b^2\sigma_x^2 - 2b\sigma_{xy}]\right) = 0 \\
b\sigma_x^2 - \sigma_{xy} &= 0 \\
&\implies b = \frac{\sigma_{xy}}{\sigma_x^2} = \frac{Cov(X,Y)}{Var(X)} = \left(\frac{\sigma_{xy}\sigma_x\sigma_y}{\sigma_x^2} = \sigma_{xy}\frac{\sigma_y}{\sigma_x}\right)
\end{aligned}$$

We want the expected val of Y, which we can calculate by finding the above.

In practice, don't know $\mathrm{Cov}[X, Y], \mathrm{Var}[X]$ so estimate b^* using MC sample

$$\hat{b} = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}$$



We want to find $\mathbb{E}(Y)$ with the knowledge of what $\mathbb{E}(X)$ is. We can then estimate the correlation/slope.

Furthermore, we now show that the optimal variance is $Var(\bar{Y}(b^\star) = Var[\bar{Y}](1-\rho_{XY}^2)$

$$\begin{aligned}
\mathbb{V}[\bar{Y}(b^\star)] &= \frac{1}{n}(\sigma_y^2 - b^{\star 2}\sigma_x^2 - 2b^\star\sigma_{XY}) \\
&= \frac{1}{n}\left(\sigma_y^2 + \left(\rho_{XY}\frac{\sigma_Y}{\sigma_X}\right)\sigma_x\sigma_Y\rho_{XY}\right) \\
&= \frac{1}{n}\left(\sigma_Y^2 + \rho_{XY}^2\sigma_y^2 - 2\rho_{XY}^2\sigma_Y^2\right) \\
&= \frac{1}{n}\sigma_Y^2(1-\rho_{XY}^2) \\
&= \mathbb{V}(\bar{Y})\cdot(1-\rho_{XY}^2)
\end{aligned}$$

- In practice, use sample estimates of Var$[\bar{Y}]$, ρ_{XY} Good control variates have high absolute correlation with option payoff (high ρ_{XY}) !

We can apply this to price a European option

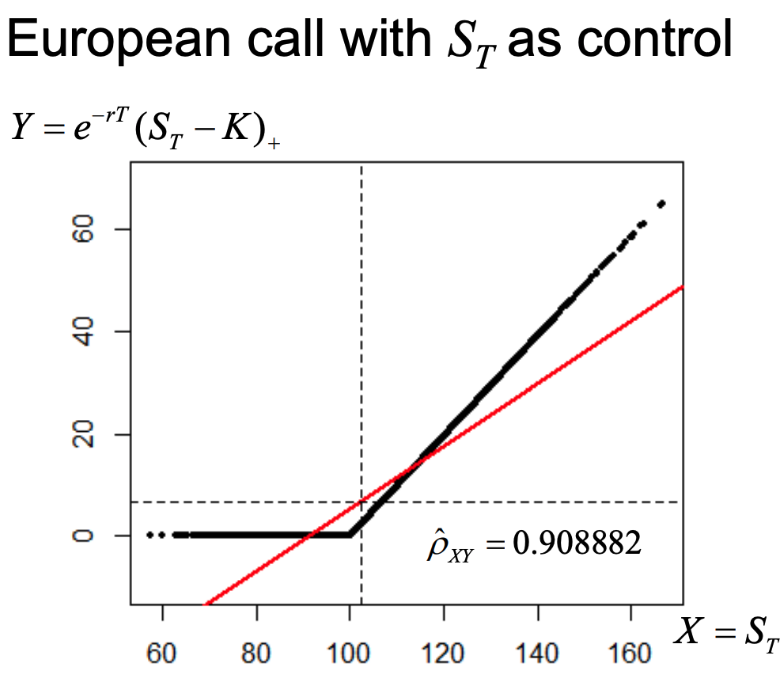
14.1.3.2 Example

Price a European option using final asset price (S_T) as control assuming GBM w/ r, σ. We know:

- $X = S_T = g(Z) = S_0 \exp\left\{\left(r - \frac{\sigma^2}{2}\right)T + \sigma\sqrt{T}\cdot Z\right\}$
- $\mathbb{E}(X) = \mathbb{E}(S_T) = S_0 e^{rT}$

We guess the correlation of control with the following:

- (Deep) In-the-money call (most likely to be exercised): $\rho_{XY} \approx 1$
  - $Cov((S_T - K)_+, S_T)$ which is most likely positively correlated
- Out-of-the-money call: $\rho_{XY} \approx 0$
  - $Cov(0, S_T) \approx 0$
- In the money put: $\rho_{XY} \approx -1$
- Out of the money put: $\rho_{XY} \approx 0$



$$\begin{pmatrix} S_0 = K = 100, T = .5 \\ r = .05, \sigma = .2, n = 10^4 \end{pmatrix}$$

	Mean	***Std. Error***
Simple MC	**6.927122**	**0.0986554**
Control Var.	**6.914076**	**0.0411446**

(Exact Black-Scholes price: 6.888729)

This produced a closer estimate than simple MC, but it is much more complex to calculate (almost double the computation, as it uses twice as many samples). If we doubled the samples for simple MC, then the SE would drop by a factor of $\frac{1}{\sqrt{n}}$

14.1.4 Importance Sampling

Idea: Attempt to reduce variance by **changing the distribution** (probability measure) from which paths (random variates) are generated.

- Change measure to give more weight to important outcomes, thereby increasing sample efficiency
- E.g. for European call, put more weight to paths with positive payoff i.e., for which we exercise)

Performance of importance sampling relies **heavily** on equivalent measure being used.

We want to estimate $a = \mathbb{E}_\phi[f(Z)] = \int_z f(z)\phi(z)\,dz$ where $\phi(z)$ is the pdf of Z (Normal in this case)

- In the Simple MC: Generate sample $Z_i \sim^{iid} \phi, \quad i = 1, \dots, n$ and use $\hat{\alpha} = \frac{\sum_{i=1}^n f(Z_i)}{n}$ which by CLT and LLN it will converge to $\mathbb{E}_\phi(f(Z))$

Assuming you ahve a same $Z_i' \sim^{iid} \psi, \quad i = 1, \dots, n$ from a new pdf ψ, you can still estimate α as follows:

$$
\begin{aligned}
\alpha = \int_z f(z)\phi(z)\,dz = \int_z f(z)\frac{\phi(z)}{\psi(z)}\psi(z)\,dz = \mathbb{E}_z\left[f(Z')\frac{\phi(Z')}{\psi(Z')}\right] \\
\implies \hat{\alpha} = \frac{1}{n}\sum_{i=1}^n f(Z_i')\frac{\phi(Z')}{\psi(Z')}
\end{aligned}
$$

Also used in some cases where you can't simulate efficiently from $\phi(Z)$, but you have a way to calculate this density. You can simulate from a distribution that has a cdf $\psi(Z)$ instead, and has a method of simulating variates from.

There is a way to simulate from a Normal distribution (which doesn't have a CDF) other than this, he did not mention exactly what, nor should it be that important.

14.1.4.1 Importance Sampling is unbiased

Proof that $\hat{\alpha'}$ is unbiased provided the simple MC estimate $\hat{\alpha}$ is unbiased.

$$
\begin{aligned}
\mathbb{E}_\psi[\hat{\alpha'}] &= \mathbb{E}_\psi\left[\frac{1}{n}\sum f(Z_i')\frac{\phi(Z_i')}{\psi(Z_i')}\right] \\
&= \frac{1}{n}\sum_{i=1}^n \mathbb{E}_\psi\left[f(Z_i')\frac{\phi(Z_i')}{\psi(Z_i')}\right] \\
&= \int_{-\infty}^{\infty} \underbrace{f(z')\frac{\phi(z')}{\psi(z')}}_{value} \underbrace{\psi(z')}_{prob}\,dz' \\
&= \int_{-\infty}^{\infty} f(z')\phi(z')\,dz' \\
&= \mathbb{E}_\phi\left[f(z)\right] \\
&= \alpha
\end{aligned}
$$

But when is this efficient (better than simple MC)?

When the variance of $\hat{\alpha}'$ is lower

$$\begin{aligned}
\mathbb{V}_{\psi}(\hat{\alpha}') &= \mathbb{V}_{\psi}\left[\frac{1}{n}\sum_{i=1}^{n} f(z')\frac{\phi(z')}{\psi(z')}\right] \\
&= \frac{1}{n}\mathbb{V}\left[f(z_i')\frac{\phi(z_i')}{\psi(z_i')}\right] \\
&= \frac{1}{n}\left\{\mathbb{E}_{\psi}\left[\left(f(z')\frac{\phi(z')}{\psi(z')}\right)^2\right] - \left(\mathbb{E}_{\psi}\left[f(z')\frac{\phi(z')}{\psi(z')}\right]\right)^2\right\}
\end{aligned}$$

Now we show that this variance is lower than $\mathbb{V}_{\phi}(\hat{\alpha})$ iff $\mathbb{E}_{\phi}\left[f^2(Z)\frac{\phi(Z)}{\psi(Z)}\right] \leq \mathbb{E}_{\phi}\left[f^2(Z)\right]$

$$\begin{aligned}
\mathbb{V}_{\psi}\left[\hat{\alpha}'\right] \leq \mathbb{V}_{\phi}[\hat{\alpha}] &\iff \frac{1}{n}\left\{\mathbb{E}_{\psi}\left[\left(f(z')\frac{\phi(z')}{\psi(z')}\right)^2\right] - \alpha^2\right\} \leq \frac{1}{n}\left\{\mathbb{E}_{\phi}\left[f^2(z)\right] - \alpha^2\right\} \\
&\iff \mathbb{E}_{\psi}\left[f^2(z')\frac{\phi^2(z')}{\psi^2(z')}\right] \leq \mathbb{E}_{\phi}\left[f^2(z)\right] \\
&\iff \int f^2(z')\frac{\phi^2(z')}{\psi^2(z')}\psi(z')\,dz' = \int f^2(z')\frac{\phi(z')}{\psi(z')}\phi(z')\,dz' \\
&\qquad = \mathbb{E}_{\phi}\left[f^2(z)\frac{\phi(z)}{\psi(z)}\right]
\end{aligned}$$

We can even show that $\mathbb{V}_{\psi}[\hat{\alpha}] = 0$ iff $\psi(z) \propto f(z)\psi(z)$ for positive f.

Then $\psi(z) = \frac{1}{c}f(z)\phi(z) \implies \int \psi(z)\,dz = 1 \implies \int \frac{1}{c}f(z)\phi(z)\,dz = 1 \implies \int f(z)\phi(z)\,dz = \mathbb{E}_{\phi}[f(Z)]$

We need the price of the option to be finite, or the integral will not be finite. If the ϕ idstirbution is completely proportional to the functions, then the normalizing value is exactly what we're looking for. This is an extreme case, although the overall idea is that we want ψ to be as similar as possible

$$\begin{aligned}
\mathbb{V}_{\psi}[\hat{\alpha}'] &= \frac{1}{n}\left\{\mathbb{E}_{\psi}\left[\left(f(z')\frac{\phi(z')}{\psi(z')}\right)^2\right] - \alpha^2\right\} \\
&= \frac{1}{n}\left\{\mathbb{E}_{\psi}\left[\left(\frac{f(z')\phi(z')}{\frac{1}{c}f(z')\phi(z')}\right)^2\right] - \alpha^2\right\} \\
&= \frac{1}{n}\left\{\underbrace{\mathbb{E}_{\psi}\left[c^2\right]}_{\alpha^2} - \alpha^2\right\} \\
&= \frac{1}{n}\left(\alpha^2 - \alpha^2\right) = 0
\end{aligned}$$

Importance sampling works best when new pdf ψ "resembles" payoff $\times$ (original pdf) $f \times \varphi$

14.1.4.2 Extending to paths

Importance sampling can be extended to multiple random variates per path - E.g. path-dependent option, with payoff $f(Z_1, \ldots, Z_m)$ a function of $\#m$ variates forming discretized path

$$E_\varphi\left[f(Z_1, \ldots, Z_m)\right] = E_\psi\left[f(Z_1', \ldots, Z_m')\frac{\varphi(Z_1', \ldots, Z_m')}{\psi(Z_1', \ldots, Z_m')}\right]$$

- If in addition, $Z_j \sim^{iid} \varphi_j \& Z_j' \sim^{iid} \psi_j$, then

$$E_\varphi\left[f(Z_1, \ldots, Z_m)\right] = E_\psi\left[f(Z_1', \ldots, Z_m')\prod_{j=1}^{m}\frac{\varphi_j\left(Z_j'\right)}{\psi_j\left(Z_j'\right)}\right]$$

14.1.5 Ex. Deep OOTM European Call

$S_0 = 50, K = 65$ Find the price of the option using simple MC

Generate final prices from the log Normal distribution

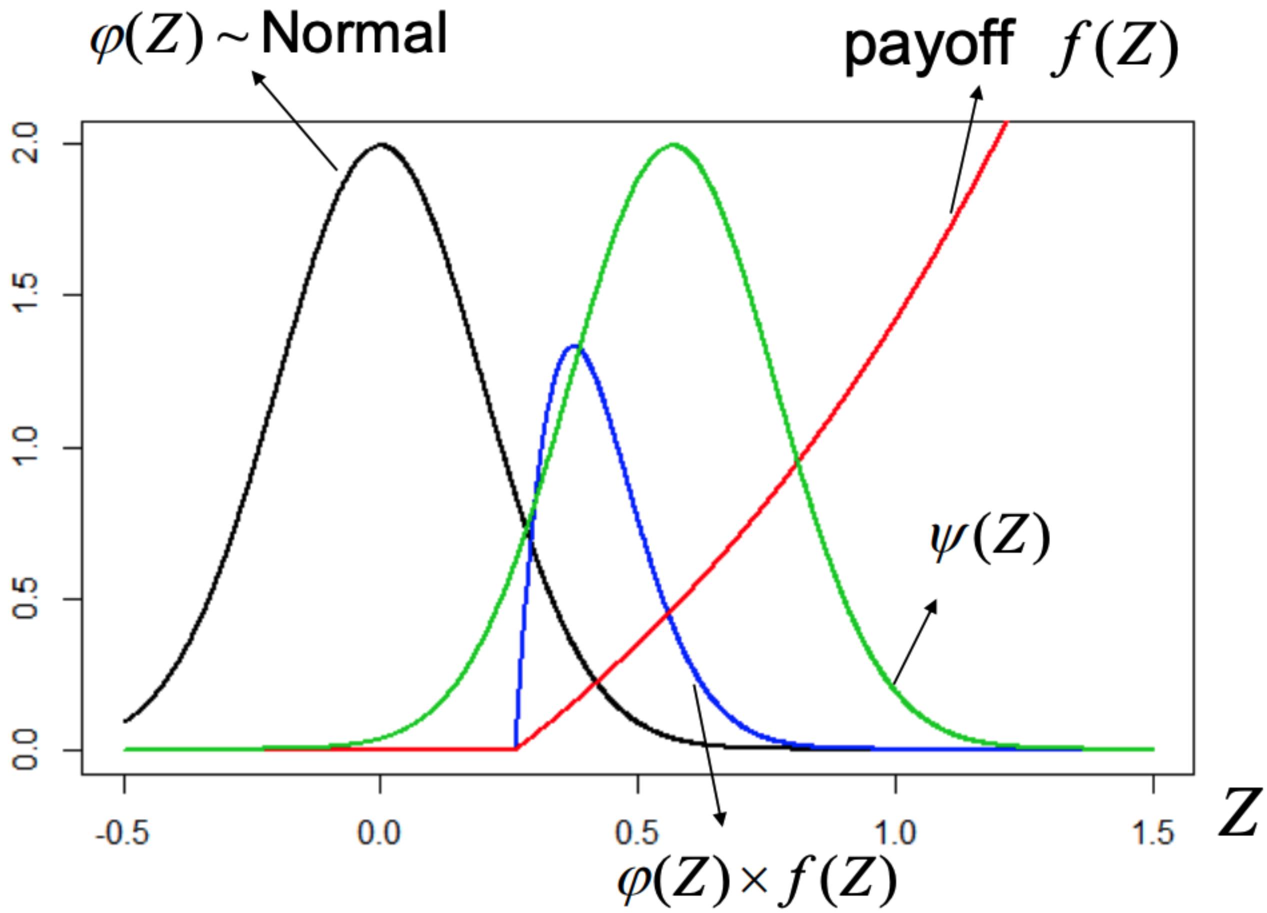
- With simple MC, generate final prices as

$$S_T = S_0 e^Z,\ \text{where } Z \sim \varphi = N\left(\left(r - \frac{\sigma^2}{2}\right)T, \sigma^2 T\right)$$

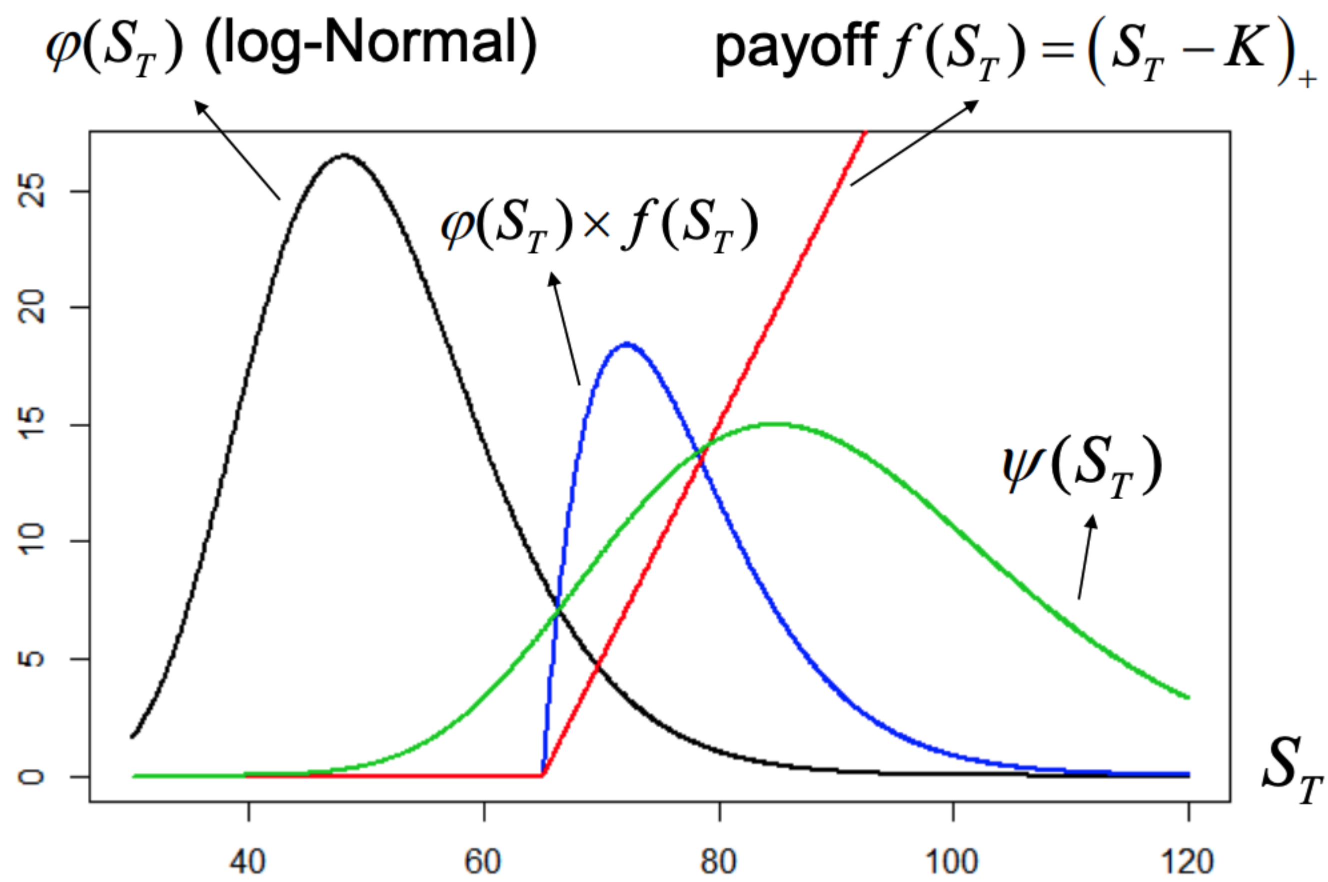
What would be a good candidate for ψ ?

$$Z' \sim \psi = N\left(\log\left(\frac{90}{50}\right) - \frac{\sigma^2}{2}T, \sigma^2 T\right) \text{ or } N\left(\log\left(\frac{30}{50}\right) - \frac{\sigma^2}{2}T, \sigma^2 T\right)$$

Only diff is the mean, but the first distrbution has a higher mean, which is what we want to simluate from as we want the final price to be closer to higher values (65)



$\varphi(Z) \sim$ Normal
payoff $f(Z)$
$\psi(Z)$
$\varphi(Z) \times f(Z)$
Z
2.0
1.5
1.0
0.5
0.0
-0.5
0.0
0.5
1.0
1.5



Most of the values simulated in the black distribution (RN measure) you would get 0 payoff, except in the tail. We want to shift the distribution so most of the simulations give us a positive pay off, such as in the green distribution. This will do much better in simulation. The blue distribution is the payoff x RN measure that we ideally want to simulate (but we can't actually get since ϕ is hard to get?)

$$\psi(Z) = N\left(\log\left(\frac{90}{S_0}\right) - \frac{\sigma^2}{2}T, \sigma^2 T\right) \Rightarrow E_\psi\left[S_T\right] = 90$$

	Mean	***Std. Error***
Simple MC	**0.618857**	**0.02670904**
Importance Sampling	**0.6153817**	**0.007132492**

(Exact Black-Scholes price: 0.6160138)

Chapter 15

Optimization

15.1 Role Optimization

Most real-world problems involve making decision, often under uncertainty. In finance, we must typically decide how to invest *over time* and *across assets*. Making good/optimal decisions typically involves some optimizations, such as mean-variance analysis or Kelly criterion.

15.2 Types of Optimization

- Different types of optimization problems:
- Straightforward (closed-form or polynomial complexity):
  - Linear, Quadratic, Convex
  - Equality/linear/convex constraints
- Difficult
  - Discrete optimization (discrete variable)
    - E.g. individisble assets, transaction costs
  - Dynamic optimization (previous decisions affect future ones)
    - Investing over time
  - Stochastic optimization (uncertainty)

15.2.1 Example

Imagine you could foresee the price of a stock. You want to make optimal use of such knowledge, assuming:

- You can only trade integer units of an asset
- Every transaction costs you a fixed amount
- You cannot short-sell the asset

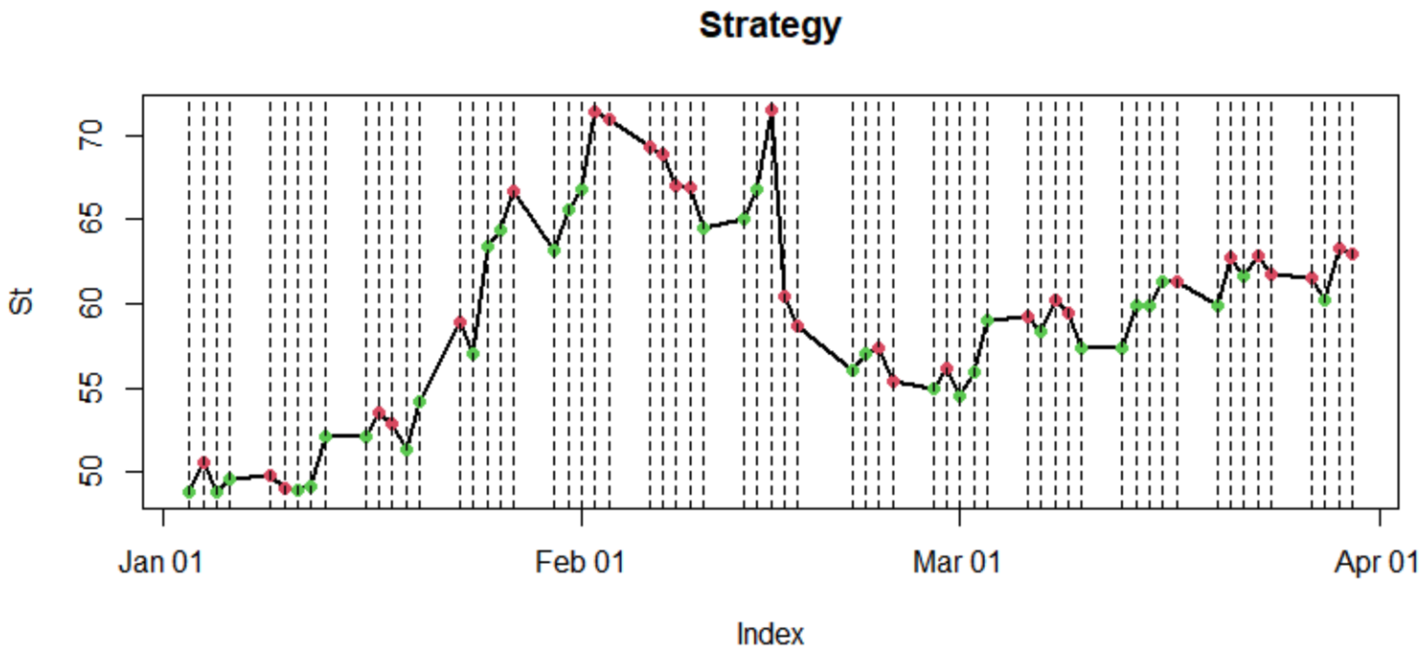
This is a **discrete, dynamic optimization problem**. There is no randomness, we have perfect knowledge. Nevertheless, the issue is not trivial.

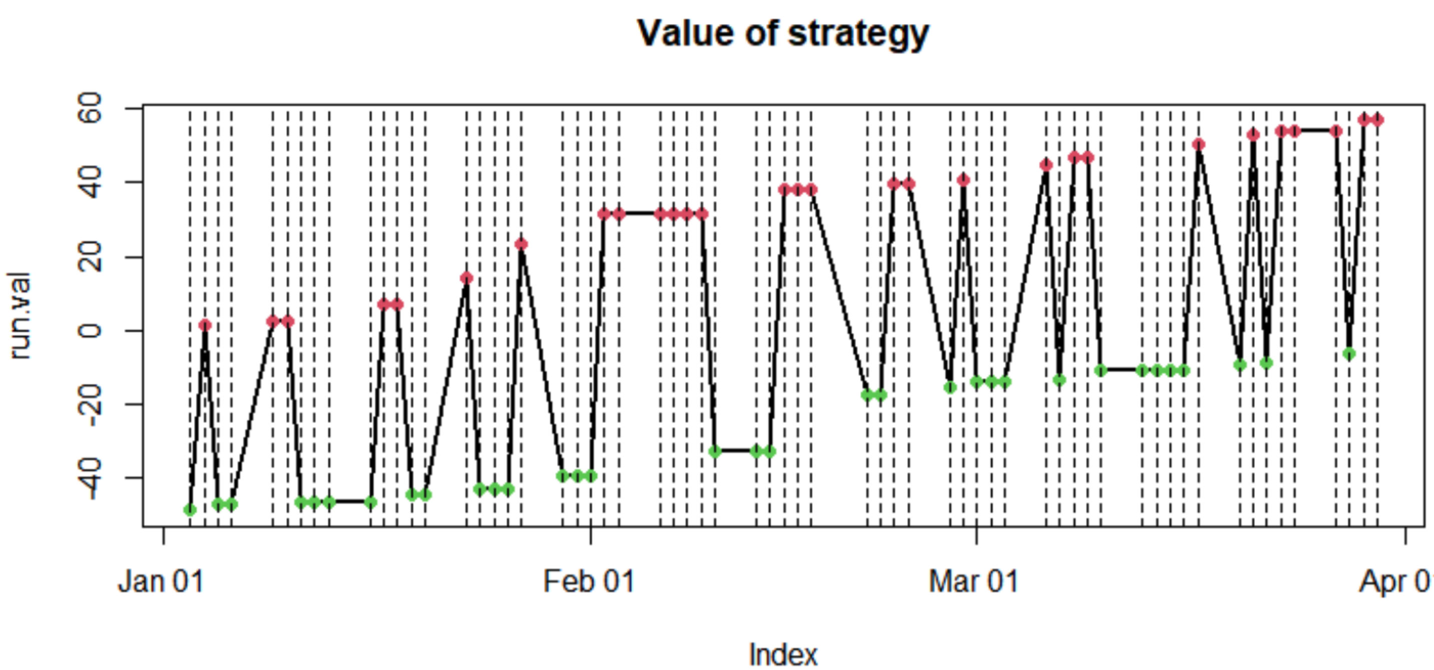
Let $S(t)$ = asset price @ t , tc = transaction cost ,
$V_{np}(t)$ = opt. value for *no position @* t
$V_{lp}(t)$ = opt. value for *long position @* t
Find evolution of value, assuming no position at t=0,>n

state \ time	t=1	t=2	t=3=n	>n
no position				
long position				

State Time	no position	long position
t = 1	$V_{np}(1) = max(0 + V_{np}(2), -S(1) - tc + V_{lp}(2))$	XXX
t=2	$V_{np}(2) = max(0 + V_np(3), -S(2) - tc + V_{lp}(3))$	$V_{lp} = max(0 + V_{lp}(3), S(2) - tc + V_{np}(3))$
t=3=n	$V_{np}(3) = 0$	$V_{lp}(3) = S(s) - tc$
t >n	0	XXX

The result of our strategy





In reality, we won't have perfect information like this, which means we need to include some randomness. This leads up to stochastic optimization

15.3 Stochastic Optimization

We'll first introduce the binomial model, and assume prices follow this tree with some probabilities.

