STAD70 Notes

Summer Yang
Taught by Prof. Sotirios Damouras

Table of contents

1 W1: Financial Data & Returns

1.1 Continuous double auction

- Real-time mechanism to match buyers & sellers and determine prices at which trades execute
- At any time, participants can place orders in the form of bids (buy) and asks (sell)
- Matching orders (bid $\geq$ ask) are executed right away, whereas outstanding orders are maintained in an order book

1.1.1 E.g. Quiz 1 question 1

Assume that an asset's limit order book at depth-5 looks as follows:

orders	5 best bids					5 best asks					
price	...	5.50	5.55	5.56	5.58	5.60	5.65	5.66	5.68	5.69	5.71	...
volume	...	150	90	80	170	30	20	80	135	25	90	...

Furthermore, assume that a market order arrives for selling 100 shares. What is the average price (per share) received by the seller of this market order?

Solution: The market order will be fulfilled by the best available bids until the required volume is achieved. This means that the seller will sell 30 shares at \$5.60 and 70 shares at \$5.58, which results in an average share price of:

$$\frac{5.6 \times 30 + 5.58 \times 70}{100} = 5.586$$

1.2 Order types

- **Limit order**: transact at no more/less than a specific price
  - If order not filled, it's kept in the order book
- **Market order**: transact immediately at current market price
  - A single order can have more than one price
- **Iceberg order**: contains both hidden and displayed liquidity
  - Splits a large order into smaller ones to maintain order anonymity

1.3 Financial data

- **Quote data**: record of bid/ask prices from order book
- **Trade data**: record of filled orders

1.3.1 Daily data

- Open/close
  - Adjusted close (used for calculating returns): adjusted for dividends and splits
- High/low
- Volume

1.3.2 Candlestick

- Green: close > open
- Red: close < open

1.3.3 Other data

- FX rates: currency prices set by global financial centers
- LIBOR rates: average interest rate that major London banks would be charged when borrowing from each other

1.4 Reliability of financial data

Financial data could be skewed by

- Fake orders: trades placed to manipulate prices w/o intention to trade
- Fake trades: trades where buyer and seller is the same party, used to increase trading activity

1.5 Returns

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

1.5.1 Log returns (assumes continuous compounding)

$$r_t = \log(1 + R_t) = \log\left(\frac{P_t}{P_{t-1}}\right) = \underbrace{\log(P_t)}_{p_t} - \log(P_{t-1}) = p_t - p_{t-1}$$

1.5.2 Dividend adjustment

Assuming dividend is reinvested, the adjusted return is

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

The dividend is added back to the price (after the price drop)

1.5.3 Split adjustment

$$R_t = \frac{P_t}{P_{t-1}/2} - 1$$
$$r_t = \log\left(P_t\right) - \log\left(P_{t-1}/2\right)$$

1.5.4 Net vs log returns

$R_t \approx r_t$ for small values of R_t (<1%)

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

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

Monthly returns

For daily net returns $R_1, R_2, ..., R_{22}$, monthly net return is:

$$R_{1-22} = (1 + R_1) \times (1 + R_2) \times \cdots \times (1 + R_{22}) - 1$$

For daily log returns $r_1, r_2, ..., r_{22}$, monthly log return is:

$$\begin{aligned} r_{1-22} &= \log(\frac{P_{22}}{P_{21}}\frac{P_{21}}{P_{20}}...\frac{P_1}{P_0}) \\ &= \log(\frac{P_{22}}{P_{21}}) + \log(\frac{P_{21}}{P_{20}}) + ... + \log(\frac{P_1}{P_0}) \\ &= r_1 + r_2 + \cdots + r_{22} \end{aligned}$$

1.6 Random walk model

Additive log returns suggest using the following to model asset prices

$$\log(\frac{P_t}{P_0}) = r_1 + r_2 + ... + r_t$$

If $\{r_t\}$ are i.i.d., then the log return process is a RW with **drift** μ and **volatility** σ

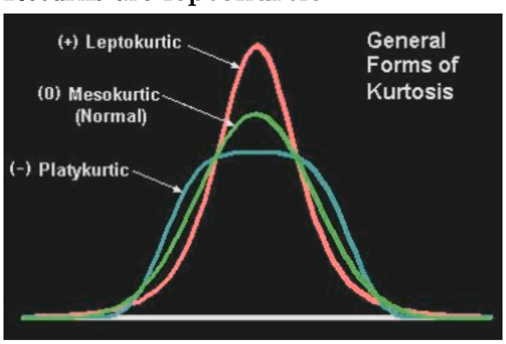
- aggregate returns over n periods has mean $n\mu$ and volatility $\sqrt{n}\sigma$

1.6.1 Exponential/geometric random walk

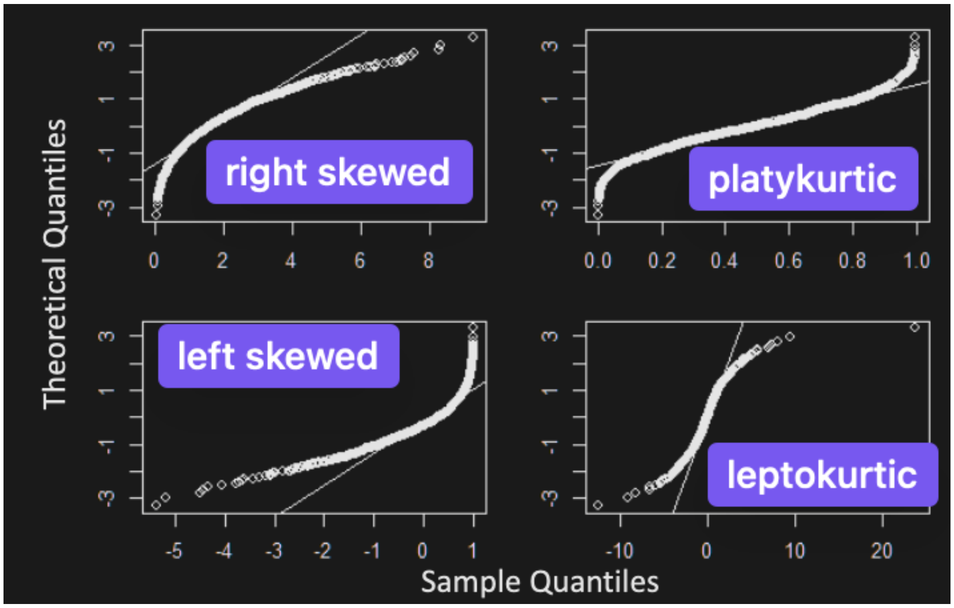
$$P_t = P_0 \exp\{r_1 + ... + r_t\}$$

1.7 Return distribution

- Most convenient assumption: normal (by CLT)
  - Not a good description of reality due to fat tails (heavier than normal)
- Skewness $= E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right]$
  - Right skewed $\iff$ positively skewed
- Kurtosis $= E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] - 3$
  - Defined as the standardized fourth central moment of a distribution minus 3, which is the kurtosis of the standard normal distribution
  - Returns are **leptokurtic**



1.7.1 E.g. Identify skewness/kurtosis from QQ plot



2 W2: Univariate Return Modelling

2.1 Normality tests

- Kolmogorov-Smirnov - Based on distance of empirical & Normal CDF
- Jarque-Bera - Based on skewness & kurtosis combined
- Shapiro-Wilk (most powerful) - Based on sample & theoretical quantiles (QQ plot)

2.2 Heavy tail distributions

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

- Exponential tails, if $f(x) \propto \exp(-x/\lambda)$
- Polynomial tails, if $f(x) \propto x^{-(1+\alpha)}$

Heavy tailed distributions are those with polynomial tails.

- α is the **tail index** controlling tail weight: smaller $\iff$ heavier tails
- for $k \geq \alpha$, moments are **infinite**: $E(X^k) = \infty$
  - Although the MGF's is infinite, the characteristic function always exists (refer to PS2 Q2)

2.2.1 Examples

Pareto

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

Cauchy: $\alpha = 1$

$$f(x) = \frac{1}{\pi(1+x^2)}$$

Student's t: $\alpha = \nu$

$$f(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})}(1+\frac{x^2}{\nu})^{-\frac{\nu+1}{2}}$$

2.2.2 E.g. Quiz 1 question 2

Let the random variable $X \sim \text{Exp}(1)$ follow exponential distribution with CDF $F_X(x) = 1 - e^{-x}, x > 0$. Define the transformed RV $Y = 1/X$, known as an inverse exponential.

(a) (6 points)

Show that the CDF of Y is $F_Y(y) = e^{-1/y}, y > 0$. (Hint: Express the CDF of Y in terms of the CDF of X.)

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(1/X \leq y) \quad (\text{ where } y > 0) \\ &= P(X \geq 1/y) = 1 - P(X \leq 1/y) = 1 - F_X(1/y) \\ &= 1 - \left(1 - e^{-1/y}\right) = e^{-1/y} \end{aligned}$$

(b) (8 points)

Find the tail index α of the inverse exponential distribution. (Hint: Find the PDF of Y using its CDF in the previous part, and look at its asymptotic behaviour as $y \to \infty$.)

$$\begin{aligned} f_Y(y) &= \frac{d}{dy} F_Y(y) = \frac{d}{dy}\left(e^{-1/y}\right) \\ &= e^{-1/y}\frac{d}{dy}(-1/y) = \frac{e^{-1/y}}{y^2} \\ \Rightarrow f_Y(y) &\sim \frac{1}{y^2}, \text{ as } y \to \infty \Rightarrow \text{ the tail index is } \alpha = 1 \end{aligned}$$

2.2.3 Theoretical justification

Let $r_1, \ldots, r_n \sim$ i.i.d. heavy tail distributions with tail index $0 < \alpha < 2$

By the generalized CLT, the aggregate return $r_{1\to n} = r_1 + \ldots + r_n \sim$ **stable distribution**

> A distribution is stable if linear combinations of independent RVs have the same distribution, up to location and scale parameters.

- All stable distributions besides the Normal have heavy tails, but not all heavy tailed distributions are stable (unstable if tail index > 2)
- Moreover, the sum of independent stable RVs also follows a stable distribution
- Thus, adding many heavy tail ($\sigma = \infty$) i.i.d. price changes, we get heavy tail returns

2.2.4 Modeling tail behaviour

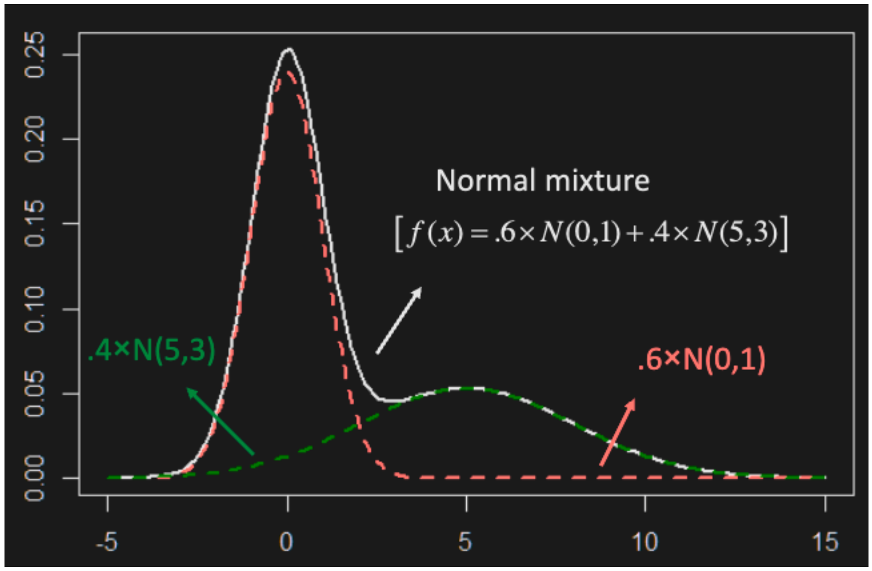
The **complementary CDF** of a heavy tail distribution behaves as:

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

To model (absolute) returns above a cutoff $r_{\min}$, use Pareto distribution $\bar{F}(r) = (r/r_{\min})^{\alpha}$, $\forall r > r_{\min}$

To estimate tail index α, use:

- Maximum Likelihood: $\hat{\alpha} = \frac{n}{\sum_{i=1}^{n} \ln(r_i/r_{\min})}$
- Pareto QQ plots (for tails, e.g. top 25% of returns):
  - Plot empirical CDF vs returns in log-log-scale
  - Estimate α using slope of best fitting line (simple linear regression)
- Student's t QQ plot (for entire distribution, not just tails)
  - Adjust for location and scale $Y = \mu + \sigma X$ where $X \sim t(\nu)$, $\begin{cases} E(X) = 0 \\ V(X) = \frac{\nu}{\nu-2} \end{cases}$
  - Estimate parameters (μ, σ, ν) using MLE
- Mixture models
  - Generate an RV from one out of a family of distributions, chosen at random according to another distribution (a.k.a. mixing distribution)
    - Easy to generate, but not easy to work with analytically
  - 2 types: discrete and continuous
    - e.g. (discrete mixing distribution) RV generated from $\begin{cases} N(0,1) & p = 60\% \\ N(5,3) & p = 40\% \end{cases}$



.

- e.g. (continuous mixing distribution) $Y = \mu + \sqrt{V} \cdot Z$ where V is a RV. This is called a normal scale mixture.
  - Examples with heavy tails:
    - (GARCH) $r_t = \mu + \sigma_t Z_t$ where the mixing process for σ_t is $\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i r_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2$
    - (T-dist) $t = Z\sqrt{\nu/W}$ where $W \sim \chi^2(df = \nu)$

E.g. using mixture models, verify that for $X \sim t(\nu)$, $\begin{cases} E(X) = 0 \\ V(X) = \frac{\nu}{\nu-2} \end{cases}$

Hint: if $W \sim \chi^2(\nu)$, then $E(1/W) = 1/(\nu - 2)$.

$$E(t) = E(Z\sqrt{\nu/W}) = \underbrace{E(Z)}_{0} E(\sqrt{\nu/W}) = 0$$

$$V(t) = E(t^2) - \underbrace{[E(t)]^2}_{0} = E\left(Z^2 \cdot \frac{\nu}{W}\right) = E(Z^2)E\left(\frac{\nu}{W}\right) = 1 \cdot \nu E\left(\frac{1}{W}\right) = \frac{\nu}{\nu-2}$$

2.3 Stylized Facts

Typical empirical asset return characteristics:

1. Absence of simple autocorrelations
2. Volatility clustering
3. Heavy tails
4. Intermittency (alternation between periodic and chaotic behaviour)
5. Aggregation changes distribution (the distribution is not the same at different time scales)
6. Gain/loss asymmetry

2.4 Extreme value theorem

2 limit results for modelling extreme events that happen with small probability

2.4.1 1st EVT (Normalized max of an iid sequence converges to the generalized extreme value distribution)

Let $X_1, X_2, \dots$ be i.i.d. RVs and $M_n = \max(X_1, \dots, X_n)$.

$\exists$ normalizing constants $a_n > 0, b_n$ s.t.

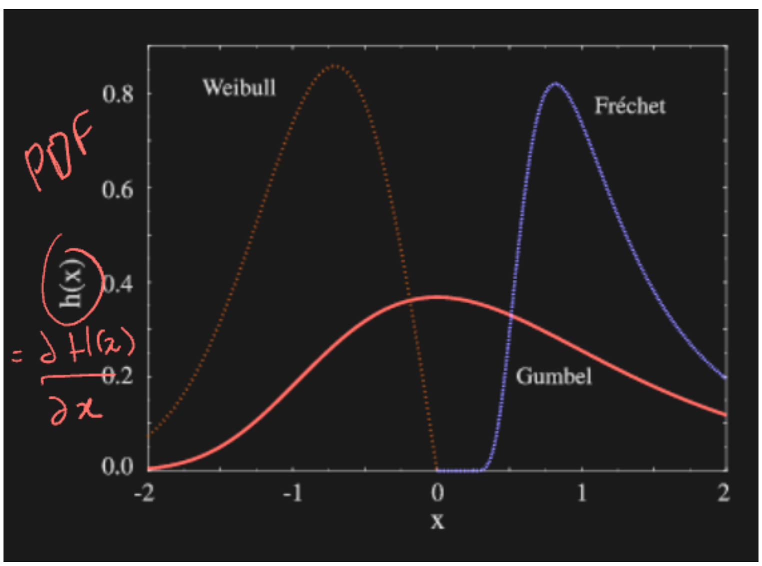
$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) = \left[F\left(a_n x + b_n\right)\right]^n \to H(x)$$

If $H(x)$ exists, it must be one of:

$$\text{Gumbel (exponential tails): } H(x) = \exp\left\{-e^{-x}\right\}, x \in \mathbb{R}$$

$$\text{Frechet (heavy tails): } H(x) = \begin{cases} 0 & x < 0 \\ \exp\left\{-x^{-\alpha}\right\} & x > 0 \end{cases}$$

$$\text{Weibull (light/finite tails): } H(x) = \begin{cases} \exp\left\{-|x|^{\alpha}\right\} & x < 0 \\ 1 & x > 0 \end{cases}$$



We can combine the three types into the **generalized extreme value distribution**

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

ξ is the shape parameter:

- > 0 for heavy tails

- $= 0$ for exponential tails
- < 0 for light tails

E.g. Show that the "normalized" max of iid Uniform (0, 1) with $a_n = \frac{1}{n}$, $b_n = 1$ converges to Weibull for $x < 0$

$$
\begin{aligned}
P(\frac{M_n - b_n}{a_n} \leq x) &= P(\frac{M_n - 1}{1/n} \leq x) = P(M_n \leq \frac{x}{n} + 1) \\
&= P(\max(U_1, ..., U_n) \leq 1 + \frac{x}{n}) \\
&= \prod_{i=1}^{n} P(U_i \leq 1 + \frac{x}{n}) \\
&= (1 + \frac{x}{n})^n = (1 - \frac{|x|}{n})^n
\end{aligned}
$$

$$\lim_{n\to\infty}(1 - \frac{|x|}{n})^n = e^{-|x|} \leftarrow \text{ CDF for Weibull with } \alpha = 1$$

2.4.2 2nd EVT (Conditional distribution converges to GPD above threshold)

For RV X with CDF $F(\cdot)$, consider its conditional distribution given that it exceeds some threshold u:

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

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

In certain cases, as $u \to x_F$, the conditional distribution converges to the same (family of) distributions called the **Generalized Pareto Distribution** (GPD)

$$
\begin{aligned}
F_u(y) \to G_{\xi,\sigma}(y) &= 1 - \left(1 + \xi\frac{y}{\sigma}\right)_+^{-1/\xi} \\
&= \begin{cases} 1 - \left(1 + \xi\frac{y}{\sigma}\right)^{-1/\xi} & \xi \neq 0 \\ 1 - e^{-y/\sigma}, & \xi = 0 \end{cases}
\end{aligned}
$$

where $\sigma > 0, y \geq 0$, and for $\xi < 0, y \leq -\sigma/\xi$

This gives:

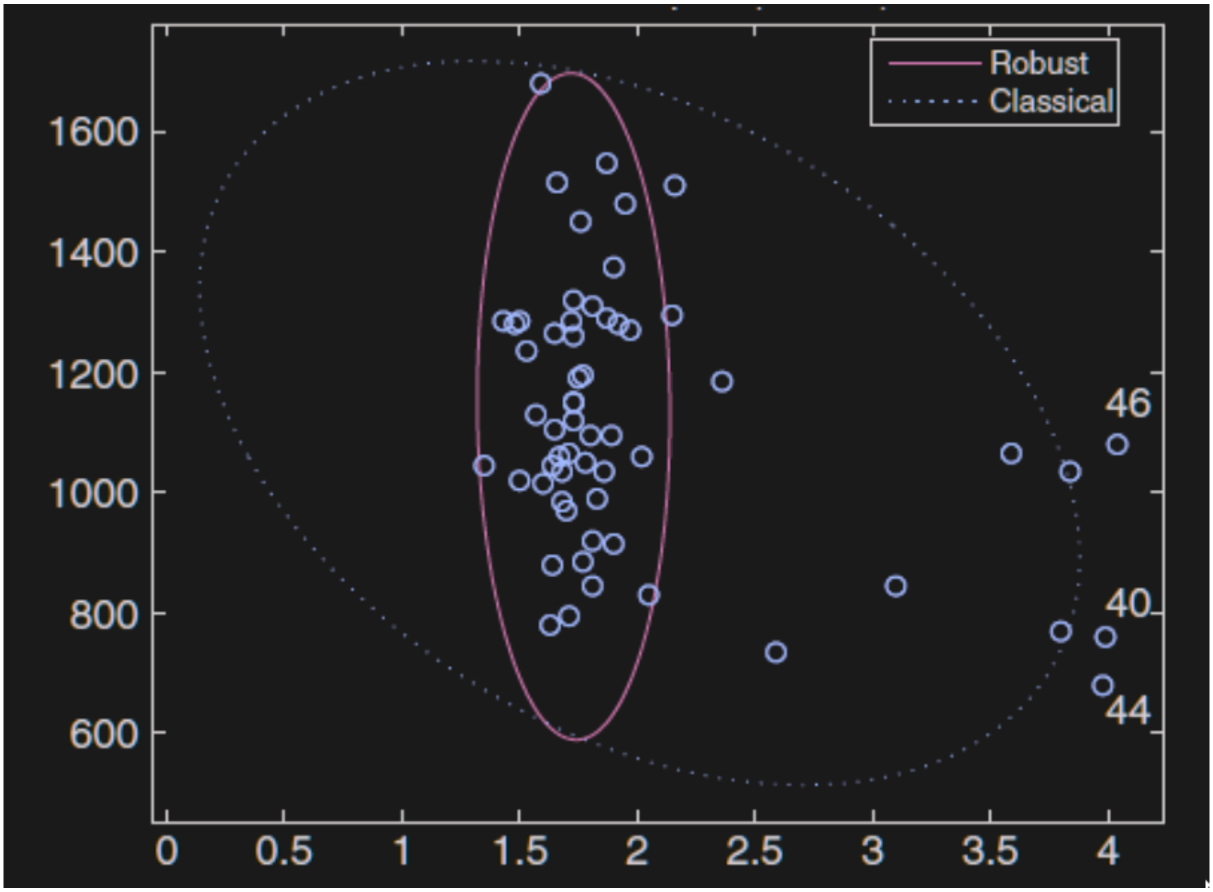
- $\xi > 0:$ heavy tails (tail index $1/\xi$)
- $\xi = 0:$ exponential distribution
- $\xi < 0:$ finite upper endpoint

3 W3: Multivariate Modeling

We can model the returns of a linear combination of assets using a constant matrix A like so

$$Cov(A^T R) = A^T Cov(R) A$$

To minimize the effect of outliers, we can use **robust estimation** - an estimation technique that is insensitive to small departures from the idealized assumptions that were used to optimize the algorithm



However, we should never remove outliers in finance. We can instead model heavy tails using the following.

3.1 Multivariate (Student's) t distribution

A more practical/realistic distribution than Normal for modelling financial returns.