We want to find the best trading strategy that maximizes the **expected P/L**..

We define:

- **State**: X_t is a RV (contains the price S_t and position)
- **Action**: a_t is a change in the state, such as buying/selling
- **Reward**: F_t is some reward function, such as cashflow

[](Notes/Obsidian-Attachments/13-Optimization-in-Finance-3.png|300]]

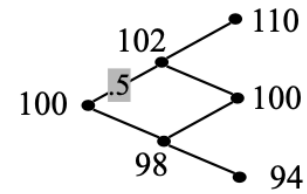
To maximize the expected reward over stochastic actions, let $V(t, X_t)$ be the optimal value function:

$$
\begin{aligned}
V(t, X_t) &= \max_{a_{t\to T}} \left\{ \mathbb{E}\left[\sum_{s=t}^{T} f(s, X_s, a_s) \mid X_t \right] \right\} \\
&= \max_{a_{t\to T}} \left\{ f(t, X_t, a_t) + \mathbb{E}\left[\sum_{s=t+1}^{T} f(s, X_s, a_s) \mid X_t \right] \right\} \\
&= \max_{a_t} \left\{ f(t, X_t, a_t) + \max_{a_{(t+1)\to T}} \left\{ \mathbb{E}\left[\sum_{s=t+1}^{T} f(s, X_s, a_s) \mid X_t \right] \right\} \right\} \\
&= \max_{a_t} \left\{ f(t, X_t, a_t) + \mathbb{E}\left[\max_{a_{(t+1)\to T}} \left\{ \mathbb{E}\left[\sum_{s=t+1}^{T} f(s, X_s, a_s) \mid X_{t+1}^{(a_t)} \right] \right\} \mid X_t \right] \right\} \\
&= \max_{a_t} \left\{ f(t, X_t, a_t) + \mathbb{E}\left[V\left(t+1, X_{t+1}^{(a_t)}\right) \mid X_t \right] \right\}
\end{aligned}
$$

> Essentially, instead of maximizing over the entire sequence of options, you use backward induction to maximize one step at a time.

15.3.1 Example

Example



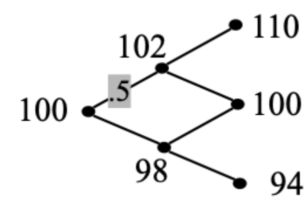
- Expected value under *all* strategies for transaction cost (tc)

Strategy (0, up, down)	Expected value $\mathbb{E}\left[\sum_{s=0}^{2} f(s,X_s,a_s) \middle| X_0\right]$
(n, n, n)	0
(n, n, l)	$-0.5-tc$
(n, l, n)	$1.5-tc$
(n, l, l)	$1-tc$
(l, n, n)	$-2tc$
(l, n, l)	$-0.5-tc$
(l, l, n)	$1.5-2tc$
(l, l, l)	$1-2tc$

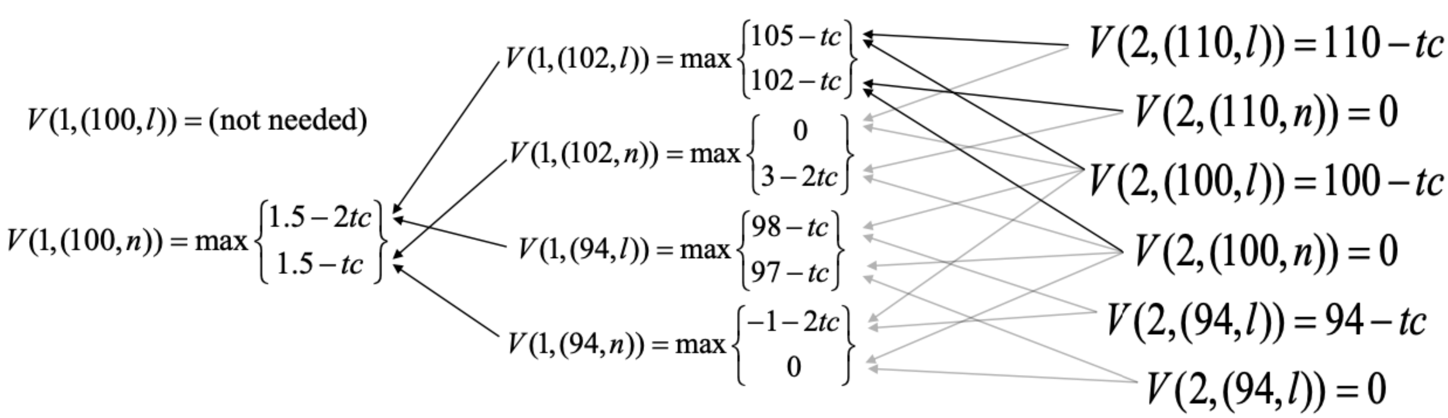
n: neutral
l: long

12

Example



- Using Dynamic Programming / Backward Induction



Chapter 16

Problem Set 1

16.1 Q1

16.1.1 Data Prep

Loading libraries required:

```
library(tidyverse)
library(ggplot2)
library(plotly)
library(lubridate)
```

Settings column names for our data (as per the demo file)

```
msg_columns <- c( "Time" , "Type" , "OrderID" ,
                  "Size" , "Price" , "TradeDirection" )
ordr_columns <- c("ASKp1" , "ASKs1" , "BIDp1",  "BIDs1")

# Levels
nlevels = 10;
# naming the columns of data frame
if (nlevels > 1)
{
    for ( i in 2:nlevels )
    {
        ordr_columns <- c (ordr_columns,paste("ASKp",i,sep=""),
                           paste("ASKs",i,sep=""),
                           paste("BIDp",i,sep=""),
                           paste("BIDs",i,sep=""))
    }
}
```

Reading in the data

```
message = read_csv(paste0(path,"/AMZN_2012-06-21_34200000_57600000_message_10.csv"),
                   col_names=msg_columns)

order_book = read_csv(paste0(path,"/AMZN_2012-06-21_34200000_57600000_orderbook_10.csv"),
                      col_names=ordr_columns)
```

```
head(message)
```

```
## # A tibble: 6 x 6
##      Time  Type  OrderID  Size   Price TradeDirection
##     <dbl> <dbl>    <dbl> <dbl>   <dbl>          <dbl>
## 1 34200.      5        0     1 2238200             -1
## 2 34200.      1 11885113    21 2238100              1
## 3 34200.      1  3911376    20 2239600             -1
## 4 34200.      1 11534792   100 2237500              1
## 5 34200.      1  1365373    13 2240000             -1
## 6 34200.      1 11474176     2 2236500              1
```

```
head(order_book)
```

```
## # A tibble: 6 x 40
##      ASKp1 ASKs1  BIDp1 BIDs1  ASKp2 ASKs2  BIDp2 BIDs2  ASKp3 ASKs3  BIDp3 BIDs3
##      <dbl> <dbl>  <dbl> <dbl>  <dbl> <dbl>  <dbl> <dbl>  <dbl> <dbl>  <dbl> <dbl>
## 1 2239500   100 2.23e6   100 2.24e6   100 2.23e6   200 2.24e6   220 2.23e6   100
## 2 2239500   100 2.24e6    21 2.24e6   100 2.23e6   100 2.24e6   220 2.23e6   200
## 3 2239500   100 2.24e6    21 2.24e6    20 2.23e6   100 2.24e6   100 2.23e6   200
## 4 2239500   100 2.24e6    21 2.24e6    20 2.24e6   100 2.24e6   100 2.23e6   100
## 5 2239500   100 2.24e6    21 2.24e6    20 2.24e6   100 2.24e6   100 2.23e6   100
## 6 2239500   100 2.24e6    21 2.24e6    20 2.24e6   100 2.24e6   100 2.24e6     2
## # ... with 28 more variables: ASKp4 <dbl>, ASKs4 <dbl>, BIDp4 <dbl>,
## #   BIDs4 <dbl>, ASKp5 <dbl>, ASKs5 <dbl>, BIDp5 <dbl>, BIDs5 <dbl>,
## #   ASKp6 <dbl>, ASKs6 <dbl>, BIDp6 <dbl>, BIDs6 <dbl>, ASKp7 <dbl>,
## #   ASKs7 <dbl>, BIDp7 <dbl>, BIDs7 <dbl>, ASKp8 <dbl>, ASKs8 <dbl>,
## #   BIDp8 <dbl>, BIDs8 <dbl>, ASKp9 <dbl>, ASKs9 <dbl>, BIDp9 <dbl>,
## #   BIDs9 <dbl>, ASKp10 <dbl>, ASKs10 <dbl>, BIDp10 <dbl>, BIDs10 <dbl>
```

Setting up the X-axis (time of day) and Y-axis (price) variables

```
# Trading hours (start & end)
startTrad = 9.5*60*60       # 9:30:00.000 in ms after midnight
endTrad = 16*60*60        # 16:00:00.000 in ms after midnight

# Define interval length
freq = 5*60;   # Interval length in ms 5 minutes

# Number of intervals from 9:30 to 4:00
no_int= (endTrad-startTrad)/freq
```

```r
# Convert prices into dollars
#    Note: LOBSTER stores prices in dollar price times 10000

# Sample rows closest to 5 minute increments
message_data = message %>%
  mutate(Price = Price / 10000,
         idx = row_number(),
         five_min_grp = ((Time - startTrad) %/% freq) * freq + startTrad) %>%
  group_by(five_min_grp) %>%
  filter(row_number() == 1) %>%
  ungroup() %>%
  arrange(idx) %>%
  # Format the Time column to be understandable,
  # needed to add 4 hours to be in right time zone
  mutate(X_Time = (Time + 4*60*60) %>%
           as.POSIXct.numeric(origin = '2012-06-21') %>%
           lubridate::round_date(unit='minute'))

order_data = order_book %>% mutate(idx = row_number())
price = order_data %>%
  dplyr::select(contains("p"),idx) %>%
  pivot_longer(!idx,
               # First column is Order Type, takes first capture group, second
               # capture group is excluded, third capture group of levels
               # becomes col names, indicated by .value
               names_to = c("Order Type", ".value"),
               names_pattern = "(...)(?:p)(\\d{1,2})") %>%
              # need to pivot longer again, but excluding Order Type and row idx
              # this pivots the levels into its own column
  pivot_longer(!c(`Order Type`,idx),
               names_to = c("Level"),
               values_to= "Price")

quantity = order_data %>%
  dplyr::select(!contains("p"), idx) %>%
  pivot_longer(!idx,
               names_to = c("Order Type", ".value"),
               names_pattern = "(...)(?:s)(\\d{1,2})") %>%
  pivot_longer(!c(`Order Type`,idx),
               names_to = c("Level"),
               values_to= "Quantity") %>%
  dplyr::select(Quantity)

order_longer = price %>% bind_cols(quantity) %>%
  mutate(Limit_Price = Price / 10000) %>% filter(idx %in% message_data$idx) %>%
  dplyr::select(-Price)

plt_data =  order_longer %>%
  left_join(message_data %>% dplyr::select(Time, Price, idx, X_Time), by="idx")
```

16.1.2 Plot

```
(plt_data %>%
  ggplot(aes(x=X_Time)) +
  geom_point(aes(y=Limit_Price, size=Quantity, color=`Order Type`)) +
  geom_line(aes(y=Price)) +
  scale_color_manual(values=c("red2", "green3")) +
  scale_x_datetime(breaks = seq(min(plt_data$X_Time), max(plt_data$X_Time),
                                by = 30 * 60), date_labels="%H:%M") +
  scale_y_continuous(labels = scales::dollar_format(prefix="$")) +
  labs(title = "Evolution of AMZN Limit Order Book at 5-minute intervals",
       y="Price($)",
       x="Time of Day")) %>%
  ggplotly()
```

16.2 Q2

R (Daily Return Data) Download the closing and adjusted closing daily prices for Apple Inc. (symbol APPL) for the previous year (Jan 1, 2022 to Dec 31, 2022). Use the the function get.hist.quote() from the tseries package.

Loading Data

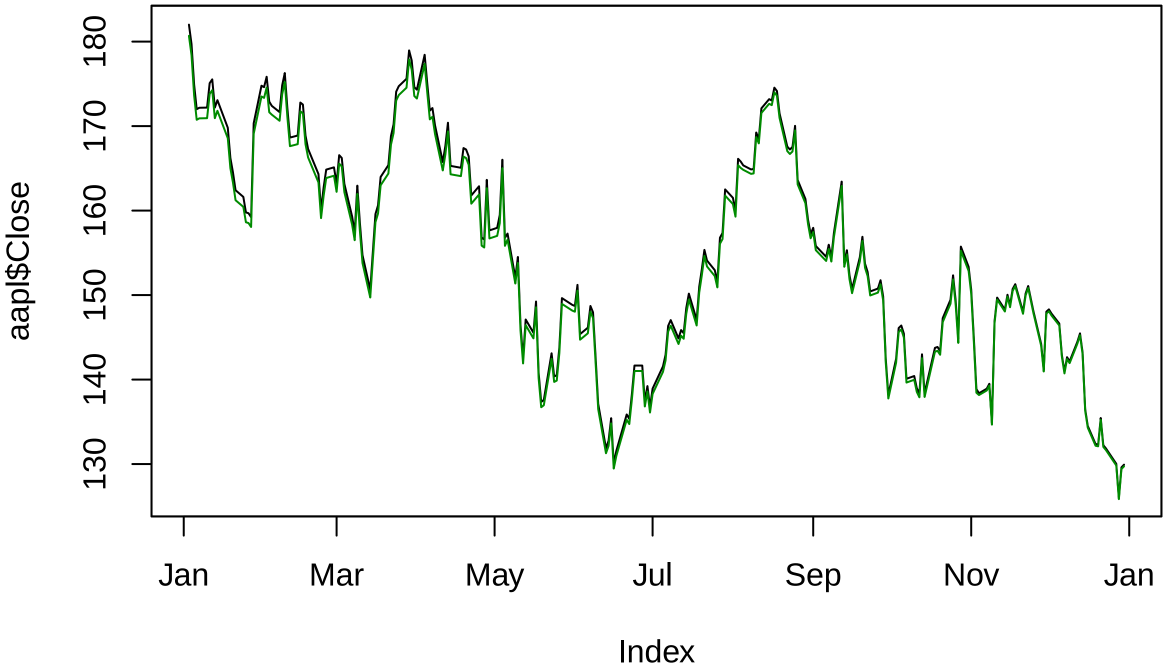
```
library(tseries)
aapl = get.hist.quote('aapl', start="2022-01-01", end="2022-12-31",
                      quote=c("Close", "Adjusted"), compression="d")
```

```
## time series starts 2022-01-03
## time series ends   2022-12-30
```

16.2.1 2.a

Plot the closing and adjusted-closing price series on the same plot. Do you see any differences?

```
plot(aapl$Close)
lines(aapl$Adjusted, col="green4")
```

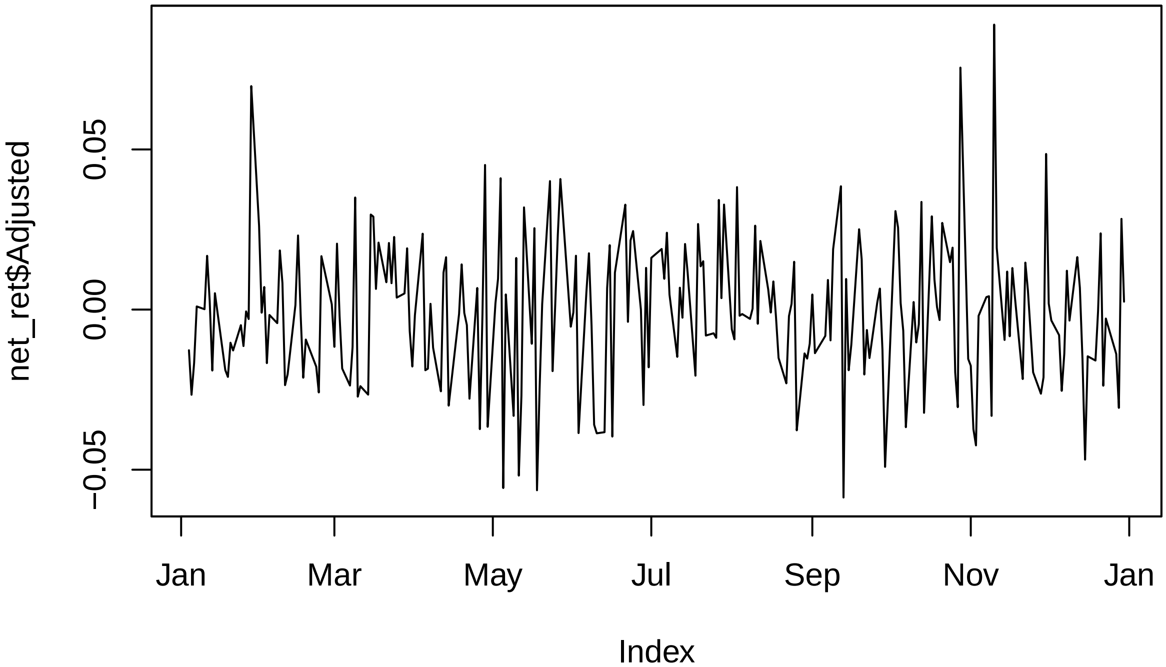


It seems there are differences in the first half of the year.

16.2.2 2.b

Calculate and plot the net returns based on the adjusted closing price.

```
# we shouldn't have a value for Jan 3rd, because that was the first trading day
# no way to calculate return from the day before. time series starts jan 4th
net_ret = aapl / stats::lag(aapl, -1) - 1
log = log(aapl) - log(stats::lag(aapl, -1))
returns = cbind(net_ret, log)
plot(net_ret$Adjusted)
```



16.2.3 2.c

Apple paid a dividend of $0.23 on Nov 4, 2022 (such events are documented here). Calculate the net return of Nov 4, using both the adjusted closing prices (designated way), and the actual closing prices, where you manually adjust for the dividend (i.e., you add the dividend amount to the Nov 4 close price before calculating the log-return). Are the two returns equal?

```r
adj_nov4_ret = net_ret["2022-11-04"]$Adjusted[[1]]

act_nov4_ret = (aapl["2022-11-04"]$Close + 0.23)[[1]] /
  aapl["2022-11-03"]$Close[[1]] - 1

log_nov4_ret = log(aapl["2022-11-04"]$Close + 0.23)[[1]] -
  log(aapl["2022-11-03"]$Close)[[1]]

# Net Return using adjusted close
adj_nov4_ret
```

```
## [1] -0.001947376
```

```
# Net Return using actual close
act_nov4_ret
```

```
## [1] -0.001944124
```

```
# Log Return using actual close
log_nov4_ret
```

```
## [1] -0.001946017
```

They are the same up to 5 significant digits.

16.2.4 2.d

Consider a simple momentum strategy whereby: - you buy APPLE stock for 1 day, if its price increased by more than 2% on the previous day, and - you (short) sell APPLE stock for 1 day, if its price decreased by more than 2% on the previous day Calculate and plot the daily and cumulative net returns of this strategy.

```
# If return at t-1 > 2%, buy APPLE for 1 day, sell next day
# If return at t-1 < 2%, short sell APPLE for 1 day, buy back next day

# Dates where AAPL went up 2%
days_plus2 = time(returns[returns$Adjusted.log > 0.02])

# Dates where AAPL went down 2%
days_minus2 = time(returns[returns$Adjusted.log < 0.02])

## records the return of the day after a >2% return
return_2d_pos = stats::lag(returns,1)[days_plus2]
return_2d_neg = stats::lag(returns,1)[days_minus2]
return_2d = rbind(return_2d_pos, return_2d_neg)

net_cum_2d_ret = cbind("Daily Returns" = return_2d$Adjusted.log,
                       "Cumulative Returns" = cumsum(return_2d$Adjusted.log))
(net_cum_2d_ret %>%
  ggplot(aes(x=time(net_cum_2d_ret))) +
  geom_line(aes(y=`Daily Returns`, color="one")) +
  geom_line(aes(y=`Cumulative Returns`,color="two")) +
  scale_color_discrete(name = "Return Type",
                       labels = c("Daily Net Return",
                                  "Cumulative Net Return")) +
  labs(title="Momentum strategy returns")) %>%
  ggplotly()
```

16.2.5 2.e

Find the maximum drawdown of the strategy, i.e. the biggest drop in cumulative net returns from their highest point. You can use the maxdrawdown() function from the tseries package.

```
maxdrawdown(net_cum_2d_ret$`Cumulative Returns`)
```

```
## $maxdrawdown
## [1] 0.3460493
##
## $from
## [1] 58
##
## $to
## [1] 247
```

16.3 Q3

A stock has an expected annual log-return of 8% and an annual volatility (i.e. st. deviation of annual log-return) of 20%. You want to model the daily log-return using a Normal Random Walk model. What should the mean and variance of the Normal distribution of the daily log-returns be, so that your model matches the annual parameters above? Assume that a calendar year has 252 trading days of daily returns.

```
# num trading days
num_days = 252
```

16.3.1 Mean of daily log-return

Question implies $log(\frac{P_{252}}{P_0}) = 0.08$ We want the daily mean, so $log(\frac{P_{252}}{P_{251}}) + ... + log(\frac{P_1}{P_0}) = 0.08$ assuming $log(\frac{P_{252}}{P_{251}}) = ... = log(P_1/P_0)$ implies $log(P_t/P_{t-1}) = 0.08/252 = 0.0003174603$

Can also use formula from class

$$E[r_{1-n}] = n\mu\mu = E[r_{1-n}]/n$$

```
daily_mean = 0.08 / 252
```

16.3.2 Volatility of daily log-return

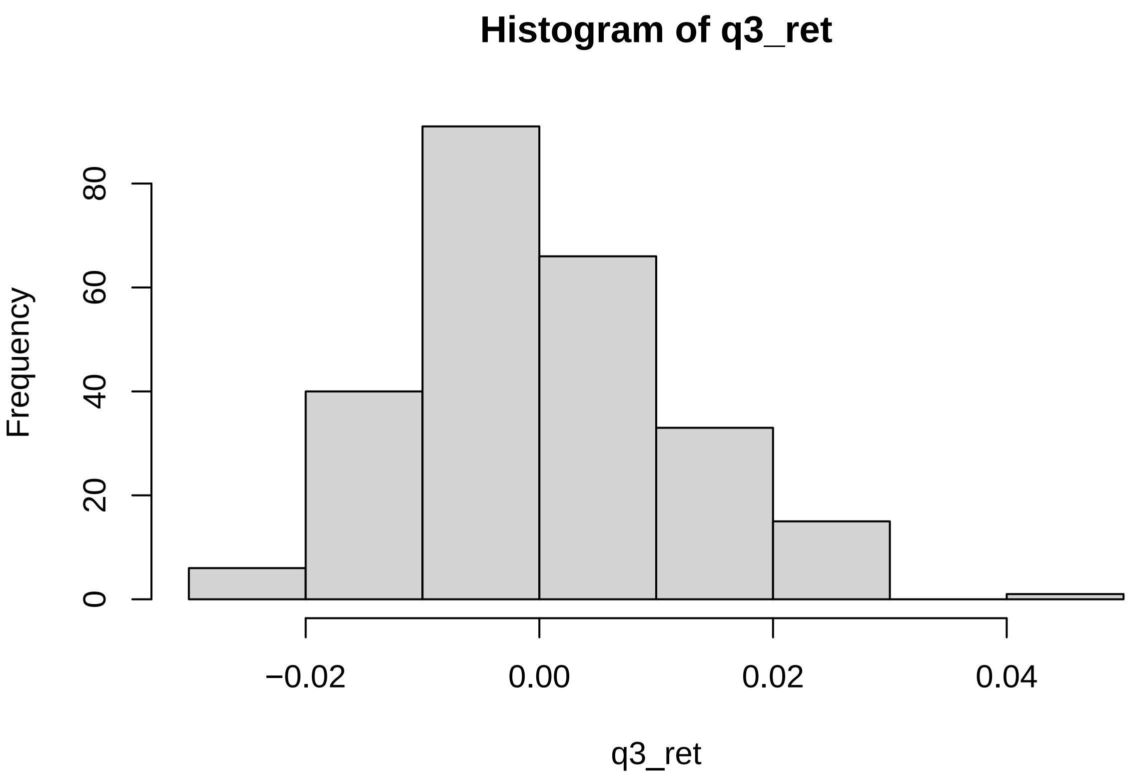
Using formula from class

$$\sqrt{Var[r_{1-n}]} = \sqrt{n}\sigma\sigma = \frac{sd[r_{1-n}]}{\sqrt{n}}$$

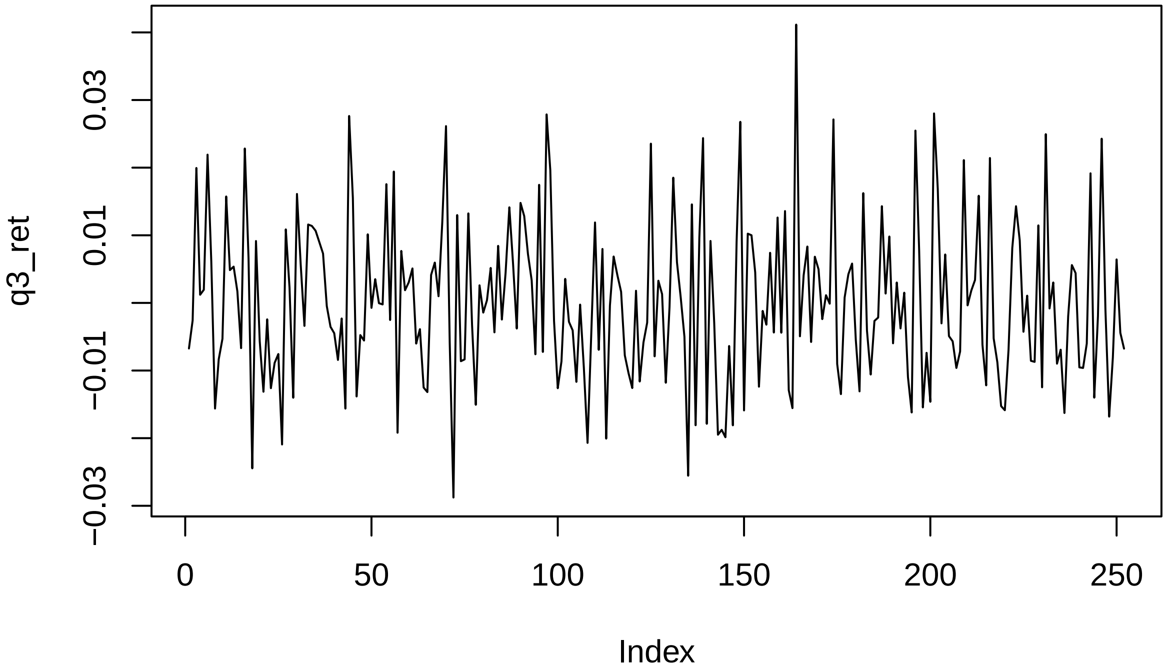
```
# volatility/variance
std_dev = 0.20
daily_std = std_dev / sqrt(num_days)
```

16.3.3 Plot

```
set.seed(123)
q3_ret = rnorm(num_days, daily_mean, daily_std)
hist(q3_ret)
```




```
plot(q3_ret, type="l")
```




```
plot(cumsum(q3_ret), type="l")
```

16.4 Q4

Assume that the daily log returns on a stock with current price $S_0 = 100$ are i.i.d. with a Normal distribution with mean `0.0003` and standard deviation `0.01`.