$\mathbf{R} = \mu + \mathbf{Z}\sqrt{\nu/W}$ where $Z \sim N(\mathbf{0}, \Lambda), W \sim \chi^2(df = \nu)$

Note that it is a Normal that gets scaled/divided by the square root of a Chi-square

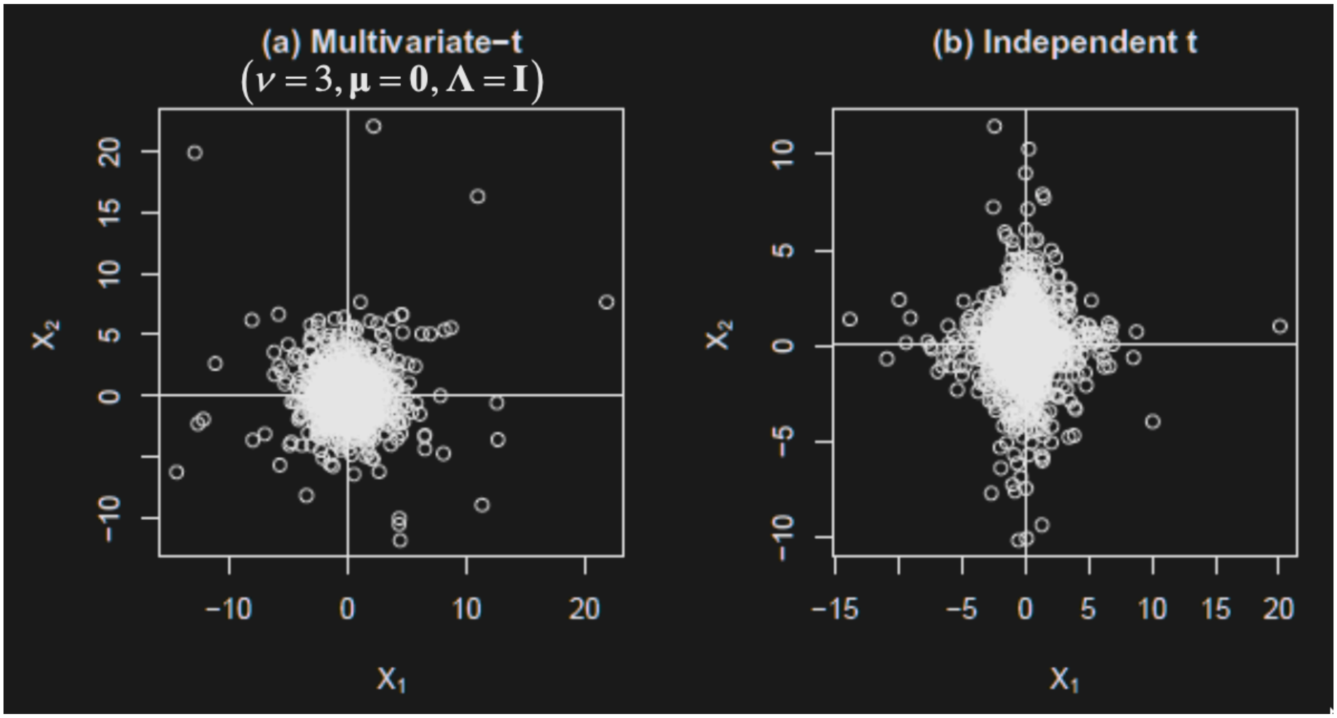
Notation: $\mathbf{R} \sim t_\nu(\mu, \Lambda)$ where $\Lambda = Cov(Z)$, **not** $Cov(R)$

$E(R) = E(Z)$, but $Cov(Z) \neq Cov(R)$

- $E(R) = E(\mu + Z\sqrt{\nu/W}) = \mu + \cancel{E(Z)} \cdot E(\sqrt{\nu/W}) = \mu$
- $Cov(R) = Cov(\mu + Z\sqrt{\nu/W}) = Cov(Z)Cov(\sqrt{\nu/W}) = \Lambda\frac{\nu}{\nu-2}$ for v>2

Marginals are t-distributed with the same degrees of freedom $\implies$ all asset returns have the same tail index α

There is **tail dependence** - extreme values are observed at the same time in all dimensions (desirable property for modelling financial returns)



The greater the tail dependence, the more points we will observe in the corners (figure on the left has tail dependence).

Linear combinations of multivariate t follow 1D t with the *same df*

$\mathbf{R} \sim t_\nu(\mu, \Lambda) \implies \mathbf{w}^T\mathbf{R} \sim t_\nu(\mathbf{w}^T\mu, \mathbf{w}^T\Lambda\mathbf{w})$ - $E(\mathbf{w}^T R) = \mathbf{w}^T E(R)$ - $Var(\mathbf{w}^T R) = \mathbf{w}^T Var(\mu + Z\sqrt{\frac{\nu}{\mathbf{w}}})\mathbf{w} = \mathbf{w}^T \underbrace{Var(Z)}_{\Lambda} Var(\sqrt{\frac{\nu}{\mathbf{w}}})\mathbf{w} = \frac{\nu}{\nu-2}(\mathbf{w}^T\Lambda\mathbf{w})$

Using the same degree of freedom is limiting. A more flexible way is to model dependencies with **copulas**.

3.2 Copula

Intuitively, copulas allow us to decompose a joint probability distribution into the following:

- their marginals (which by definition have no correlation)
- a function which couples them together

thus allowing us to specify the correlation separately. The copula is that *coupling function.* (joint = copula + marginals)

Formally, a copula is a multivariate CDF with Uniform(0, 1) marginals

$C(u_1, ..., u_d) \in [0,1], \forall u_1, ..., u_d \in [0,1]$

- $C(0, ..., 0) = 0$
- $C(1, ..., 1) = 1$
- $C(..., u_{i-1}, 0, u_{i+1}, ...) = 0$
- $C(1, ..., 1, u_i, 1, ..., 1) = u_i$

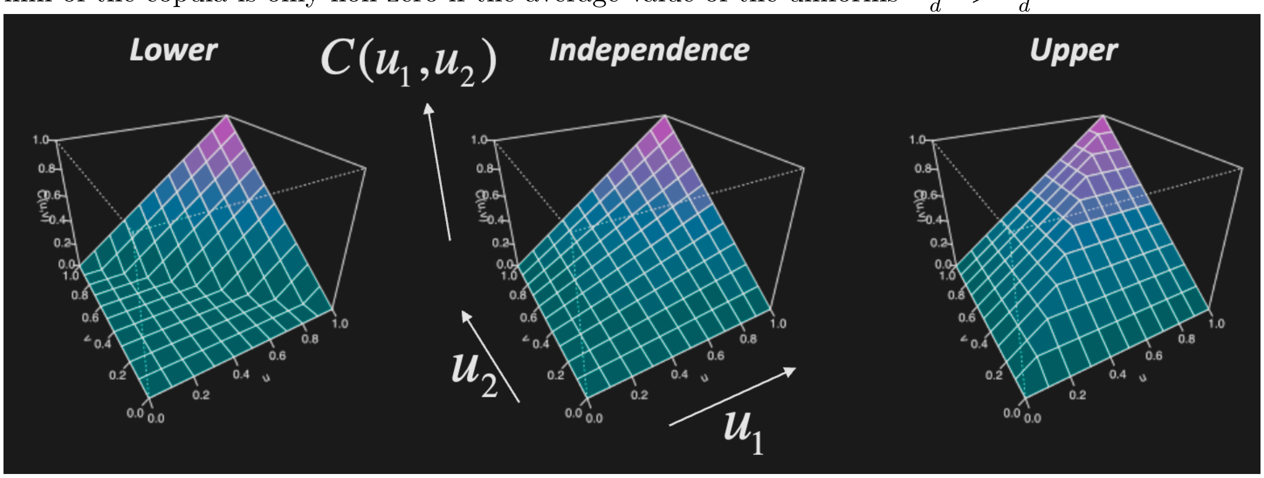
3.2.1 Independence copula

$C_{\text{indep}}\ (u_1, \ldots, u_d) = u_1 \times \cdots \times u_d$

3.2.2 Fréchet-Hoeffding theorem (Copula bounds)

Any copula is bounded like so

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

- The lower bound is 1 minus number of uniforms plus the values of the uniforms. Observe that the min of the copula is only non-zero if the average value of the uniforms $\frac{\sum u_i}{d} > \frac{d-1}{d}$
- 

3.2.3 Sklar's Theorem

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

$$F(x_1, ..., x_d) = C(F_1(x_1), ..., F_d(x_d))$$

The inverse is also true: any copula combined with marginal CDFs gives a multivariate CDF

If we let $u_i = F_i(x_i) \implies x_i = F_i^{-1}(u_i) \implies C(u_1, ..., u_d) = F(F_i^{-1}(u_i), ..., F_d^{-1}(u_d))$

So, for continuous CDF $F(x_1, .., x_d)$ with marginals $F_i(x_i)$, the copula is given by

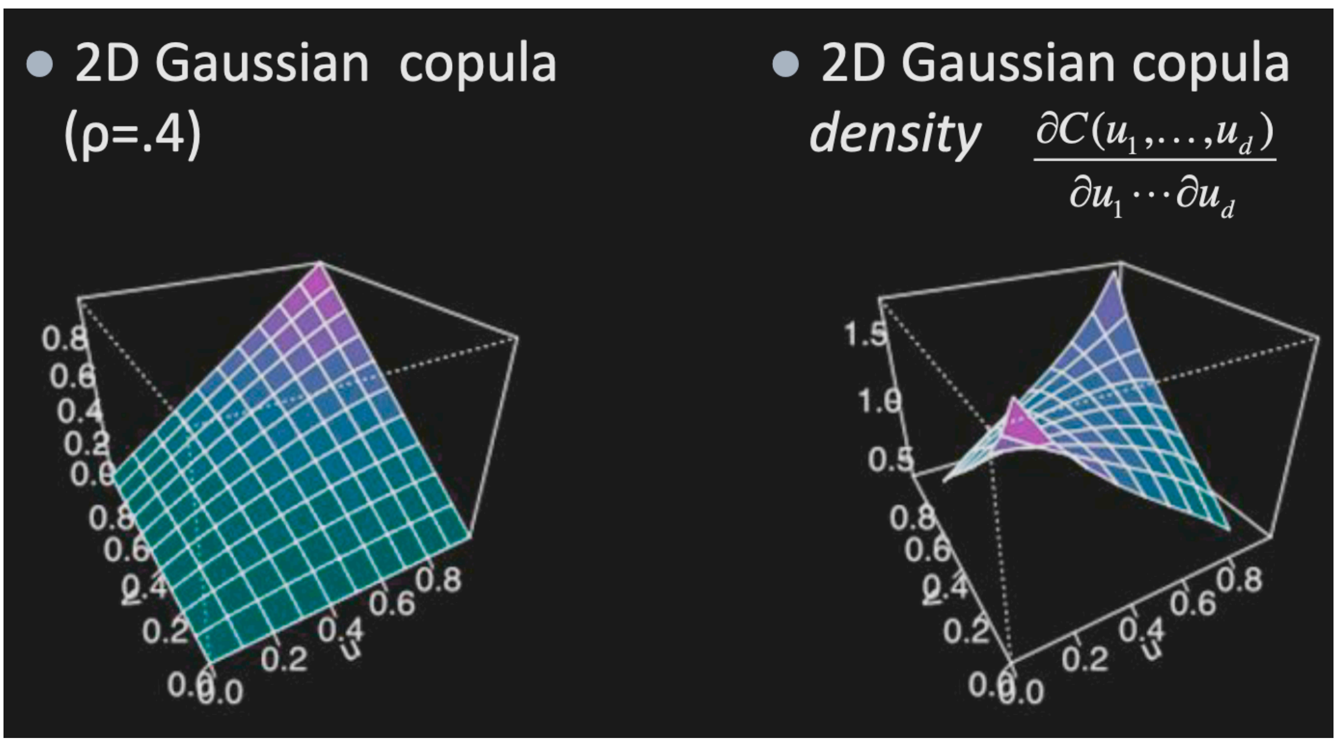
$$C(u_1, ..., u_d) = F(F_1^{-1}(u_1), ..., F_d^{-1}(u_d))$$

E.g. If $X \sim F$, then $F(X) \sim Unif(0, 1)$, and $F^{-1}(Unif) = X \sim F$

3.2.4 Gaussian Copula

Suppose $\mathbf{X} \sim N_d(,)$ with correlation matrix . Its copula is given by

$$C_\rho(u_1, ..., u_2) = \Phi_{\mu,\Sigma}(\Phi^{-1}_{\mu_1,\sigma_1^2}(u_1), ..., \Phi^{-1}_{\mu_d,\sigma_d^2}(u_d))$$



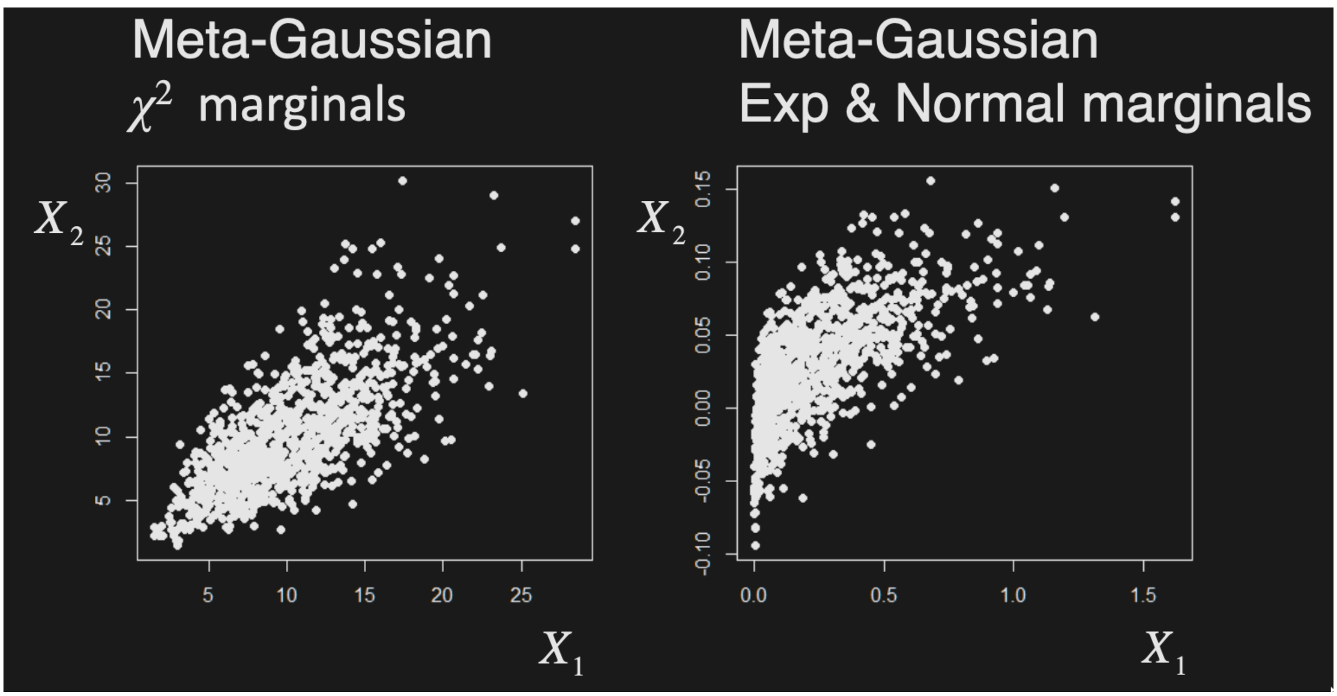
For the independent copula, the derivative would be a plane.

Note: The Gaussian copula only depends on ρ, *not* on the individual means and variances (μ_i's and σ's). Shown below.

$$\begin{aligned}
C_{Gaussian}(u_1, \dots, u_d) &= \Phi_{\mu,\Sigma}(\Phi^{-1}_{\mu_1,\sigma_1^2}(u_1)..., \Phi^{-1}_{\mu_2,\sigma_2^2}(u_2)) \\
&= \Phi_{\mu,\Sigma}\left(\mu_1 + \sigma_1 \Phi^{-1}_{0,1}(u_1), \dots, \mu_d + \sigma_d \Phi^{-1}_{0,1}(u_d)\right) \\
&= \Phi_{0,\phi}\left(\frac{[\mu_1 + \sigma_1 \Phi^{-1}_{0,1}(u_1)] - \mu_1}{\sigma_1}, \dots, \frac{[\mu_d + \sigma_d \Phi^{-1}_{0,1}(u_d)] - \mu_d}{\sigma_d}\right) \\
&= \Sigma_{0,\phi}(\Phi^{-1}_{0,1}(u_1), \dots, \Phi^{-1}_{0,1}(u_d))
\end{aligned}$$

Meta-Gaussian distributions

Multivariate distributions with a Gaussian copula



3.2.5 Simulation

Copulas can be created from known distributions. To simulate data from a distribution with copula C and marginals F_i:

1. Generate (dependent) uniforms
$$(U_1, ..., U_d) \sim C$$
2. Generate target variates from marginals
$$X_i = F_i^{-1}(U_i) \forall i$$

E.g. To generate uniforms from Gaussian copula:

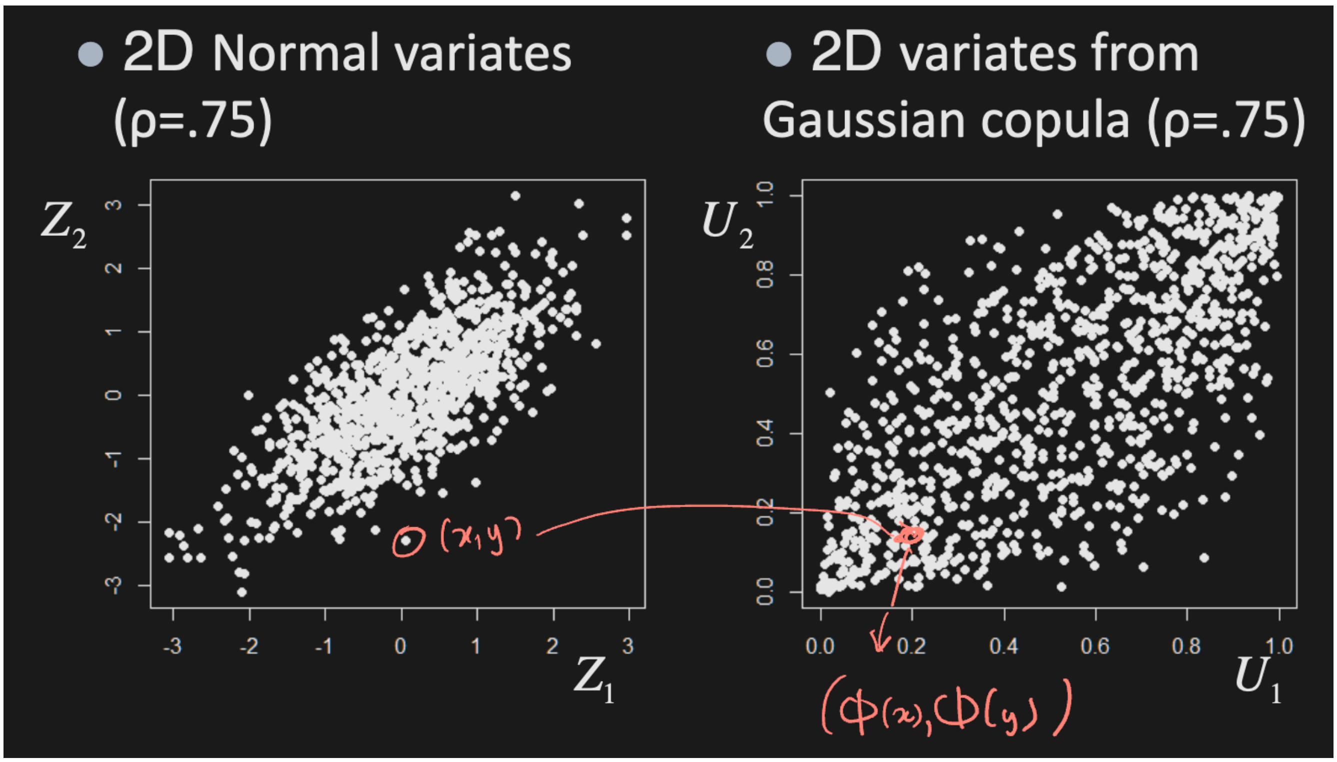
1. Generate multivariate normals with correlation

$$\mathbf{Z} \sim N_d(0, \rho)$$

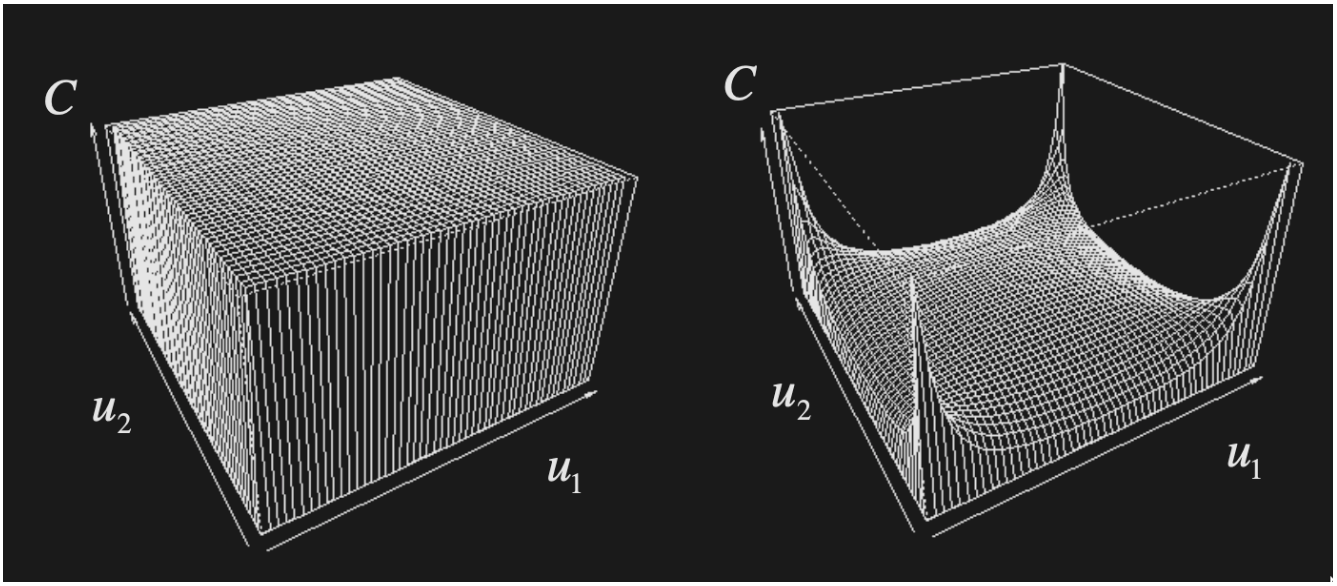
2. Calculate uniforms as their marginal CDF's

$$U_i = \Phi(Z_i)$$

3. Then, use these uniforms with any other marginals

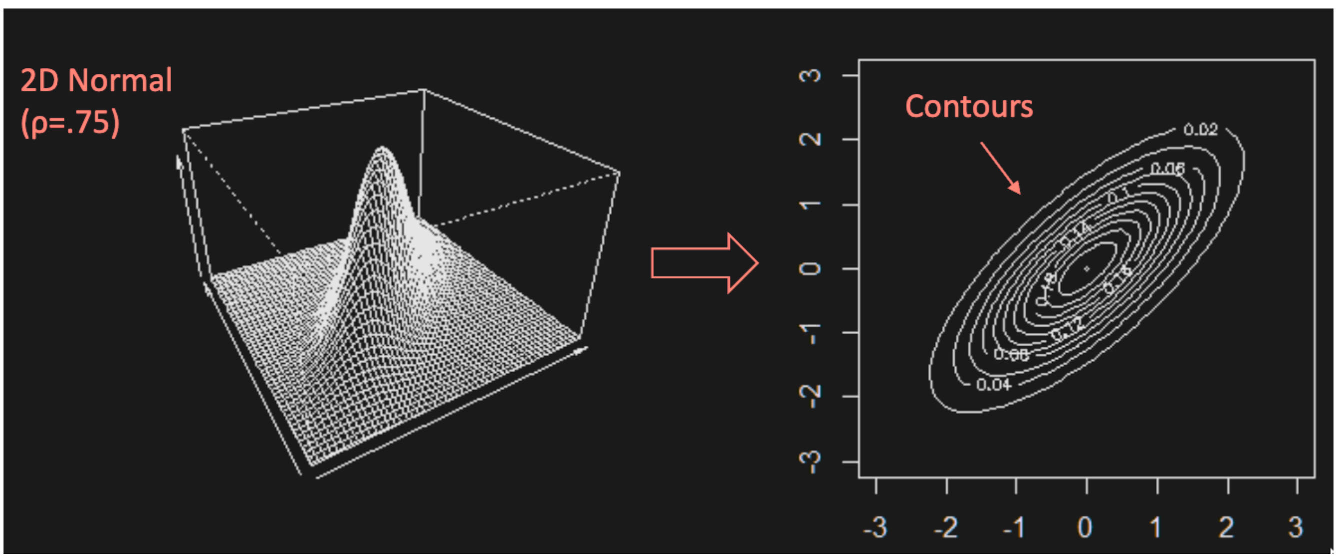


For $\rho = 0$, the pdf of a Gaussian copula vs a t copula looks like



3.3 Elliptical copula

Normal and t distributions both have a dependence structure that is said to be elliptical (due to their elliptical contours)