16.4.1 4.a

Calculate the probability that the price will be above \$120 after 50 days.

```
mean = 0.0003
std = 0.01
s_0 = 100
days = 50

# using the normal distribution, just need to find params for the 50 log-returns
# and then find prob above 120 (ie. return is 20%)
q4_50_day_mean = days * mean
q4_50_day_std = std * sqrt(days)

# probability
```

```
prob_4a = pnorm(.20, q4_50_day_mean, q4_50_day_std, lower.tail = FALSE)
prob_4a
```

```
## [1] 0.004444485
```

16.4.2 4.b

Calculate the probability that the price will increase (i.e., have +ve return) in 30 or more of the next 50 days. Note: You will have to use (pnorm(), pbinom()) to find the numerical values of these probabilities.

```
prob_4b = pbinom(30, 50, pnorm(0, mean, std, lower.tail = FALSE))
prob_4b
```

```
## [1] 0.9177322
```

Chapter 17

Problem Set 2

17.1 Q1

An MLE can be derived from finding the derivative of the log $L(\alpha)$, $l(\alpha)$ function, where

$$L(\alpha) = \prod_{i=1}^{n}\left[\frac{\alpha x_i^{-(\alpha+1)}}{\ell^{-\alpha}}\right]$$

$$\log L(\alpha) = l(\alpha) = \sum_{i=1}^{n}\left[\log(\alpha) - (\alpha+1)\log(x_i) + \alpha\log(\ell)\right]$$

$$\begin{aligned}
l'(\alpha) &= \sum_{i=1}^{n}\left(\frac{1}{\alpha} - [\log(x_i) + \log(\ell)]\right) = 0 \\
0 &= \frac{n}{\alpha} - \sum_{i=1}^{n}(\log(x_i) - \log(\ell)) \\
\frac{n}{\alpha} &= \sum_{i=1}^{n}(\log(x_i) - \log(\ell)) \\
\frac{1}{\alpha} &= \frac{1}{n}\left[\sum_{i=1}^{n}\log\left(\frac{x_i}{\ell}\right)\right]
\end{aligned}$$

$$\therefore \hat{\alpha} = \left(\frac{1}{n}\left[\sum_{i=1}^{n}\log\left(\frac{x_i}{\ell}\right)\right]\right)^{-1}$$

17.2 Q2

17.2.1 Part A

Heavy tail distributions are those with polynomial tails which follow by

$$f(x) \propto x^{-(1+\alpha)}$$

where smaller $\alpha \to$ heavier tails

We say that distribution F is heavy tailed if the integral on the positive half of $\mathbb{R}^+$ is infinite for some t

$$\begin{aligned}
m_X(t) &= \mathbb{E}(e^{tX}) \\
&= \sum_{k=1}^{\infty} \frac{t^k}{k!}\mathbb{E}[X^k] \\
&= \left[1 + t\mathbb{E}(X) + \frac{t^2\mathbb{E}(X^2)}{2!} + \frac{t^3\mathbb{E}(X^3)}{3!} + ... + \frac{t^n\mathbb{E}(X^n)}{n!} + ...\right]
\end{aligned}$$

But for any n^{th} moment where $n > \alpha$, $\mathbb{E}(X^n) = \infty$. This implies that the mfg of any heavy tail distribution (which by definition has $\alpha > 0$) will be infinite.

17.2.2 Part B

$$\begin{aligned}
\phi_X(t) &= \mathbb{E}[e^{itX}] = \mathbb{E}[\cos(tX) + i\sin(tX)] \\
&= \mathbb{E}[\cos(tX)] + i\mathbb{E}[\sin(tX)] \\
\mathbb{E}[\sin(tX)] &= \int_{-\infty}^{\infty} \sin(tX)f(x)\,dx \\
&= \int_{-\infty}^{0} \sin(tX)f(x)\,dx + \int_{0}^{\infty} \sin(tX)f(x)\,dx \\
&= -\int_{0}^{\infty} \sin(tX)f(x)\,dx + \int_{0}^{\infty} \sin(tX)f(x)\,dx \\
&= 0 \\
\therefore \phi_X(t) &= \mathbb{E}[\cos(tX)] \text{ which only takes real values}
\end{aligned}$$

17.2.3 Part C

$$\begin{aligned}
\phi_X(t) &= \mathbb{E}[e^{itX}] \\
\phi_Y(t) &= \mathbb{E}[e^{itY}] \\
\phi_Z(t) &= \mathbb{E}[e^{itZ}] \\
&= \mathbb{E}[e^{it(X+Y)}] \\
&= \mathbb{E}[e^{itX}e^{itY}] \\
&= \mathbb{E}[e^{itX}]\mathbb{E}[e^{itY}] \quad \text{As they are independent} \\
&= e^{-c|t|^\alpha} \cdot e^{-d|t|^\alpha} \\
&= e^{-(c+d)|t|^\alpha}
\end{aligned}$$

Which shows that $\phi_Z(t)$ shares the same form as X and Y.

17.2.4 Part D

$$
\begin{aligned}
\phi_X(t) &= \mathbb{E}[e^{itX}] = e^{-|t|} \\
\phi_{\bar{X}}(t) &= \mathbb{E}[e^{it\bar{X}}] \\
&= \mathbb{E}\left[e^{it\frac{1}{n}(\sum_{i=1}^{n} X_i)}\right] \\
&= \mathbb{E}\left[e^{it\frac{1}{n}X_1} e^{it\frac{1}{n}X_2} \dots e^{it\frac{1}{n}X_n}\right] \\
&= \mathbb{E}\left[e^{it\frac{1}{n}X_1}\right] \mathbb{E}\left[e^{it\frac{1}{n}X_2}\right] \dots \mathbb{E}\left[e^{it\frac{1}{n}X_n}\right] \\
&= [e^{-|t|/n}]^n \\
&= e^{-|t|/n \cdot n} = e^{-|t|} \quad \text{Which is exactly the } t(1) \text{ distribution}
\end{aligned}
$$

17.2.5 Part E

According to theory, $\mathbb{E}[X^k] = \infty$ for any $k \geq \alpha$ where α is the tail index.

If the tail index is greater than 2, then the second order moment exists, implying both the mean and variance both exist/are finite.

By definition, a sum of stable distributions would converge to a stable distribution as shown by the characteristic functions above, but according to the Central Limit Theorem, any sum of iid random variables with finite mean and variance will produce a Normal distribution.

17.3 Q3

$$
\begin{aligned}
Z &= \frac{X}{Y} \\
X &\sim Exp(1) \quad Y \sim Exp(1) \\
\bar{F}_X(x) &= 1 - (1 - e^{-x}) = e^{-x} \quad x > 0 \\
\bar{F}_Y(y) &= 1 - (1 - e^{-y}) = e^{-y} \quad y > 0
\end{aligned}
$$

Show Z follows a heavy-tailed distribution, and find the tail index:

Z follows a heavy-tailed distribution if it has polynomial tails ie $Z \propto x^{-(1+\alpha)}$

We will integrate w.r.t $F_Y(y)$, as we condition on the fact that $y > 0$. We can also convert the probability to be of $\mathbb{P}(X < \alpha y)$

so that it becomes a function of y, which will work in our integral.

$$\begin{aligned}
\bar{F}_Z(z) = \mathbb{P}(Z > z) = \mathbb{P}\left(\frac{X}{Y} > z, x > 0, y > 0\right) \\
&= \int_{y=0}^{\infty} \mathbb{P}\left(\frac{X}{Y} > z | Y = y\right) dF_y(y) \\
&= \int_{y=0}^{\infty} \underbrace{\mathbb{P}(X > zy)}_{e^{-zy}} \underbrace{f_Y(y)}_{e^{-y}} \, dy \\
&\quad \text{as } Y = y \\
&= \int_{y=0}^{\infty} e^{-y(z+1)} \, dy \\
&= \left[\frac{e^{-y(z+1)}}{-(z+1)}\right]_{y=0}^{\infty} \\
&= 0 + \frac{1}{z+1} = (z+1)^{-1} \quad z > 0 \\
\implies f_Z(z) = (z+1)^{-2} \propto x^{-(1+\alpha)} \text{ Where } \alpha = 1
\end{aligned}$$

$\therefore Z$ is heavy tailed with tail index $\alpha = 1$

17.4 Q4

Libraries required:

```
# install.packages("yahoofinancer")
library(yahoofinancer)
library(tidyverse)
library(MASS)
```

Obtain stock data from Yahoo using `yahoofinancer` methods:

```
lto <- Ticker$new('L.TO')

lto_adj_close = lto$get_history(start = '2015-01-01', end = "2022-12-31",
                                interval = '1d') %>% dplyr::select(date, adj_close)
```

17.4.1 Part A

Calculate daily net returns, plot the returns, sample auto-correlation plot, and Normal QQ plot.

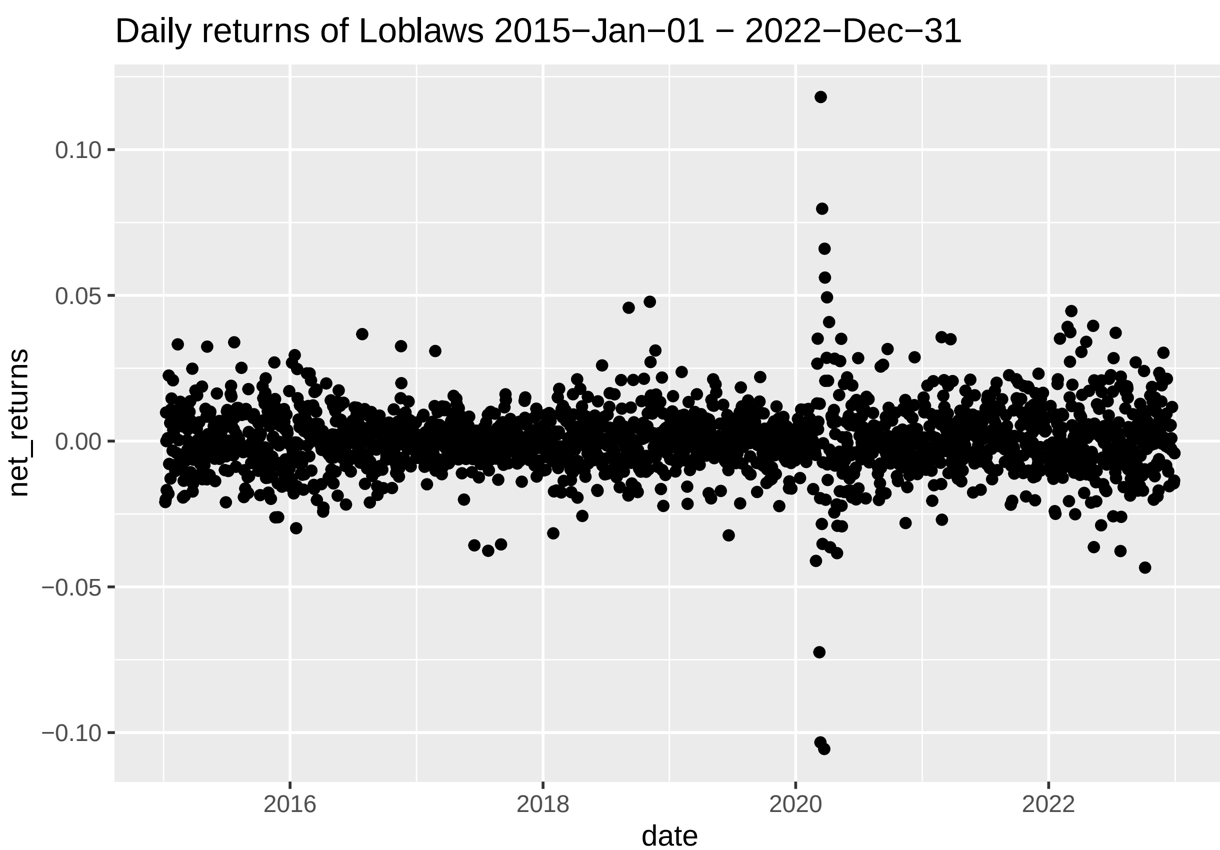
```
head(lto_adj_close)
```

```
##                  date adj_close
## 1 2015-01-02 14:30:00  44.02381
## 2 2015-01-05 14:30:00  43.10342
```

```
## 3 2015-01-06 14:30:00  42.23925
## 4 2015-01-07 14:30:00  42.65377
## 5 2015-01-08 14:30:00  42.65377
## 6 2015-01-09 14:30:00  41.93713
```

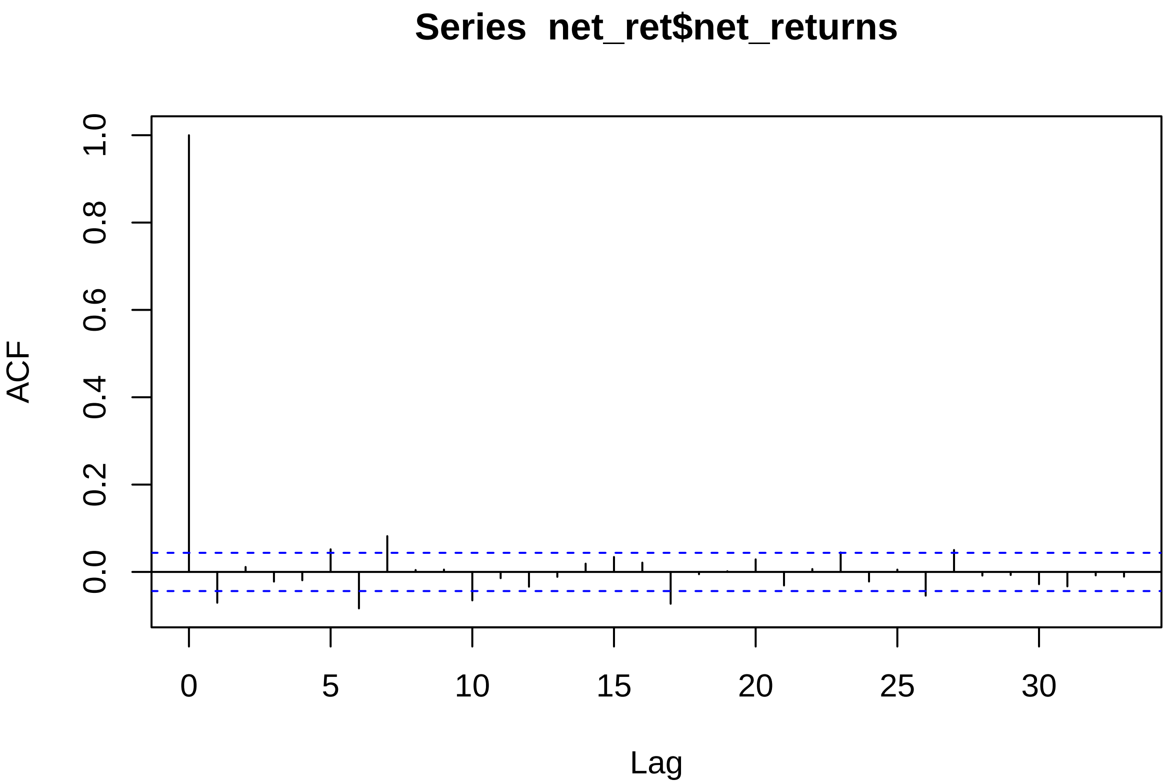
```
net_ret = lto_adj_close %>%
  mutate(net_returns =
           (adj_close - lag(adj_close, 1)) / lag(adj_close, 1)) %>%
  filter(!is.na(net_returns))
```

```
net_ret %>% ggplot(aes(x = date, y = net_returns)) +
  geom_point() +
  labs(title="Daily returns of Loblaws 2015-Jan-01 - 2022-Dec-31")
```



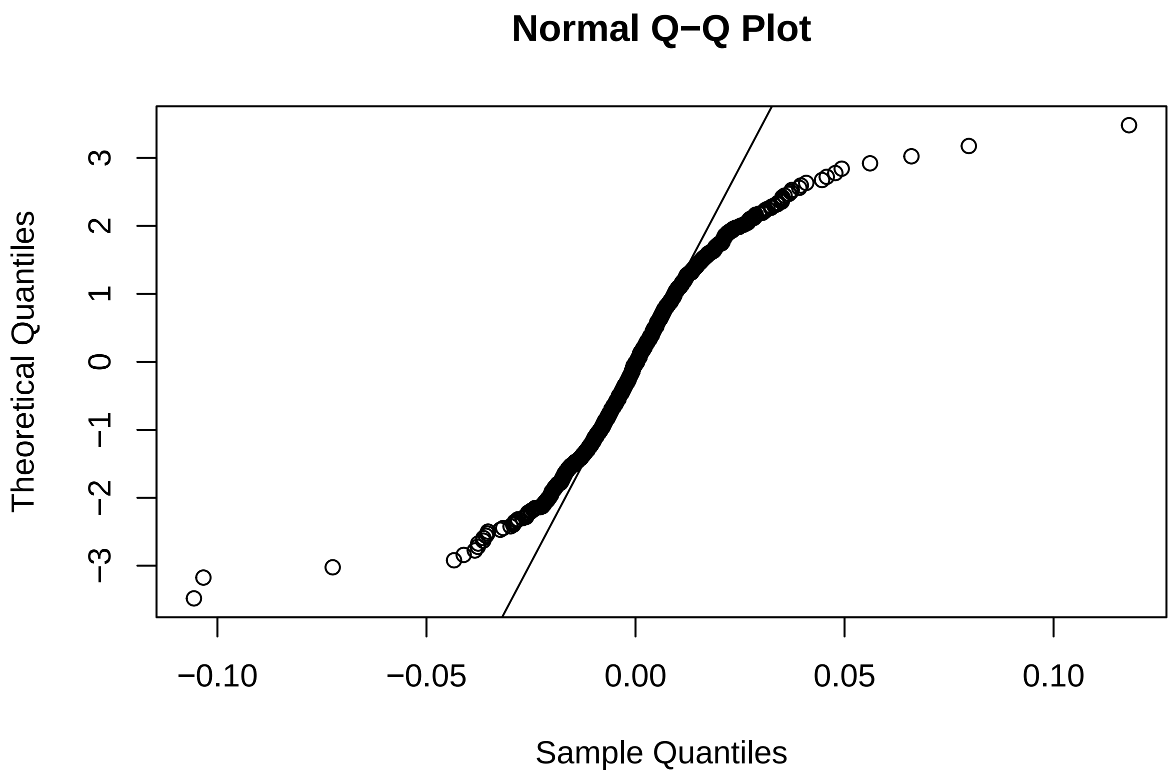
The plot of the daily net returns demonstrate volatility clutering right after the start of 2020.

```
acf(net_ret$net_returns)
```



There are no simple auto-correlations (current returns are not correlated with past returns)

```
qqnorm(net_ret$net_returns,datax = TRUE)
qqline(net_ret$net_returns,datax = TRUE)
```



Based on the QQ plot, the distribution seems leptokurtic as the sample quantiles are more distributed than the theoretical, implying heavy tails.

This supports the stylized fact that returns follow heavy tail distributions.

17.4.2 Part B

```
fitted = fitdistr(net_ret$net_returns, "t")

mu = fitted$estimate[1]
sigma = fitted$estimate[2]
v = fitted$estimate[3]
var = sigma^2 * (v / (v-2))


sample_mu = mean(net_ret$net_returns)
sample_var = var(net_ret$net_returns)
```

We have a fitted mean μ of 3.2240243×10^{-4} and variance σ^2 of 1.4634806×10^{-4}. This is compared to the sample mean μ of 5.6615315×10^{-4} and variance σ^2 of 1.3827256×10^{-4}.

These differences can occur due to the fitted t-distribution smoothing away more extreme values.

17.4.3 Part C

We will now compare two different return distribution approaches in an practical investment setting. Assume you invest all of your wealth in L.TO for 4 years ($4 \times 252 = 1008$ days). Simulate 5,000 iterations of 1008 daily returns from the following models:

17.4.3.1 Part i

```
set.seed(123)
n = 5000
m = 1008

# create a matrix of 5000 simulations of 1008 daily returns
R_mat = matrix(rnorm(n*m, sample_mu, sqrt(sample_var)), nrow = n, ncol = m)

R_final = apply(1+R_mat, 1, cumprod) - 1
sim_norm_returns = R_final[,m]

# max drawdown
mdd = function(R){ return(tseries::maxdrawdown(R)$maxdrawdown) }
MDD = apply(R_final, 1, mdd)
```

17.4.3.2 Part ii

```
# using a t distribution
# create a matrix of 5000 simulations of 1008 daily returns
R_mat_t = matrix(rnorm( n*m, mu, sigma ) * sqrt( v / rchisq( n*m , df = v) ), nrow = n, ncol = m)

R_final_t = apply(1+R_mat_t, 1, cumprod) - 1
sim_t_returns = R_final_t[,m]
MDD_t = apply(R_final_t, 1, mdd)
```

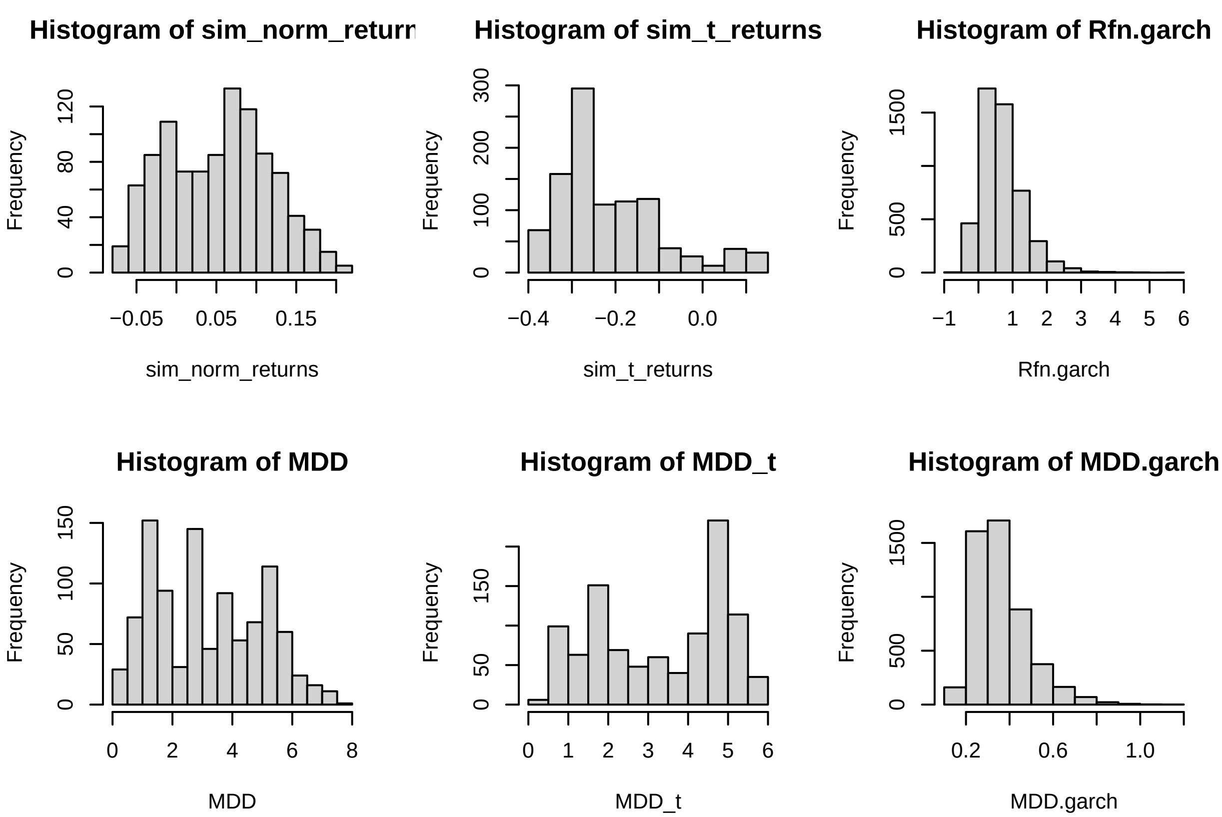
17.4.3.3 Part iii

```
library(fGarch)
```

```
## NOTE: Packages 'fBasics', 'timeDate', and 'timeSeries' are no longer
## attached to the search() path when 'fGarch' is attached.
##
## If needed attach them yourself in your R script by e.g.,
##         require("timeSeries")
```

```
garch_model=garchFit(~garch(1,1), data=net_ret$net_returns, trace = FALSE) # fit GARCH(1,1) model
GARCH.param=garch_model@fit$coef # fitted coefficients
GARCH.spec=garchSpec(model=list(mu=GARCH.param['mu'], # define GARCH model specification for simulation
                                omega=GARCH.param['omega'],
                                alpha=GARCH.param['alpha1'],
                                beta=GARCH.param['beta1'] ))
R.garch = matrix( 0, nrow = n, ncol =m )
for(i in 1:n){
  R.garch[i,]=as.numeric(garchSim(GARCH.spec,m))
}
Rcm.garch = t( apply( 1+R.garch, MARGIN = 1, FUN = cumprod) - 1 )
Rfn.garch = Rcm.garch[,m]
MDD.garch = apply(Rcm.garch, MARGIN = 1, FUN = mdd )
```

```
par(mfrow=c(2,3))
hist(sim_norm_returns)
hist(sim_t_returns)
hist(Rfn.garch)
hist(MDD)
hist(MDD_t)
hist(MDD.garch)
```



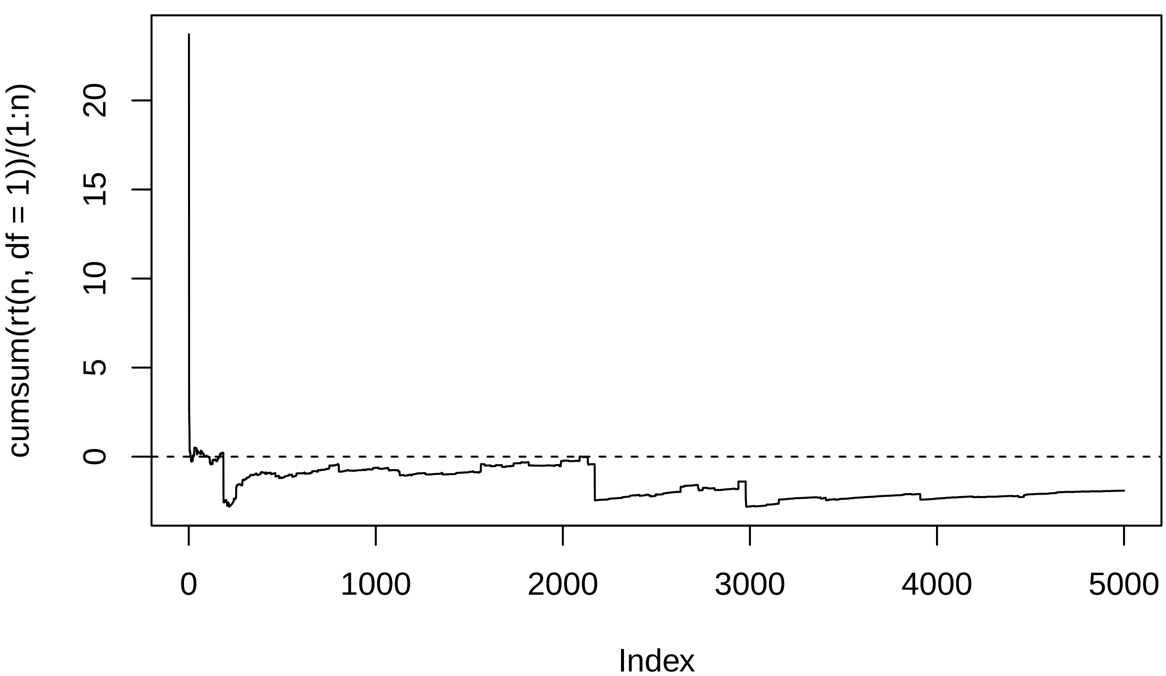
The final returns for t-distr. and GARCH gave lower negative values, but are comparable to those of Normal. In terms of maximum drawdown, the GARCH values have a smaller range, but are otherwise also comparable. Not that for for both metrics we combine/compound many returns, so the effects of heavy tails is suppressed.

I definitely fucked up somewhere in these simulations...don't know where though.