Symmetry of covariance matrix $\iff$ same dependence strength for positively and negatively correlated values

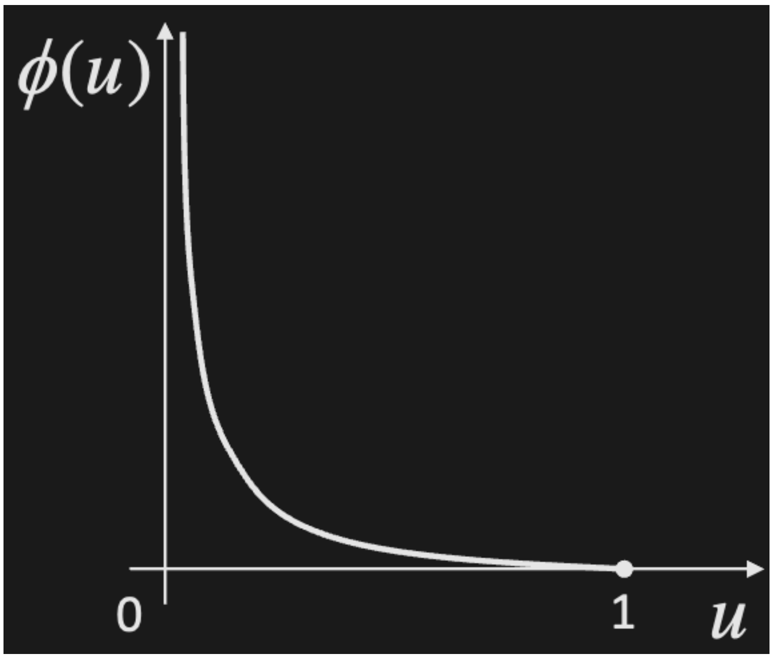
3.4 Archimedean copula

Family of copulas with the following form

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

where ϕ is called the generator function with the following properties:

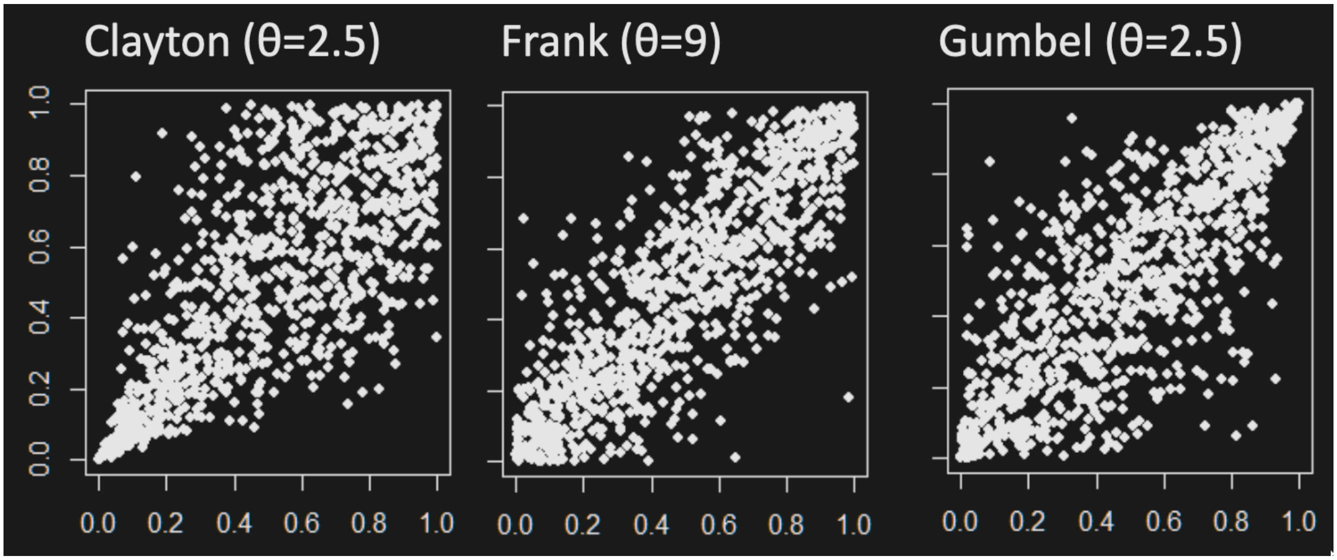
- ϕ is continuous and convex
- $\phi : [0,1] \to [0,\infty]$
- $\phi(0) = \infty, \phi(1) = 0$



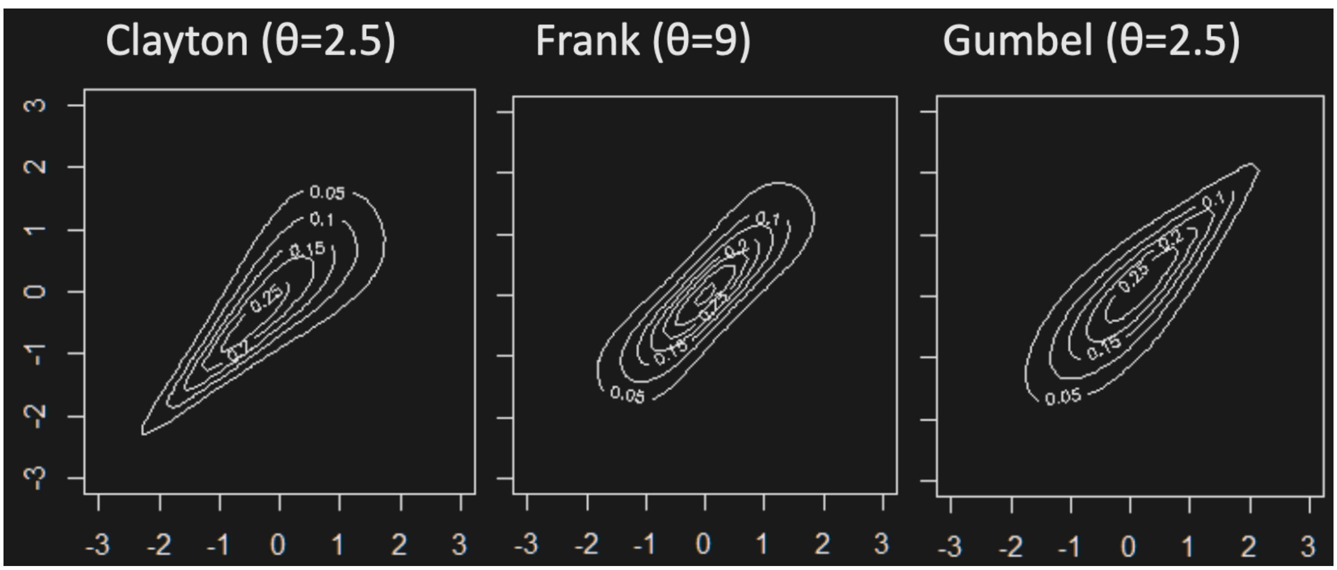
There are infinitely many choices for ϕ, but the most common ones are:

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

2D Archimedean copula random variates:



Contours of 2D pdf's with Archimedean copulas and standard normal marginals:



Although Archimedean copulas can model dependence *asymmetries*, there are limitations in $\geq$ 3D

- The copula value is constant for any permutation of coordinates $u_1, \ldots, u_d$

$$C(u_1, \ldots, u_d) = \phi^{-1}(\phi(u_1) + \cdots + \phi(u_d))$$

- All pairs of coordinates have the same dependence, which is *not* the case for elliptical copulas

Alternative: vine copulas, which allow for both asymmetry and differences in pairwise dependence.

3.5 Fitting copulas

For given copula and marginals, we can use MLE to fit multivariate distribution parameters to data, but the number of parameters can be very high.

Instead, use pseudo-MLE to break problem down into marginals and copula:

- Estimate marginal params for each dimension and calculate uniforms

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

- Then estimate copula using ML on uniforms

4 W4: Portfolio Theory

Assumptions:

- Static multivariate return distribution
- Investors have same views on mean & variance
- Investors want minimum risk for maximum return
- Investors measure risk by portfolio's variance
- No borrowing or short-selling restrictions
- No transaction costs

4.1 Two asset portfolio

The portfolio return is $R_p = w_1 R_1 + w_2 R_2$

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

We can model it like so

$$R_p = \mathbf{w}^T \cdot \mathbf{R} = \begin{pmatrix} w_1 & w_2 \end{pmatrix} \begin{pmatrix} R_1 \\ R_2 \end{pmatrix}$$

where $\mathbf{R} \sim N_2(\mu, \Sigma) \implies R \sim N_1(\mu_p, \sigma_p^2)$ with

$$\begin{aligned}
\mu_p &= E(R_p) = E(\mathbf{w}^T\mathbf{R}) = \mathbf{w}^T E(\mathbf{R}) = \mathbf{w}^T \\
\sigma_p^2 &= V(R_p) = V(\mathbf{w}^T\mathbf{R}) = \mathbf{w}^T V(\mathbf{R})\mathbf{w} = \mathbf{w}^T\Sigma\mathbf{w} \\
&= \begin{bmatrix} w_1 & w_2 \end{bmatrix} \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \\
&= w_1^2\sigma_1^2 + 2w_1w_2\sigma_{12} + w_2^2\sigma_2^2
\end{aligned}$$

Let $w_1 = w$, $w_2 = 1 - w$, then $\sigma_p^2 = w^2\sigma_1^2 + 2w(1-w)\sigma_{12} + (1-w)^2\sigma_2^2$

To minimize, differentiate w.r.t. w and set to 0: $\frac{\partial}{\partial w}\sigma_p^2 = 0$

$$\begin{aligned}
2w\sigma_1^2 + 2(1-w)\sigma_{12} - 2w\sigma_{12} - 2(1-w)\sigma_2^2 &= 0 \\
w(\sigma_1^2 - \sigma_{12}) + (1-w)(\sigma_{12} - \sigma_2^2) &= 0 \\
w &= \frac{-\sigma_{12} + \sigma_2^2}{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}
\end{aligned}$$

4.1.1 E.g. Quiz 2 question 2

Consider a market consisting of 2 assets with bivariate Normal net returns:

$$R = \left[\begin{array}{c} R_1 \\ R_2 \end{array}\right] \sim \mathrm{N}\left(\mu = \left[\begin{array}{c} 0.05 \\ 0.1 \end{array}\right], \Sigma = 0.04\left[\begin{array}{cc} 1 & 1 \\ 1 & 1 \end{array}\right]\right)$$

Note that the returns have the same variance and are perfectly positively correlated.

(a) (6 points)

Show that any portfolio consisting of the two assets will have variance $\mathbb{V}\left[R_p\right] = \mathbb{V}\left[wR_1 + (1-w)R_2\right] = 0.04, \quad \forall w \in \mathbb{R}$

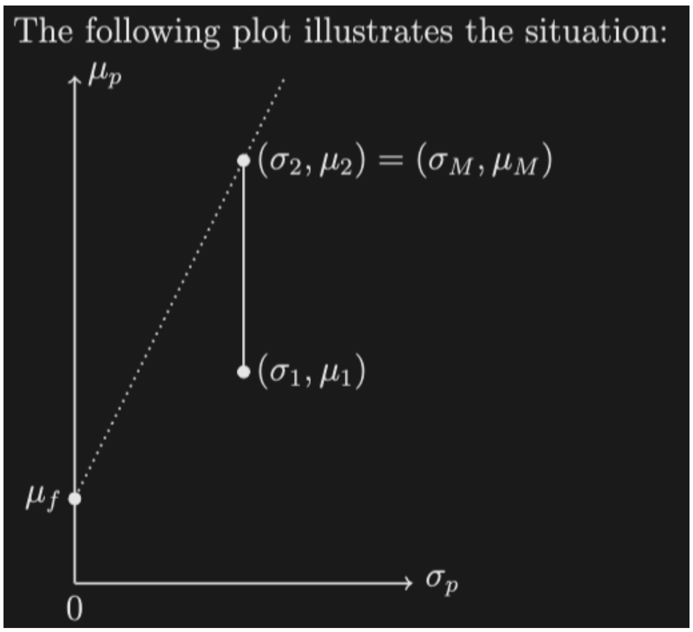
$$\begin{aligned}
\mathbb{V}\left[R_p\right] &= w^\top\Sigma w \\
&= 0.04\left[\begin{array}{cc} w & (1-w) \end{array}\right]\left[\begin{array}{cc} 1 & 1 \\ 1 & 1 \end{array}\right]\left[\begin{array}{c} w \\ (1-w) \end{array}\right] \\
&= 0.04[w + (1-w) \quad w + (1-w)]\left[\begin{array}{c} w \\ (1-w) \end{array}\right] \\
&= 0.04\left[w^2 + 2w(1-w) + (1-w)^2\right] \\
&= 0.04\left[xx^2 + (2\psi\alpha - 2w^2) + (1 - 2wx + xb^2)\right] = 0.04
\end{aligned}$$

(b) (6 points)

Now restrict the set of feasible portfolios to those without short-selling (i.e., $w \in [0,1]$). Assuming the risk-free interest rate is $\mu_f = 0.02$, find the Sharpe ratio of the tangency/market portfolio for this restricted model. (Hint: the answer can be found geometrically from the (σ_p, μ_p) risk-return diagram.)

Since every portfolio has the same variance, the feasible set without short-selling is the vertical line at $\sigma_p = \sqrt{0.04} = 0.2$, from returns $\mu_p \in [0.05, 0.1]$. Obviously, the only optimal portfolio, which is also the tangency portfolio, consists of only the higher-return asset (R_2). The slope of the line with the risk-free return, i.e., the Sharpe ratio, is thus

$$\frac{\mu_2 - \mu_f}{\sigma_2} = \frac{0.1 - 0.02}{0.2} = \frac{0.08}{0.2} = 0.4$$



4.2 Multiple asset portfolio

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

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

A portfolio with weights $\mathbf{w} = [w_1, \ldots, w_n]^T$ s.t. $\sum_{i=1}^n w_i = \mathbf{w}^T\mathbf{1} = 1$ has

$$R_p \sim N_1(\mu_p, \sigma_p^2)$$

To find the min variance portfolio with given expected return μ_p, we solve the following quadratic optimization problem with linear constraints

$$\min_{\mathbf{w}} \left\{\mathbf{w}^\top \Sigma \mathbf{w}\right\}, \text{ subject to } \begin{cases} \mathbf{w}^\top \mu = \mu_p \\ \mathbf{w}^\top \mathbf{1} = 1 \end{cases}$$

The set of such portfolios forms a parabola in mean-variance space, containing attainable portfolios.

4.2.1 Minimum variance portfolio weights

We can use Lagrange multipliers to find the minimum variance portfolio weights:

> Lagrange Multipliers are used to find the local max/min subject to equality constraints
>
> 1 constraint (2 variables): example
>
> $$\mathcal{L}(x, y, \lambda) = f(x, y) - \lambda g(x, y)$$
> $$\nabla_{x,y,\lambda}\mathcal{L}(x, y, \lambda) = 0 \Longleftrightarrow \begin{cases} \nabla_{x,y} f(x, y) = \lambda \nabla_{x,y} g(x, y) \\ g(x, y) = 0 \end{cases}$$
>
> M constraints (n variables):
>
> $$\mathcal{L}\left(x_1, \ldots, x_n, \lambda_1, \ldots, \lambda_M\right) = f\left(x_1, \ldots, x_n\right) - \sum_{k=1}^{M} \lambda_k g_k\left(x_1, \ldots, x_n\right)$$
> $$\nabla_{x_1,\ldots,x_n,\lambda_1,\ldots,\lambda_M}\mathcal{L}\left(x_1, \ldots, x_n, \lambda_1, \ldots, \lambda_M\right) = 0 \Longleftrightarrow \begin{cases} \nabla f(\mathbf{x}) - \sum_{k=1}^{M} \lambda_k \nabla g_k(\mathbf{x}) = 0 \\ g_1(\mathbf{x}) = \cdots = g_M(\mathbf{x}) = 0 \end{cases}$$

Objective function (Lagrangian):

$$L(\mathbf{w}, \lambda) = \mathbf{w}^T\ \mathbf{w} - \lambda(\mathbf{w}^T\mathbf{1} - \mathbf{1})$$

Differentiate and set to 0:

$$\begin{aligned} \frac{\partial \mathbf{L}}{\partial \mathbf{w}} &= 0 \\ 2\ \mathbf{w} - \lambda\mathbf{1} &= 0 \\ \mathbf{w} &= \frac{\lambda}{2}\ ^{-1}\mathbf{1} \end{aligned}$$

Solve for lambda:

$$\mathbf{w}^T\mathbf{1} = \mathbf{1} \implies \frac{\lambda}{2} \cdot \mathbf{1}^T\Sigma^{-1}\mathbf{1} = 1 \implies \lambda = \frac{2}{\mathbf{1}^T\Sigma^{-1}\mathbf{1}}$$

Plugging lambda into $\mathbf{w}$, we get

$$\mathbf{w}^* = \frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}^T\Sigma^{-1}\mathbf{1}}$$

It is the row sums of Σ^{-1} divided by the sum of all its elements.

4.2.2 Risk-free asset

Consider splitting an investment into portfolio (μ_p, σ_p) & risk-free asset, with weights w_p and $1 - w_p$

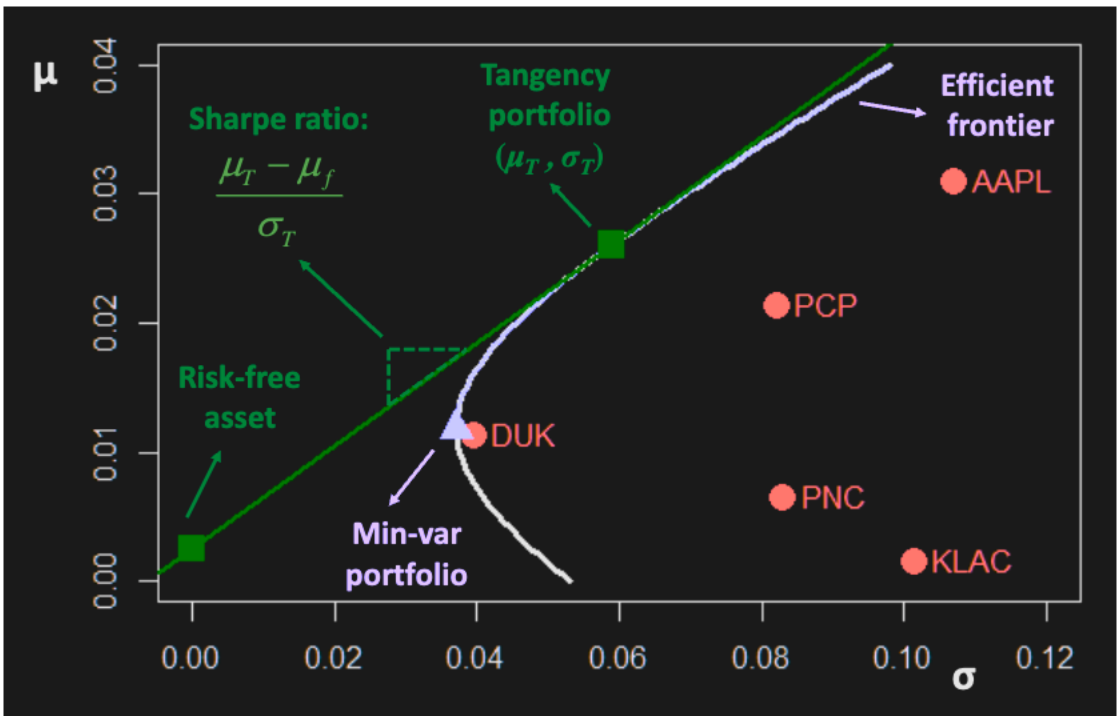
A risk-free asset has constant return $R_f = \mu_f > 0, \sigma_f = 0$

The return is given by $R = w_p R_p + (1 - w_p) R_f$, with

- $E(R) = w_p E(R_p) + (1 - w_p) E(R_p) = w_p \mu_p + (1 - w_p) \mu_f$
- $V(R) = V(w_p R_p + (1 - w_p) R_f) = w_p^2 V(R_p) = w_p^2 \sigma_p^2$

For a set of assets including risk-free ones, the best investments lie on the line tangent to the efficient frontier - they are combinations of the tangency portfolio and risk free assets.

- The tangency portfolio is the efficient frontier portfolio that belongs to the tangent line.
- The slope of the line is the Sharpe ratio.



To find the tangency portfolio, maximize Sharpe ratio.

$$\max \left\{ \frac{\mu_p - \mu_f}{\sigma_p} \right\} = \max_{\mathbf{w}} \left\{ \frac{\mathbf{w}^\top \mu - \mu_f}{\sqrt{\mathbf{w}^\top \Sigma \mathbf{w}}} \right\}, \text{ subject to } \mathbf{w}^\top \mathbf{1} = 1$$

Tangency portfolio weights (solution to above) are given by

$$\mathbf{w}_T = \frac{\Sigma^{-1} (\mu - \mu_f)}{\mathbf{1}^\top \Sigma^{-1} (\mu - \mu_f)}$$

4.3 CAPM (Capital asset pricing model)

If every investor follows mean-variance analysis & the market is in equilibrium, then:

1. Every investor holds some portion of the same tangency portfolio
2. The entire financial market is composed of the same mix of risky assets
3. Tangency portfolio is simply the market value-weighted index

4.3.1 Market portfolio

Since composition of the tangency portfolio is equivalent to that of the market portfolio, its weights are just

$$w_i = \frac{S_i O_i}{\sum_{i=1}^{N} S_i O_i}$$

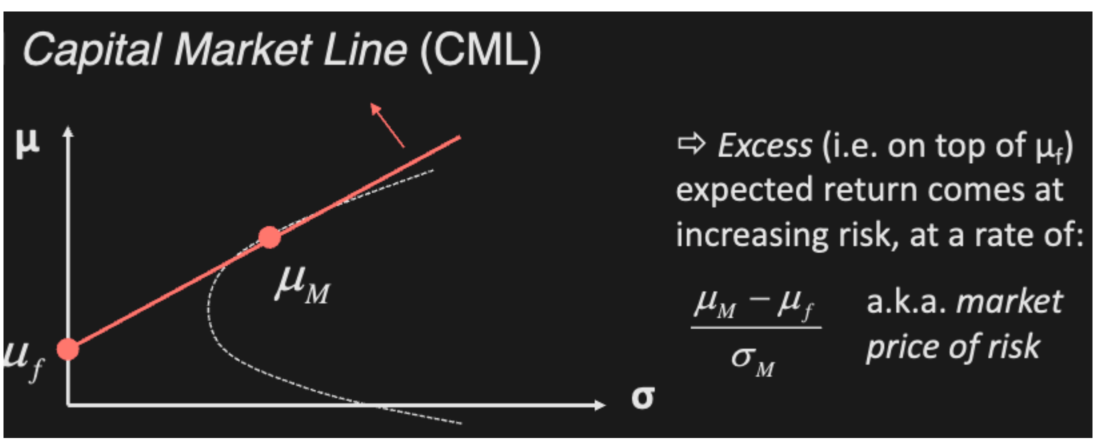
where S_i = price of asset i, O_i = # shares outstanding

4.3.2 Capital market line

Every mean-variance efficient portfolio (μ_p, σ_p) lies on the capital market line:

$$\mu_p = \mu_f + \frac{\mu_M - \mu_f}{\sigma_M}\sigma_p$$

where μ_f is the risk free rate, and (μ_M, σ_M) is the market portfolio



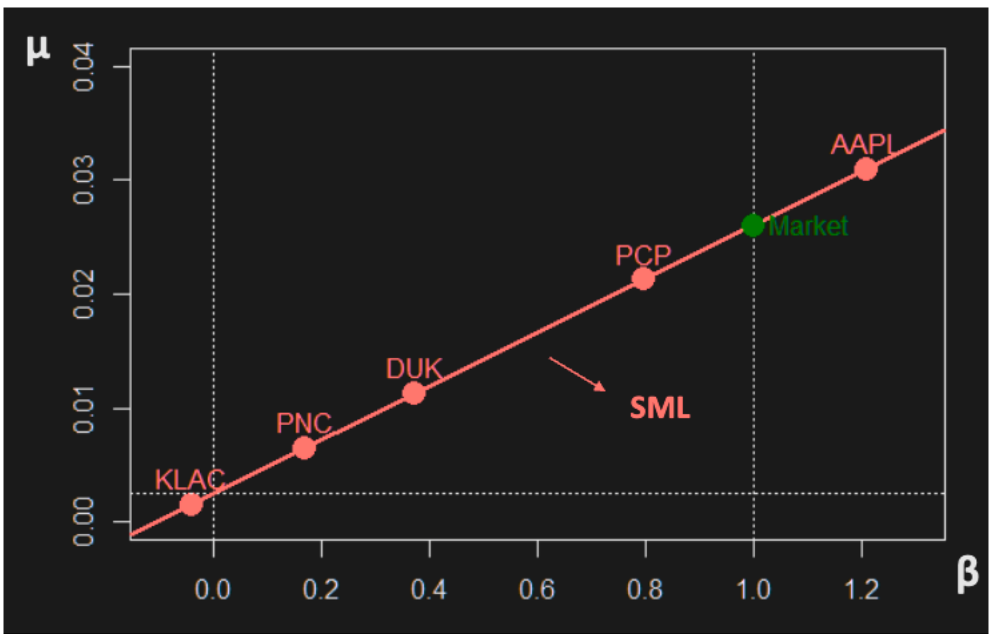
4.3.3 Security market Line

CAPM implies the following relationship between risk and expected return for all assets/portfolios (not just efficient ones)

$$\begin{aligned}\mu_i &= \mu_f + \beta_i(\mu_M - \mu_f) \quad \text{where } \beta_i = \frac{\sigma_{iM}}{\sigma_M^2} = \frac{Cov(R_i, R_M)}{\sigma_M^2} \\ &= \mu_f + \beta_i \underbrace{\left(\frac{\mu_M - \mu_f}{\sigma_M}\right)}_{\text{Sharpe Ratio}} \sigma_M\end{aligned}$$

Implications:

- At equilibrium, an asset's return depends only on its relation to the market portfolio.
  - β_i measures the extent to which an asset's return is related to the market. Higher $|\beta| \iff$ higher risk and reward.
- Investors are only rewarded with higher returns for taking on market/systematic risk



Derivation: max Sharpe ratio using 1st order conditions

$$\max_{\mathbf{w}} \left\{\frac{\mu_M - \mu_f}{\sigma_M}\right\} = \max_{\mathbf{w}} \left\{\frac{\mathbf{w}^\top \mu - \mu_f}{\sqrt{\mathbf{w}^\top \Sigma \mathbf{w}}}\right\}, \text{ subject to } \mathbf{w}^\top \mathbf{1} = 1$$

$$f(\underline{w}) = \frac{\underline{w}^T\underline{\mu} - \mu_f}{\sqrt{\underline{w}^T\Sigma\,\underline{w}}} = \frac{\underline{w}^T\cdot(\underline{\mu} - \underline{1}\cdot\mu_f)}{\sqrt{\underline{w}^T\Sigma\,\underline{w}}} \text{ is maximized @ } w^* \Rightarrow \frac{\partial f}{\partial \underline{w}} = 0 \Rightarrow$$

$$\Rightarrow \frac{(\underline{\mu} - \underline{1}\mu_f)\sqrt{\underline{w}^T\Sigma\,\underline{w}} - \frac{1}{2}\cdot\big(\overbrace{\underline{w}^T\Sigma\cdot\underline{w}}^{\sigma_M^2}\big)^{-1/2}\cdot 2\Sigma\underline{w}\cdot\big(\overbrace{\underline{w}^T\cdot(\underline{\mu} - \underline{1}\mu_f)}^{\mu_M}\big)}{\underline{w}^T\Sigma\cdot\underline{w}} =$$

$$= \frac{(\underline{\mu} - \underline{1}\cdot\mu_f)\cdot\sigma_M - \frac{1}{\sigma_M}\cdot\Sigma\underline{w}\cdot(\mu_M - \mu_f)}{\cancel{\sigma_M^2}} = 0 \Rightarrow$$

$$\Rightarrow (\underline{\mu} - \underline{1}\mu_f) = \frac{1}{\sigma_M^2}\cdot\Sigma\cdot\underline{w}\cdot(\mu_M - \mu_f) \Rightarrow$$

$$\Rightarrow \begin{bmatrix} \mu_1 - \mu_f \\ \vdots \\ \mu_n - \mu_f \end{bmatrix} = \frac{(\mu_M - \mu_f)}{\sigma_M^2}\cdot\begin{bmatrix} Cov(R_1, R_M) \\ \vdots \\ Cov(R_n, R_M) \end{bmatrix} = \frac{(\mu_M - \mu_f)}{\sigma_M^2}\cdot\begin{bmatrix} \sigma_{1,M} \\ \vdots \\ \sigma_{n,M} \end{bmatrix}$$

$$\text{b/c } Cov(\underline{R}, R_M) = Cov(\underline{R}, \underline{w}^T\cdot\underline{R}) = \underline{w}^T\cdot\underbrace{Cov(\underline{R}, \underline{R})}_{=\Sigma}$$

$$\hookrightarrow = \underline{w}^T\cdot\Sigma$$

E.g. Consider N assets with iid $N(\mu, \sigma^2)$ returns and risk free return $\mu_f < \mu$. Find market portfolio weights and SML.

$\forall \mathbf{w}$ s.t. $\mathbf{w}^T\mathbf{1} = 1$, $\mathbf{w}^T\mu\mathbf{1} = \mu$

Since the market portfolio is the min variance portfolio, we have

$$\mathbf{w}^* = \frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}^T\Sigma^{-1}\mathbf{1}} = \frac{\frac{1}{\sigma^2}\mathbf{I}\mathbf{1}}{\frac{1}{\sigma^2}\mathbf{1}^T\mathbf{I}\mathbf{1}} = \frac{\mathbf{1}}{N} \implies w_i = \frac{1}{N},\ \forall i = 1, \ldots, N$$

So the minimum variance is

$$\mathbf{w}^T \Sigma \mathbf{w} = \left(\frac{\mathbf{1}}{N}\right)^2 \mathbf{1}^T(\sigma^2 \mathbf{I})\mathbf{1} = \sigma^2 \frac{1}{N^2} N = \frac{\sigma^2}{N}$$

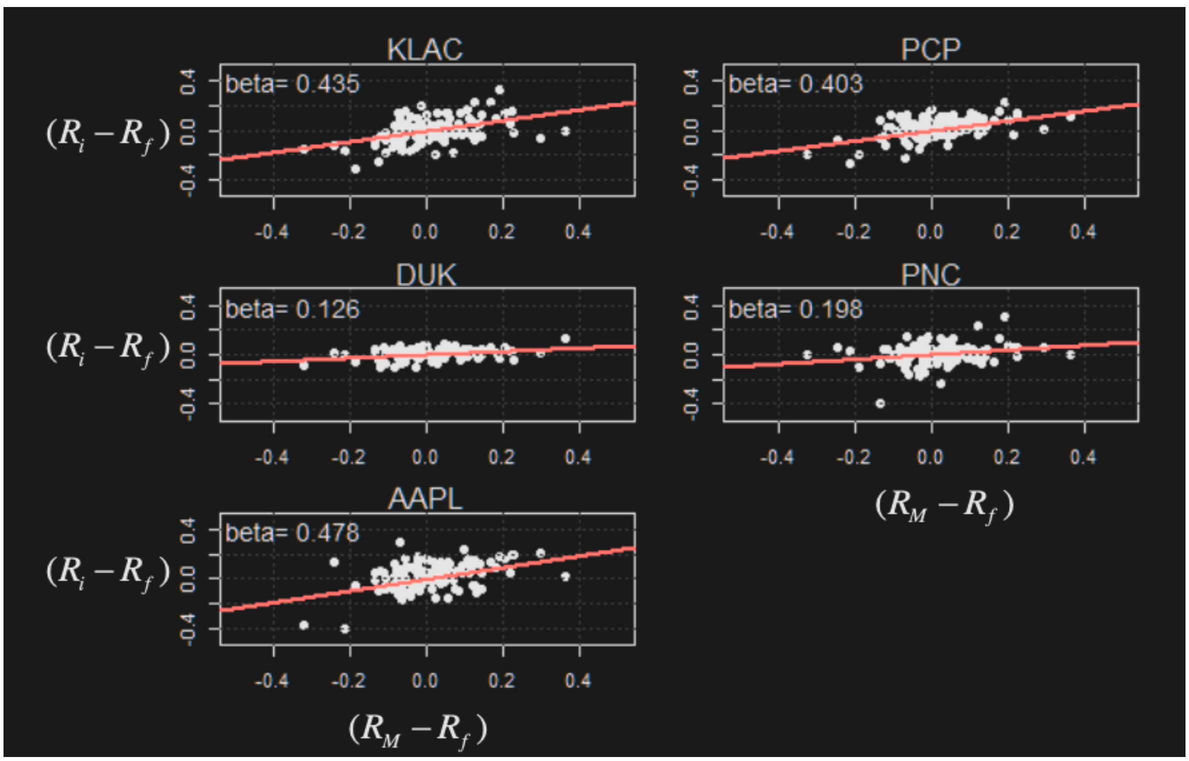
4.3.4 Security characteristic line

To find β_i's empirically, regress $(R_i - R_f)$ on $(R_M - R_f)$

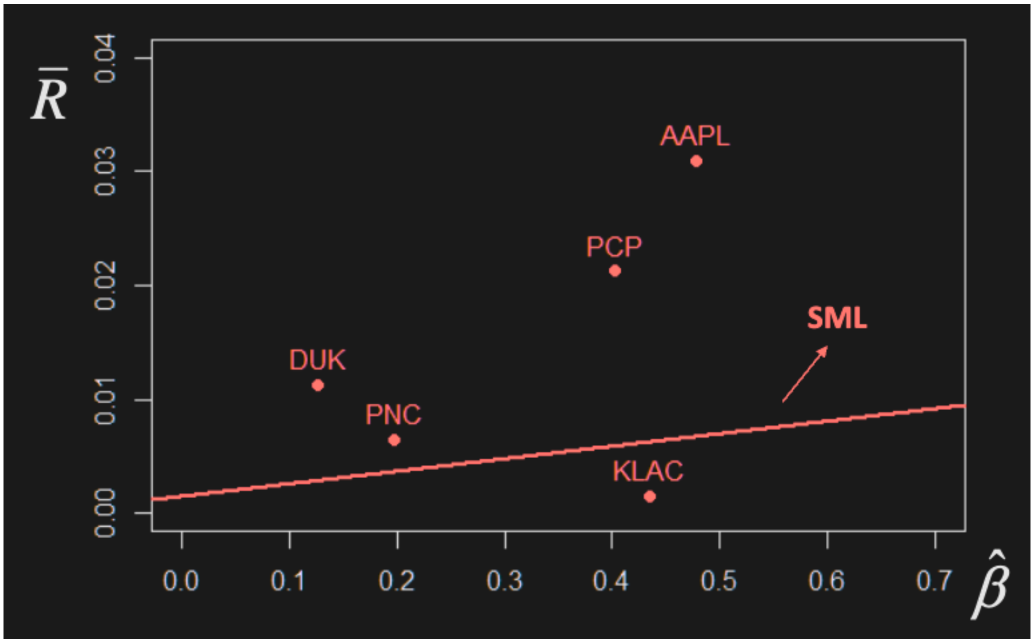
$$(R_{i,t} - R_{f,t}) = \beta_i (R_{M,t} - R_{f,t}) + \epsilon_t, \qquad \text{where } \epsilon_t \sim N(0, \sigma^2_{\epsilon,i})$$

- 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)

The slope of the SCL is the beta estimate:



Mean return vs estimated betas:



Now consider adding an intercept, α

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

The mean and variance are given by

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

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

α_i measures the excess increase in asset return on top of that explained by the β_i. The bigger the alpha, the higher the outperformance (compared to the market portfolio).

4.4 Legacy of CAPM

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.

CAPM is wrong, but 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

4.5 Performance Evaluation

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

Sharpe ratio: (excess return per unit risk)

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

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

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

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.

5 W5: Factor Models

Main implication of CAPM: the market is the single factor driving asset returns

To improve performance, use more factors that drive asset returns

5.1 Factor Models

3 types:

1. Macroeconomic: Factors are observable economic and financial time series
2. Fundamental: Factors are created from observable asset characteristics
3. Statistical: Factors are unobservable, and extracted from asset returns

All 3 types follow some form of

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

- $R_i(t)$ 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}

In matrix form:

$$\begin{bmatrix} R_1(t) \\ \vdots \\ R_N(t) \end{bmatrix} = \begin{bmatrix} \beta_{1,0} \\ \vdots \\ \beta_{N,0} \end{bmatrix} + \begin{bmatrix} \beta_{1,1} & \cdots & \beta_{1,p} \\ \vdots & \ddots & \vdots \\ \beta_{N,1} & \cdots & \beta_{N,p} \end{bmatrix} \begin{bmatrix} F_1(t) \\ \vdots \\ F_p(t) \end{bmatrix} + \begin{bmatrix} \varepsilon_1(t) \\ \vdots \\ \varepsilon_p(t) \end{bmatrix}$$

$$\Leftrightarrow$$

$$\mathbf{R}(t) = \beta_0 + \beta^\top \mathbf{F}(t) + \varepsilon(t)$$

Assumptions:

- Asset specific errors ϵ_i are uncorrelated with common factors

$$Cov(\epsilon(t), \mathbf{F}(t)) = \mathbf{0}$$

- The factors $F_j(t)$ are stationary, with moments

$$E(\mathbf{F}(t)) = \mu_F, \mathbb{V}(\mathbf{F}(t)) = \Sigma_F$$

- Errors are serially and contemporaneously uncorrelated across assets

$$\begin{aligned} E(\epsilon) &= \mathbf{0} \\ \mathbb{V}(\epsilon(t)) &= \text{diag}[\left\{\sigma^2_{\epsilon_i}\right\}_{i=1,\dots,N}] = \Sigma_\epsilon \\ Cov[\epsilon(t), \epsilon(s)] &= \mathbf{0} \end{aligned}$$

Find moments of model $\mathbf{R}(t) = \beta_0 + \beta^T \mathbf{F}(t) + \epsilon(t)$

$$\begin{aligned} \mu_R &= E(\beta_0 + \beta^T F(t) + \epsilon(t)) \\ &= \beta_0 + \beta^T E(F(t)) + E(\epsilon(t)) \\ &= \beta_0 + \beta^T \mu_F \\ \Sigma_R &= V(\beta_0 + \beta^T F(t) + \epsilon(t)) \\ &= V(\beta^T F(t)) + V(\epsilon(t)) \\ &= \beta^T V(F(t))\beta + \Sigma_\epsilon \\ &= \beta^T \Sigma_F \beta + \Sigma_\epsilon \end{aligned}$$

Find moments of portfolio with $\mathbf{w} = [w_1 \ \dots \ w_N]^T$, i.e. $R = \mathbf{w}^T \mathbf{R}$

$$\begin{aligned} \mu &= E(w^T R) = w^T E(R) = w^T(\beta_0 + \beta\mu_F) \\ \sigma^2 &= V(w^T R) = w^T V(R) w = w^T(\beta^T \Sigma_F \beta + \Sigma_\epsilon) w \end{aligned}$$

5.2 Time Series Regression Models

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

We can estimate betas & risks (variances) for one asset at a time. For each $i = 1, \dots N$, fit regression model:

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

over observations $t = 1, \dots, T$

Most models will always include some proxy for the overall economy (e.g. the market). The following is a famous example.

5.2.1 Fama-French 3 Factor Model

3 factors

- Excess Market Return (XMT)
  - Same as in CAPM.
- Small Minus Big (SMB)
  - Captures the size (market cap) of the company/stock.
- High Minus Low (HML)
  - High = value stock; low = growth stock.
  - Measured using book-to-market ratio.

We can use the factor model to estimate 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.

5.3 Statistical Factor Models

Factors are unknown and unobserved

- Need to estimate both β and F
- Problem is ill-posed $\implies$ need constraints

5.3.1 Assumptions

- Asset specific errors ϵ_i are uncorrelated with common factors

$$Cov(\epsilon(t), \mathbf{F}(t)) = \mathbf{0}$$

- The factors $F_j(t)$ are orthogonal, with moments

$$E(\mathbf{F}(t)) = \mathbf{0}$$
$$\mathbb{V}(\mathbf{F}(t)) = I$$

- Errors are serially and contemporaneously uncorrelated across assets

$$\begin{aligned} E(\epsilon) &= \mathbf{0} \\ \mathbb{V}(\epsilon(t)) &= \text{diag}[\left\{\sigma_{\epsilon_i}^2\right\}_{i=1,\dots,N}] = \Sigma_\epsilon \\ Cov[\epsilon(t), \epsilon(s)] &= \mathbf{0} \end{aligned}$$

Resulting moments of returns

$$\mu_R = E[R] = E[\beta_0 + \beta^T F + \epsilon] = \beta_0 + \beta^T \underbrace{E[F]}_{0} + \underbrace{E[\epsilon]}_{0} = \beta_0$$
$$\Sigma_R = \mathbb{V}[R] = \mathbb{V}[\beta_0 + \beta^T F + \epsilon] = \beta^T \mathbb{V}[F]\beta + \Sigma_\epsilon = \beta^T (I \cdot \sigma_F)\beta + \Sigma_\epsilon$$

5.3.2 E.g. Quiz 3 question 2

Consider the general statistical factor model $\underbrace{R_t}_{n\times 1} = \underbrace{\mu}_{n\times 1} + \underbrace{\beta^\top}_{n\times m} \underbrace{F_t}_{m\times 1} + \underbrace{\epsilon_t}_{n\times 1}$, with n variables and $m < n$ factors.

(a) (6 points)

List all the conditions on the 1st and 2nd order moments (i.e., means, variance, and covariances) of the random vectors F_t and ϵ_t, so that the model is identifiable (i.e., there is a unique representation for given moments of R_t).

For i.i.d. data over time, we need

$$\begin{aligned} \mathbb{E}[F_t] &= \mathbb{E}[\epsilon_t] = \mathbf{0} \\ \mathbb{V}[F_t] &= I \\ \mathbb{V}[\epsilon_t] &= \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n^2 \end{bmatrix} \\ \text{Cov}[F_t, \epsilon_t] &= 0 \end{aligned}$$

(b) (6 points)

Specify a statistical factor model with the smallest possible number of factors that can represent R_t with moments $\mathbb{E}[R_t] = \mathbf{0}$ and $\mathbb{V}[R_t] = \begin{bmatrix} 2 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 2 \end{bmatrix}$. We just need 1 factor as follows:

$$
\begin{aligned}
\mu &= \mathbf{0}_{4\times 1} \\
\beta^\top &= \mathbf{1}_{4\times 1} \\
\mathbb{V}[\epsilon_t] &= I_{4\times 4} \\
\Rightarrow \mathbb{V}[R_t] &= \beta^\top \beta + \mathbb{V}[\epsilon_t] \\
&= \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \\
&= \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 2 \end{bmatrix}
\end{aligned}
$$

5.4 Principal component analysis

PCA: constructing a set of variables (components) that capture most of the variability given a set of N assets

It can be thought of as a linear transformation of original variables.

For a random vector $\mathbf{R} = \begin{bmatrix} R_1 & \cdots & R_N \end{bmatrix}'$ with covariance Σ_R (correlation ρ_R), the PC's are linear combinations of $(R_1, \ldots, R_N)$

$$
\begin{gathered}
F_1 = \gamma_1^\top \mathbf{R} = \gamma_{11} R_1 + \cdots + \gamma_{1N} R_N \\
\vdots \\
F_N = \gamma_N^\top \mathbf{R} = \gamma_{N1} R_1 + \cdots + \gamma_{NN} R_N
\end{gathered}
$$

such that:

- $F_1, \ldots, F_N$ are uncorrelated
- Each component has maximum variance

Problem definition

We want to find components F_i, F_j (i.e. find coefficient vectors γ_i) s.t.