17.5 Q5

17.5.1 Part A

```
n = 5000

# simulate from t distribution with df=1
plot( cumsum( rt(n , df = 1) ) / (1:n), type = "l");
abline(0,0, lty =2)
```

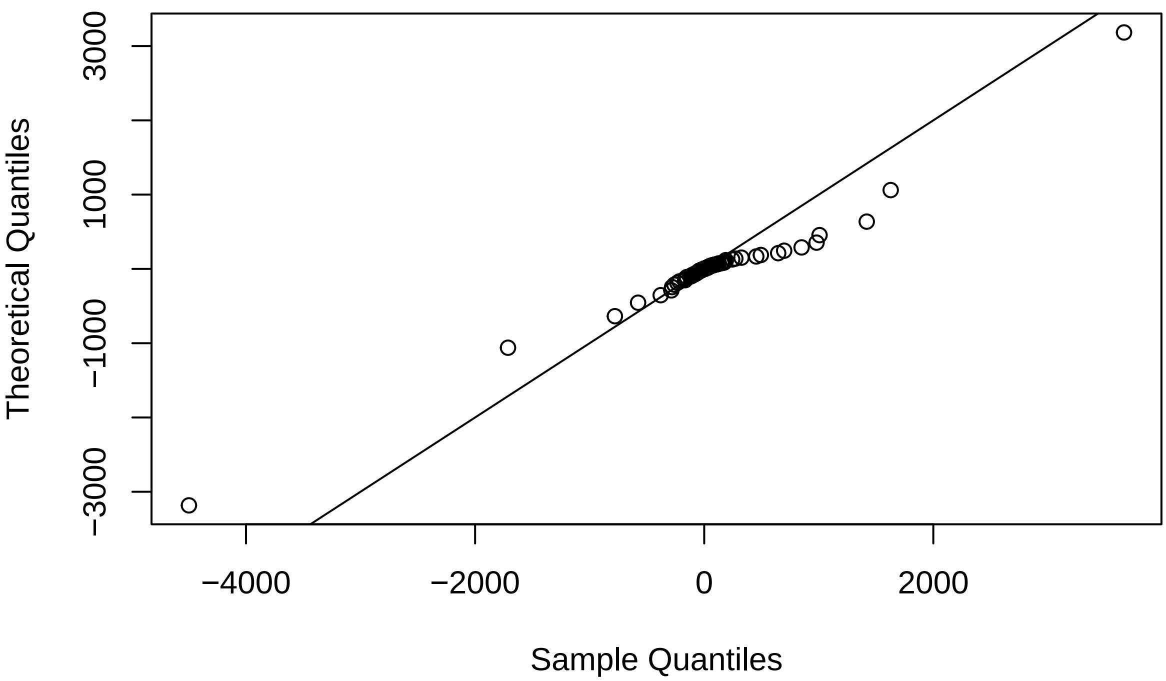


If the WLLN held for this distribution, you would expect to see the average converge to 0.

The cauchy distribution will not converge to 0 because of the extreme values we have sampled will always cause large shifts in the cumulative sum.

17.5.2 Part B

```
x = rt(n , df = 1)
y = rt(n , df = 1)
w = x+y/2

# sort w so we can create a qq plot
w_sorted = sort(w)
cauchy_quantiles = qt(ppoints(n), df = 1)
plot(w_sorted, cauchy_quantiles, xlab="Sample Quantiles", ylab="Theoretical Quantiles")
abline(0,1)
```

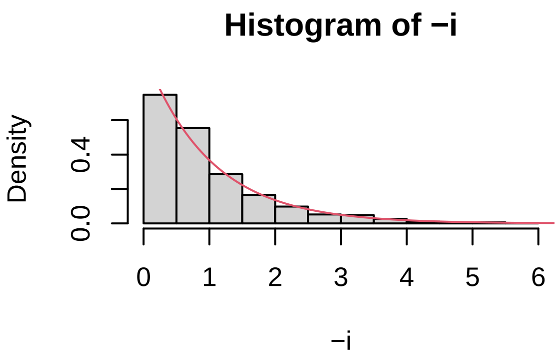


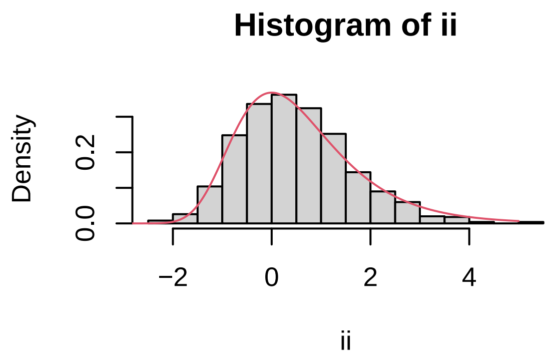
On average, the sample and theoretical quantiles agree. With heavy tailed distributions like the Cauchy, we can expect some (but few) extreme values like the outliers we see here.

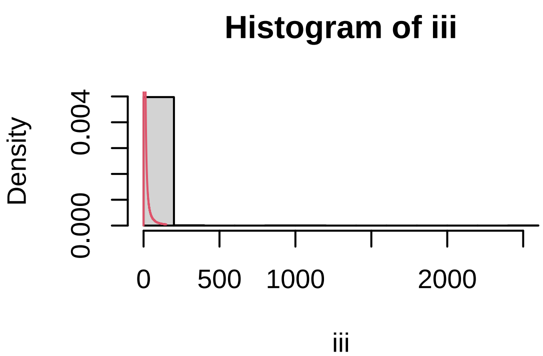
17.5.3 Part C

A simulation experiment demonstrating the first EVT (convergence of the max of RVs goes to one of three distributions)

```r
m = 1000
n = 100

partc = function(func, an, bn, ...){
  mn = c()
  for(i in 1:m){
    mn[i] = max(func(n,...))
  }
  yj = (mn - bn)/an
  return(yj);
}

i = partc(runif, 1/n, 1)
ii = partc(rnorm, 1/(n*dnorm(qnorm(1-1/n))), qnorm(1-1/n))
iii = partc(rt, n/pi, 0, df =1)

# overlay theoretical distribution
par(mfrow=c(2,2))
hist(-i, probability = T)
x=seq(0,10,.01)
lines(x, dexp(x), col = 2)

hist(ii, probability = T)
x=seq(-5,5,.01)
lines(x, exp( - x - exp(-x)), col = 2)

hist(iii, probability = T)
x=seq(0,150,.01)
lines(x, 1/x^2 * exp(- 1/x), col = 2)
```





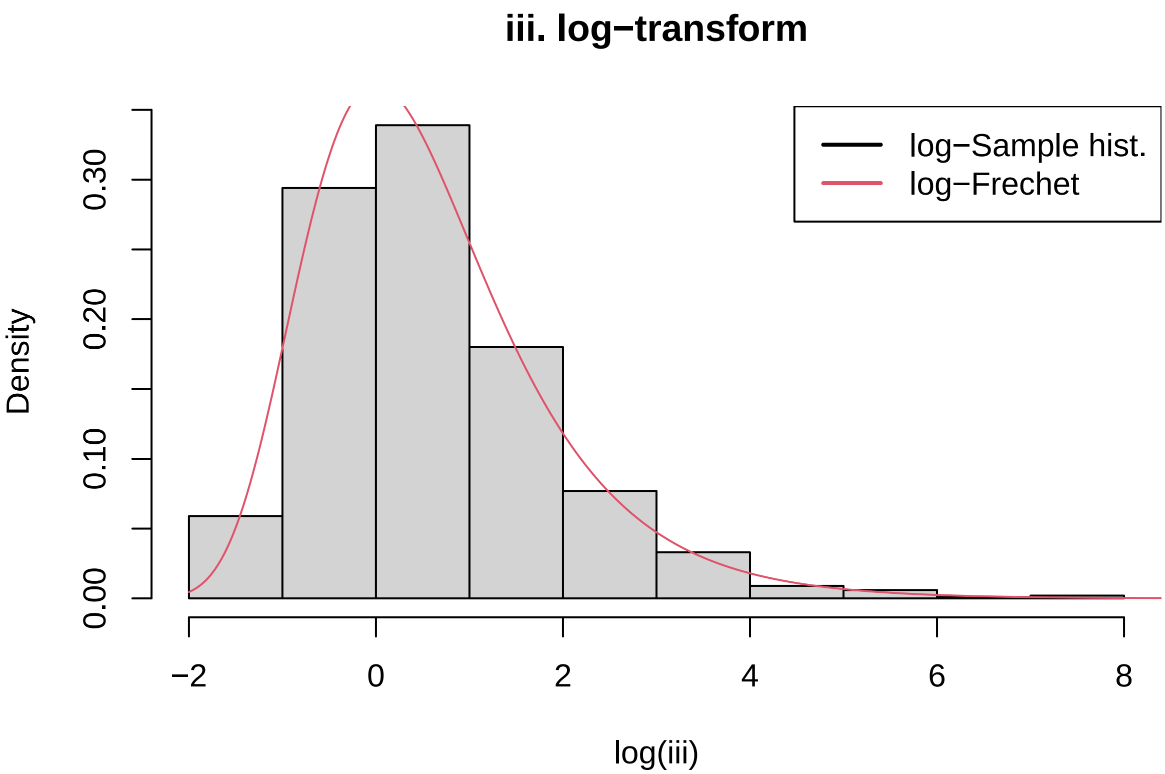


From answer key: >In cases i. and ii. the convergence to the theoretical distribution is evident, but in iii. the histogram is not as informative because of the extreme values involved. For the last case, one can plot the histogram & density of the log-transformed normalized maxima, which actually show the convergence. (Note: the log-transformation of the Frechet($\alpha = 1$) is the Gumbel distribution.)

I think it is incorrect, as I think the exponential transform of the Frechet($\alpha = 1$) is the Gumbel distribution.

The solutions continue on to show:

```
hist( log(iii), probability = T, main = "iii. log-transform" )
x=seq(-2,15,.01)
lines(x, exp( -x-exp(-x)), col = 2)
legend( "topright", lwd = 2, col = 1:2, c("log-Sample hist.","log-Frechet"))
```


iii. log−transform
log−Sample hist.
log−Frechet
Density
0.00
0.10
0.20
0.30
−2
0
2
4
6
8
log(iii)

Chapter 18

Problem Set 3

18.1 Q1

Easy copula question about certain configurations of the marginal distributions. In the first case, we have two exactly equal uniforms and the second case is when one uniform is the exact complement of the other.

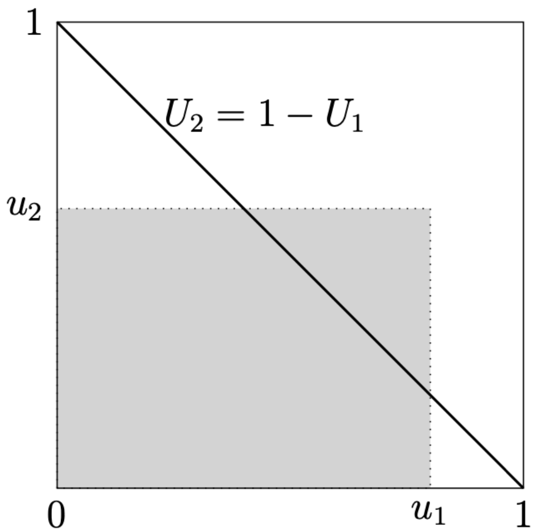
18.1.1 Part A

$$\begin{aligned}
U_1 &= U_2 \sim U(0,1) \\
\implies C(u_1, u_2) &= \mathbb{P}(U \leq u_1, U \leq u_2) \\
&= \mathbb{P}(U \leq \min(u_1, u_2)) \\
&= \min(u_1, u_2) \quad \forall\ u_1, u_2 \in [0,1] \quad \text{by def of Uniform} \\
&= \bar{C}(u_1, u_2)
\end{aligned}$$

18.1.2 Part B

$$\begin{aligned}
U_1 &= 1 - U_2 \\
C(u_1, u_2) &= \mathbb{P}(U_1 \leq u_1, U_2 \leq u_2) \\
&= \mathbb{P}(1 - U \leq u_1, U \leq u_2) \\
&= \mathbb{P}(U \geq 1 - u_1, U \leq u_2) \\
&= \mathbb{P}(1 - u_1 \leq U \leq u_2) \\
&= \begin{cases} 0 & 1 - u_1 \geq u_2 \implies 1 \geq u_1 + u_2 \\ u_2 - (1 - u_1) = u_1 + u_2 - 1 & 1 - u_1 \leq u_2 \implies 1 \leq u_1 + u_2 \end{cases} \\
&= \max(u_1 + u_2 - 1, 0) \\
&= \underline{C}(u_1, u_2)
\end{aligned}$$

Note that for both cases, the joint 2D distribution is degenerate, i.e. the RVs U_1, U_2 take values in a *lower-dimensional* space. In the first case, the RVs take values in the (1D) identity line of the (2D) unit square. In the second case they take values in the line with slope -1 an intercept 1 over the unit square, as in the plot below:



Degenerate distributions arise when some RV(s) can be expressed as deterministic functions of the others.

18.2 Q2

This question concerns Archimedean copula's, where instead of uniforms, it uses some convex generator function ϕ.

$$\phi(u) = \ln\left(\frac{1-\theta(1-u)}{u}\right) \quad \theta \in [-1, 1)$$

18.2.1 Part A

Closed form expression for the resulting bivariate copula:

First we need to find the inverse, where instead of y being a function of u, we have u is some function of y:

$$\begin{aligned}
y = \phi(u) &= \ln\left(\frac{1-\theta(1-u)}{u}\right) \\
e^y &= \frac{1-\theta(1-u)}{u} \\
ue^y &= 1-\theta(1-u) \\
ue^y &= 1-\theta+\theta u \\
u(e^y-\theta) &= 1-\theta
\end{aligned}$$

$$\implies \phi^{-1}(y) = u = \frac{1-\theta}{e^y-\theta}$$

Now we can plug in the two variables:

$$\begin{aligned}
C(u_1,u_2) &= \phi^{-1}(\phi(u_1)+\phi(u_2)) \\
&= \phi^{-1}\left(\ln\left(\frac{1-\theta(1-u_1)}{u_1}\right)+\ln\left(\frac{1-\theta(1-u_2)}{u_2}\right)\right) \\
&= \phi^{-1}\left(\ln\left(\frac{1-\theta(1-u_1)}{u_1}\cdot\frac{1-\theta(1-u_2)}{u_2}\right)\right) \\
&= \frac{1-\theta}{\left(\frac{1-\theta(1-u_1)}{u_1}\cdot\frac{1-\theta(1-u_2)}{u_2}\right)-\theta}
\end{aligned}$$

Multiply top and bottom by $u_1 \cdot u_2$

$$\begin{aligned}
&= \frac{(1-\theta)\cdot u_1\cdot u_2}{[((1-\theta)+\theta u_1)\cdot((1-\theta)+\theta u_2)]-\theta\cdot u_1\cdot u_2} \\
&= \frac{(1-\theta)\cdot u_1\cdot u_2}{[((1-\theta)^2+(1-\theta)\theta u_1+(1-\theta)\theta u_2)+\theta^2 u_1u_2]-\theta\cdot u_1\cdot u_2} \\
&= \frac{(1-\theta)\cdot u_1\cdot u_2}{(1-\theta)\,[(1-\theta)+\theta u_1+\theta u_2]+\theta^2u_1u_2-\theta\cdot u_1\cdot u_2} \\
&= \frac{(1-\theta)\cdot u_1\cdot u_2}{(1-\theta)\,[(1-\theta)+\theta u_1+\theta u_2]-(1-\theta)(\theta u_1u_2)} \\
&= \frac{u_1u_2}{1-\theta+\theta u_1+\theta u_2-\theta u_1u_2} \\
&= \frac{u_1u_2}{1-\theta(1-u_1-u_2+u_1u_2)} \\
&= \frac{u_1u_2}{1-\theta(1-u_1)(1-u_2)}
\end{aligned}$$

18.2.2 Part B

$$\begin{aligned}
M_n &= \max\{U_1, \dots, U_n\} \\
F_{M_n}(m) &= P(M_n \leq m) = P\left(\prod_{i=1}^{n}(U_i \leq m)\right) \\
&= P(U_1 \leq m, \dots, U_n \leq m) \\
&= C(U_1 \leq m, \dots, U_n \leq m) \\
&= \phi^{-1}(\phi(m) + \dots + \phi(m)) \\
&= \phi^{-1}(n\phi(m)) \\
&= \frac{1-\theta}{e^{n\phi(m)} - \theta} \\
&= \frac{1-\theta}{\exp\{n \ln\left(\frac{1-\theta(1-m)}{m}\right)\} - \theta} \\
&= \frac{1-\theta}{\left(\frac{1-\theta(1-m)}{m}\right)^n - \theta}
\end{aligned}$$

18.3 Q3

We want to show that the Archimedean copula with $\phi(u) = -\log(u)$ is exactly the independence copula.

$$\begin{aligned}
y = \phi(u) &= -\log(u) = \log(u^{-1}) \\
e^y &= u^{-1} \\
e^{-y} &= u = \phi^{-1}(y)
\end{aligned}$$

> Professor's solution has $\phi^{-1} = e^y$ which makes is key to showing equality to the independence copula. I have yet to understand why that is the inverse generator function.

$$\begin{aligned}
C(u_1, \dots, u_n) &= \phi^{-1}(\phi(u_1) + \dots + \phi(u_n)) \\
&= \exp(\log(u_1) + \dots + \log(u_n)) \\
&= \exp(\log(u_1 \times \dots \times u_n)) \\
&= u_1 \times \dots \times u_n \\
&= C_{indep}
\end{aligned}$$

It is easy to show that this holds for any logarithm with any base.

18.4 Q4

A convex combination of k joint CDFs is itself a joint CDF (finite mixture), but is a convex combination of k copula functions a copula function itself?

The result can be shown using induction. A convex combination of two copula's $C() = wC_1() + (1-w)C_2$ is also a copula. We are also given that a convex combination of k joint CDFs is also a CDF, which means that any convex combination of Copula's will still be some CDF. We only need to show that the marginals are Uniform(0,1) in order to show that it will be a copula.

For 1D Uniform CDFs $F_1(u) = F_2(u) = u \quad \forall u \in [0, 1]$ then

$$F(u) = wF_1(u) + (1 - w)F_2(u) = wu + (1 - w)u = u \quad \forall\ u \in [0, 1]$$

Chapter 19

Problem Set 4

Q1-Q3 are from the SDAFE textbook, 16.11 Q4-Qsmth are from 16.10

19.1 Q1.1

Suppose that there are two risky assets, A and B, with expected returns equal to 2.3% and 4.5%, respectively. Suppose that the standard deviations of the returns are $\sqrt{6}$% and $\sqrt{11}$% and that the returns on the assets have a correlation of 0.17.

19.1.1 Part A

What portfolio of A and B achieves a 3 % rate of expected return?

$$\begin{aligned}
\mathbb{E}(wA + (1-w)B) &= 0.03 \\
&= w\mathbb{E}(A) + (1-w)\mathbb{E}(B) \\
&= w \cdot .023 + (1-w) \cdot .045 \\
0.03 &= w(.023 - .045) + .045 \\
w &= \frac{.03 - .045}{.023 - .045} \\
w &= 0.6818
\end{aligned}$$

A portfolio made up of 68% of asset A and 32% of asset B will achieve a 3% rate of expected return.

19.1.2 Part B

What portfolios of A and B achieve a 5.5% standard deviation of return? Among these, which has the largest expected return?

$$\begin{aligned}
\mathbb{V}(R_p) &= \mathbb{V}(wA + (1-w)B) \\
&= w^2\mathbb{V}(A) + (1-w)^2\mathbb{V}(B) + 2 \cdot w(1-w)Cov(A,B) \\
&= w^2 0.06 + (1-w)^2 0.11 + w(1-w)2 \cdot 0.17 \cdot \sqrt{0.06 \cdot 0.11} \\
&= w^2 0.06 + (1-2w+w^2)0.11 + (w-w^2)2 \cdot 0.17 \cdot \sqrt{0.06 \cdot 0.11} \\
&= w^2(6 + 11 - 2.7621) + w(2.7621 - 2 \cdot 11) + 11 \\
0.055^2 &= 14.2379w^2 - 19.2379w + 11 \\
0 &= 14.2379w^2 - 19.2379w + 11 - 5.5 \\
w &= 0.4108 \text{ and } 0.9404
\end{aligned}$$

As the portfolio with weight 41% in A will give asset B more weight, and because asset B has a higher expected return, this portfolio will have the largest expected return.

19.2 Q1.2

Suppose there are two risky assets, C and D, the tangency portfolio is 65% C and 35% D, and the expected return and standard deviation of the return on the tangency portfolio are 5% and 7%, respectively. Suppose also that the risk-free rate of return is 1.5%. If you want the standard deviation of your return to be 5%, what proportions of your capital should be in the risk-free asset, asset C, and asset D?

$$\begin{aligned}
\text{Let } w_f &= \text{weight of risk-free asset} \\
\mathbb{V}(R_p) &= \mathbb{V}(\underbrace{w_f r_f}_{0} + (1-w_f)r_{tangency}) \\
5^2\% &= (1-w_f)^2 \cdot 7\%^2 \\
\frac{5\%}{7\%} &= 1 - w_f \\
0.7142 &= 1 - w_f \\
\implies w_c &= .7142 \cdot .65 = 0.4642 \\
w_d &= .7142 \cdot .35 = 0.25 \\
w_f &= 0.2857
\end{aligned}$$

But the solutions states: 2/7 = 0.2857 in risk-free, 0.4643 in C, and 0.2500 in D

> I have no idea how. Can't spot my mistake : I'm an idiot, remember to square std dev to get variance and double check your work

19.3 Q1.3

19.3.1 Part A

Suppose that stock A shares sell at \$75 and stock B shares at \$115. A portfolio has 300 shares of stock A and 100 of stock B. What are the weights w and 1−w of stocks A and B in this portfolio?

$$w = \frac{75 \cdot 300}{75 \cdot 300 + 115 \cdot 100} = 0.66176$$

$$1 - w = 0.33824$$

19.3.2 Part B

More generally, if a portfolio has N stocks, if the price per share of the jth stock is P_j, and if the portfolio has n_j shares of stock j, then find a formula for w_j as a function of $n_1, \ldots, n_N$ and $P_1, \ldots, P_N$.

$$w_j = \frac{n_j \cdot P_j}{\sum_{i=1}^{N} n_i P_i}$$

19.4 Q1.4

Let R_P be a return of some type on a portfolio and let $R_1, \ldots, R_N$ be the same type of returns on the assets in this portfolio. Is $R_P = w_1 R_1 + \cdots + w_N R_N$ true if R_P is a net return? Is this equation true if R_P is a gross return? Is it true if R_P is a log return? Justify your answers.

$$\begin{gathered}
\text{Gross Returns:} \\
R'(t) = \frac{S(t)}{S(t-1)} \\
\text{Net Returns:} \\
R(t) = R'(t) - 1 = \frac{S(t) - S(t-1)}{S(t-1)} \\
\text{Log Returns:} \\
r(t) = \log(R'(t)) = \log\{S(t)\} - \log\{S(t-1)\}
\end{gathered}$$

We have shown in class that this equation holds for net returns. For gross returns:

$$\begin{aligned}
R'(t) = R(t) + 1 &= w_1 R_1(t) + \cdots + w_N R_N(t) + 1 \\
&= w_1 R_1(t) + \cdots + w_N R_N(t) + (w_1 + \cdots + w_N) \\
&= w_1 (R_1(t) + 1) + \cdots + w_N (R_N(t) + 1) \\
&= w_1 R_1'(t) + \cdots + w_N R_N'(t)
\end{aligned}$$

So it holds for gross returns as well.

$$\begin{aligned}
r(t) = \log(R'(t)) &= \log(w_1 R_1'(t) + \cdots + w_N R_N'(t)) \\
&= \log(w_1 R_1'(t)) \times \cdots \times \log(w_N R_N'(t)) \\
&= [\log(w_1) + \log(R_1'(t))] \times \cdots \times [\log(w_N) + \log(R_N'(t))] \\
&= [\log(w_1) + r_1(t)] \times \cdots \times [\log(w_N) + \log(r_N(t))] \\
&\neq w_1 r_1(t) + \cdots + w_N R_N(t)
\end{aligned}$$

So it does not hold for log returns.

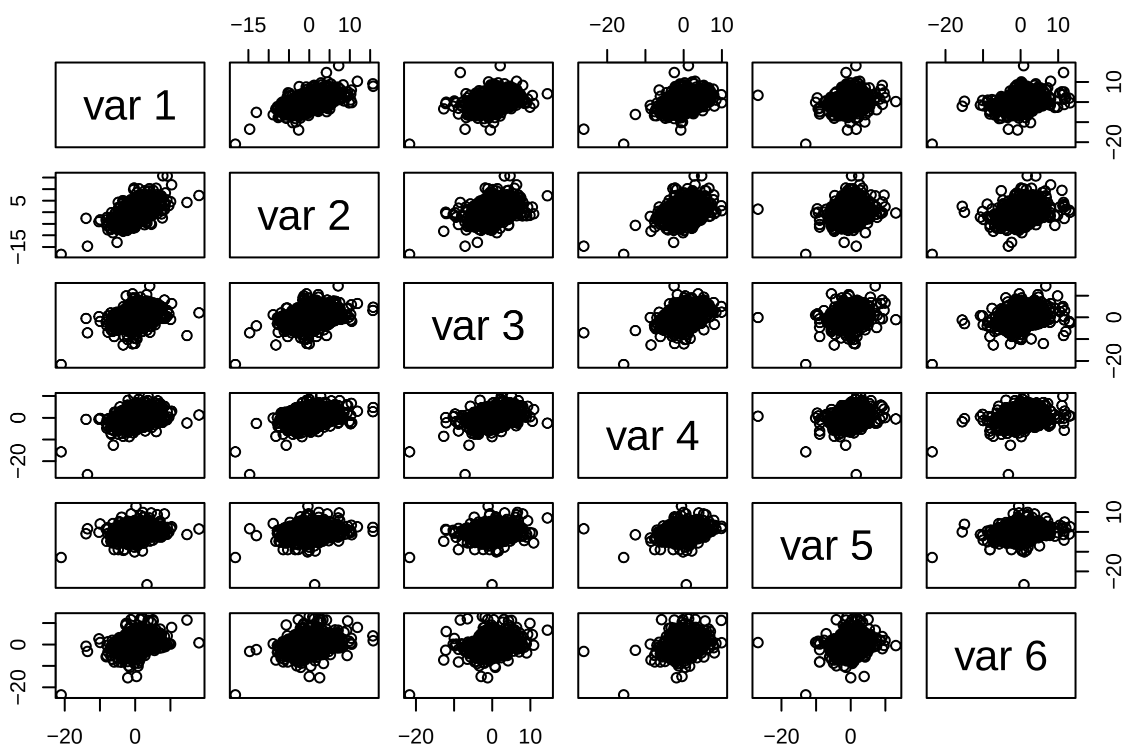
The following is from 16.10 1-3

19.5 Q2.1

Write an R program to find the efficient frontier, the tangency portfolio, and the minimum variance portfolio, and plot on “reward-risk space” the location of each of the six stocks, the efficient frontier, the tangency portfolio, and the line of efficient portfolios. Use the constraints that -0.1 wj 0.5 for each stock. The first constraint limits short sales but does not rule them out completely. The second constraint prohibits more than 50 % of the investment in any single stock. Assume that the annual risk-free rate is 3 % and convert this to a daily rate by dividing by 365, since interest is earned on trading as well as nontrading days.