- $F_i = \gamma_i^\top \mathbf{X}$ maximizes $\operatorname{Var}(F_i) = \gamma_i^\top \Sigma_R \gamma_i$ subject to $\gamma_i^\top \gamma_i = 1$
- $Cov(F_i, F_j) = \gamma_i^\top \Sigma_R \gamma_j = 0$, for any $j < i$

Solution

Given by eigen-decomposition of Σ_{R}

$$\Sigma_R = \lambda_1 \mathbf{e}_1 \mathbf{e}_1^\top + \cdots + \lambda_N \mathbf{e}_N \mathbf{e}_N^\top = \mathbf{P}\Lambda\mathbf{P}^\top$$

where

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

- P is an orthogonal matrix, i.e. $P^{-1} = P^T$
- $\lambda_1 \geq \cdots \geq \lambda_N \geq 0$

Principal components: $F_j = e_j^T \mathbf{R} = e_{j_1} R_1 + \cdots + e_{jN} R_N$, $\forall j = 1, \ldots, N$

$$\mathbf{F} = \mathbf{P}^T \mathbf{R} = \begin{bmatrix} - & e_1^T & - \\ & \vdots & \\ - & e_N^T & - \end{bmatrix} \mathbf{R}$$

Find $Var(\mathbf{F})$

$$\begin{aligned} \mathbb{V}(\mathbf{F}) &= \mathbb{V}(\mathbf{P}^T \mathbf{R}) \\ &= \mathbf{P}^T \mathbb{V}(\mathbf{R}) \mathbf{P} \\ &= \mathbf{P}^T \Sigma_R \mathbf{P} \\ &= \mathbf{P^T}(\mathbf{P}\ \mathbf{P^T})\mathbf{P} \\ &= \Lambda \end{aligned}$$

Find the loading of R_i on F_j (the beta)

$$\begin{aligned} R_i &= \sum_{j=1}^{N} \beta_{i,j} F_j \quad \text{ where } \beta_{i,j} = \frac{Cov(R_j, F_j)}{Var(F_j)} \\ Cov(\mathbf{R}, \mathbf{F}) &= Cov(\mathbf{R}, \mathbf{P}^T \mathbf{R}) \\ &= Cov(\mathbf{R}, \mathbf{R}) \mathbf{P} \\ &= \Sigma_R \mathbf{P} \\ &= (P \Lambda P^T) P \\ &= P \Lambda \\ &= P \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_N \end{bmatrix} \\ Var(F_j) &= \lambda_j \end{aligned}$$

Total variance of all PC's = variance of original variable

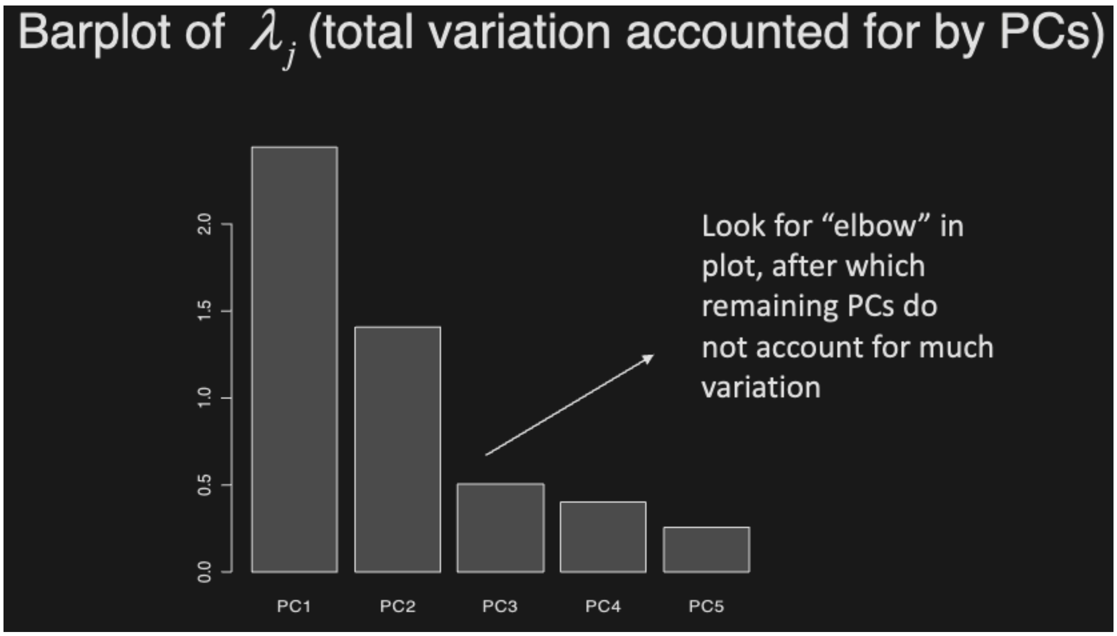
$$\begin{aligned} tr(\Lambda) &= tr(\Sigma) \\ \lambda_1 + \cdots + \lambda_p &= \sigma_1^2 + \ldots \sigma_N^2 \end{aligned}$$

Proportion of total variance explained by each PC is

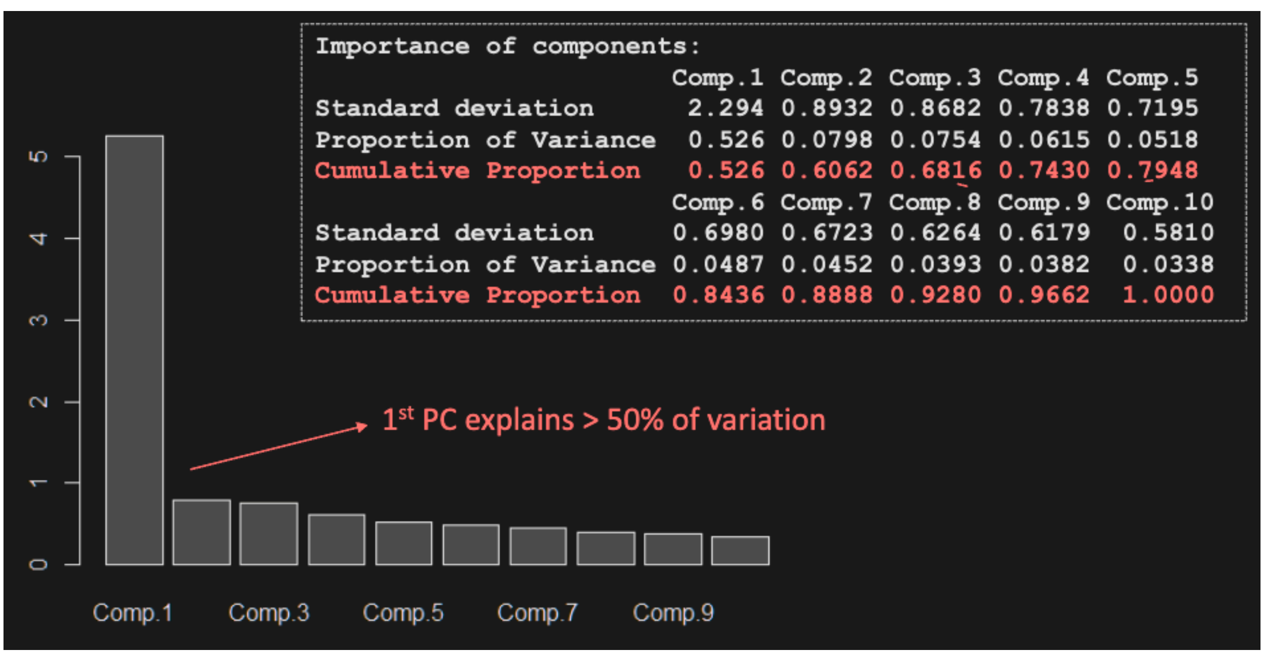
$$\frac{\lambda_j}{\lambda_1 + \cdots + \lambda_N}$$

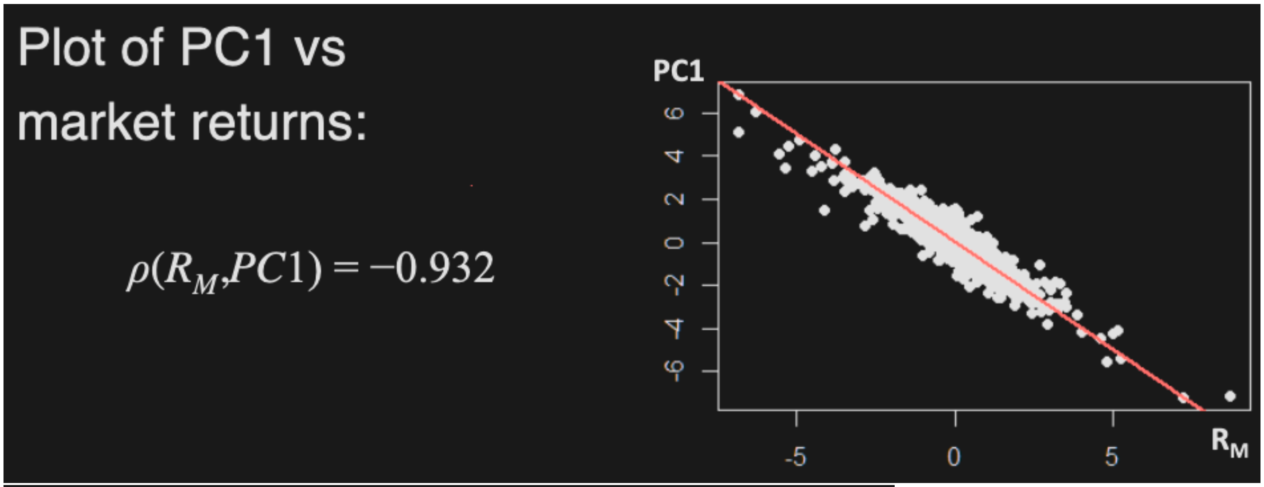
How do we choose the number of PCs?

- We can use a scree plot:



- E.g. In this example, one PC already explains much of the variation





	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
XMT	-0.937	0.0538	-0.0779	0.0512	0.000822
SMB	-0.182	-0.1621	0.0676	0.0772	-0.091280
HML	-0.270	0.0883	-0.2319	0.2694	-0.101209

	Comp.6	Comp.7	Comp.8	Comp.9	Comp.10
XMT	-0.00324	-0.01111	-0.0205	0.0523	-0.0613
SMB	-0.12035	-0.03198	0.0889	0.0107	-0.0200
HML	0.00789	0.00613	0.0092	0.0524	-0.2075

PCA can be used to identify components that explain overall variation of data, but it does not always give meaningful PC's - PC's are just transformations that capture the most variability, they do not explain how data was generated.

For a proper data-generating model, use Factor Analysis:

5.5 Factor Analysis

Assuming $\Sigma_F = \mathbf{I}$, the return variance becomes $\Sigma_R = \underbrace{\beta^T \beta}_{communality} + \underbrace{\Sigma_\epsilon}_{uniqueness}$

We need to estimate β and variances $\sigma^2_{\epsilon_i}$ using maximum likelihood.

A rotation of β (scaling it with orthogonal matrix P) has no effect on the model:

$$\Sigma_R = \left(\beta^\top \mathbf{P}\right)\left(\mathbf{P}^\top \beta\right) + \Sigma_\varepsilon = \beta^\top \left(\mathbf{P}\mathbf{P}^\top\right) \beta + \Sigma_\varepsilon = \beta^\top \beta + \Sigma_\varepsilon$$

We need further constraints on β. A common constraint is to rank factors by explained variability, similar to PCA.

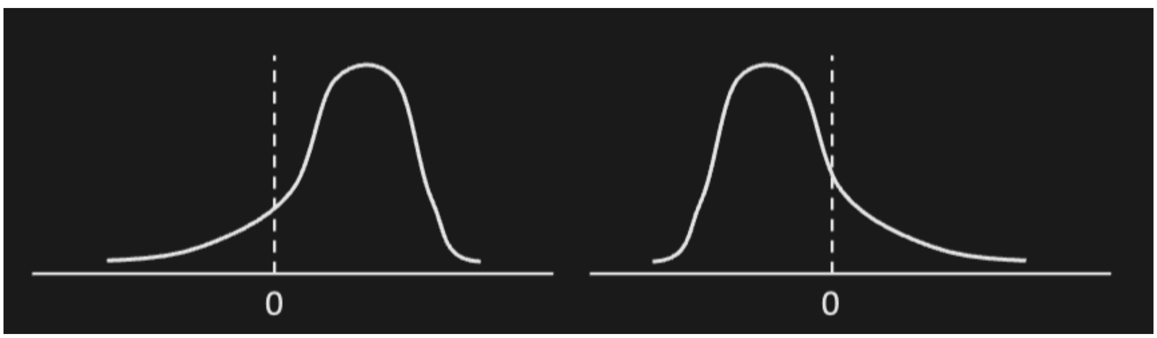
6 W6: Risk Management

6.1 Types of risks

- Market risk: due to changes in market prices
- Credit risk: counterparty doesn't honour obligations
- Liquidity risk: lack of asset tradability
- Operational risk: from organization's internal activities (e.g. legal, fraud, or human error risk)

6.2 Risk measures

- There exists different notions of risk (losing money, bankruptcy, not achieving desired return), but in practice risk measures are used to determine the amount of cash to be kept on reserve
- Return volatility is not a good risk measure. The following distributions have the same σ, but their risk profiles are very different



  - LHS: average return is positive but it has a fat left tail, so returns could be large negative values
  - RHS: average return is negative, but no chance of getting very large negative values; hence less risky

6.3 Value at Risk (VaR)

The VaR is the amount that covers losses with probability $1 - \alpha$.

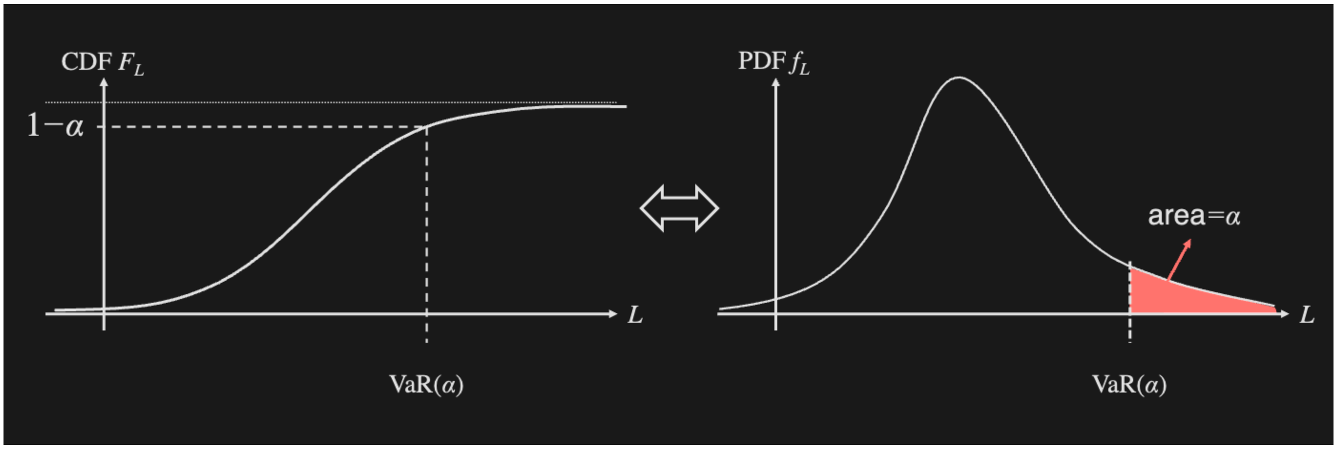
Let L be the loss of an investment over time period T. ($L = -R$, where R is revenue).

The VaR is defined as the $1 - \alpha$ quantile of L for some $\alpha \in (0, 1)$:

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

For a continuous RV with CDF F_L, it is defined as:

$$VaR_\alpha = F_L^{-1}(1 - \alpha)$$



E.g. Consider asset with $N(\mu = 0.03, \sigma^2 = 0.04)$ annual log-returns. Find the 95% confidence level annual VaR for a \$1000 investment in this asset.

Want to find VaR(α) s.t.

$$\begin{aligned}
P(L > VaR) &= P(-R > VaR) \\
&= P(R < -VaR) \\
&= P(S_T - S_0 < -VaR) \\
&= P(S_T < S_0 - VaR) \\
&= P(S_0 e^X < S_0 - VaR) \\
&= P\left(X < \log\left(\frac{S_0 - VaR}{S_0}\right)\right) \\
&= P\left(\frac{X - 0.03}{.2} < \frac{\log\left(1 - \frac{VaR}{S_0}\right) - 0.03}{.2}\right) \\
0.05 &= P\left(Z < \frac{\log\left(1 - \frac{VaR}{S_0}\right) - 0.03}{.2}\right) \\
\implies z = -1.645 &= \frac{\log\left(1 - \frac{VaR}{S_0}\right) - 0.03}{.2} \\
VaR &= (1 - \exp\{-1.645 \cdot 0.20 + 0.03\}) \cdot S_0 \\
&= 258.44
\end{aligned}$$

Formula for VaR given log returns:

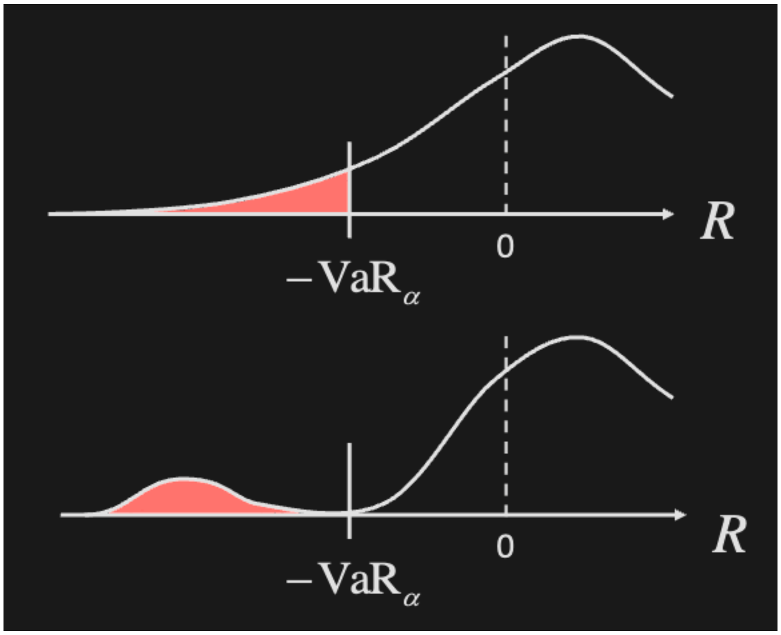
$$\text{VaR} = S_0(1 - \exp\{\mu + z\sigma\})$$

6.3.1 Limitations

VaR can be misleading as it hides tail risk and discourages diversification.

However, it is still widely used due to the Basel framework (banking regulations).

As an example, the following have the same VaR but vastly different risk

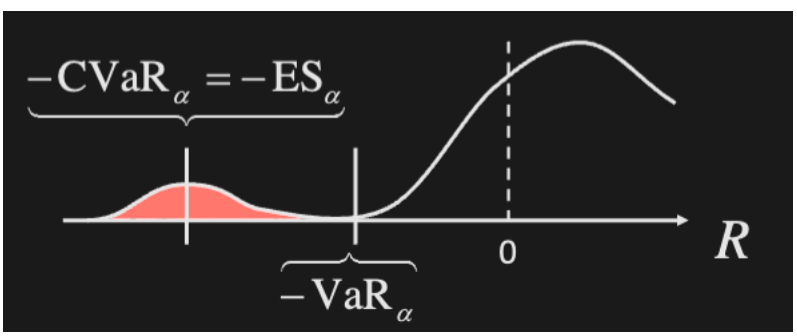


Solution: use conditional VaR / expected shortfall

6.4 Conditional VaR / Expected Shortfall

Defined as the expected value (or average) of losses *beyond* VaR

$$\frac{1}{\alpha}\int_0^\alpha VaR(u)du = E(L|L \geq VaR_\alpha)$$



6.5 Examples

6.5.1 VaR & CVaR of a Normal Variable

If R~N(0, 1), find ES at confidence level α

Let Z_α denote the top α-quantile of the standard normal

Normal pdf:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}$$

Standard normal pdf:

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\}$$

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

More generally, for $L \sim N(\phi, \sigma^2)$

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

6.5.2 E.g. Quiz 3 question 1 - VaR & CVaR of a Uniform Variable

Assume that the loss L of an investment follows continuous Uniform $(-60, 40)$ distribution.

(a) (4 points)

Find the Value-at-Risk (VaR) at confidence level $1 - \alpha = 95\%$.

This is the 95% quantile of the Uniform$(-60, 40)$ loss distribution, which is

$$\text{VaR}_{\alpha=.05} = -60 + 100 * (0.95) = 35$$

(b) (4 points)

Find the Conditional-VaR/Expected Shortfall at the same level.

$$E(L|L \geq 35) = 35 + \frac{40 - 35}{2} = \frac{35 + 40}{2} = 37.5$$

The conditional distribution of the loss above the previous VaR will also be Uniform in $(35, 40)$, therefore the CVaR/ES is the midpoint of that interval.

6.6 Risk measure properties

Let $\rho(L)$ denote a risk measure for an investment with loss L.

A *coherent* risk measure must satisfy the following properties:

1. **Normalized** (the risk of holding no assets is 0)
   - $\rho(0) = 0$
2. **Translation invariance** (adding loss c to portfolio increases risk by c)
   - $\rho(L + c) = \rho(L) + c$, $\forall c \in \mathbb{R}$
3. **Positive homogeneity**
   - $\rho(bL) = b\rho(L)$, $\forall b > 0$
4. **Monotonicity**
   - $L_1 \geq L_2 \implies \rho(L_1) \geq \rho(L_2)$
5. **Sub additivity** (due to diversification)
   - $\rho(L_1 + L_2) \leq \rho(L_1) + \rho(L_2)$

E.g. Show that VaR and CVaR are translation invariant and positively homogeneous

Let $L' = bL + c$, then

$$\begin{aligned}
\mathrm{VaR}_\alpha(L) &= \inf\{x : P(L > x) \leq \alpha\} \\
\mathrm{VaR}_\alpha(L') &= \inf\{x' : P(L' > x') \leq \alpha\} \\
&= \inf\{x' : P(bL + c > x') \leq \alpha\} \\
&= \inf\left\{x' : P\left(L > \frac{x' - c}{b}\right) \leq \alpha \right\} \\
&= \inf\{bx + c : P(L > x) \leq \alpha\} \\
&= b \inf\{x : P(L > x) \leq \alpha\} + c \\
&= b\,\mathrm{VaR}(L) + c
\end{aligned}$$

$$
\begin{aligned}
\mathrm{CVaR}_\alpha(L) &= \frac{1}{\alpha}\int_0^\alpha \mathrm{VaR}_u(L)\,du \\
\mathrm{CVaR}_\alpha(L') &= \frac{1}{\alpha}\int_0^\alpha b\,\mathrm{VaR}_u(L) + c\,du \\
&= b\left(\frac{1}{\alpha}\int_0^\alpha VaR_u(L)\,du\right) + c \\
&= b\cdot \mathrm{CVaR}(L) + c
\end{aligned}
$$

E.g. Consider 2 risky zero-coupon bonds priced at \$95 per \$100 face value. If each one has 4% independent default probability, show that $\mathrm{VaR}_{5\%}$ is not sub-additive.

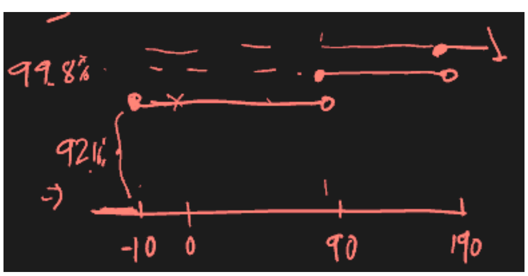
Distribution of L_1 or L_2:

$$
L_i = \begin{cases} -5, & p = 96\% \\ 95, & p = 4\% \end{cases}
$$

$$
\begin{aligned}
\mathrm{VaR}_{5\%}(L_i) &= \inf\{x : P(L > x) \leq 5\%\} \\
&= \inf\{x : P(L \leq x) \geq 95\%\} \\
&= -5
\end{aligned}
$$

Distribution of $L_1 + L_2$:

$$
L_1 + L_2 = \begin{cases} -5-5 = -10, & p = .96^2 = 92.15\% \\ -5+95 = 90, & p = 2(.96)(.04) = 7.68\% \\ 95+95 = 190, & p = (.04)^2 = 0.16\% \end{cases}
$$



$$
\mathrm{VaR}_{5\%}(L_1+L_2) = \inf\{x : P(L_1+L_2 \leq x) \geq 95\%\} = 90
$$

which is greater than

$$
\mathrm{VaR}_{5\%}(L_1) + \mathrm{VaR}_{5\%}(L_2) = -5-5 = -10
$$

This shows that under VaR, owning both bonds is riskier than owning them separately. VaR is thus *incoherent* at the 5% level (it hides tail risk). At 3%, it would be coherent.

E.g. Show that $\text{CVaR}_{5\%}$ is sub-additive.

$$\text{CVaR}_{5\%}(L_1) = \frac{1}{5\%}\int_0^{5\%} \text{VaR}_u(L_1)\,du = \frac{1}{5\%}(95\cdot 4\% + (-5)\cdot 1\%) = 73$$
$$\text{CVaR}_{5\%}(L_1+L_2) = \frac{1}{5\%}\int_0^{5\%} \text{VaR}_u(L_1+L_2)\,du = \frac{1}{5\%}(190\cdot 16\% + 90\cdot 4.84\%) = 93.2$$

We see that $\text{CVaR}_{5\%}(L_1+L_2) = 93.2 \le \text{CVaR}_{5\%}(L_1) + \text{CVaR}_{5\%}(L_2) = 2\cdot 73$

6.7 Entropic VaR

EVaR is a coherent alternative to VaR based on the **Chernoff bound**, which is attained by applying Markov's inequality to e^{tX}. It is an exponentially decreasing upper bound on the tail of a RV based on its MGF.

Markov inequality: for a positive RV X, we have

$$P(X \ge c) \le \frac{E(X)}{c},\ \forall c > 0$$

For loss RV L with MGF $M_L(z) = E(e^{zL}) < \infty$, $\forall z > 0$, we have

$$P(L \ge c) = P(e^{zL} \ge e^{zc}) \le \frac{M_L(z)}{e^{zc}}$$

Bound this by α and solve for c:

$$M_L(z)e^{-zc} \le \alpha \implies c = z^{-1}\ln(\frac{M_L(z)}{\alpha})$$

Thus, EVaR is defined as

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

6.7.1 EVaR of a Normal Variable

The MGF of a Normal variable $L \sim N(\mu, \sigma^2)$ is

$$M_L(Z) = e^{\mu z + \frac{1}{2}\sigma^2 z^2} = E(e^{zL}),\ \forall z$$

EVaR is the infimum of the following:

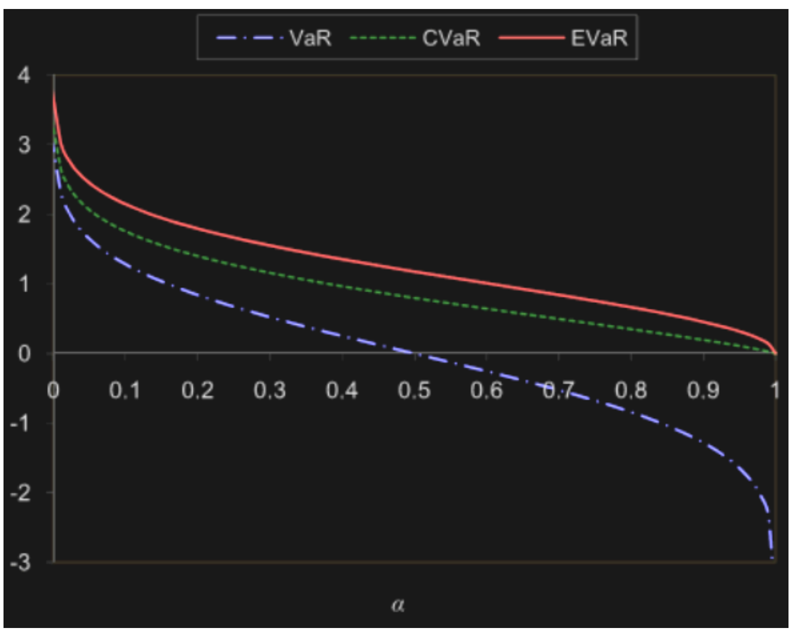
$$\begin{aligned} z^{-1}\ln\left(\frac{M_L(z)}{\alpha}\right) &= \frac{1}{z}\ln\left(\frac{e^{\mu z + \frac{1}{2}\sigma^2 z^2}}{\alpha}\right) \\ &= \frac{1}{z}\left(\mu z + \frac{\sigma^2 z^2}{2} - \ln\alpha\right) \\ &= \mu + z\frac{\sigma^2}{2} - \frac{\ln\alpha}{z} = f(z) \end{aligned}$$

To find infimum (minimum) over z>0, differentiate and set to 0:

$$
\begin{aligned}
0 &= \frac{\sigma^2}{2} + \ln\alpha\left(\frac{1}{z^2}\right) \\
z^* &= \frac{\sqrt{-2\ln\alpha}}{\sigma}
\end{aligned}
$$

So we have

$$
\begin{aligned}
\mathrm{EVaR}_\alpha(L) &= \inf_{z>0}\{f(z)\} = f(z^*) \\
&= \mu + \frac{\sqrt{-2\ln\alpha}}{\sigma}\frac{\sigma^2}{2} - \frac{\ln\alpha}{\frac{\sqrt{-2\ln\alpha}}{\sigma}} \\
&= \mu + \frac{\sqrt{-2\ln\alpha}}{\sigma}\frac{\sigma^2}{2} + \sigma\frac{\ln\alpha^{-1}}{\sqrt{2\ln\alpha^{-1}}} \\
&= \mu + \sigma\sqrt{-\frac{\ln\alpha}{2}} + \sigma\sqrt{\frac{\ln\alpha^{-1}}{2}} \\
&= \mu + 2\sigma\sqrt{\frac{\ln\alpha^{-1}}{2}} \\
&= \mu + \sigma\sqrt{2\ln\alpha^{-1}}
\end{aligned}
$$



6.8 Calculating risk measures

3 ways:

- Parametric modeling
- Historical simulation

- Monte Carlo simulation

Other risk management techniques: stress-testing (worst-cast scenario) and extreme value theory (EVT)

~85% of large banks use historical simulation, the remaining use MC simulation

6.8.1 Parametric modeling

Fitting a distribution to revenues/returns and calculating VaR or CVaR/ES based on distribution

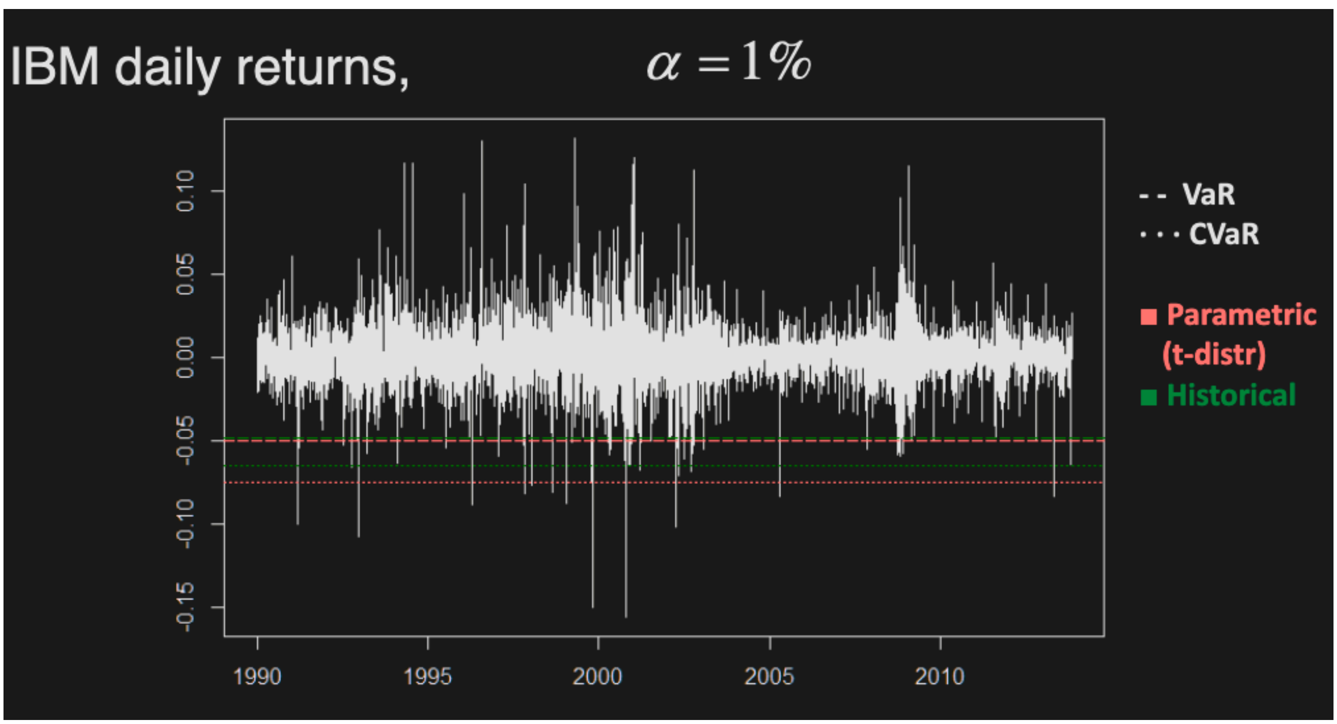
E.g. Assuming net returns of an investment follow a normal distribution, then for an initial capital S_0, the parametric VaR and CVaR at confidence $1-\alpha$ are

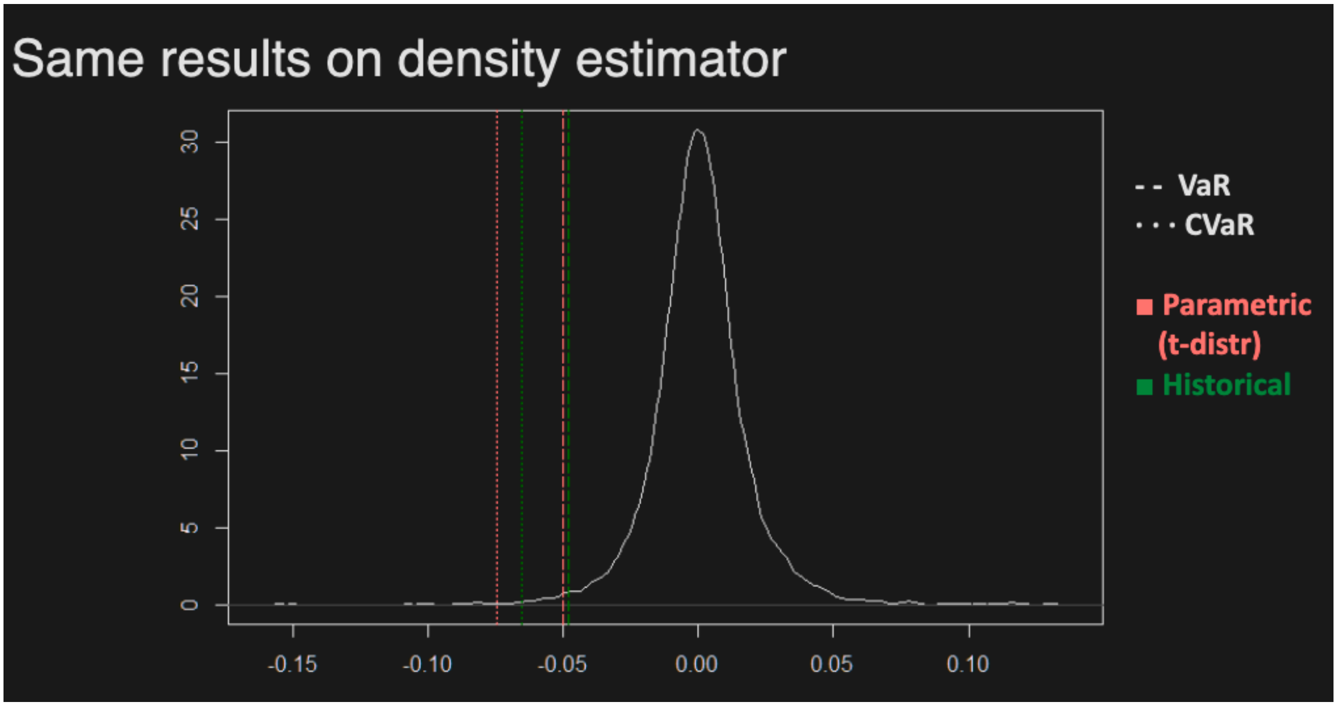
$$\mathrm{VaR}_\alpha = -S_0 \times \{\hat{\mu} + \hat{\sigma}\Phi^{-1}(\alpha)\}$$
$$\mathrm{CVaR}_\alpha = -S_0 \times \left\{\hat{\mu} + \hat{\sigma}\frac{\phi\left(\Phi^{-1}(\alpha)\right)}{\alpha}\right\}$$

where $\hat{\mu}, \hat{\sigma}$ are sample estimates, and Φ, ϕ are standard normal cdf, pdf.

6.8.2 Historical simulation

Instead of assuming a specific distribution, it uses the empirical distribution of returns estimated by historical data.





6.8.3 Monte Carlo simulation

Even if returns are parametrically modelled, their resulting distribution is often intractable.

E.g. Consider a portfolio of 2 assets - one with normal returns, one with t-distributed returns. The distribution the portfolio return is not explicitly known.

We can simulate returns from such a model and treat simulated values as historical returns.

6.9 Time Series Models

Static models assume independence over time (but allow dependence across assets)

6.9.1 RiskMetrics Model

A simple time series model using the exponentially weighted moving average for return volatility

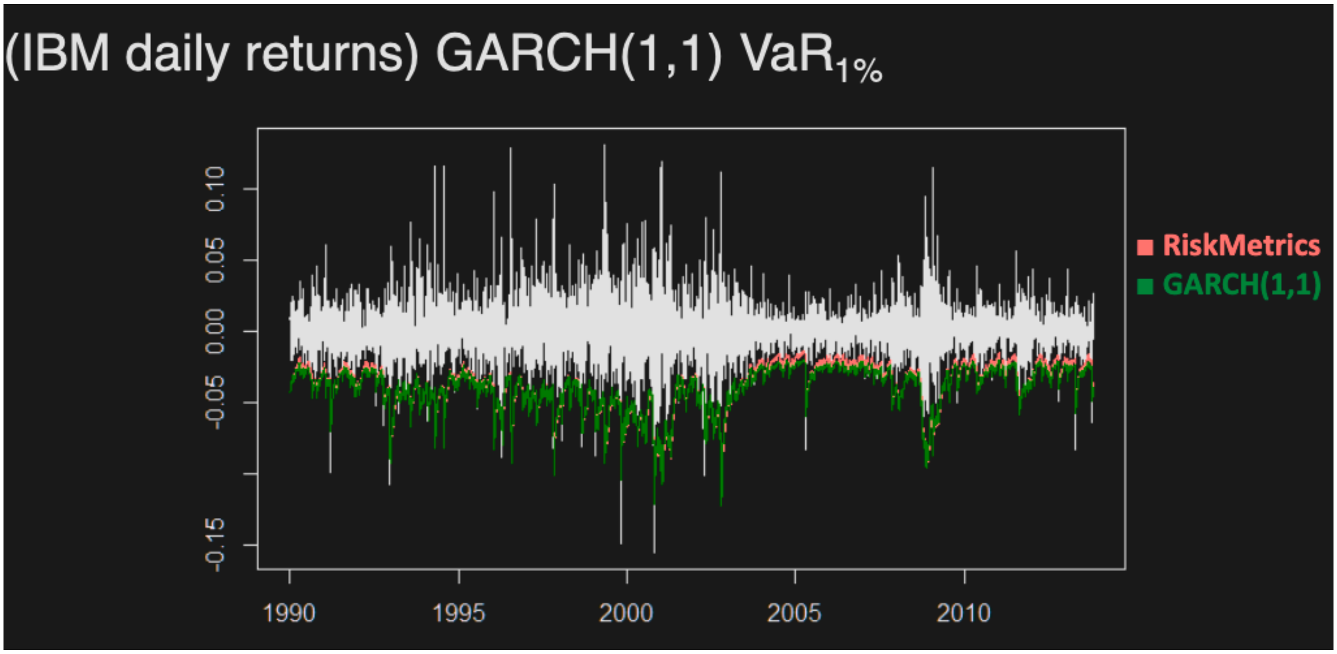
$\sigma_t^2 = \lambda\sigma_{t-1}^2 + (1-\lambda)r_{t-1}^2$ where $r_t \sim N(0, \sigma_t^2)$

Typically, use $\lambda = .94$ for daily returns

6.9.2 GARCH(p, q) Model

$$r_t = \mu + \sigma_t \epsilon_t$$

where ϵ_t are iid and $\sigma_t^2 = \alpha_0 + \sum_{j=1}^{p} \alpha_j r_{t-j}^2 + \sum_{k=1}^{q} \beta_k \sigma_{t-k}^2$



7 W7: Betting Strategies

If we have a sequence of gambles where we have a positive expected payoff, how do we wager our bets for optimal results? We will look at a few different strategies below.

Setup. Consider a sequence of independent & identical gambles with

- Let p = P(win) $\in [0.5, 1]$
- For each \$1 placed, the payoff is $\begin{cases} 1 & p \\ -1 & q \end{cases}$

Starting with initial wealth V_0, assume you bet a constant amount X at each step. Find expected wealth $E(V_n)$ after n steps (ignoring ruin: $V_t \leq 0$ for some $t > 0$)

Define indicator RV of winning i-th bet: $I_i = \begin{cases} 1 & p \\ 0 & q \end{cases}$

$$\begin{aligned}
V_1 &= V_0 + XI_1 - X(1-I_1) = V_0 + X(2I_1 - 1)\\
V_2 &= V_1 + XI_2 - X(1-I_2) = V_1 + X(2I_2 - 1) = V_0 + X(2(I_1+I_2) - 2)\\
&\vdots\\
V_n &= V_0 + X\left(2\underbrace{\sum_{i=1}^{n} I_i}_{\sim Binomial(n,p)} - n\right)\\
E(V_n) &= V_0 + X\left(2E\left(\sum_{i=1}^{n} I_i\right) - n\right)\\
&= V_0 + X(2np - n)\\
&= V_0 + nX\underbrace{(2p-1)}_{>0}
\end{aligned}$$

Notice that if $p < \frac{1}{2}$, $E(V_n)$ could be negative. Otherwise, we can expect to have some positive wealth at time n which increases linearly. The variance increases quadratically.

Assume we bet \$1 at each step. Start with $V_0 = M$. Find the probability of eventual ruin, i.e. $V_n = 0$ for some n.

Let $\pi_i = P(\text{eventual ruin for } V_0 = i)$, $\forall i \geq 1$ and $\pi_0 = 1$

$$\pi_i = \pi_{i+1} \cdot p + \pi_{i-1} \cdot q,\ \forall i \geq 1$$

Assume solution of the form $\pi_i = y^i$

$$y^i = py^{i+1} + qy^{i-1} \overset{i=1}{\implies} y = py^2 + q \implies py^2 - y + q = 0$$

Solve quadratic:

$$\begin{aligned}
y &= \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}\\
&= \frac{-(-1) \pm \sqrt{(-1)^2 - 4pq}}{2p}\\
&= \frac{1 \pm \sqrt{1 - 4p(1-p)}}{2p}\\
&= \frac{1 \pm \sqrt{1 - 4p + 4p^2}}{2p}\\
&= \frac{1 \pm \sqrt{(2p-1)^2}}{2p}\\
&= \frac{1 \pm (2p-1)}{2p}\\
&= \begin{cases} 1 & \text{trivial sol'n}\\ \frac{1-2p+1}{2p} = 1 - \frac{p}{p} = \frac{q}{p} \end{cases}
\end{aligned}$$

If $y = 1$, then $\pi_i = y^i = 1$, $\forall i \geq 0$. This is a trivial solution (probability of ruin = 1 at all times).

So $y = \frac{q}{p}$, and the probability of eventual ruin is $\pi_i = y^i = (\frac{q}{p})^i$, $\forall i \geq 0$

Assume we bet everything (entire wealth) at each step. What is our expected wealth $E(V_n)$ after n steps, *not* ignoring ruin?

$$
\begin{aligned}
V_n &= \begin{cases} V_0 \cdot 2^n & p^n \\ 0 & 1 - p^n \end{cases} \\
E(V_n) &= 2^n V_0 \cdot p^n + 0 \cdot (1 - p^n) = V_0 (2p)^n
\end{aligned}
$$

Note that as $n \to \infty, E(V_n) \to \infty$ since 2p>1. Wealth will grow exponentially.

Assume we bet a fixed fraction f of wealth at each step. What is our expected wealth $E(V_n)$ after n steps?

$$
\begin{aligned}
V_i &= \begin{cases} V_{i-1}(1+f) & p \\ V_{i-1}(1-f) & q \end{cases} \\
V_n &= V_{n-1}(1+f)^{I_n}(1-f)^{1-I_n} \\
&= V_0(1+f)^{\sum_{i=1}^n I_i}(1-f)^{n-\sum_{i=1}^n I_i} \\
E(V_n) &= E(V_0(1+f)^w(1-f)^{n-w}) \qquad \leftarrow \text{w} = |\text{wins}| \\
&= V_0(1-f)^n \underbrace{E\left(\frac{1+f}{1-f}\right)^w}_{\left(q+p\frac{1+f}{1-f}\right)^n}
\end{aligned}
$$

This step uses the PGF (**probabilistic generating function**) of Binomial(n, p):

$$G_w(z) = E(z^w) = (q + pz)^n$$

Continuing:

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

We have exponential growth and a low probability of ruin.

7.1 Kelly Criterion

Bet fraction of wealth that maximizes expected log return (or equivalently log of V_n, or geometric average of returns).

Note: by Jensen's inequality, maximizing log wealth != maximizing wealth, i.e.

$$E(\log V_n) \neq \log(E(V_n))$$

7.1.1 What is the optimal value of the fraction?

$$\begin{aligned}
E\left(\log \frac{V_n}{V_0}\right) &= E\left(\log\left(\frac{\cancel{V_0}(1+f)^w(1-f)^{n-w}}{\cancel{V_0}}\right)\right) \\
&= E(w\log(1+f) + (n-w)\log(1-f)) \\
&= \log(1+f)E(w) + \log(1-f)(n-E(w)) \\
&= \log(1+f)np + \log(1-f)(n-np) \\
&= \log(1+f)np + \log(1-f)nq \\
&= G(f)
\end{aligned}$$

Now maximize G(f) w.r.t. f and set to 0

$$\begin{aligned}
\frac{dG(f)}{df} &= \frac{d}{df}(\log(1+f)np + \log(1-f)nq) \\
0 &= n\left(\frac{1}{1+f}p - \frac{1}{1-f}q\right) \\
\frac{p}{1+f} &= \frac{q}{1-f} \\
p(1-f) &= q(1+f) \\
p-q &= f(p+q) \\
f^* &= p-q = 2p-1
\end{aligned}$$

The optimal fraction is the difference between P(win) and P(lose).

7.1.2 E.g. Quiz 4 question 1

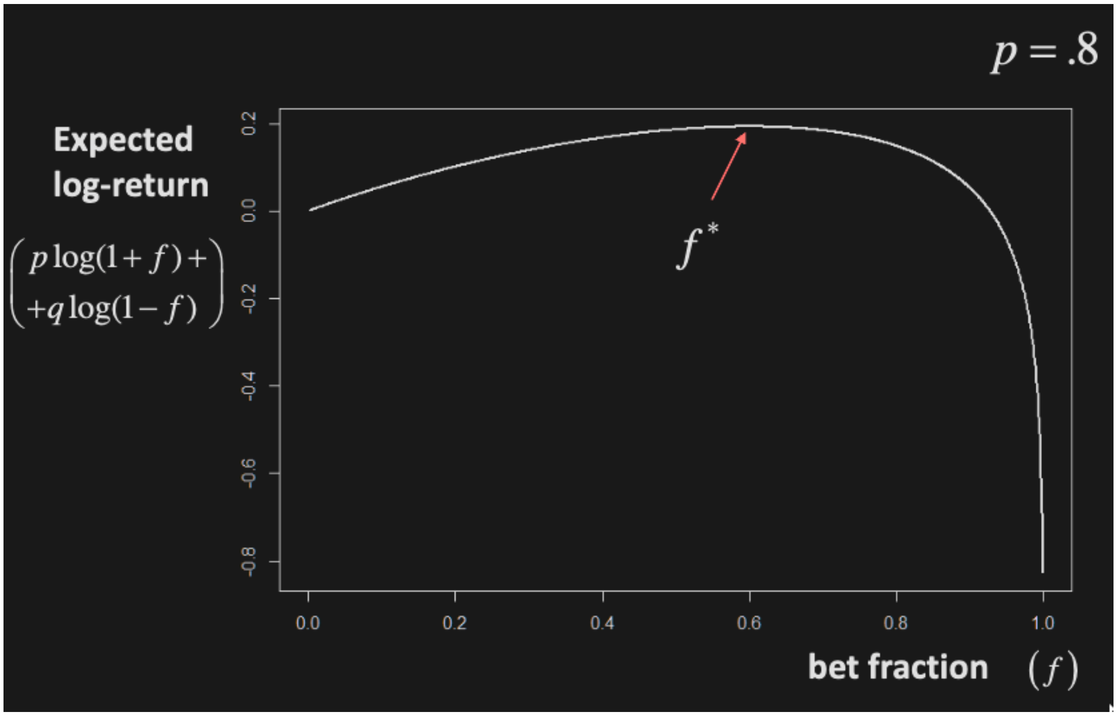
Consider a bet with equal odds (i.e., win/lose \$1 for each \$1 you bet) and probability of winning $p = 0.52$.