Taken straight from the answer key

```
library(readr)
dat = read.csv("ProblemSets/datasets/Stock_Bond.csv", header = T)
prices = cbind(dat$GM_AC, dat$F_AC, dat$CAT_AC, dat$UTX_AC,
  dat$MRK_AC, dat$IBM_AC)
n = dim(prices)[1]
returns =  100 * (prices[2:n, ] / prices[1:(n-1), ] - 1)
pairs(returns)
```

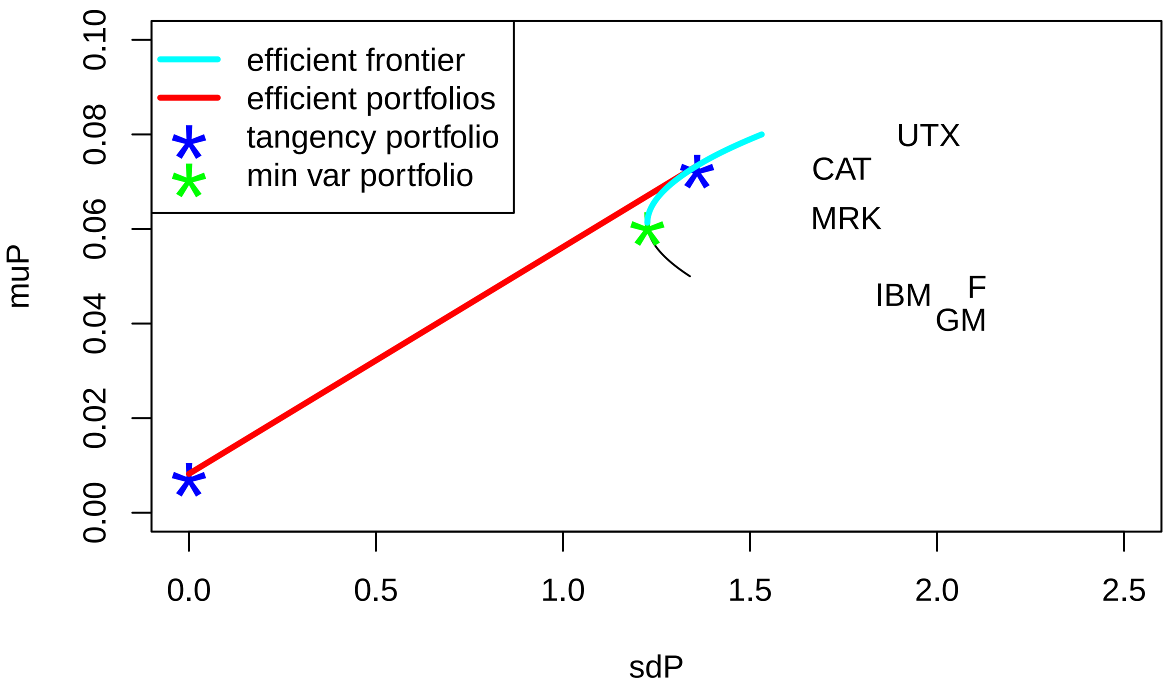


```
mean_vect = colMeans(returns)
cov_mat = cov(returns)
sd_vect = sqrt(diag(cov_mat))
```

```
M = length(mean_vect)
library(quadprog)
Amat = cbind(rep(1,M),mean_vect,diag(1,nrow=M),-diag(1,nrow=M))
muP = seq(min(mean_vect)+.02,max(mean_vect)-.02,length=10)
muP = seq(.05,0.08,length=300)
sdP = muP
weights = matrix(0,nrow=300,ncol=M)
for (i in 1:length(muP))
{
result =
solve.QP(Dmat=cov_mat,dvec=rep(0,M), Amat=Amat,
c(1,muP[i],rep(-.1,M),rep(-.5,M)), meq=2)
sdP[i] = sqrt(2*result$value)
weights[i,] = result$solution
}
plot(sdP,muP,type="l",xlim=c(0,2.5),ylim=c(0,.1))
mufree = 3/365
points(0,mufree,cex=3,col="blue",pch="*")
sharpe =( muP-mufree)/sdP
ind = (sharpe == max(sharpe)) # locates the tangency portfolio
weights[ind,] # weights of the tangency portfolio
```

```
## [1] -0.091181044 -0.002910879  0.335318542  0.383714329  0.319484849
## [6]  0.055574204
```

```
lines(c(0,sdP[ind]),c(mufree,muP[ind]),col="red",lwd=3)
points(sdP[ind],muP[ind],col="blue",cex=3,pch="*")
ind2 = (sdP == min(sdP))
points(sdP[ind2],muP[ind2],col="green",cex=3,pch="*")
ind3 = (muP > muP[ind2])
lines(sdP[ind3],muP[ind3],type="l",xlim=c(0,.25),
ylim=c(0,.3),col="cyan",lwd=3)
text(sd_vect[1],mean_vect[1],"GM")
text(sd_vect[2],mean_vect[2],"F")
text(sd_vect[3],mean_vect[3],"UTX")
text(sd_vect[4],mean_vect[4],"CAT")
text(sd_vect[5],mean_vect[5],"MRK")
text(sd_vect[6],mean_vect[6],"IBM")
legend("topleft",c("efficient frontier","efficient portfolios",
                   "tangency portfolio","min var portfolio"),
       lty=c(1,1,NA,NA),
       lwd=c(3,3,1,1),
       pch=c("","","*","*"),
       col=c("cyan","red","blue","green"),
       pt.cex=c(1,1,3,3)
)
```



19.6 Q2.2

If an investor wants an efficient portfolio with an expected daily return of 0.07%, how should the investor allocate his or her capital to the six stocks and to the risk-free asset? Assume that the investor wishes to use the tangency portfolio computed with the constraints $-0.1 \leq w_j \leq 0.5$, not the unconstrained tangency portfolio.

```
options(digits=3)
# it divides by the risk free rate converted to daily but theres a typo,
# should be 3/365 (unless this is the amt of trading days, but that would be 252)
# omega = (.07 - muP[ind]) / (3/265 - muP[ind])
omega = (.07 - muP[ind]) / (3/365 - muP[ind])
omega
```

```
## [1] 0.0518
```

```
1-omega
```

```
## [1] 0.948
```

```
(1-omega)*weights[ind]
```

```
## [1] -0.08645 -0.00276  0.31794  0.36382  0.30292  0.05269
```

```
omega + sum((1-omega)*weights[ind])
```

```
## [1] 1
```

19.7 Q2.3

Does this data set include Black Monday?

> Again, straight from the answer key.

Yes, Black Monday was October 19, 1987 and data go from January 2, 1987 to Sept 1, 2006. Black Monday is the 202th day in the original data set or the 201st day of returns. If you look in the spread sheet you will see huge price declines that day. The returns that day were:

```
returns[201,]
[1] -21.0 -18.2 -21.7 -15.7 -13.0 -23.5
```

19.8 Q3

From answer key:

```
library(tseries)
library(zoo)

tickers = c("ENB.TO","CP.TO","RCI-A.TO","TD.TO","L.TO")
N = length(tickers)

P=vector("list", N) # list for holding prices

for (i in 1:N) {
  cat("Downloading ", i, " out of ", N , "\n")
  P[[i]] = get.hist.quote(instrument = tickers[i],
    start = as.Date("2010-01-01"),
    end=as.Date("2020-12-01"),
    compression="m", quote = "AdjClose",
    retclass = "zoo", quiet = T)
}
```

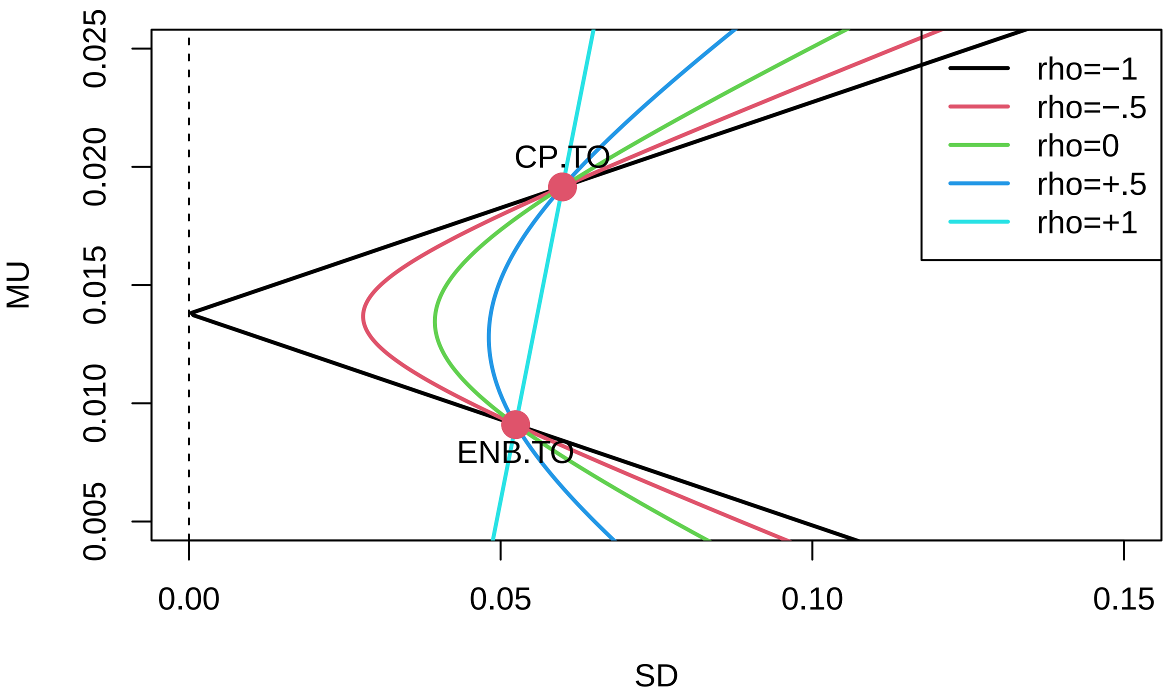
```
## Downloading  1  out of  5
## Downloading  2  out of  5
## Downloading  3  out of  5
## Downloading  4  out of  5
## Downloading  5  out of  5
```

```
# net returns
R = sapply(P, FUN=function(x){ as.numeric(diff(x) / stats::lag(x, -1)) } )
colnames(R) = tickers # assign names
```

19.8.1 Part A

Use only the first two stocks (ENB.TO and CP.TO), and calculate the sample means and (individual) sample variances of their returns. Consider the following hypothetical values for their correlation : $\rho = -1,\ -0.5,\ 0,\ +0.5,\ +1$. For each value of ρ, calculate their corresponding 2D variance-covariance matrix and plot the risk-return profiles of portfolios combining the two assets with weights [w, $(1 - $w)], $\forall w \in [-1, 2]$. Plot all profile curves on the same (μ_p, σ_p)-space, using a different color for each value of ρ.

```
R2=R[,1:2]
MU = colMeans(R2)
SD = sqrt(diag(var(R2)))
par(mfrow=c(1,1), mar=c(5, 4, 4, 2) + 0.1)
plot(SD,MU,pch=16,cex=2, col=2, xlim=c(0,.15), ylim=c(0.005,.025))
abline(v=0, lty=2); abline(h=0, lty=2)

w = seq(-2,+3,.01); W = cbind(w, 1-w) # weights
MU.p = W %*% MU    # portfolio means

rho=c(-1, -.5, 0, .5, 1) # correlations
for(i in 1:5){
  COR=matrix(c(1,rho[i],rho[i],1),2,2)
  COV= COR * (SD%*%t(SD))
  SD.p = sqrt(rowSums((W %*% COV)*W)) # portfolio st.dev.
  lines(SD.p, MU.p, type='l', lwd=2, col=i);
}

points(SD,MU,pch=16,cex=2, col=2)
text(SD, MU, pch=16, colnames(R2), pos=c(1,3))
legend('topright', lwd=rep(2,5),col=1:5,
       c("rho=-1","rho=-.5","rho=0","rho=+.5","rho=+1")  )
```

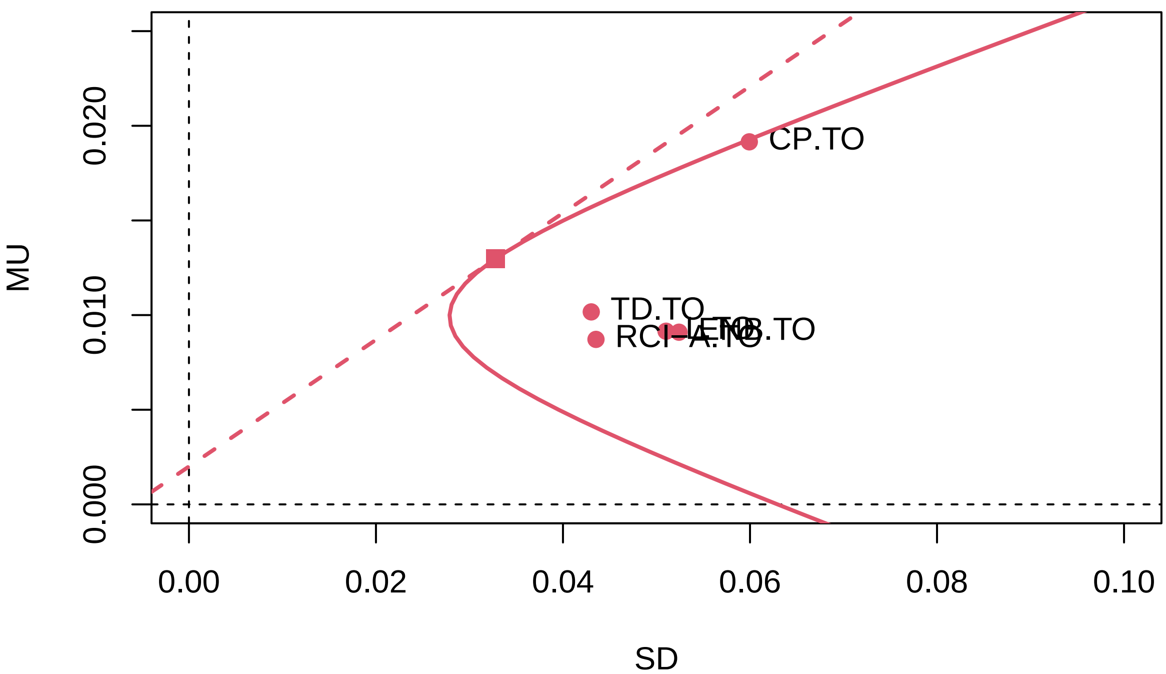


19.8.2 Part B

Consider all 5 stocks together now, and use the sample mean and sample variance-covariance matrix of their returns. Plot the efficient frontier and the capital market line on the same (μ_p, σ_p)-space and report the tangency portfolio weights. (Hint: adapt the code from Example 16.6 on p. 476 of SDAFE.)

```
library(quadprog)

COV=cov(R)
MU=colMeans(R)
SD=sqrt(diag(COV))
N=dim(R)[2]

plot(SD, MU, pch=16, cex=1.2, col= 2, xlim=c(0,.1), ylim=c(0,.025))
abline(v=0, lty=2); abline(h=0, lty=2)
text(SD, MU, tickers, cex=1, pos=4)

Amat = cbind(rep(1,N),MU)
mu.p = seq( -.005, .05,length=100)
sd.p = mu.p;
```

```
for (i in 1:length(mu.p))
{
  bvec=c(1,mu.p[i])
  out=solve.QP(Dmat=2*COV,dvec=rep(0,N),Amat=Amat,bvec=bvec,meq=2)
  sd.p[i] = sqrt(out$value)
}
lines(sd.p,mu.p,type="l", lwd=2, col=2) # plot least variance portfolios

mu.f = .002 # monthly risk-free interest rate

COV.i=solve(COV)
W.tang=COV.i%*%(MU-mu.f) / sum( COV.i%*%(MU-mu.f) )
mu.tang=sum(W.tang*MU)
sd.tang=sqrt(sum( (COV %*% W.tang) * W.tang ) )
points( sd.tang, mu.tang, pch=15, cex=1.3, col=2)
sharpe=(mu.tang-mu.f)/sd.tang
abline(mu.f,sharpe,lwd=2,lty=2,col=2)
```



The Sharpe ratio is 0.335, and the tangency portfolio weights are:

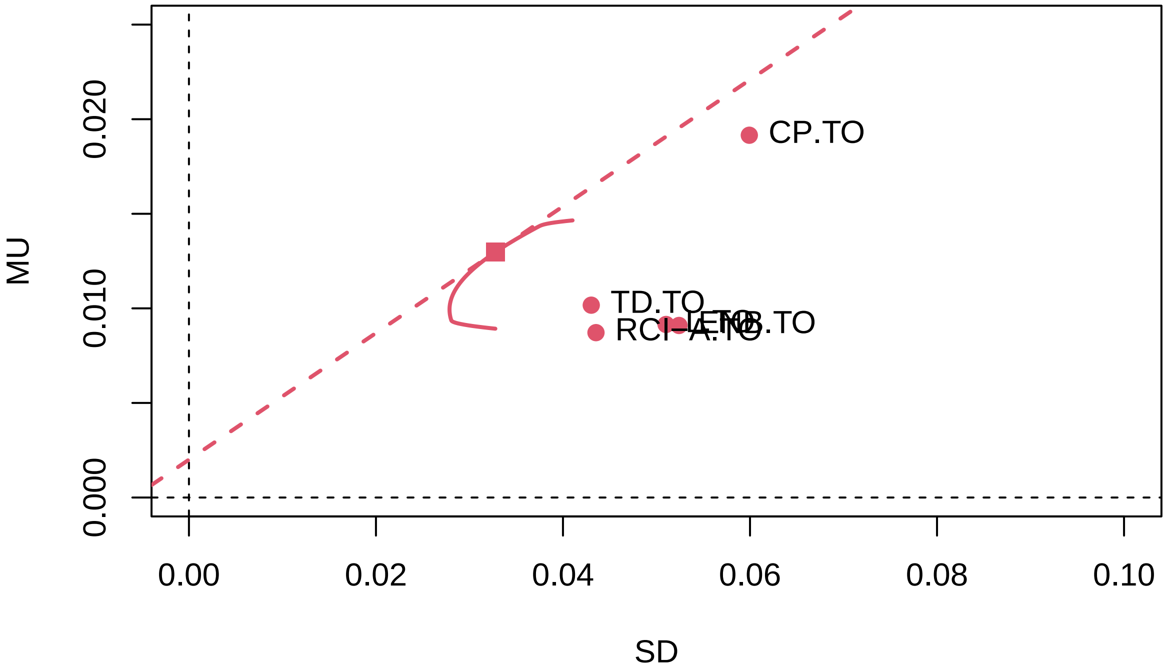
```
round( t(W.tang), 4 )
```

```
##      ENB.TO CP.TO RCI-A.TO TD.TO  L.TO
## [1,]  0.135 0.362    0.106 0.248 0.149
```

19.8.3 Part C

Repeat the previous part b. (i.e. plot the efficient frontier and capital market line, and report the tangency portfolio weights) with the restriction that all weights are within the bounds $0 \leq w_i \leq 0.5, \quad \forall i = 1, \ldots, 5$. (Hint: adapt the code from Example 16.7 on p. 479 of SDAFE.)

```
plot(SD, MU, pch=16, cex=1.2, col= 2,  xlim=c(0,.1), ylim=c(.0,.025))
abline(v=0, lty=2); abline(h=0, lty=2)
text(SD, MU, tickers, cex=1, pos=4)

Amat = cbind(rep(1,N),MU,diag(1,nrow=N),-diag(1,nrow=N))
mu.pot = seq( min(MU), max(MU),length=300) # potential mean returns

mu.p = NULL # initialize portofio standard error
sd.p = NULL # initialize portofio standard error
W.p = NULL # initialize portofio weights

for (i in 1:length(mu.pot))
{
  bvec=c(1,mu.pot[i],rep(0,N),rep(-0.5,N))
  #check whether potential mean return can be achieved with given constraints
  out=tryCatch( solve.QP(Dmat=2*COV,dvec=rep(0,N),Amat=Amat,bvec=bvec,meq=2), error=function(e) NULL)
  #if mean return is achievable, save its st.dev. & portfolio weights
  if(!is.null(out)){ #  if
    mu.p=c(mu.p, mu.pot[i])
    sd.p=c(sd.p, sqrt(out$value))
    W.p=rbind(W.p, out$solution)
  }
}
lines(sd.p,mu.p,type="l", lwd=2, col=2) # plot least variance portfolios
colnames(W.p)=tickers

sharpe=( mu.p-mu.f)/sd.p
ind.tang=which.max(sharpe)
W.tang=W.p[ind.tang,]
sd.tang = sd.p[ind.tang]
mu.tang = mu.p[ind.tang]
points( sd.tang, mu.tang, pch=15, cex=1.3, col=2)
abline( c(mu.f, sharpe[ind.tang]),lwd=2, col=2, lty=2)
```



The set of feasible portfolios will be a subset of that of the unconstrained problem. Note that the *constrained* efficient frontier is not a parabola any more. But the constrained tangency portfolio is the same as the unconstrained one (since the since the unconstrained tangency portfolio weights were all within [0,.5]), and have (approximately) the same Sharpe ratio 0.335.

Note that the constrained tangency portfolio weights are (approximately) the same as the unconstrained ones:

```
round( W.tang, 4 )
```

```
##   ENB.TO    CP.TO RCI-A.TO    TD.TO     L.TO
##    0.135    0.362    0.106    0.248    0.149
```

Chapter 20

Problem Set 5

20.1 Q1

Assume a market of N assets with returns following the 1-factor CAPM model

$$R_i = \beta R_M + \varepsilon_i, \quad i = 1, \dots, N$$

where $R_M \sim \mathrm{N}\left(\mu_M, \sigma_M^2\right)$ and $\varepsilon_i \sim^{i.i.d.} \mathrm{N}\left(0, \sigma_\varepsilon^2\right), \quad \forall i$. Therefore, the model assumes all assets have the same systematic risk $(\beta^2\sigma_M^2)$ and the same idiosyncratic risk (σ_ε^2).

20.1.1 Part A

(a) Find the weights of the minimum-variance portfolio.

$$\mathbf{R} = \begin{bmatrix} R_1 \\ \vdots \\ R_N \end{bmatrix} = \begin{bmatrix} \beta \\ \vdots \\ \beta \end{bmatrix} R_M + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_N \end{bmatrix}$$

$$\begin{aligned}
\Sigma = \mathbb{V}(\mathbf{R}) &= \mathbb{V}(\beta \underline{1} R_M + \epsilon_i) \\
&= \beta^2 \underline{1} \mathbb{V}(R_M) \underline{1}^T + \mathbb{V}(\epsilon_i)\mathbf{I} \\
&= \beta^2 \sigma_M^2 \begin{bmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{bmatrix} + \begin{bmatrix} \sigma_\epsilon^2 & 0 & \dots & 0 \\ 0 & \sigma_\epsilon^2 & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_\epsilon^2 \end{bmatrix} \\
&= \begin{bmatrix} \beta^2\sigma_M^2 + \sigma_\epsilon^2 & \dots & \beta^2\sigma_M^2 \\ \vdots & \ddots & \vdots \\ \beta^2\sigma_M^2 & \dots & \beta^2\sigma_M^2 + \sigma_\epsilon^2 \end{bmatrix}
\end{aligned}$$

The minimum variance portfolio weights as calculated in class are:

$$w^* = \frac{\Sigma^{-1}\underline{1}}{\underline{1}^T \Sigma^{-1} \underline{1}}$$

By symmetry of the variance matrix, all the weights must be equal, implying $w = \frac{1}{N}\mathbf{1}$. Each row sum of the matrix is also equivalent, as intuitively all the assets follow the same distribution and have the same systematic risk. It also means that changing the order of the assets in the vector **R** does not change the covariance matrix.

20.1.2 Lemma for special inverse covariance matrices

Given that $a, b > 0$

$$\Sigma = a\mathbf{1}\mathbf{1}^T + bI = \begin{bmatrix} a & \dots & a \\ \vdots & \ddots & \vdots \\ a & \cdots & a \end{bmatrix} + \begin{bmatrix} b & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & b \end{bmatrix}$$

then

$$\Sigma^{-1} = c\mathbf{1}\mathbf{1}^T + dI = \begin{bmatrix} c & \dots & c \\ \vdots & \ddots & \vdots \\ c & \cdots & c \end{bmatrix} + \begin{bmatrix} d & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & d \end{bmatrix}$$

where $c = -\frac{1}{b}\frac{a}{Na+b}$ and $d = \frac{1}{b}$

20.1.3 Part B

(b) Show that the minimum variance can never be smaller than $\beta^2\sigma_M^2$, no matter how many assets we have (i.e., no matter how large N is).

$$\begin{aligned} \mathbb{E}\left[R_{mv}\right] &= \mathbb{E}\left[w^T R\right] = w^T E[R] = \frac{1}{N}\mathbf{1}^T\left(\beta\mu_M\mathbf{1}\right) \\ &= \beta\mu_M\frac{\mathbf{1}^T\mathbf{1}}{N} = \beta\mu_M \quad (\text{ since } \mathbf{1}^T\mathbf{1} = N) \end{aligned}$$

$$\begin{aligned} \mathbb{V}\left[R_{mv}\right] &= \mathbb{V}\left[w^T R\right] = w^T\mathbb{V}[R]w = \frac{1}{N^2}\mathbf{1}^T\left(\beta^2\sigma_M^2\mathbf{1}\mathbf{1}^T + \sigma_\varepsilon^2 I\right)\mathbf{1} = \\ &= \frac{\beta^2\sigma_M^2}{N^2}\left(\mathbf{1}^T\mathbf{1}\right)\left(\mathbf{1}^T\mathbf{1}\right) + \frac{\sigma_\varepsilon^2}{N^2}\left(\mathbf{1}^T\mathbf{1}\right) = \beta^2\sigma_M^2 + \frac{\sigma_\varepsilon^2}{N} > \beta^2\sigma_M^2, \forall N \geq 1 \end{aligned}$$

Thus, the minimum variance portfolio will always have variance at least $\beta^2\sigma_M^2$. This problem illustrates that you cannot "diversify away" systemic risk (i.e., risk from common factors) the same way you can with idiosyncratic risk.

20.2 Q2

Consider the following 2 -factor model with 3 assets:

$$\begin{bmatrix} R_1 \\ R_2 \\ R_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} F_1 \\ F_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix} \Leftrightarrow R = \beta^\top F + \varepsilon$$

where

$$\mathbb{V}[F] = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix} = \sigma^2 I_2, \quad \mathbb{V}[\varepsilon] = \begin{bmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{bmatrix} = \sigma^2 I_3, \quad \operatorname{Cov}[\varepsilon, F] = 0$$

Find the minimum variance portfolio weights for this model. (Hint: You can use R to invert the matrix.)

From class,

$$w^\star = \frac{\Sigma^{-1}\underline{1}}{\underline{1}^T\Sigma^{-1}\underline{1}}$$

So we first need to find Σ.

$$\begin{aligned}
\Sigma &= \mathbb{V}(^{\mathbf{T}}\mathbf{F} +) \\
&= {}^{T}\mathbb{V}(\mathbf{F}) + \mathbb{V}() \\
&= \sigma^2{}^{\mathbf{T}}\mathbf{I} + \sigma^2\mathbf{I_3} \\
&= \sigma^2 \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} + \sigma^2 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\
&= \sigma^2 \begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{bmatrix} + \sigma^2 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\
&= \sigma^2 \begin{bmatrix} 2 & 1 & 0 \\ 1 & 3 & 1 \\ 0 & 1 & 2 \end{bmatrix} = \begin{bmatrix} 2\sigma^2 & 1\sigma^2 & 0 \\ 1\sigma^2 & 3\sigma^2 & 1\sigma^2 \\ 0 & 1\sigma^2 & 2\sigma^2 \end{bmatrix}
\end{aligned}$$

Solving with R, we get Σ^{-1}

$$\Sigma^{-1} = \begin{bmatrix} 0.625 & -0.25 & 0.125 \\ -0.250 & 0.50 & -0.250 \\ 0.125 & -0.25 & 0.625 \end{bmatrix}$$

So the weights become:

$$w^\star = \frac{\begin{bmatrix} 0.5 & 0.0 & 0.5 \end{bmatrix}}{1} = \begin{bmatrix} 0.5 & 0.0 & 0.5 \end{bmatrix}$$

which means that we split our portfolio equally between the 1st and 3rd assets. Notice that the R_1 and R_3 are uncorrelated, so splitting (or "diversifying") the portfolio between these uncorrelated assets will give the highest reduction in variance.

These questions encompass q5-15 on pg 489 from SDAFE 18.8

20.3 Q3

Suppose one has a sample of monthly log returns on two stocks with sample means of 0.0032 and 0.0074, sample variances of 0.017 and 0.025, and a sample covariance of 0.0059. For purposes of resampling, consider these to be the "true population values." A bootstrap resample has sample means of 0.0047 and 0.0065, sample variances of 0.0125 and 0.023, and a sample covariance of 0.0058.

20.3.1 Part A

Using the resample, estimate the efficient portfolio of these two stocks that has an expected return of 0.005; that is, give the two portfolio weights.

$$\begin{aligned}
0.005 &= \mathbb{E}(wS_1 + (1-w)S_2) = w \cdot 0.0047 + (1-w) \cdot 0.0065 \\
0.005 - 0.0065 &= w(0.0047 - 0.0065) \\
w &= \frac{0.005 - 0.0065}{(0.0047 - 0.0065)} \\
&= 0.8333 \\
1 - w &= 0.1667
\end{aligned}$$

For some reason, the answers state: (0.0047)w + (0.0065)(1 − w) = 0.005 so that the estimated efficient portfolio is 57.14% stock 1 and 42.86% stock 2.

20.3.2 Part B

What is the estimated variance of the return of the portfolio in part (a) using the resample variances and covariances?

$$\begin{aligned}
\mathbb{V}(wS_1 + (1-w)S_2) &= w^2\mathbb{V}(S_1) + (1-w)^2\mathbb{V}(S_2) + 2w(1-w)Cov(S_1, S_2) \\
&= w^2.0125 + (1-w)^2 \cdot 0.023 + 2w(1-w)0.0058 \\
&= 0.01092
\end{aligned}$$

20.3.3 Part C

What are the actual expected return and variance of return for the portfolio in (a) when calculated with the true population values (e.g., with using the original sample means, variances, and covariance)?

$$\begin{aligned}
\mathbb{E}(R_{act}) &= 0.8333 \cdot 0.0032 + .16667 \cdot 0.0074 = 0.0038 \\
\mathbb{E}(R_{act}) &= .8333^2 \cdot 0.0032 + 0.16667^2 \cdot 0.0074 + 2 \cdot 0.8333 \cdot 0.16667 \cdot 0.0059 \\
&= 0.00406
\end{aligned}$$

20.4 Q4

Straight from answer key

For this problem you will use regression to identify the composition of various mutual funds.

20.4.1 Part A

Download the adjusted daily closing prices from Jan 1 2020 to Dec 31 2022 for the 5 mutual funds below (use tseries::get.hist.quote() for each ticker):

- FCNTX: Fidelity Contrafund
- PIODX: Pioneer A
- AIVSX: American Funds Invmt Co of Amer A
- PRBLX: Parnassus Core Equity Investor
- VFIAX: Vanguard 500 Index Admiral

Note that each of these funds has at least 90% of their weight in the US stocks market. You can actually check the composition of the investment over different stock sectors from Yahoo Finance, under the fund's holdings tab; e.g. for FCNTX at https://finance.yahoo.com/quote/FCNTX/holdings.

```
library(tseries)
N = 5
funds = c("FCNTX", "PIODX", "AIVSX", "PRBLX", "VFIAX")

P = list()
for(i in 1:N){
  P[[i]] = tseries::get.hist.quote(funds[i],
                        start = as.Date("2020-01-01"),
                        end = as.Date("2022-12-31"),
                        quote="AdjClose",
                        compression = "d",
                        quiet = T)
}
R_mut=lapply(P, FUN = function(x){ diff(x) /stats::lag(x,-1) }) # calculate MF returns
m_funds = matrix(unlist(R_mut), ncol=N)
colnames(m_funds) = funds
```

20.4.2 Part B

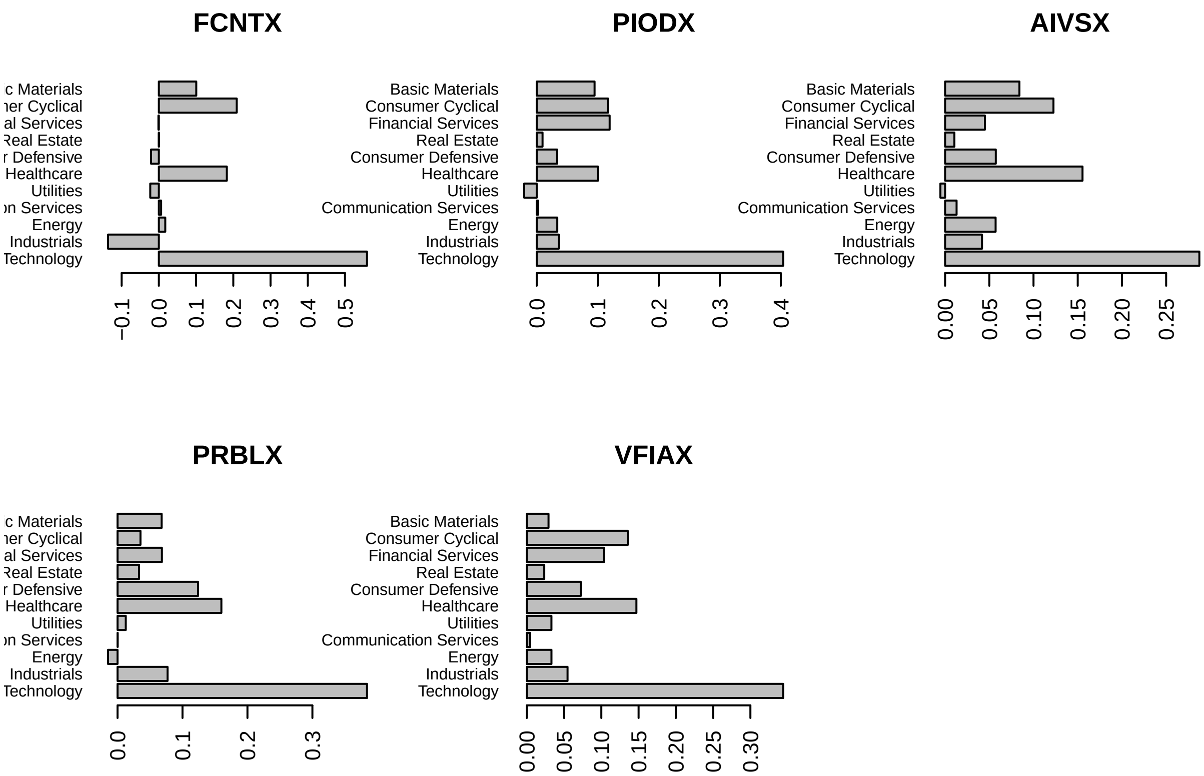
Assume you do not have any information about the investment strategy of the funds. Download the daily prices and calculate returns of the following EFTs, which track different sectors of the economy: - XLB: Basic Materials - XLY: Consumer Cyclical - XLF: Financial Services - VNQ: Real Estate - XLP: Consumer Defensive - XLV: Healthcare - XLU: Utilities - XTL: Communication Services - XLE: Energy - XLI: Industrials - XLK: Technology Regress each of the mutual fund returns on the above ETF returns and create barplots of the estimated beta coefficients. Do these accurately reflect the allocation over the different sectors (as described in Yahoo Finance)?

```
etfs = c("XLB", "XLY", "XLF", "VNQ", "XLP", "XLV", "XLU", "XTL","XLE", "XLI", "XLK")
sectors = c("Basic Materials","Consumer Cyclical","Financial Services",
            "Real Estate","Consumer Defensive", "Healthcare", "Utilities",
            "Communication Services", "Energy", "Industrials", "Technology")
M = length(etfs)
length(sectors)
```

```
## [1] 11
```

```
ETFS = list()
for(i in 1:M){
  ETFS[[i]] = tseries::get.hist.quote(etfs[i],
                        start = as.Date("2020-01-01"),
                        end = as.Date("2022-12-31"),
                        quote="AdjClose",
                        compression = "d",
                        quiet = T)
}
R_etf = lapply(ETFS, FUN=function(x){ diff(x) / stats::lag(x,-1)})
```

```r
m_etfs = matrix(unlist(R_etf), ncol = M)
colnames(m_etfs) = etfs

par(mfrow=c(2,3))
out=list();
for(i in 1:N){
  out[[i]]=lm(m_funds[,i] ~ m_etfs)
  weights = out[[i]]$coef[-1]
  barplot( rev(weights), names.arg = rev(sectors), main = funds[i],
          horiz = TRUE, las = 2, cex.names = .8)
}
```



The barplots of the regression coefficients (betas) roughly follow the sector weightings for each fund. Nevertheless, they are not always close in actual value (e.g. in some cases the betas are negative, even though weightings are positive). The differences can be due to the fact that we use ETFs as *proxies* for a sector, but the actual holding of the fund within the sector might be different. Moreover, there will be estimation error in our regression model, which is only based on the last year's returns.

Nevertheless, it is quite impressing that we can (approximately) identify the strategy of a fund, without knowing anything beyond its past returns. This approach works because of the linear formula for net portfolio returns:

$$R_p = w_1 R_1 + \cdots + w_N R_N$$

Regressing portfolio returns on other assets, we can estimate the weights, assuming the portfolio composition is constant.

20.4.3 Part C

Compare the performance of the mutual funds to that of a portfolio of ETFs by reporting the value of Jensen's alpha (based on the regressions from the previous part) and its corresponding p-value.

```
alpha=p.val=rep(0,N)
for(i in 1:N){
  alpha[i]=out[[i]]$coefficients[1]*250 # annualized Jensen alpha
  p.val[i]=summary(out[[i]])$coefficients[1,4] # first coef (alpha)'s p-val
}
cbind(alpha, p.val)
```

```
##         alpha p.val
## [1,] -0.03766 0.251
## [2,] -0.01396 0.557
## [3,] -0.02970 0.144
## [4,] -0.01708 0.458
## [5,] -0.00494 0.609
```

All the funds' alphas are negative, although their p-values are not very small. A likely cause for this is that funds charge a *fee* which consistently eats up some of the returns of their constituent assets. ETFs have typically lower fees than mutual funds, but our regression does not account for transaction costs (it is more costly to buy multiple assets than a single one), so the comparison is more nuanced.

Note that you can find Jensen alphas and other performance measures (e.g., Sharpe & Treynor Ratios) for assets in Yahoo! Finance under the risk tab. These metrics are based on the CAPM/Market factor model, by regressing the asset's returns on a proxy for the market return (e.g., S&P500).

Chapter 21

Problem Set 6

21.1 Q1

Find a closed form expression for the VaR at confidence level $(1-\alpha)$ of the following continuous loss distributions.

21.1.1 Part A

Pareto distribution with CDF$F_L(x) = 1-(x/m)^{-\beta}, x > m$, with shape parameter (tail index) $\beta > 0$.

$$\begin{aligned}\mathbb{P}(L \leq VaR_\alpha) &= 1-\alpha \\ 1-\left(\frac{VaR_\alpha}{m}\right)^{-\beta} &= 1-\alpha \\ \alpha^{-1/\beta}m &= VaR_\alpha\end{aligned}$$

Part B Gumbel distribution with CDF $F_L(x) = \exp\left\{-\exp\left\{-\frac{x-\mu}{\sigma}\right\}\right\}, \quad \forall x \in \mathbb{R}$, with mean parameter $\mu \in \mathbb{R}$ and scale parameter $\sigma > 0$.

$$\begin{aligned}\mathbb{P}(L \leq VaR_\alpha) &= 1-\alpha \\ \exp\left\{-\exp\left\{-\frac{VaR_\alpha-\mu}{\sigma}\right\}\right\} &= 1-\alpha \\ VaR_\alpha &= -\ln(\ln((1-\alpha)^{-1}))\cdot\sigma+\mu \\ &= \mu+\sigma\ln\left(\frac{1}{\ln\left(\frac{1}{1-\alpha}\right)}\right)\end{aligned}$$

Part C Fréchet distribution with CDF $F_L(x) = \exp\left\{-(x/\sigma)^{-\beta}\right\}, \quad \forall x > 0$, with shape parameter $\beta > 0$ and scale parameter $\sigma > 0$.

$$\begin{aligned}\mathbb{P}(L \leq VaR_\alpha) &= 1-\alpha \\ \exp\left\{-\left(\frac{VaR_\alpha}{\sigma}\right)^{-\beta}\right\} &= 1-\alpha \\ VaR_\alpha &= \sigma\ln\left(\frac{1}{1-\alpha}\right)^{-1/\beta}\end{aligned}$$

21.2 Q2

For a loss RV L with continuous distribution, show that the integral definition of Conditional Value-at-Risk (CVaR), a.k.a. Expected Shortfall (ES), is equivalent to the conditional expectation:

$$\frac{1}{\alpha}\int_0^\alpha \mathrm{VaR}_u\, du = \mathbb{E}\left[L \mid L \geq \mathrm{VaR}_\alpha\right]$$

(Hint: Recall that for absolutely continuous L with CDF $F(\cdot)$, the VaR is given by the inverse CDF: $\mathrm{VaR}_\alpha = F^{-1}(1-\alpha)$. Perform the change of variable $x = F^{-1}(1-u)$, using the fact that the derivative of the inverse of F is $\left[F^{-1}(x)\right]' = 1/F'\left(F^{-1}(x)\right) = 1/f\left(F^{-1}(x)\right)$, where $f = F'$ is the PDF.)

$$\begin{aligned}
\text{Let } x = F^{-1}(1-u), &\implies VaR_u = 1 - x \\
\frac{d}{du}x = [F^{-1}(1-u)]' &= \underbrace{-\frac{1}{F'(F^{-1}(1-u))} = -\frac{1}{f(x)}}_{\text{by definition}} \\
du &= f(x)dx \\
\frac{1}{\alpha}\int_0^\alpha VaR_u\, du &= \frac{1}{\alpha}\int_0^\alpha F^{-1}(1-u)\, du \\
&= \frac{1}{\alpha}\int_{F^{-1}(1-0)}^{F^{-1}(1-\alpha)} x f(x)\, dx \\
&= \int_\infty^{VaR_\alpha} x\frac{f(x)}{\alpha}\, dx \\
&= \int_\infty^{VaR_\alpha} x\underbrace{\frac{f(x)}{\mathbb{P}(L \geq VaR_\alpha)}}_{=\text{cond. PDF } f_{L|L\geq VaR_\alpha}}\, dx
\end{aligned}$$

Since we are now integrating from VaR_α to ∞, f(x) is still some pdf spanning $[-\infty, \infty]$, so we need to scale it appropriately by the total probability we are in the space we're integrating, ie $[VaR_\alpha, \infty]$.

$$\begin{aligned}
CVaR_\alpha &= \int_{VaR_\alpha}^\infty x f_{L|L\geq VaR_\alpha}(x)\, dx \\
&= \mathbb{E}(L|L \geq VaR_\alpha)
\end{aligned}$$

21.3 Q3

(Exercise 2 from 19.13 of SDAFE) Assume that the loss distribution has a polynomial tail with tail index $\alpha = 3.1$. If $VaR_{5\%} = 252$, what is $VaR(0.005)$? (Hint: Read section 19.6)

We need to use the complementary CDF of returns to obtain the CDF of losses.

A loss distribution with polynomial tail would look something like:

$$\bar{F}(x) = P(X > x) \propto \int_x^\infty s^{-(\alpha+1)}\, ds \propto [-s^{-\alpha}]_{s=x}^\infty = x^{-\alpha}$$

If we plug in $x = VaR_{5\%} = 5\%$ then we would have

$$P(X > VaR_{5\%}) = 5\% \implies 5\% \propto (VaR_{5\%})^{-\alpha} \implies VaR_{5\%} \propto \frac{1}{0.05^{1/\alpha}}$$

Given that $VaR_{5\%} = 252$, then

By definition,

$$P(R \leq y) \sim \int_{-\infty}^{y} f(u)\, du = \frac{A}{a} y^{-a} \text{ as } y \to -\infty$$

Which somehow means that:

$$\frac{P(R < y_0)}{P(R < -y_1)} \approx \left(\frac{y_0}{y_1}\right)^{-a}$$

$$\frac{VaR_p}{VaR_q} = \left(\frac{P(L \geq p)}{P(L \geq q)}\right)^{-1/a} = \left(\frac{P(L \geq q)}{P(L \geq p)}\right)^{1/a}$$

Using this definition, we have:

$$\frac{P(L \geq VaR_{0.005})}{P(L \geq VaR_{0.05})} = \left(\frac{VaR_{0.005}}{VaR_{0.05}}\right)^{-a}$$

$$\left(\frac{0.005}{0.05}\right)^{-1/a} = \frac{VaR_{0.005}}{252}$$

$$\left(\frac{0.05}{0.005}\right)^{1/a} = \frac{VaR_{0.005}}{252}$$

$$VaR_{0.005} = 10^{1/a} \cdot 252 = 10^{1/3.1} \cdot 252$$

21.4 Q4

Consider the example with the two risky zero-coupon bonds priced at \$95 per \$100 face value, where each has 4% default probability independently of the other.

21.4.1 Part A

Calculate the $\alpha = 5\%$ Entropic Value-at-Risk (EVaR) for one of these bonds. Note that you will need to use numeric minimization, e.g. optimize() in R, to find EVaR.

$$EVaR = inf_{z>0} \left\{ \ln\left(\frac{M_L(z)}{\alpha}\right) / z \right\}$$

The (marginal) loss distribution of each bond ($L_{1\ or\ 2}$) is given by the PMF

$$p_L(\ell) = \mathbb{P}(L = \ell) = \begin{cases} 0.04, & \ell = 95 - 0 = 95 \\ 0.96, & \ell = 95 - 100 = -5 \end{cases}$$

with MGF (not series expansion, simply expected formula as we have a simple PMF)

$$M_L(z) = \mathbb{E}\left[e^{zL}\right] = 0.04e^{95z} + 0.96e^{-5z}$$

The EVaR at α is given by

$$\begin{aligned} EVaR_\alpha &= \inf_{z>0} \left\{ \ln \left(M_L(z)/\alpha \right)/z \right\} \\ &= \inf_{z>0} \left\{ \ln \left\{ \left(0.04e^{95z} + 0.96e^{-5z} \right)/0.05 \right\}/z \right\} \end{aligned}$$

Running this minimization w.r.t. z in R,

```
fn = function(z){ log( ( 0.04 * exp( 95*z ) + 0.96 * exp(-5*z) )
               / 0.05 ) / z }
optimise(fn, c(0,1) )
```

> z must lie between $[0, 1]$, as the MGF of any value greater than 1 results in $\mathbb{E}[e^{zL}] \to \infty$ since e^{zL} would become an increasing function.

gives

```
$minimum
[1] 0.06690106

$objective
[1] 92.10402
```

Which leaves us with $EVaR_{0.05}(L) = 92.10402$.

21.4.2 Part B

Calculate the EVaR of a portfolio of two of these bonds, and show that it is sub-additive.

The loss distribution for the sum of the two bonds $(L_1 + L_2)$ is

$$p_{L_1+L_2}L(\ell) = \mathbb{P}\left(L_1 + L_2 = \ell\right) = \begin{cases} (0.04)^2 = 0.0016, & \ell = 95 + 95 = 190 \\ 2(0.96)(0.04) = .0768, & \ell = 95 - 5 = 90 \\ (0.96)^2 = .9216, & \ell = -5 - 5 = -10 \end{cases}$$

with MGF

$$M_{L_1+L_2}(z) = \mathbb{E}\left[e^{z(L_1+L_2)}\right] = 0.0016e^{190z} + 0.0768e^{90z} + 0.09216e^{-10}$$

Running this minimization w.r.t. z in R

```
fn = function(z){ log( ( 0.0016 * exp( 190*z ) +
       0.0768 * exp(90*z) + 0.9216 * exp(-10*z)) / 0.05 ) / z }
optimise(fn, c(0,1) )

$minimum
[1] 0.03841828
$objective
[1] 122.0294
```

This shows that

$$\begin{gathered} EVaR_{0.05}(L_1 + L_2) \leq EVaR_{0.05}(L_1) + EVaR_{0.05}(L_2) \\ 122.0294 \leq 2 \times 92.10402 = 184.208 \end{gathered}$$

21.5 Q5

Consider a loss distribution with fat upper tail and some tail index $\beta > 0$.

21.5.1 Part A

Can you find an EVaR for such distributions? Justify your answer.

> You cannot, as to calculate EVaR you need the MGF, and the MGF is not defined for fat tailed distributions. Specifically, moment i, where $i \geq \beta$ are infinite.

21.5.2 Part B

If the tail index is $\beta = 1$ (e.g., a Cauchy distribution), can you find CVaR/ES for such distributions? Justify your answer.

> Because $\beta = 1$, that means the first moment and beyond are infinite. As CVaR calculates the expected value of the distribution given our losses are already larger than some number, this calculation will not be possible with tail index 1.
>
> We also know that conditional distributions like the one used in CVaR are proportional to the original distribution, which means that the mean is also proportional to infinite.

21.6 Q6

(Exercise 5 from §19.13 of SDAFE) Suppose the risk measure is VaR for some α. Let P_1, P_2 be two portfolios whose returns have a joint (2D) normal distribution with means μ_1, μ_2, standard deviations σ_1, σ_2, and correlation ρ. Suppose the initially investments are S_1, S_2. Show that $\mathrm{VaR}\left(P_1 + P_2\right) \leq \mathrm{VaR}_\alpha\left(P_1\right) + \mathrm{VaR}_\alpha\left(P_2\right), \forall \alpha < 1/2$, i.e. that VaR is sub-additive in this case, for $\alpha < 1/2$.

> Copied all this cuz no way i'd figure this out myself...

Solution: Since $P_i \sim N\left(\mu_i, \sigma_i^2\right)$ with initial investment amounts S_i for $i = 1, 2$, we have that $\mathrm{VaR}_\alpha\left(P_i\right) = -S_i \times \left(\mu_i + \sigma_i z_\alpha\right) = -S_i\mu_i - S_i\sigma_i z_\alpha, \forall i = 1, 2$, where $z_\alpha = \Phi^{-1}(\alpha)$ is the standard Normal quantile function (i.e., $\Phi(\cdot)$ is the standard Normal CDF). Thus,

$$\mathrm{VaR}_\alpha\left(P_1\right) + \mathrm{VaR}_\alpha\left(P_2\right) = -\left(S_1\mu_1 + S_2\mu_2\right) - \left(S_1\sigma_1 + S_2\sigma_2\right) z_\alpha$$

> Just from the VaR definition, now we look at the portfolio

Looking at the combined portfolio, with weights $w_i = \frac{S_i}{S_1+S_2}, \quad i = 1, 2$ proportional to the investment amounts in each stock, we have:

$$w_1P_1 + w_2P_2 \sim N\left(w_1\mu_1 + w_2\mu_2, w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\sigma_1\sigma_2\rho\right)$$

> The portfolio distribution by statistics

The resulting VaR is

$$\begin{aligned}
&\mathrm{VaR}_{\alpha}\left(P_{1}+P_{2}\right)= \\
&=-\left(S_{1}+S_{2}\right)\times\left(\left(w_{1}\mu_{1}+w_{2}\mu_{2}\right)+\sqrt{w_{1}^{2}\sigma_{1}^{2}+w_{2}^{2}\sigma_{2}^{2}+2w_{1}w_{2}\sigma_{1}\sigma_{2}\rho}\times z_{\alpha}\right) \\
&=-\left(S_{1}\mu_{1}+S_{2}\mu_{2}\right)-\sqrt{S_{1}^{2}\sigma_{1}^{2}+S_{2}^{2}\sigma_{2}^{2}+2S_{1}S_{2}\sigma_{1}\sigma_{2}\rho}\times z_{\alpha}
\end{aligned}$$

Again using the definition of VaR

But for any $\rho \in [-1,+1]$, we have:

$$\begin{aligned}
S_{1}^{2}\sigma_{1}^{2}+S_{2}^{2}\sigma_{2}^{2}+2S_{1}S_{2}\sigma_{1}\sigma_{2}\rho &\leqslant S_{1}^{2}\sigma_{1}^{2}+S_{2}^{2}\sigma_{2}^{2}+2S_{1}S_{2}\sigma_{1}\sigma_{2} \\
&=\left(S_{1}\sigma_{1}+S_{2}\sigma_{2}\right)^{2} \\
\Rightarrow \sqrt{S_{1}^{2}\sigma_{1}^{2}+S_{2}^{2}\sigma_{2}^{2}+2S_{1}S_{2}\sigma_{1}\sigma_{2}\rho} &\leqslant S_{1}\sigma_{1}+S_{2}\sigma_{2} \quad (\text{ for } \alpha<.5\rightarrow z_{\alpha}<0) \\
\Rightarrow -\sqrt{S_{1}^{2}\sigma_{1}^{2}+S_{2}^{2}\sigma_{2}^{2}+2S_{1}S_{2}\sigma_{1}\sigma_{2}\rho}\times z_{\alpha} &\leqslant -\left(S_{1}\sigma_{1}+S_{2}\sigma_{2}\right)\times z_{\alpha}
\end{aligned}$$

$$\begin{aligned}
\Rightarrow -\left(S_{1}\mu_{1}+S_{2}\mu_{2}\right)z_{\alpha}-\sqrt{S_{1}^{2}\sigma_{1}^{2}+S_{2}^{2}\sigma_{2}^{2}+2S_{1}S_{2}\sigma_{1}\sigma_{2}\rho}\times z_{\alpha} &\leqslant \\
-\left(S_{1}\mu_{1}+S_{2}\mu_{2}\right)\times z_{\alpha}-\left(S_{1}\sigma_{1}+S_{2}\sigma_{2}\right)\times z_{\alpha}
\end{aligned}$$

$$\Rightarrow \mathrm{VaR}_{\alpha}\left(P_{1}+P_{2}\right)\leqslant \mathrm{VaR}_{\alpha}\left(P_{1}\right)+\mathrm{VaR}_{\alpha}\left(P_{2}\right)$$

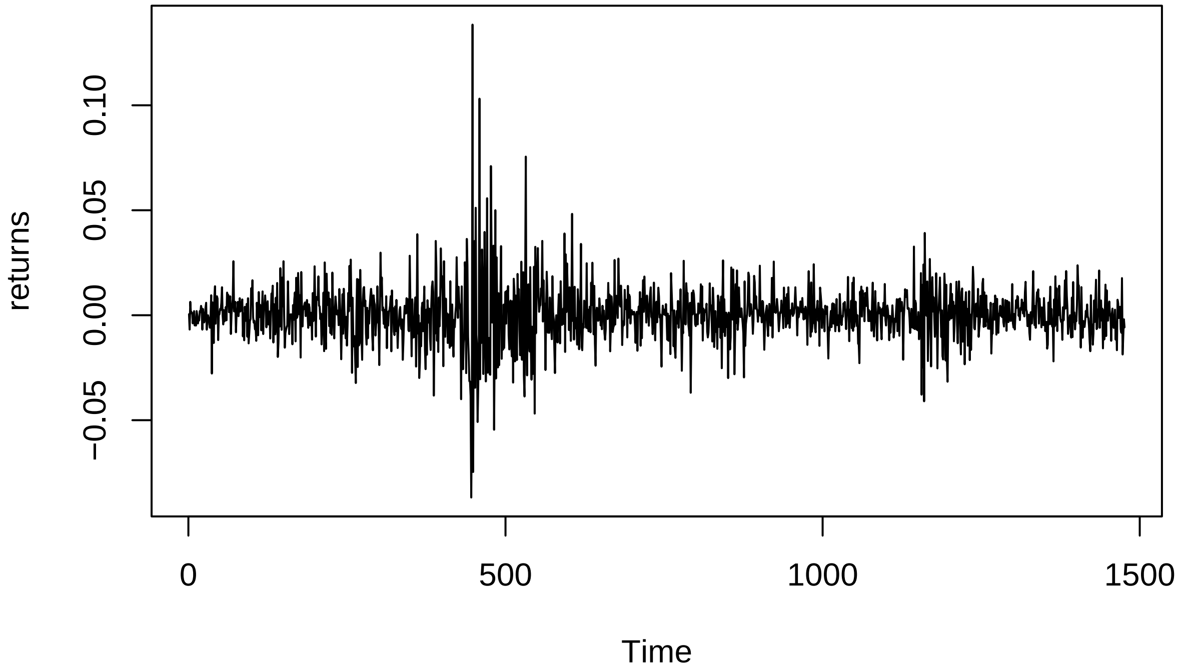
21.7 Q7

For an investment of \$4,000, what are estimates of $VaR^t(0.05)$ and $ES^t(0.05)$?