(a) What is the Kelly-optimal proportion (f^*) of wealth to wager at each bet?

For even odds, the Kelly-optimal fraction is equal to the magnitude of the "edge", i.e. $f^* = p - q = 2p - 1 = 0.52 - 0.48 = 0.04 = 4\%$.

(b) Starting with initial wealth $V_0 = 100$, write a formula for your wealth after winning 5 and losing 4 bets with the above strategy.

$$
\begin{aligned}
V_9 &= V_0\,(1+f^*)^5\,(1-f^*)^4 \\
&= 100(1.04)^5(0.96)^4
\end{aligned}
$$



7.1.3 What is the geometric average of the returns as $n \to \infty$?

Denoting the growth rate from $i-1$ to i with $r_i = \frac{V_i}{V_{i-1}}$, the geometric average is

$$
\begin{aligned}
\sqrt[n]{r_1 \times \cdots \times r_n} &= \left[\prod_{i=1}^{n} (1+f)^{I_i}(1-f)^{1-I_i}\right]^{1/n} \\
&= (1+f)^{\sum_{i=1}^{n} I_i/n}(1-f)^{\sum_{i=1}^{n}(1-I_i)/n} \\
&= (1+f)^{W_n/n}(1-f)^{1-W_n/n}
\end{aligned}
$$

where $W_n \sim Binom(n,p)$

By SLLN, $\frac{W_n}{n} \to p$ as $n \to \infty$, so the geometric average $\to (1+f)^p(1-f)^q$

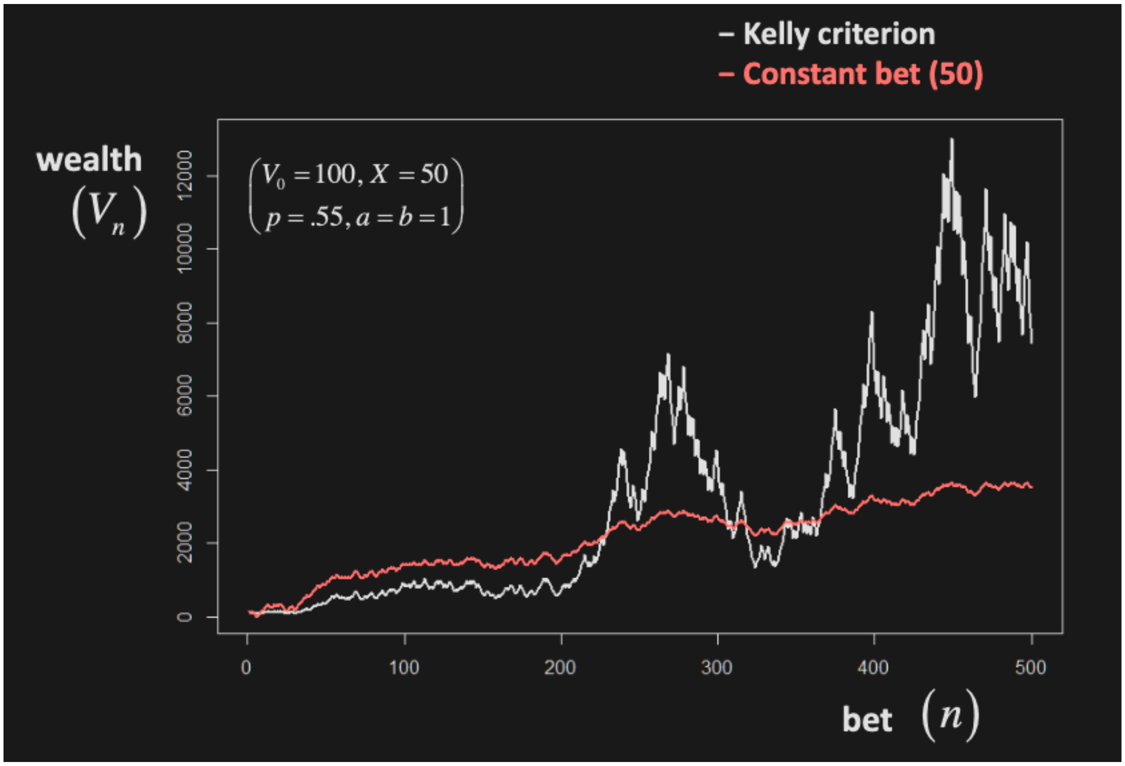
7.1.4 General Setup

Now consider a general sequence of bets, where \$1 bet +\$a if win and −\$b if lose. (In previous examples, a = b = 1, and the bet is favourable, i.e.$pa > qb$.)

The Kelly criterion optimal fraction to bet is: (Proved in PS 7.1b)

$$f^* = \frac{pa - qb}{ab}$$

In the following example, f = 0.55 - 0.45 = 0.1



7.1.5 Investing Example

Now consider the following. We have a…

- risk free asset with return r_f
- risky asset with return R: $E(R) = \mu, \mathbb{V}(R) = \sigma^2$

We invest fraction f of wealth into risky asset & remaining (1–f) into risk-free asset.

Apply Kelly Criterion to find f that maximizes logarithm of wealth:

$$\begin{aligned}
V_1 &= fV_0 \cdot (1+R) + (1-f)V_0 \cdot (1+r_f) \\
&= V_0[\cancel{f} + fR + 1 \cancel{-f} + r_f - fr_f] \\
&= V_0[1 + r_f + f(R - r_f)]
\end{aligned}$$

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

Use Taylor expansion for $x_0 = \log(1 + r_f + f(R_t - r_f))$ around $1 + r_f$

$g'(x_0 + \delta) \approx g(x_0) + g'(x_0)\delta + \frac{1}{2}g''(x_0)\delta^2$

Applying this, we get

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

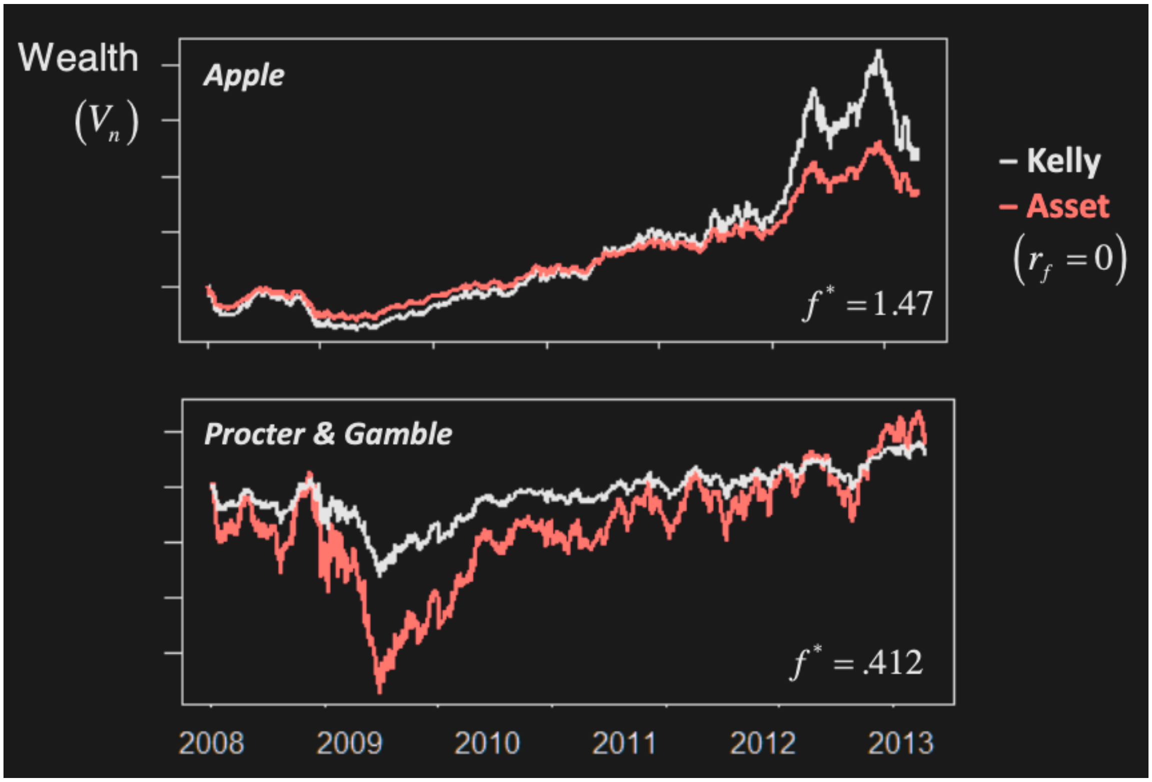
$$\begin{aligned}
E(\log(1 + r_f + f \cdot (R_t - r_f))) &\approx \log(1 + r_f) + \frac{1}{1 + r_f} f \cdot \underbrace{E(R_t - r_f)}_{\mu - r_f} + \frac{1}{2}\left(-\frac{1}{(1 + r_f)^2}\right) f^2 \cdot \underbrace{E[(R_t - r_f)^2]}_{(\sigma^2 + \mu^2) + r_f^2 - 2\mu r_f} \\
&\approx \log\left(1 + r_f\right) + f\frac{(\mu - r_f)}{(1 + r_f)} - f^2 \frac{\sigma^2 + \left(\mu - r_f\right)^2}{2\left(1 + r_f\right)^2} \\
&= G(f)
\end{aligned}$$

Differentiate w.r.t. f and set to 0:

$$\begin{aligned}
\frac{\partial}{\partial f} G(f) &= 0 \\
\frac{\mu - r_f}{1 + r_f} - f\frac{\sigma^2 + (\mu - r_f)^2}{(1 + r_f)^2} &= 0 \\
f^* &= \left(1 + r_f\right) \frac{(\mu - r_f)}{\sigma^2 + (\mu - r_f)^2}
\end{aligned}$$

Since $\sigma^2 \gg \mu - r_f$, we have

$$f^* \approx \left(1 + r_f\right) \frac{(\mu - r_f)}{\sigma^2}$$



7.1.6 Theoretical properties

In the long term ($n \to \infty$) with probability 1, a strategy based on Kelly criterion:

1. Maximizes limiting exponential growth rate of wealth

2. Maximizes median of final wealth
   - Half of distribution is above median & half below it

3. Minimizes the expected time required to reach a specified goal for the wealth

7.1.7 Criticism

- Can have considerable wealth volatility (b/c of multiplicative bet amounts)
- Does not account for the uncertainty in probability of winning

  – Many practitioners use fractional or partial Kelly, i.e. using smaller than Kelly fraction (e.g., f*/2)

- In practice, investing horizons are not infinite and there are many other considerations (e.g. transaction costs, short-selling limits etc)

8 W8: Statistical Arbitrage

Statistical Arbitrage (StatArb) refers to trading strategies that utilize the "statistical mispricing" of related assets

StatArb strategies are typically short term and market neutral, involving long & short positions simultaneously

Examples of StatArb strategies:

- Pairs trading
- Index Arbitrage
- Volatility Arbitrage
- Algorithmic & High Frequency Trading

8.1 Pairs Trading

- Original & most well-known StatArb technique developed by Morgan Stanley quants
- Profit not affected by overall market movement (market neutral)
- Contrarian strategy profits from price convergence of related assets

8.1.1 Main idea

1. Select pair of assets "moving together", based on certain criteria
2. If prices diverge beyond certain threshold, buy low sell high
3. If prices converge again, reverse position and profit

8.1.2 Example

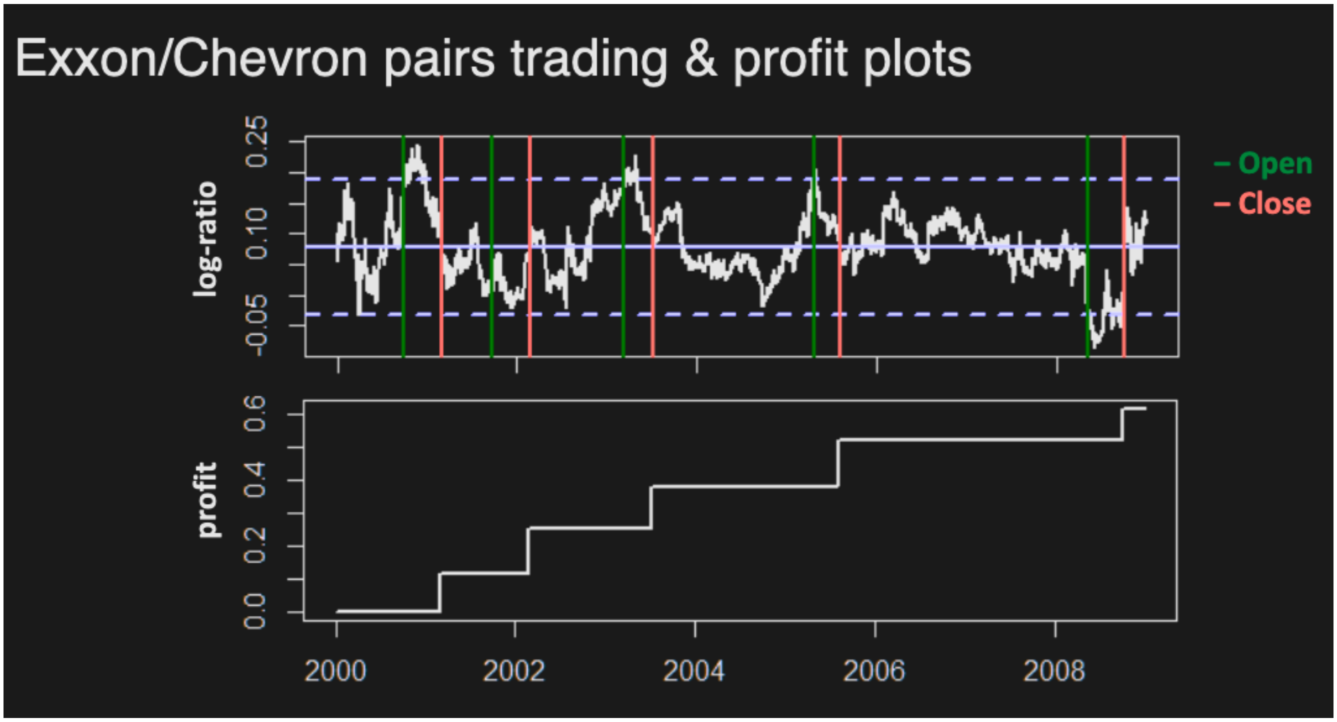
Let L = price of lower asset, H = price of higher asset.

- Open the position when prices diverge: buy \$1 of low asset ($\frac{1}{L_0}$ units), sell \$1 of high asset ($\frac{1}{H_0}$ units)
  - cost = $\frac{1}{L_0}L_0 - \frac{1}{H_0}H_0 = 0$
- Close the position when prices converge
  - profit = $\frac{1}{L_0}L_c - \frac{1}{H_0}H_c$ (c stands for closing)

Profitability is determined by **asset price ratios** (hence the use of log ratios for modelling):

$$\frac{H_c}{H_0} - \frac{L_c}{L_0} < or > 0$$
$$\frac{H_c}{H_0} < or > \frac{L_c}{L_0}$$
$$\log\left(\frac{H_c}{L_c}\right) < or > \log\left(\frac{H_0}{L_0}\right)$$

The strategy is market neutral, i.e. profitability is not affected by market movement - Assets typically have common market betas



8.1.3 What can go wrong?

Prices may not converge.

8.1.4 Factors to consider

1. Which pairs to trade
2. When to open trade
3. What amounts to buy/sell
4. When to close trade
5. When to bail out of trade

Most of these decisions involve trade-offs, so how do we select pairs to trade?

- Profitable pairs must have log-ratio with strong mean reversion
  - Note: Mean reversion is *not* the same as simply having constant mean

8.1.5 Mean Reversion

Suggests log-ratio process $\{X_t\}$ is stationary

- $E(X_t) = \mu$, $\forall t$
- $Var(X_t) = \sigma^2 < \infty$, $\forall t$
- $Cov(X_t, X_s) = Cov(X_{t+r}, X_{s+r})$, $\forall r, s, t$

- Autocorrelation function $\rho(h)$, $\forall h = 0, 1, ...$ describes linear dependence at lag $h = |t - s|$

Stationarity ensures process will revert back to its mean within reasonable time.

E.g. Let $X_t = \log(\frac{H_t}{L_t}) \sim^{iid} N(0, \sigma^2)$, $\forall t = 1, 2, ...$

If $X_0 = 2\sigma$, what is the expected time until $X_T \leq 0$? I.e. until $\log H_t - \log L_t < 0$

On any day t, $P(X_t \leq 0) = \frac{1}{2}$

Let T = # days until $\{X_t \leq 0\}$ for the first time. T is called hitting time, it is equal to # trials until 1st success (if $X \leq 0$), so $T \sim Geom(p = \frac{1}{2})$ which has prob mass function

$$p_T(t) = \left(\frac{1}{2}\right)^t, \forall t \geq 1$$

The expected time is hence

$$E(T) = \frac{1}{p} = 2$$

E.g. 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*.

Let $T_c = \{$first time standard BM with $W_0 = 0\}$ hits level c Let $M_t = \max\{W_u; 0 \leq u \leq t\} \implies M_t \sim |W_t|$

This means:

$$P(T_c \leq t) = \ P(M_t \geq c) = P(|W_t| \geq c) = 2\Phi\left(-\frac{c}{\sqrt{t}}\right)$$

PDF of T_c is given by $f(t) = \frac{d}{dt}P(T_c \leq t)$

$$\begin{aligned}
f(t) &= \frac{d}{dt}\left[2\Phi\left(-\frac{c}{\sqrt{t}}\right)\right] \\
&= 2\phi\left(-\frac{c}{\sqrt{t}}\right) \cdot \frac{d}{dt}\left(-\frac{c}{\sqrt{t}}\right) \\
&= 2\phi\left(-\frac{c}{\sqrt{t}}\right) \cdot \left(\frac{1}{2}\frac{c}{\sqrt{t^3}}\right) \\
&= \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\frac{c^2}{t}}\frac{c}{\sqrt{t^3}} \\
E(T_c) &= \int_0^\infty t \cdot f(t)\, dt = \int_0^\infty t \cdot \frac{c}{\sqrt{2\pi t^3}}e^{-\frac{c^2}{2t}}\, dt \\
&= \frac{c}{\sqrt{2\pi}}\int_0^\infty \frac{1}{\sqrt{t}}e^{-c^2/2t}\, dt \\
&\geq \frac{c'}{\sqrt{2\pi}}\int_0^a \frac{1}{\sqrt{t}}\, dt + \int_a^\infty \frac{1}{\sqrt{t}}e^{-c^2/2t}\, dt \to \infty \\
\implies E(T_c) &= \infty
\end{aligned}$$

8.1.6 E.g. Quiz 4 question 2

Consider the following prices of two assets on different days:

stock \ day	1	2	3	4	5
P	35.63	35.59	35.49	35.43	35.41
S	33.74	32.72	31.69	32.45	34.98

Calculate the pairs-trading strategy profit of going long \$100 of S & short \$100 of P on day 1 , and unwinding the position on day 5 .

Solution: On day 1, we buy/long $100/S_1$ shares of S and short-sell $100/P_1$ shares of P. Since the strategy has 0 set-up cost (long & short \$100), the profit is just the payoff when you unwind the position:

$$\begin{aligned} \text{profit} &= S_5 \times (\text{ shares long }) - P_5 \times (\text{ shares short }) \\ &= S_5 \times 100/S_1 - P_5 \times 100/P_1 \\ &= 100 \times (34.98/33.74 - 35.41/35.63) \quad (= 4.29262) \end{aligned}$$

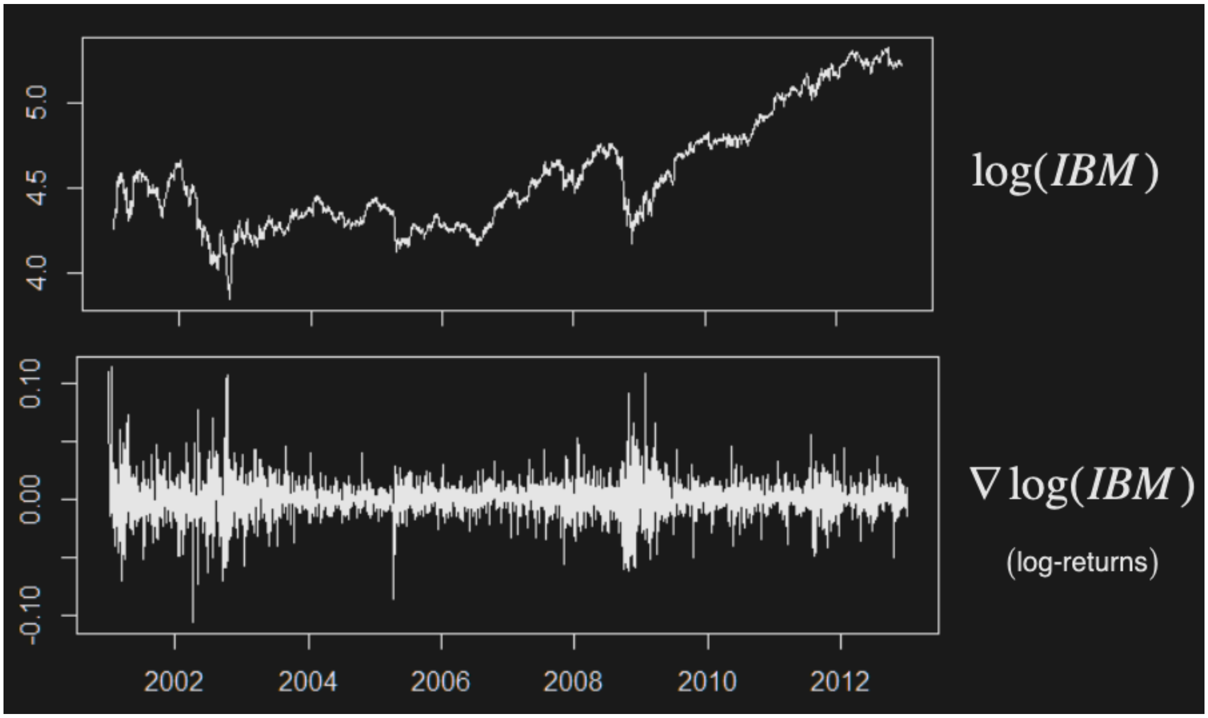
8.2 Integrated Series

A non-stationary time series $\{X_t\}$ whose difference $\{\nabla X_t = X_t - X_{t-1}\}$ is stationary

Asset log prices are *not* stationary - will need to apply differencing

Although $r_t = \log(\frac{S_t}{S_{t-1}})$ follows a stationary process, $\log S_t = \log S_0 + \sum_{i=1}^{t} r_i$ is a random walk

Example: IBM stock price before and after differencing



8.2.1 Cointegration

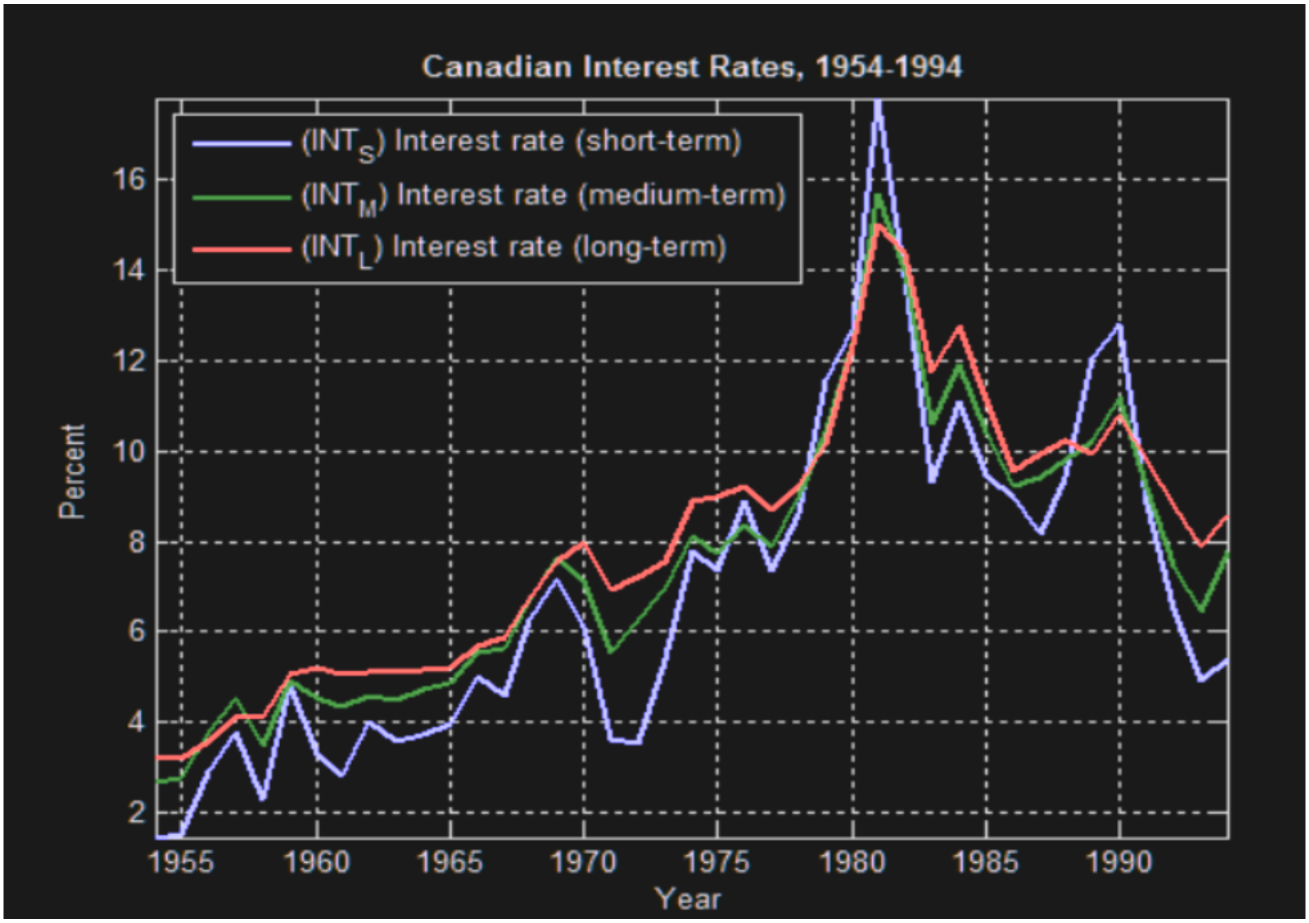
Two integrated series $\{X_t, Y_t\}$ are *cointegrated* if there exists a linear combination of them that is stationary.