Now, fit a ARMA(0,0)+GARCH(1,1) model to the returns and calculate one step forecasts.

```
CokePepsi = read.table("ProblemSets/datasets/CokePepsi.csv", header=T)
```

```
price = CokePepsi[,1]
returns = diff(price)/lag(price)[-1]
ts.plot(returns)
```




```
S=4000
alpha = 0.05
library(MASS)
res = fitdistr(returns,'t')
mu = res$estimate['m']
lambda = res$estimate['s']
nu = res$estimate['df']
qt(alpha, df=nu)
```

```
## [1] -2.33
```

```
dt(qt(alpha, df=nu), df=nu)
```

```
## [1] 0.0465
```

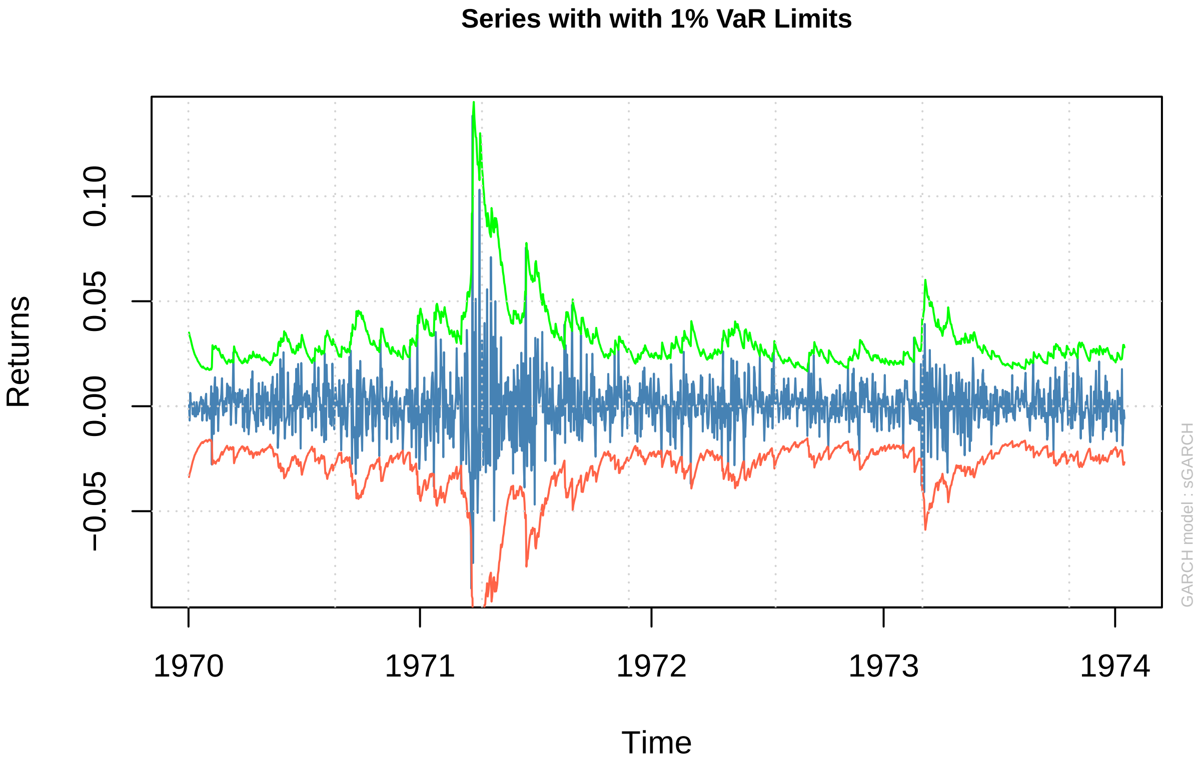
```
library(fGarch) # for qstd() function
library(rugarch)
garch.t = ugarchspec(mean.model=list(armaOrder=c(0,0)),
variance.model=list(garchOrder=c(1,1)),
distribution.model = "std")
KO.garch.t = ugarchfit(data=returns, spec=garch.t)
show(KO.garch.t)
```

```
##
## *---------------------------------*
## *          GARCH Model Fit        *
## *---------------------------------*
##
## Conditional Variance Dynamics
## -----------------------------------
## GARCH Model  : sGARCH(1,1)
## Mean Model   : ARFIMA(0,0,0)
## Distribution : std
##
## Optimal Parameters
## ------------------------------------
##         Estimate  Std. Error  t value Pr(>|t|)
## mu      0.000675    0.000240  2.81396 0.004894
## omega   0.000003    0.000004  0.79633 0.425839
## alpha1  0.093766    0.035130  2.66910 0.007606
## beta1   0.892222    0.038780 23.00727 0.000000
## shape   5.890859    1.097148  5.36925 0.000000
##
## Robust Standard Errors:
##         Estimate  Std. Error  t value Pr(>|t|)
## mu      0.000675    0.000257   2.6235 0.008704
## omega   0.000003    0.000016   0.1717 0.863675
## alpha1  0.093766    0.149882   0.6256 0.531575
## beta1   0.892222    0.171120   5.2140 0.000000
## shape   5.890859    2.499085   2.3572 0.018413
##
## LogLikelihood : 4596
##
## Information Criteria
## ------------------------------------
##
## Akaike       -6.2209
## Bayes        -6.2030
## Shibata      -6.2210
## Hannan-Quinn -6.2143
##
## Weighted Ljung-Box Test on Standardized Residuals
## ------------------------------------
##                         statistic p-value
## Lag[1]                     0.8041  0.3699
## Lag[2*(p+q)+(p+q)-1][2]    0.8718  0.5417
## Lag[4*(p+q)+(p+q)-1][5]    2.1466  0.5842
## d.o.f=0
## H0 : No serial correlation
##
## Weighted Ljung-Box Test on Standardized Squared Residuals
## ------------------------------------
##                         statistic p-value
## Lag[1]                     0.9099  0.3401
```

```
## Lag[2*(p+q)+(p+q)-1][5]    5.1826  0.1391
## Lag[4*(p+q)+(p+q)-1][9]    6.7262  0.2232
## d.o.f=2
##
## Weighted ARCH LM Tests
## ------------------------------------
##             Statistic Shape Scale P-Value
## ARCH Lag[3]    0.1194 0.500 2.000  0.7297
## ARCH Lag[5]    0.8386 1.440 1.667  0.7814
## ARCH Lag[7]    1.9592 2.315 1.543  0.7260
##
## Nyblom stability test
## ------------------------------------
## Joint Statistic:  15.8
## Individual Statistics:
## mu     0.05684
## omega  2.24878
## alpha1 0.28907
## beta1  0.12949
## shape  0.14088
##
## Asymptotic Critical Values (10% 5% 1%)
## Joint Statistic:          1.28 1.47 1.88
## Individual Statistic:     0.35 0.47 0.75
##
## Sign Bias Test
## ------------------------------------
##                    t-value    prob sig
## Sign Bias            1.040 0.29849
## Negative Sign Bias   1.760 0.07857   *
## Positive Sign Bias   1.041 0.29825
## Joint Effect         4.374 0.22380
##
##
## Adjusted Pearson Goodness-of-Fit Test:
## ------------------------------------
##   group statistic p-value(g-1)
## 1    20     20.23       0.3807
## 2    30     26.52       0.5976
## 3    40     35.17       0.6455
## 4    50     41.55       0.7662
##
##
## Elapsed time : 0.489
```

```
plot(KO.garch.t, which = 2)
```

```
##
## please wait...calculating quantiles...
```




```
pred = ugarchforecast(KO.garch.t, data=returns, n.ahead=1) ; pred
```

```
##
## *------------------------------------*
## *       GARCH Model Forecast         *
## *------------------------------------*
## Model: sGARCH
## Horizon: 1
## Roll Steps: 0
## Out of Sample: 0
##
## 0-roll forecast [T0=1974-01-15 19:00:00]:
##        Series   Sigma
## T+1 0.0006751 0.01038
```

```
fitted(pred) ; sigma(pred)
```

```
##     1974-01-15 19:00:00
## T+1            0.000675
```

```
##     1974-01-15 19:00:00
## T+1              0.0104
```

```
nu = as.numeric(coef(KO.garch.t)[5])
q = qstd(alpha, mean = fitted(pred), sd = sigma(pred), nu = nu) ; q
```

```
##     1974-01-15 19:00:00
## T+1             -0.0158
```

```
sigma(pred)/sqrt( (nu)/(nu-2) )
```

```
##     1974-01-15 19:00:00
## T+1             0.00844
```

```
qt(alpha, df=nu)
```

```
## [1] -1.95
```

```
dt(qt(alpha, df=nu), df=nu)
```

```
## [1] 0.0688
```

```
mu = as.numeric(res$estimate['m'])
lambda = as.numeric(res$estimate['s'])
nu = as.numeric(res$estimate['df'])
qt(alpha, df=nu)
```

```
## [1] -2.33
```

```
# [1] -2.292
```

```
dt(qt(alpha, df=nu), df=nu)
```

```
## [1] 0.0465
```

```
# [1] 0.048
```

```
Finv = mu + lambda * qt(alpha, df=nu)
VaR = -S * Finv
options(digits=4)
VaR
```

```
## [1] 75.32
```

```
# [1] 75.31
```

```
den = dt(qt(alpha, df=nu), df=nu)
ES = S * (-mu + lambda*(den/alpha) * (nu+qt(alpha, df=nu)^2 )/(nu-1))
ES
```

```
## [1] 124
```

```
# [1] 122.1
```

Chapter 22

W2021 Midterm

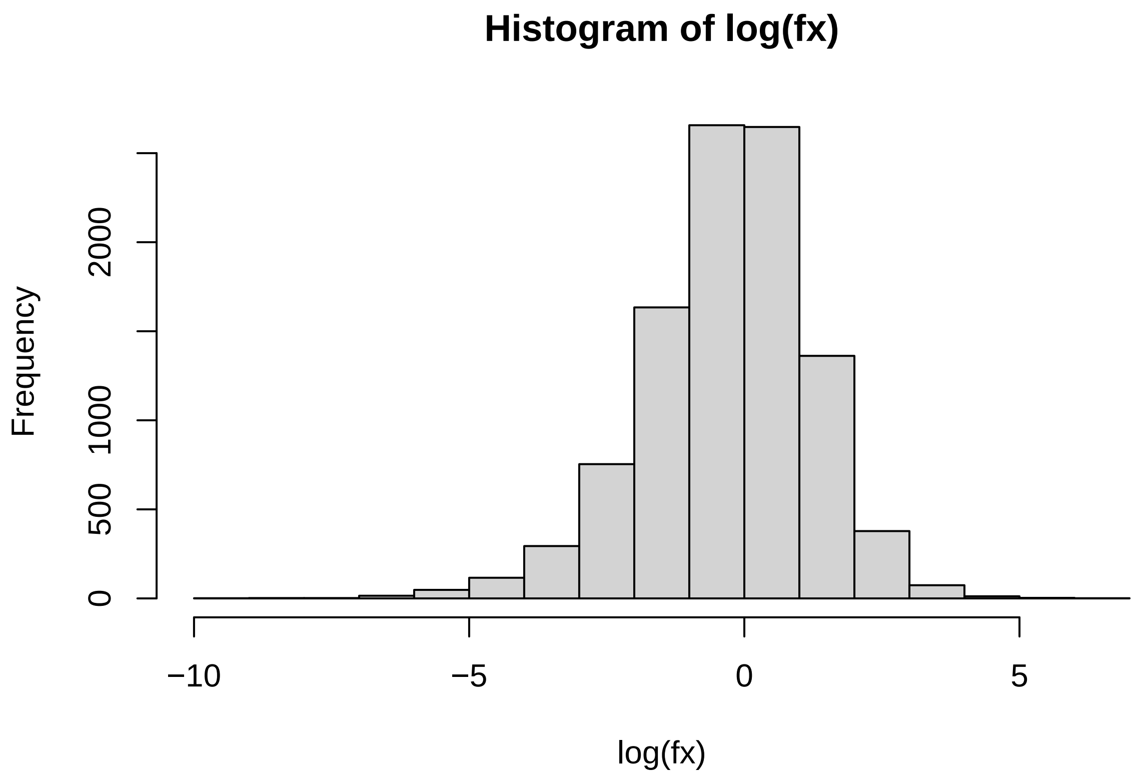
22.1 Q2

For this question you will generate GPD random variates and verify that their conditional excess distribution is also GPD.

22.1.1 Part A

(a) [10 points] Simulate $n = 10,000$ random variates from a GPD($\gamma = .5, \sigma = 1$) distribution using the inverse CDF method from Q1.(d), and plot the histogram of their log-values.

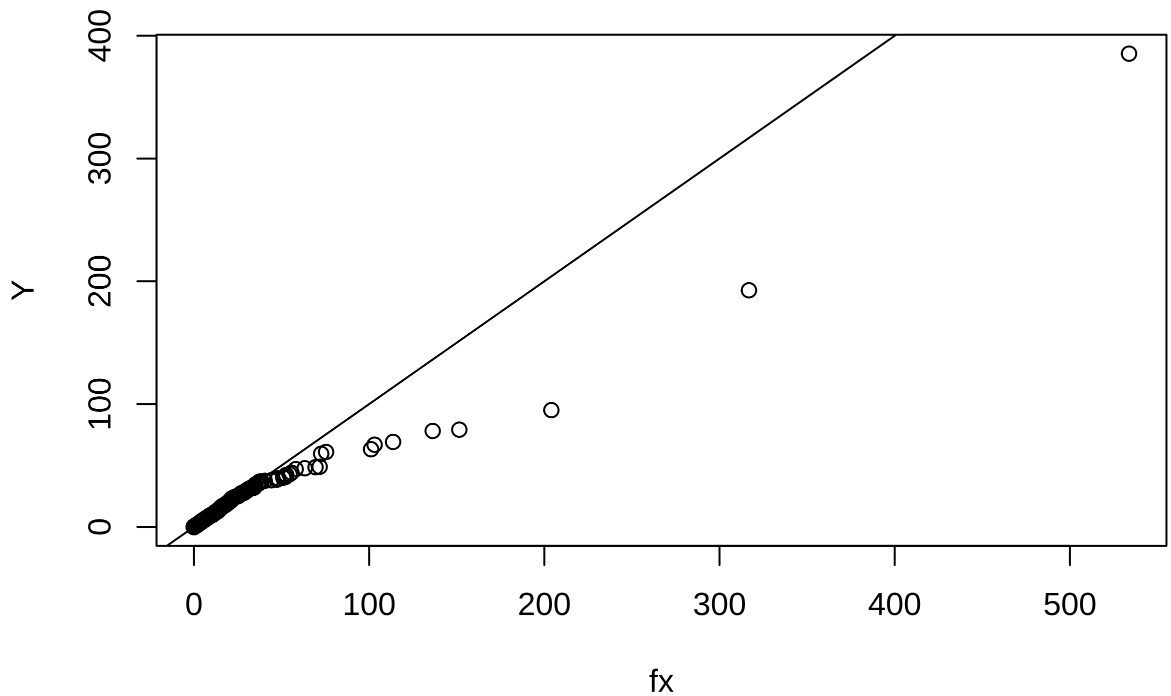
```
u = runif(10000)
fx = 1/0.5*((1-u)^(-0.5)-1)
hist(log(fx))
```



22.1.2 Part B

(b) [15 points] Simulate another 10,000 random variates from GPD($\gamma = .5, \sigma = 1$), but now use the mixture method from Q1.(e). Create a QQ-plot of the two sets of variates (from this and the previous part0), and comment whether they seem to come from the same distribution? (Hint: Use qqplot (X, Y), where X, Y are the values you generated.)

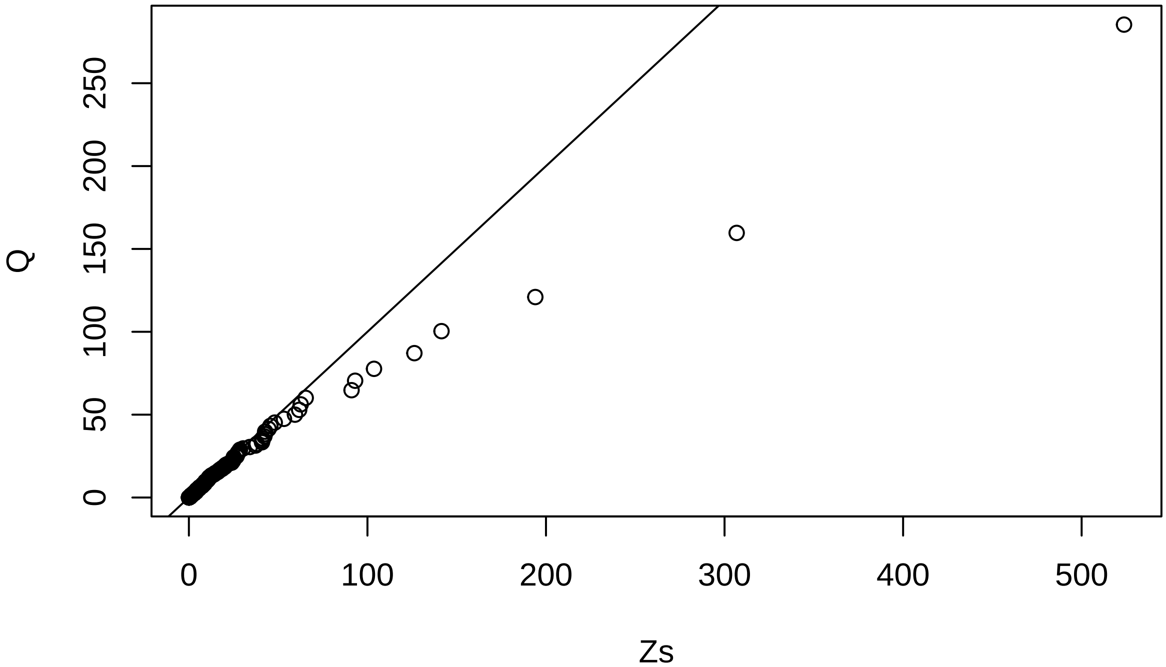
```
rg = rgamma(10000,shape=2, rate = 2)
Y = rexp(10000, rate=rg)
qqplot(fx, Y)
abline(a=0,b=1)
```



22.1.3 Part C

(c) [15 points] Take the values from Q2.(a) and calculate their excedances above 10, i.e. $Z = X - 10$ for values of X that are greater than 10. From Q1.(b), we know that Z must follow $GPD(\gamma, \sigma + 10\gamma)$. Verify that the exceedances follow this distribution, by creating a QQ-plot of Z-values versus their theoretical quantiles.

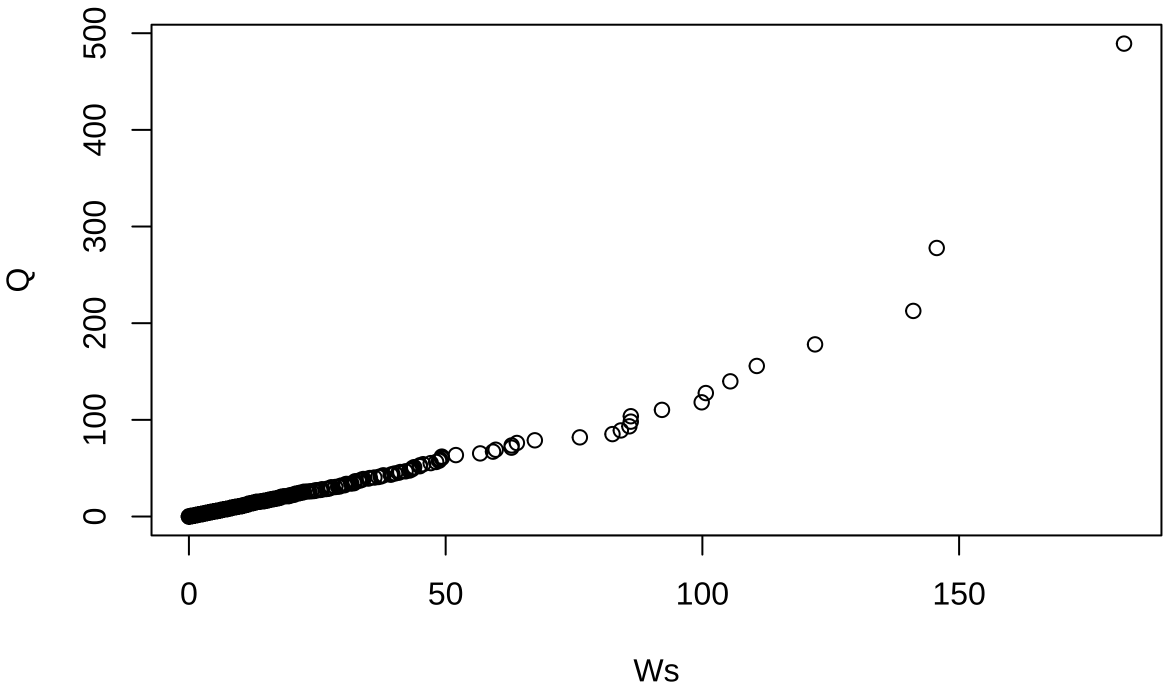
```
sigma = 1
gamma = 0.5
u = 10
Z = (fx - u)[fx-u>0]

Zs = sort(Z) #need sorting for qqplot
nZ= length(Zs)
sigma_u = sigma + gamma * u
Q = sigma_u /gamma * ( (1 - ppoints(nZ) )^(-gamma) - 1 )
plot( Zs, Q ); abline(0,1)
```



22.1.4 Part D

(d) [10 points] Finally, verify through simulation that conditional exceedances of fattail distributions with tail index α approach a GPD with $\gamma = 1/\alpha$. Generate $n = 10,000$ values from a $t(df = 4)$ distribution (i.e. $\alpha = 4$) and take their absolute value (to avoid wasting values by symmetry). Calculate the exceedances above 10 again, and create their QQ-plot versus quantiles from a GPD($\gamma = 1/\alpha, \sigma = 1$). (Note: the points in your QQ-plot should lie close to a straight line, but the slope does not need to be equal to 1 because of the arbitrary GPD scale $\sigma = 1$).

```
# t = abs(rt(10000,4))

W = abs( rt( 100000, 2 ) ) # why 2?

Ws = (W - u)[W-u>0]

Ws = sort(Ws)
nW= length(Ws)
Q = (1+u) * ( (1 - ppoints(nW) )^(-gamma) - 1 )
qqplot( Ws, Q )
```


Q
0
100
200
300
400
500
0
50
100
150
Ws

Chapter 23

W2020 Midterm

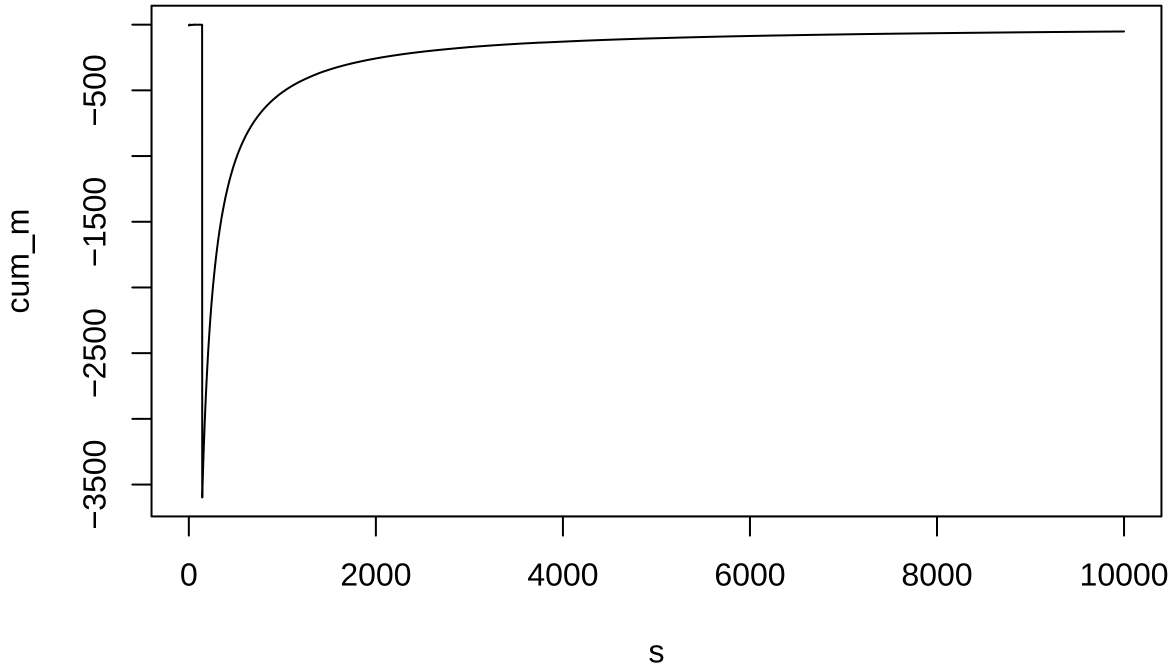
23.1 Q3

You have to perform a simulation experiment to confirm that the t(1) or Cauchy distribution is stable, but not all heavy-tailed distributions are.

23.1.1 Part A

[4 points] Simulate 10, 000 Cauchy random variates and plot their cumulative mean versus the sample size, as in the plot below.

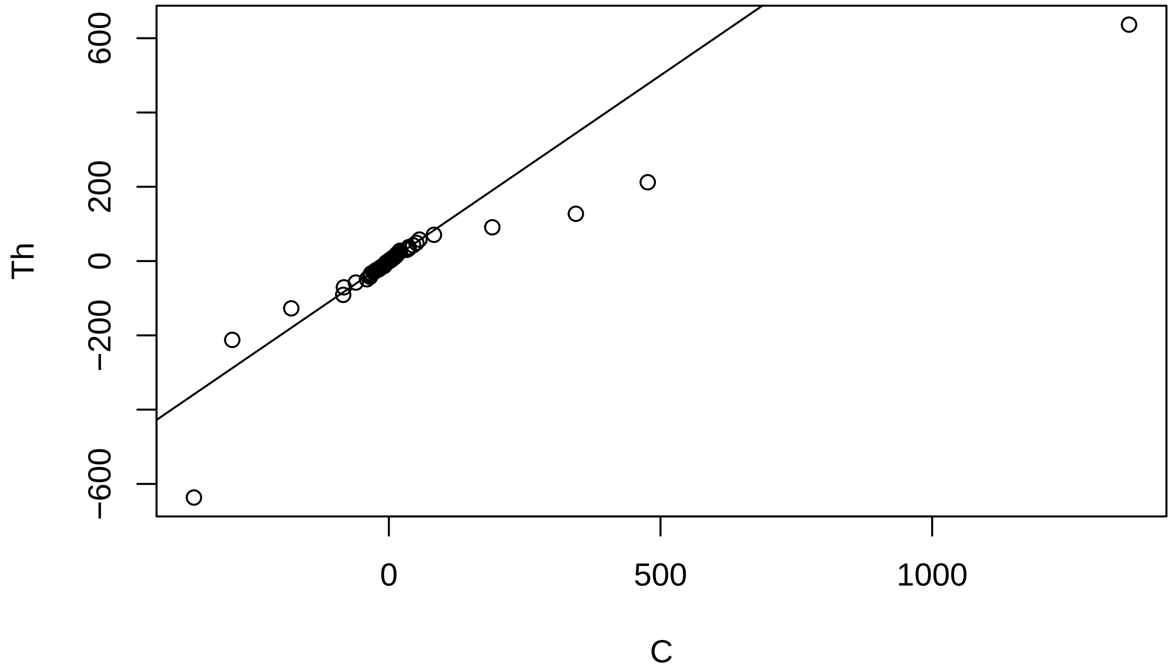
```
n = 10000
s = 1:n
X = rcauchy(n)
cum_m = cumsum(X)/s
plot(x=s, y = cum_m, type="l")
```



23.1.2 Part B

[8 points] Repeat the following experiment N = 1, 000 times: simulate n = 100 Cauchy variates and calculate their mean (in the end you should have 1,000). From question 1, we know these means should be also Cauchy distributed. Verify this by creating a QQ-plot of the simulated values versus their theoretical quantiles, and comment on the fit. Hint: Use qcauchy and ppoints() to generate theoretical quantiles from the Cauchy distribution.)

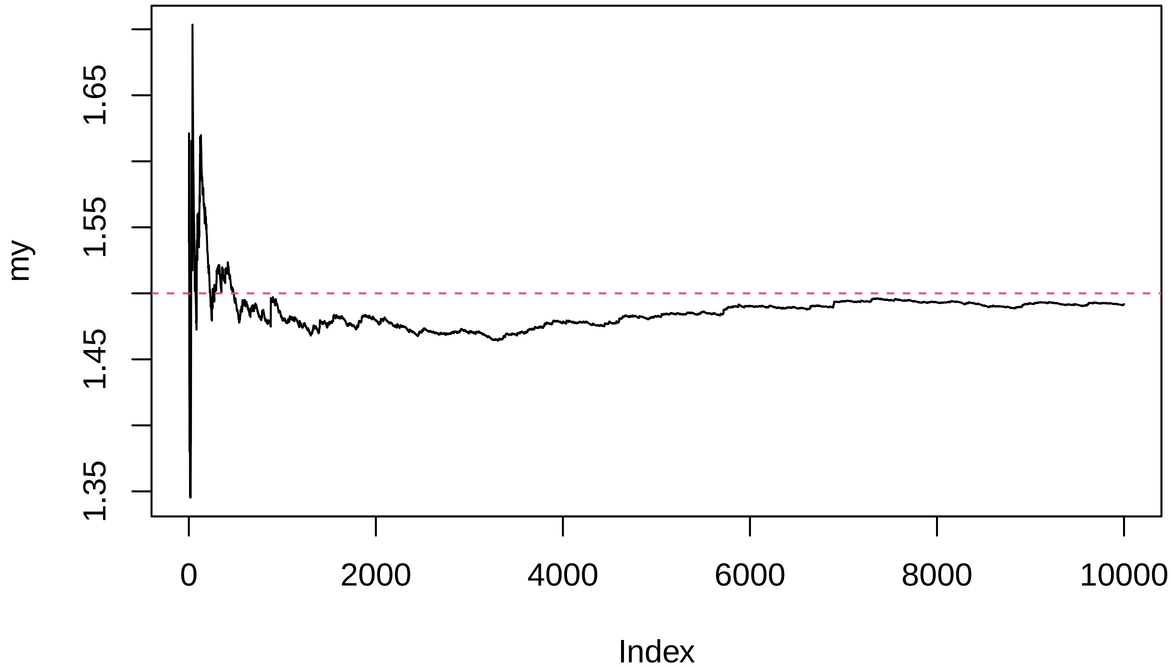
```
library(tidyverse)
N=1000
n = 100
C = rcauchy(n*N) %>% matrix(ncol=N) %>% colSums()/n %>% sort()
Th = qcauchy(ppoints(N))
qqplot(C, Th); abline(0,1)
```



23.1.3 Part C

[8 points] Now consider the Pareto distribution, with CDF FX(x) $= 1-(1/x)^{\alpha}, \forall x > 1$. Use the inverse CDF method to generate 10,000 Pareto random variates with tail index $\alpha = 3$, and plot their cumulative mean vs sample size, as in part (a).

```
u = runif(10000)
y = (1 - u)^(-1/3)
my = cumsum(y)/(1:10000)
plot(my, type = 'l' )
abline(h=1.5, col = 2, lty=2)
```



23.2 Q4

Consider the 3 -factor model for the returns of two assets:

$$\left[\begin{array}{c} R_1 \\ R_2 \end{array}\right]=\left[\begin{array}{c} .03 \\ .04 \end{array}\right]+\left[\begin{array}{ccc} .1 & .2 & .3 \\ .3 & .2 & .1 \end{array}\right]\left[\begin{array}{c} F_1 \\ F_2 \\ F_3 \end{array}\right]+\left[\begin{array}{c} e_1 \\ e_2 \end{array}\right] \Leftrightarrow R=\mu+\beta^{\top} F+e$$

where $F \sim N_3(\mathbf{0}, I)$ and $e \sim N_2\left(\mathbf{0}, \Sigma_e\right)$ with $\Sigma_e=\left[\begin{array}{cc} .1 & 0 \\ 0 & .2 \end{array}\right]$ and $F \perp e$

23.2.1 Part A

[5 points] Calculate the variance-covariance matrix of the returns $\Sigma_R=\mathbb{V}[R]$.

```
beta = matrix( c(1,2,3,3,2,1)/10, nrow = 2, byrow = TRUE )
S_e = diag( c(.1,.2) )
(S_R = beta %*% t(beta) + S_e)
```

```
##      [,1] [,2]
```

```
## [1,] 0.24 0.10
## [2,] 0.10 0.34
```

23.2.1.1 Part B

[5 points] Calculate the mean and variance of the return of the minimum-variance portfolio.

```
Si = solve(S_R)
w = rowSums(Si)/sum(Si)
mu = c(.03, .04)

mvp_mu = sum(w*mu)
(mvp_s =  t(w) %*% S_R %*% w) # also equal to 1/sum(Si)
```

```
##        [,1]
## [1,] 0.1884
```

23.2.2 Part C

[10 points] Assume the risk-free rate is .015, and find the mean and variance of the tangency portfolio.

```
rf = .015
w_tp = Si %*% (mu - rf) / sum( Si %*% (mu - rf) )

sum( w_tp * mu ) # tangency portfolio mean
```

```
## [1] 0.03634
```

```
t(w_tp) %*% S_R %*% w_tp # tangency portfolio variance
```

```
##        [,1]
## [1,] 0.2152
```

23.2.3 Part D

[10 points] Assume you could eliminate one of the factors (i.e. set $\mathbb{V}[F_i] = 0$). Which factor would you choose to eliminate in order to further reduce the variance of the minimum- variance portfolio? (justify your answer with a calculation & comparison)

```
mvp_vars = rep(0,3)

for(i in 1:3){
  beta_t = beta[,-i]
  S_t = beta_t %*% t(beta_t) + S_e
  Si_t = solve(S_t)
  mvp_vars[i] = 1/sum(Si_t)
}

which.min(mvp_vars)
```

```
## [1] 3
```

Chapter 24

Cheatsheet

24.1 Probability

$$\text{Variance: } \mathbb{V}(Z) = \mathbb{E}(Z^2) - \mathbb{E}(Z)^2$$

$$\text{Lin Com of RV: } Z = wA + (1-w)B$$
$$\mathbb{E}(Z) = \theta \implies w = \frac{\theta - \mu_B}{\mu_A - \mu_B}$$
$$Var(Z) = w^2 Var(A) + (1-w)^2 Var(B) - 2 \underbrace{Cov(A,B)}_{Cor(A,B)\cdot\sigma(A)\cdot\sigma(B)}$$

MLE Estimator for parameter θ in $f(x)$:

$$L(\theta) = \prod_{i=1}^{n} f(x;\theta)$$
$$l(\theta) = \sum_{i=1}^{n} f(x;\theta)$$
$$\text{set } \frac{\partial l}{\partial \theta} = l'(\theta) = 0, \text{solve for } \theta$$

Moment Generating Functions Characteristic functions

$$\text{MGF:} m_X(t) = \mathbb{E}(e^{tX}) = \sum_{k=1}^{\infty} \frac{t^k}{k!} \mathbb{E}[X^k] = \int_{-\infty}^{\infty} e^{tx} f_X(x)\, dx$$
$$\text{Char Fnc: } \phi_X(t) = \mathbb{E}(e^{itX}) = \mathbb{E}[\cos(tX) + i\sin(tX)]$$
$$\text{Let } X \sim \text{Stable, sym about 0} \implies CF(X) = \phi(t) = e^{-c|t|^{\alpha}}$$

Normal distribution

$$\phi(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}$$

Facts:

- Summing i.i.d RVs with **infinite** variance → Heavy-tailed Distributions

- Summing i.i.d RVs with **finite** variance $\rightarrow$ Normal Distributions
- Summing i.i.d **STABLE** random variables $\rightarrow$ Stable Distribution
- Summing i.i.d heavy-tail ($0 < \alpha < 2$) random variables $\rightarrow$ Stable Distribution
- Finite mixtures: It suffices to show (due to induction) that a convex combination of 2 copulas ($C() = wC_1() + (1-w)C_2, \quad \forall w \in [0$ is also copula. We are given that a convex combination of two CDFs is a CDF, and this result holds for any number of dimensions. The last piece is to show that the marginals are Uniform $(0,1)$. Since the initial copulas have Uniform $(0,1)$ marginals, it is straightforward to see that that mixture copula will also have Uniform $(0,1)$ marginals: for 1D Uniforms CDFs $F_1(u) = F_2(u) = u, \quad \forall u \in [0,1]$, we have $F(u) = wF_1(u) + (1-w)F_2(u) = wu + (1-w)u = u, \quad \forall u \in [0,1]$

24.2 Math

$$\int_a^b \ln(x)\, dx = [x\ln(x) - x]_{x=a}^b$$

24.3 Series

24.3.1 Integrated series

A variable that is not stationary but it's differences are, is called an integrated series. $\{X_t\}$ is not stationary but $\{\nabla X_t = X_t - X_{t-1}\}$ is.

- E.g. If log returns are stationary, then the log prices are integrated

24.3.2 Cointegration

Consider two integrated series $\{X_t, Y_t\}$ which behave as random walks, but if they seem to have some constant (stationary) relationship when linearly combined then they are called **cointegrated**.

$$\exists\ \alpha\ s.t.\ X_t + \alpha Y_t \sim Stationary$$

24.3.3 Brownian Motion

Brownian Motion (BM) forms the building block of continuous stochastic models Standard BM $\{W_t\}$ is such that

$$W_0 = 0 \quad \& \quad (W_t - W_s) \mid W_s \sim N(0, t-s)$$
$$Cov(W_s, W_t) = min(s,t)$$
$$(W_s - W_t) \sim N(0, (s-t)) \quad \forall t < s$$

Using the inverse CDF method to generate random variates:

$$X \sim Pareto \implies F(x) = 1 - \left(\frac{1}{x}\right)^\alpha = u \in [0,1]$$
$$P(X \le x) = P(1 - \left(\frac{1}{U}\right)^\alpha \le x) = P(U \le (1-x)^{-1/\alpha})$$
$$\implies F^{-1}(u) = (1-u)^{-1/\alpha} = x$$

```
u = runif(n)
y = (1 - u)^(-1/3)
```

24.3.4 Arithmetic BM/RW

Arithmetic BM (ABM) $\{X_t\}$ with drift μ & volatility σ is

$$X_0 = 0 \quad \& \quad (X_t - X_s) \mid X_s \sim N\left(\mu(t-s), \sigma^2(t-s)\right)$$

- In form of Stochastic Differential Equation (SDE)

$$dX_t = \mu dt + \sigma dW_t \Leftrightarrow X_t - X_0 = \mu t + \sigma\left(W_t - W_0\right)$$

24.3.5 Geometric Brownian Motion

A transformation of the arithmetic brownian motion. We use this to avoid negative values.

Process $\{S_t\}$ whose logarithm follows ABM

$$\log\left(S_t\right) - \log\left(S_0\right) = \log\left(\frac{S_t}{S_0}\right) = X_t \sim N\left(\mu t, \sigma^2 t\right) \Leftrightarrow$$
$$\Leftrightarrow S_t = S_0 \exp\left\{X_t\right\} \sim S_0 \times \log \text{Normal}\left(\mu t, \sigma^2 t\right)$$

As the log of $\exp\{X_t\} = X_t$ and $X_t \sim Normal(\mu t, \sigma^2 t)$

24.4 Distributions

24.4.1 Multivariate Normal

$$\mathbf{X} = \left[\begin{array}{c}\mathbf{X}_1 \\ \hdashline \mathbf{X}_2\end{array}\right] \sim \mathbf{N}\left(\boldsymbol{\mu} = \left[\begin{array}{c}\boldsymbol{\mu}_1 \\ \hdashline \boldsymbol{\mu}_2\end{array}\right], \boldsymbol{\Sigma} = \left[\begin{array}{c:c}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \hdashline \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{array}\right]\right)$$

Property	Formula
Marginals:	$\mathbf{X}_1 \sim \mathbf{N}\left(\mu_1, \Sigma_{11}\right)$
Linear combinations:	$\mathbf{a} + \mathbf{B}^\top \mathbf{X} \sim N\left(\mathbf{a} + \mathbf{B}^\top \mu, \mathbf{B}^\top \Sigma \mathbf{B}\right)$

Property	Formula
Conditionals:	$\mathbf{X}_1 \mid (\mathbf{X}_2 = \mathbf{x}) \sim \mathrm{N}\left(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}\left(\mathbf{x}\mu_2\right), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)$

24.4.2 Heavy-tailed distribution:

- has an infinite MGF
- Heavy tailed distributions can also have infinite moments (including the mean!)
- $$\mathbb{E}(X^k) = \infty \text{ for } k \geq \alpha$$

24.4.3 Stable Distribution:

- If the sum of two individual identical ones gives the same distribution (up to location & scale parameters), than distribution is Stable
- Ex. Sum of two i.i.d Normals is Normal $\implies$ Normal Distribution is stable
- All stable distributions (besides the Normal) have heavy tails with tail index = α in the characteristic function

24.4.4 Cauchy distribution (t distribution with 1 df):

$$CF(t) = \phi_X(t) = e^{-|t|}$$

24.4.5 Gamma distribution:

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$$
$$F(x) = \frac{1}{\Gamma(\alpha)} \gamma(\alpha, \beta x)$$
$$\mathbb{E}(x) = \frac{\alpha}{\beta} \quad \mathbb{V}(x) = \frac{\alpha}{\beta^2}$$
$$\int_0^\infty \lambda^{\alpha-1} e^{-(\beta+x)\lambda} d\lambda = \frac{\Gamma(\alpha)}{(\beta+x)^\alpha}$$

24.4.6 Exponential distribution:

$$X \sim Exp(\lambda)$$
$$f(x) = \lambda e^{-\lambda x}$$
$$F(x) = 1 - e^{-\lambda x}$$
$$\mathbb{E}(x) = \frac{1}{\lambda} \quad \mathbb{V}(x) = \frac{1}{\lambda^2}$$

Example of integral of a mixture model:

$$Z = \frac{X}{Y}$$
$$F_Z(z) = P(Z \le z) = P(X/Y \le z)$$
$$= \int_{domain\ Y} P(X/Y < z|Y = y) dF_Y(y) = \int P(X < yz) f_Y(y)\, dy$$

24.4.7 1st Theorem: Fisher-Tippet-Gnedenko

$$H(x) = \begin{cases} Gumbel & \exp\{-e^{-x}\} \quad\quad x \in \mathbb{R} \\ Frechet & \begin{cases} 0 & x < 0 \\ \exp\{-x^{-\alpha}\} & x > 0 \end{cases} \\ Weibull & \begin{cases} \exp\{-|x|^{\alpha}\} & x < 0 \\ 1 & x > 0 \end{cases} \end{cases}$$

24.4.8 Generalized Extreme Value (GEV) Distribution

$$H(x) = \exp\left\{ -\left(1 + \xi \frac{x-\mu}{\sigma}\right)_{+}^{-1/\xi} \right\}$$
$$\text{Where } \mu = \text{location}$$
$$\sigma = \text{scale}$$
$$\xi = \text{shape parameters} \begin{cases} \xi > 0 & \text{heavy tails (Frechet)} \\ \xi = 0 & \text{exponential tails (Gumbel)} \\ \xi < 0 & \text{short/light tails (Weibull)} \end{cases}$$

24.4.9 Normal scale mixture

$Y = \mu + \sqrt{V} \cdot Z$ Where V is a RV with non-negative mixing distribution and represents a random sd of Y.

Example (probably don't need to memorize...) which is used when simulating returns, as seem in PS2 Q4c part ii and iii.

$$\text{t-dist } t = Z\sqrt{v/W} \quad \text{where } W \sim \chi^2(df = v)$$
$$\text{GARCH model } r_t = \mu + \sigma_t Z_t \quad \text{where } \sigma_t^2 = \omega + \sum_{i=1}^{p} a_i r_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2$$

24.5 Copula

[[4-Multivariate-Return-Modelling]] https://richardye101.github.io/STAD70/copulas.html

$$\text{Independence Copula: } C_{indep}(u_1, \dots u_d) = u_1 \times \cdots \times u_d$$
$$\text{Archimedean Copula: } C_{arch}(u_1, \dots u_d) = \phi^{-1}(\phi(u_1) + \cdots + \phi(u_d))$$
$$\text{Given multi/uni var CDF } \phi_\rho, \phi\text{: } C_\rho(u_1, \dots, u_d) = \Phi_\rho(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d))$$

$$\text{Copula Properties}$$
$$\underline{C}\,(u_1, \dots, u_d) \le C\,(u_1, \dots, u_d) \le \bar{C}\,(u_1, \dots, u_d)$$
$$\text{where } \begin{cases} \underline{C}\,(u_1, \dots, u_d) = \max\left\{1 - d + \sum_{i=1}^d u_i, 0\right\} = \max\left\{1 - \left(\sum_{i=1}^d 1 - u_i\right), 0\right\} \\ \bar{C}\,(u_1, \dots, u_d) = \min\{u_1, \dots, u_d\} \end{cases}$$

If you have a copula and the marginal CDF's, you can obtain the multivariate CDF of all the marginals. The inverse is true, where you can take a multivariate CDF and come up with a copula to represent the dependency between the marginal distributions, and the marginal distributions themselves. $F(x_1, \dots x_d) = C(F_1(x_1), \dots, F_d(x_d))$

24.6 Portfolio Theory

- To achieve a μ_p in a two asset portfolio, the weight

$$w = \frac{\theta - \mu_B}{\mu_A - \mu_B}$$

-

$$\text{Gross Returns:}$$
$$R'(t) = \frac{S(t)}{S(t-1)}$$
$$\text{Net Returns:}$$
$$R(t) = R'(t) - 1 = \frac{S(t) - S(t-1)}{S(t-1)}$$
$$\text{Log Returns:}$$
$$r(t) = \log(R'(t)) = \log\{S(t)\} - \log\{S(t-1)\}$$

- Min-variance portfolio weights are

$$\vec{w} = \frac{\Sigma^{-1}\vec{1}}{\vec{1}^T\Sigma^{-1}\vec{1}}$$

if we have a variance matrix, we can calculate the minimum weights for each asset using:

```
weights = rowSums(S_inv)/sum(S_inv)
```

24.6.1 Sharpe Ratio

$$\underbrace{\left(\frac{\mu_M - \mu_f}{\sigma_M}\right)}_{\text{Sharpe Ratio}}$$

24.7 Factor Model

24.7.1 Assumptions

- Factors $F_j(t)$ are stationary, with moments:

$$E[\mathbf{F}(t)] = \mu_F \quad \& \quad \mathrm{Var}[\mathbf{F}(t)] = \Sigma_F$$

- Asset-specific errors $\varepsilon_i(t)$ are uncorrelated with common factors:

$$\mathrm{Cov}[(t), \mathbf{F}(t)] = \mathbf{0}$$

- Errors are serially & contemporaneously uncorrelated across assets

$$\mathrm{Var}[\varepsilon(t)] = \mathrm{diag}\left[\left\{\sigma^2_{\varepsilon_i}\right\}_{i=1,\ldots,N}\right] = \Sigma_\varepsilon \quad \& \, \mathrm{Cov}[\varepsilon(t), \varepsilon(s)] = \mathbf{0}$$

24.7.2 In R

Fitting a factor model to a list of returns of assets using `factanal(x, factors=2)`, we can extract the factor loadings **B** and uniqueness **V** which can be used to calculate the sample correlation matrix along with the sample variance matrix $\mathbf{SS}^T$

$$Var = (\mathbf{BB}^T + diag(\mathbf{V}))(\mathbf{SS}^T)$$

24.8 Risk Mgmt

24.8.1 VaR

Calculating the VaR_α at a $(1-\alpha)$ confidence level:

$$P(L \leq VaR_\alpha) = 1 - \alpha = F(VaR_\alpha)$$
$$\text{Essentially Isolate } VaR_\alpha\text{: } VaR_\alpha = F^{-1}(1-\alpha)$$

If L follows a normal distribution, then $VaR_\alpha(L) = -S_i \times (\mu_i + \sigma_i z_\alpha)$ where S_i is the initial investment, and $z_\alpha = \Phi^{-1}(\alpha)$.

24.8.2 Conditional VaR (CVaR)

is the mean loss in the upper α area of the loss distribution.

$$-\frac{1}{\alpha}\int_0^\alpha VaR_u(L)\,du = \mathbb{E}(L|L \geq VaR_\alpha)$$

24.8.3 Entropic VaR (EVaR)

EVaR defined as: $\mathrm{EVaR}_\alpha = \inf_{z>0}\left\{z^{-1}\ln\left(M_L(z)/\alpha\right)\right\}$
can use R to find the infemum (value of z such that the function is minimized)

```
fn = function(z){ log( ( 0.04 * exp( 95*z ) + 0.96 * exp(-5*z) )
               / 0.05 ) / z }
optimise(fn, c(0,1) )
```

24.8.4 Risk Measure Properties

> ρ here is VaR, CVaR, or EVaR

For this measure ρ to reasonably quantify risk, it must: 1. be normalized $\rho(0) = 0$ (risk of holding no assets is 0) 2. Translation invariance: $\rho(L + c) = \rho(L) + c \ \forall c \in \mathbb{R}$ - adding a loss c to the portfolio increases risk by exactly c 3. Positive Homogeneity: $\rho(bL) = b\rho(l)$ - scaling portfolio returns also will scale risk 4. Monotonicity: $L_1 \geq L_2 \implies \rho(L_1) \geq \rho(L_2)$ - The ordering of the random variables is almost surely $P(L_1 \geq L_2) = 1$ 5. Sub-additivity: $\rho(L_1 + L_2) \leq \rho(L_1 + L_2)$ - Only when the two losses are perfectly correlated, then equal - The risk of two combined portfolios cannot exceed the sum of the two portfolio risks

24.9 Betting measures

Wealth after betting:

$$\begin{aligned} V_n &= V_0 \prod_{t=1}^{n} (1 + fa)^{I_t} (1 - fb)^{1 - I_t} \\ &= V_0 (1 + fa)^{\sum_{t=1}^{n} I_t} (1 - fb)^{\sum_{t=1}^{n} (1 - I_t)} \\ &= V_0 (1 + fa)^{W_n} (1 - fb)^{n - W_n} \end{aligned}$$

Kelly criterion (which maximizes expected log wealth):

$$\text{Win amt:} a \quad \text{Lose amt:} b \quad P(win) = p \quad P(lose) = q$$
$$\text{Kelly Crit: } f^\star = \frac{ap - bq}{ab} \quad \text{or } 2p - 1 \text{ in the simple case}$$

24.10 Arbitrage

24.10.1 Pairs Trading

We trade if we have:

$$\begin{aligned} \text{Profitable if: } & \frac{P1_c}{P1_o} - \frac{P2_c}{P2_o} \neq 0 \\ \implies & \frac{P1_c}{P1_o} < \text{ or } > \frac{P2_c}{P2_o} \\ \text{Where: } & q_1 = \frac{1}{P1_o}, q_2 = \frac{1}{P2_o} \end{aligned}$$

> Our strat is profitable if in the end, we trade P1 and P2 with money left over, in the same quantities we entered the positions with. (Could buy/sell either side)

If we assume the prices of the stocks have a stationary linear relationship $P_t - \lambda S_t$. Furthermore, assume that you open a trade when $P_o - \lambda S_o > 0$, and you close it when $P_c - \lambda S_c = 0$. This is profitable given the conditions with net profit of $(P_o - \lambda S_o) - (P_c - \lambda S_c) > 0$.

24.11 Simulation

In Monte Carlo, there may always be some time point that we did not simulate, where the barrier could have been crossed and the value of the an option, ex $C_{U\&O}$ would have been worthless. Hence the **monte carlo simulation will always overestimate the value of an barrier option, because it underestimates the maximum.**

24.11.1 Simple Monte Carlo estimator

$$\bar{Y} = \frac{1}{2n}\sum_{i=1}^{2n} Y_i \quad \text{where } Y_i = f(Z_i), Z_i \sim N(\mu, \sigma^2)$$
$$Var(\bar{Y}) = \mathbb{E}[f^2(z)]$$

Generating paths of options: We can generate values of M_T directly by generating a standard brownian motion W_T and setting $M_T = |W_T|$. We then estimate the probability that $|W_T|$ crosses 1, and as we generate more Normal RVs (W_T), the probability will converge.

24.11.2 Procedure to generate maxima

Procedure for simulating maxima of arithmetic BM: 1. Generate $X_T \sim N(\mu T, \sigma^2 T)$ 2. Generate $U \sim \text{Uniform}(0,1)$ 3. Calculate $M_T \mid X_T = \frac{X_T + \sqrt{X_T^2 - 2\sigma^2 T \log(U)}}{2}$

24.11.3 Risk Neutral Pricing

The arbitrage-free price of any European derivative with payoff $G_T = f(S_T)$ is given by discounted expectation w.r.t. RN measure

$$G_0 = \mathbb{E}\left[e^{-rT} G_T\right] = \mathbb{E}\left[e^{-rT} f(S_T)\right]$$

The risk-neutral pricing measure:

$$\mathbb{E}(S) = S_0 e^{rt}$$
$$\mathbb{E}\left(\frac{S_t}{e^{rt}}\right) = S_0 \quad \text{or more generally } \mathbb{E}\left(\frac{S_t}{e^{rt}} | S_s\right) = \frac{S_s}{e^{rt}}$$

24.11.4 Black Scholes European Call

$$C = S_0 \Phi(d_1) - e^{-rT} K \Phi(d_2)$$
$$\text{Where:} \quad d_1 = \frac{\ln(S_0/K) + \left(2 + \frac{1}{2}\sigma^2\right) T}{\sigma\sqrt{T}}, \quad d_2 = d_1 - \sigma\sqrt{T}$$

At the money call:

$$C = S_0\left[\Phi(d_1) - e^{-rT}\Phi(d_2)\right] = S_0\left[\Phi\left(\frac{\left(r + \frac{1}{2}\sigma^2 T\right)}{\sigma\sqrt{T}}\right) - e^{-rT}\Phi\left(\frac{\left(r - \frac{1}{2}\sigma^2\right) T}{\sigma\sqrt{T}}\right)\right]$$

Pricing Derivatives

$$\text{Joint CDF of W and max(W): } P(W_T \leq x, M_T \geq y) = \underbrace{P(W_T \geq 2y - x,\ \overbrace{M_T \geq y}^{\text{always true}})}_{\text{By relfection principle}}$$

$$\text{W is Normally distributed: } P(W_t \leq x) = \Phi(x)$$

$$\text{Rayleigh Dist: } P(M_T \leq m | X_T = b) = 1 - \exp\left\{-2\frac{m(m-b)}{\sigma^2 T}\right\} \quad \forall m \geq (0 \cup B)$$

24.12 Variance Reduction

24.12.1 Anti-thetic Variable

$$\text{Anti-thetic Variable: } \bar{Y}_{AV} = \frac{1}{2n}\left(\sum_{i=1}^{n} Y_i + \sum_{i=1}^{n} \tilde{Y}_i\right) = \frac{1}{n}\sum_{i=1}^{n}\frac{Y_i + \tilde{Y}_i}{2}$$
$$Y_i = f(Z_i), \tilde{Y}_i = \tilde{f}(-Z_i)$$

$$\bar{Y}_{AN} \text{ is sample mean of iid } RV_i : \frac{Y_i + \tilde{Y}_i}{2}$$
$$\text{by CLT } \bar{Y_{AN}} \sim^{approx} N\left(\mathbb{E}\left[\frac{Y_i + \tilde{Y}_i}{2}\right], \frac{1}{n}\mathbb{V}\left[\frac{Y_i + \tilde{Y}_i}{2}\right]\right)$$

The payoff is only worthwhile if $f(Z)$ is not an even function, as if it was, then $\frac{Y_i + \tilde{Y}_i}{2}$ would simply be Y_i and that wouldn't help us decrease variance.

24.12.2 Stratification

This is inspired by statistical sampling. Idea: Split RV domain into equi-probable strata and draw equal number of variates from within each one.

$$\bar{Y}_{Str} = \frac{1}{m}\sum_{j=1}^{m} \bar{Y}^{(j)}, \text{ where } \bar{Y}^{(j)} = \frac{1}{n}\sum_{i=1}^{n} f\left(Z_i^{(j)}\right)$$
$$Z_i^{(j)} \sim^{\text{iid}} N\left(0, 1 \mid Z_i^{(j)} \in A_j\right), j = 1, \dots, m$$
$$\mathbb{E}(\bar{Y}_{Str}) = \sum_{j=1}^{m} \mathbb{E}\left[f(Z) | Z \in A_j\right] \cdot P(A \in A_j)$$
$$\mathbb{V}(\bar{Y}_{str}) = \frac{1}{mn}\left\{\mathbb{E}\left[f^2(z) - \frac{1}{m}\sum_{j=1}^{m}\mu_j^2\right]\right\} \leq \underbrace{\mathbb{V}\left[\bar{Y}\right] = \frac{1}{mn}\left\{\mathbb{E}\left[f^2(z) - \mu^2\right]\right\}}_{\text{Simple Random Sample}} \leq \underbrace{\mathbb{E}[f^2(z)]}_{\text{MC Est Var}}$$
$$\iff \frac{1}{m}\sum_{j=1}^{m}\mu_j^2 \geq \mu^2$$
$$\text{Where } \mu = \frac{1}{m}\sum_{j=1}^{m}\mu_j = \sum_{j=1}^{m}\underbrace{\mathbb{E}\left[f(z) | z \in A_j\right]}_{\mu_j}\underbrace{P(A_j)}_{1/m}$$

W21 Midterm Part e, integrating gamma:

$$\begin{aligned}
\bar{F}_X(x) &= P(X > x)(\text{ Law of Total Prob. }) \\
&= \int_0^\infty \overbrace{P(X > x \mid \Lambda = \lambda)}^{\sim \text{Exp}(\lambda)} \quad \overbrace{f_\Lambda(\lambda)}^{\sim \text{Gamma}(\alpha,\beta)} \quad d\lambda \\
&= \int_0^\infty e^{-\lambda x} \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} d\lambda \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty \lambda^{\alpha-1} e^{-(\beta+x)\lambda} d\lambda \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha)}{(\beta+x)^\alpha} \\
&= \left(\frac{\beta}{\beta+x}\right)^\alpha \\
&= 1 - 1 + \left(\frac{\beta+x}{\beta}\right)^{-\alpha} \\
&= 1 - \quad \overbrace{[1 - (1 + x/\beta)^{-\alpha}]}^{\text{CDF or GPD } (\gamma=1/\alpha, \sigma=\beta/\alpha)}
\end{aligned}$$