Consider a vector of time series x_t. If each element becomes stationary after differencing, but a linear combination $\alpha' x_t$ is already stationary, then x_t is said to be co-integrated with α, which is the co-integrating vector.

There may be several such co-integrating vectors so that α becomes a matrix. Interpreting $\alpha' x_t = 0$ as a long run equilibrium, co-integration implies that **deviations from equilibrium are stationary, with finite variance**, even though the series themselves are non-stationary and have infinite variance.

For pairs trading, we want to find assets which are cointegrated (their log difference is mean reverting, and thus stationary)

E.g. ST, MT, and LT interest rates are co-integrated - they move together but behave as random walks individually



E.g. Let $\{W_t\}$ be random walk, and $\begin{cases} X_t = W_t + \varepsilon_t \\ Y_t = W_t + \eta_t \end{cases}$ where $\varepsilon_t, \eta_t \sim^{\text{iid}} N(0, \sigma^2)$

Show that $\{X_t, Y_t\}$ are cointegrated

First, we need to show X_t, Y_t are integrated (not stationary, with stationary 1st order differences).

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

Which is the same case for Y_t

Next, show cointegration by showing $X_t - X_{t-1}$ is stationary.

$$\nabla X_t = X_t - X_{t-1} = W_t + \epsilon_t - W_{t-1} - \epsilon_{t-1} = \underbrace{(W_t - W_{t-1})}_{\nu_t \sim WN(0,\sigma^2)} + \epsilon_t - \epsilon_{t-1}$$

$$E(\nabla X_t) = E(\nu_t) + E(\epsilon_t) - E(\epsilon_{t-1}) = 0 \implies \text{ not stationary}$$

$$V(\nabla X_t) = V(\nu_t) + V(\epsilon_t) + V(\epsilon_{t-1}) = \sigma^2 + 2\sigma_\epsilon^2$$

$$Cov(\nabla X_t, \nabla X_s) = \gamma(t-s)$$

If |t-s| > 1, $Cov(\nabla X_t, \nabla X_s) = Cov(\nu_t + \epsilon_t - \epsilon_{t-1}, \nu_s + \epsilon_s - \epsilon_{s-1}) = 0$

If |t-s| $= 1$, $Cov(\nabla X_t, \nabla X_s) = Cov(\nu_t + \epsilon_t - \epsilon_{t-1}, \nu_{t+1} + \epsilon_{t+1} - \epsilon_t) = Cov(\epsilon_t, -\epsilon_t) = -\sigma_\epsilon^2$

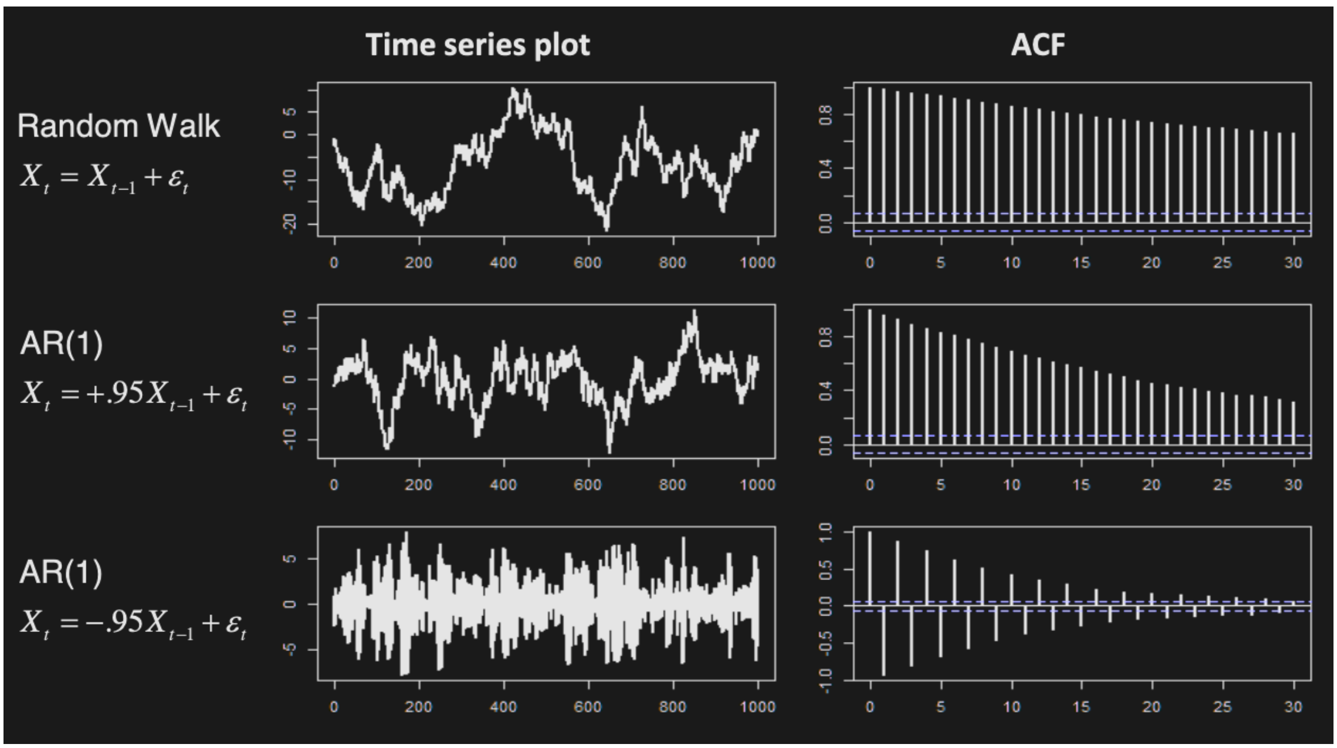
∇X_t is thus stationary $\implies X_t \sim I(1)$ is integrated of order 1 (and similarly for Y_t)

$X_t - Y_t = w_t + \epsilon_t - (w_t + n_t) = \epsilon_t - n_t \overset{iid}{\sim} N(0, \sigma_\epsilon^2 + \sigma_n^2)$ is stationary, so (X_t, Y_t) are cointegrated.

8.2.2 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 $\begin{cases} H_0 : \beta = 1 \\ H_1 : \beta < 1 \end{cases}$



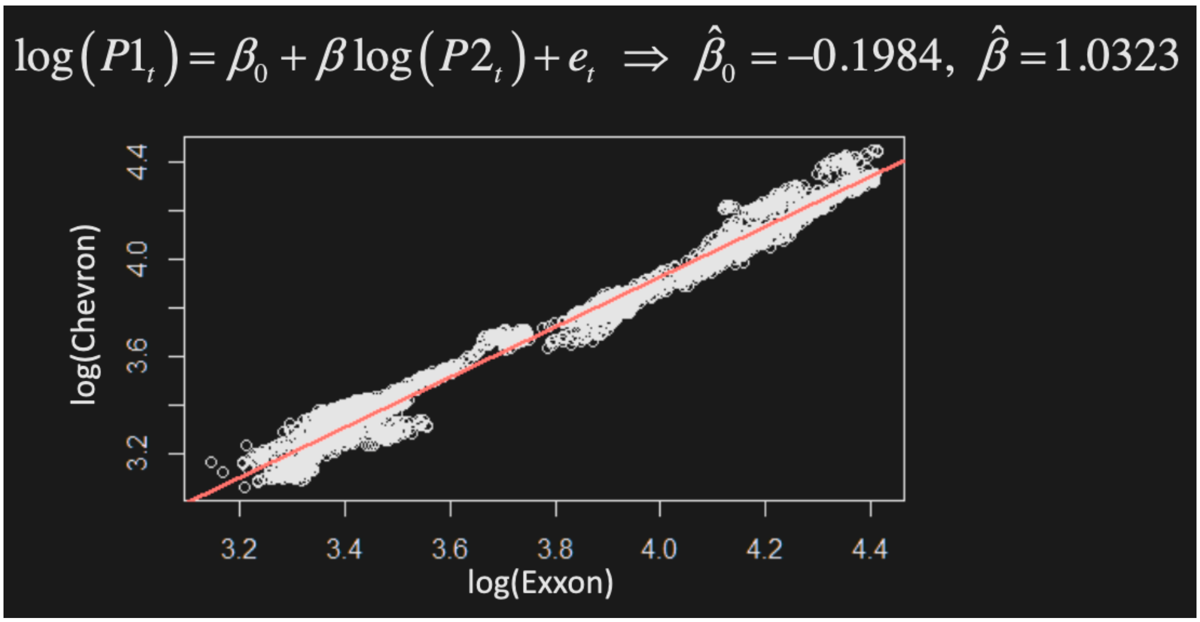
Model	Test statistic	P-value
$X_t = X_{t-1} + \varepsilon_t$	−1.9027	0.6195
$X_t = +.95X_{t-1} + \varepsilon_t$	−5.6161	<<.01
$X_t = -.95X_{t-1} + \varepsilon_t$	−232.4851	<<.01

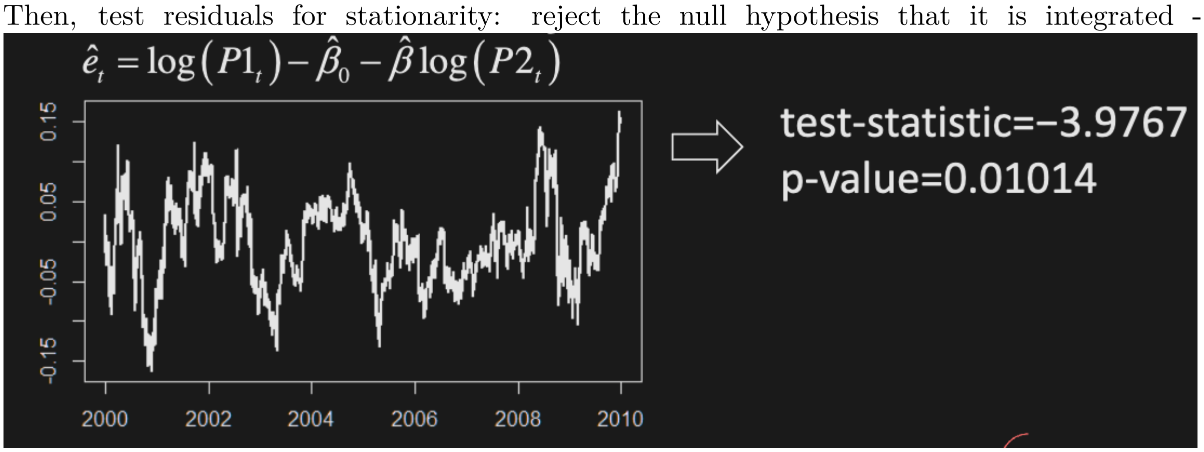
For random walk, we fail to reject the null hypothesis that X_t is integrated.

Issue: we don't know which linear combination to check for stationarity

Two-step method

- Estimate linear relationship between variables

- Test resulting difference series (residuals) for stationarity
- Example: regress Chevron on Exxon log-prices
  - 

  - Then, test residuals for stationarity: reject the null hypothesis that it is integrated -


- Problems:
  - Regressing P1 and P2 can give different results than regressing the other way around.
  - There is estimation error for residuals.

- Can be used to address spurious regression
  - Results of random walk (integrated series) regressions are NOT reliable

* Consider 2 independent random walks $\{W_t, V_t\}$
* When you regress $W_t = \beta_0 + \beta V_t + e_t, t = 1, \dots, n$ you are NOT guaranteed that $\hat{\beta} \to 0$ as the sample size $n \to \infty$ (i.e. not consistent)

Vector Error Correction models (VECM)

Combined treatment of dynamics & cointegration, using Vector AutoRegressive (VAR) models

8.3 Index Arbitrage

- Indices measure value/performance of financial markets
  - 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
- Financial products based on indices essentially offer a sophisticated version of multivariate cointegration
  - For an index of N assets $\{S_i\}_{i=1}^N$ w/ weights $\{w_i\}_{i=1}^N$, the index level is $I(t) = \sum_{i=1}^N w_i \times S_i(t)$
  - It has a co-integrating relationship with $F(t)$, an instrument tracking index (e.g. futures)

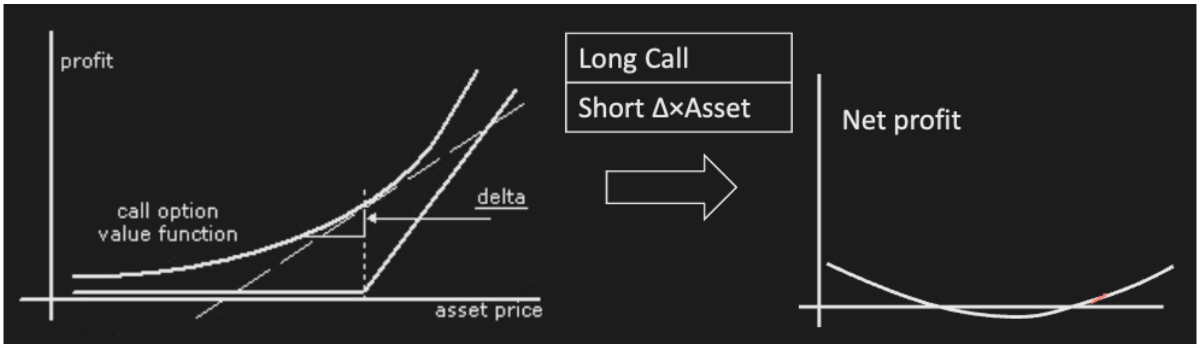
$$F(t) - I(t) = F(t) - \sum_{i=1}^{N} w_i \times S_i(t) \sim \text{ stationary}$$

8.4 Volatility Arbitrage

- VolArb is implemented with derivatives, primarily options
- The higher the volatility, the higher the option price

Consider European options:

- For Black-Scholes formula, the only unobserved input is volatility σ, which has to be estimated
- Implied volatility σ_ii s the input that makes Black-Scholes price equal to observed market price
  - not estimated from underlying asset dynamics
- If volatility will increase in the future, beyond what current options prices warrant (implied vol), some possible strategies are:
  - straddles (long at the money call and put)
  - strangles (long out of the money call and put)
  - delta-hedged long call or put
- Delta-neutral strategies eliminate effects of asset movement

- 

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

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

9 W9: Monte Carlo Simulation

9.1 Numerical Option Pricing

3 basic numerical option pricing methods:

1. Binomial trees
2. Finite difference (based on Black Scholes PDE)
3. Monte Carlo simulation (based on SDE for asset prices & risk neutral valuation)

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

9.2 Multivariate Normal Properties

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

9.2.1 Marginals

$$\mathbf{X}_1 \sim \mathrm{N}\left(\mu_1, \Sigma_{11}\right)$$

9.2.2 Linear combinations

$$\mathbf{a} + \mathbf{B}^\top \mathbf{X} \sim \mathbf{N}\left(\mathbf{a} + \mathbf{B}^\top \mu, \mathbf{B}^\top \Sigma \mathbf{B}\right)$$

9.2.3 Conditionals

$$\mathbf{X}_1 \mid (\mathbf{X}_2 = \mathbf{x}) \sim \mathbf{N}\left(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}\left(\mathbf{x} - \mu_2\right), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)$$

Notice how $\Sigma_{12} = 0 \iff \mathbf{X}_1 | (\mathbf{X}_2 = x) \sim N(\mu_1, \Sigma_{11})$

9.3 Brownian Motion

$W_T \sim N(0, T)$ forms the building block of continuous stochastic models

> **Recall Ito Processes from STAC70:**
>
> A (one-dimensional) Itô process is a stochastic process $\{X_t\}_{t \geq 0}$ of the form
>
> $$X_t = x_0 + \int_0^t a_s ds + \int_0^t b_s dB_s$$
>
> where $\{a_t\}_{t \geq 0}$ and $\{b_t\}_{t \geq 0}$ are adapted processes such that the integrals are defined. Equivalently, we can write this as
>
> $$dX_t = a_t dt + b_t dB_t$$
>
> The Brownian motion $\{B_t\}_{t \geq 0}$ is an Itô process. (Pick $a_t \equiv 0$ and $b_t \equiv 1$.)
>
> The general Brownian motion with (constant) drift $\mu \in \mathbb{R}$ (and constant volatility $\sigma > 0$) is an Itô process. (Pick $a_t \equiv \mu$ and $b_t \equiv \sigma$.)
>
> $$X_t = x_0 + \mu t + \sigma B_t$$

9.3.1 Standard Brownian Motion

$\{W_t\}$ with the following properties:

- $W_0 = 0$
- $(W_t - W_s) | W_s \sim N(0, t - s)$

9.3.2 Arithmetic Brownian Motion

$\{X_t\}$ with drift μ and volatility σ and the following properties:

- $X_0 = 0$
- $(X_t - X_s)|X_s \sim N(\mu(t-s), \sigma^2(t-s))$

$\implies X_t|X_s = x \sim N(x + \mu(t-s), \sigma^2(t-s))$

SDE form:

$$\begin{aligned} X_t - X_0 &= \int_0^t \mu ds + \int_0^t \sigma \, dW_s \\ &= \mu t + \sigma(W_t - W_0) \\ dX_t &= \mu dt + \sigma dW_t \end{aligned}$$

E.g. For $dX_t = \mu dt + \sigma dW_t$, find distribution of $X_t|X_s = x$

For $s < t$

$$\begin{aligned} \begin{bmatrix} X_s \\ X_t \end{bmatrix} &\sim N(\mu \begin{bmatrix} s \\ t \end{bmatrix}, \sigma^2 \begin{bmatrix} s & s \\ s & t \end{bmatrix}) \\ \implies \begin{bmatrix} X_t \\ X_s \end{bmatrix} &\sim N(\mu \begin{bmatrix} t \\ s \end{bmatrix}, \sigma^2 \begin{bmatrix} t & s \\ s & s \end{bmatrix}) \\ \implies X_t|X_s = x &\sim N\left(\mu t + \sigma^2 s \frac{1}{\sigma^2 s}(x - \mu s), \sigma^2\left(t - s\frac{1}{s}s\right)\right) \\ &\sim N(x + \mu(t-s), \sigma^2(t-s)) \\ Cov(W_s, W_t) &= Cov(W_s, W_s + (W_t - W_s)) \\ &= Cov(W_s, W_s) + \cancel{Cov(W_s, (W_t - W_s))} \\ &= Var(W_s) \\ &= s \end{aligned}$$

For $s \in (0, t)$ (**Brownian Bridge**)

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

9.3.3 Geometric Brownian Motion

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

$$\begin{aligned} S_t &= S_0 e^{\log\left(\frac{S_t}{S_0}\right)} \quad \text{where } \log\left(\frac{S_t}{S_0}\right) \sim N(\mu t, \sigma^2 t) \\ &\sim S_0 \log N(\mu t, \sigma^2 t) \end{aligned}$$

SDE form:

$$\begin{aligned} d\log(S_t) &= \mu dt + \sigma dW_t \\ dS_t &= \left(\mu + \frac{\sigma^2}{2}\right) S_t dt + \sigma S_t dW_t \end{aligned}$$

9.4 Risk Neutral Pricing

A **risk-neutral (RN) measure** or **equivalent martingale measure (EMM)** is a probability measure under which discounted asset prices are martingales.

Martingale: a stochastic process with the property $E(X_{n+1}|X_1, ..., X_n) = X_n$

Assuming GBM for asset $\{S_t\}$ and risk-free interest rate r, there exists a probability measure such that

$$\begin{aligned} dS_t &= rS_t dt + \sigma S_t dW_t \\ S_t &\sim S_0 \times \log N\left(\left(r - \frac{\sigma^2}{2}\right) t, \sigma^2 t\right) \end{aligned}$$

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

$$G_0 = \mathbb{E}\left[e^{-rT} G_T\right] = \mathbb{E}\left[e^{-rT} f(S_T)\right]$$

E.g. Show that under RN measure, $E(S_t) = S_0 e^{rt}$. More generally, $E(S_t/e^{rt}|S_s) = S_s/e^{rt}$.

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

Use the Normal MGF:

If $X \sim N(\mu, \sigma^2)$, then $m_X(z) = e^{\mu z + \frac{1}{2}\sigma^2 z^2}$

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

E.g. Find price of forward contract $F_{0,T}$ (no dividends)

$$G_T = f(S_T) = (S_T - F_{0,T})$$

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

By risk neutral pricing, $G_0 = E(e^{-rT} G_T)$

$$\begin{aligned} 0 &= \mathbb{E}(e^{-rT}(S_T - F_{0,T})) \\ 0 &= \mathbb{E}(S_T) - F_{0,T} \\ F_{0,T} &= \mathbb{E}(S_T) = e^{rT}\mathbb{E}(\underbrace{e^{-rT}S_T}_{\sim mgle}) = e^{rT}S_0 \end{aligned}$$

9.5 Estimating Expectations

If $\mathbb{E}\left[e^{-rT} f\left(S_T\right)\right]$ 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 (i = iterations, not time)
- By Law of Large Numbers (SLLN)

$$\hat{G}_0 = \frac{1}{n}\sum_{i=1}^{n} e^{-rT} f\left(S_i(T)\right) \to \mathbb{E}\left[e^{-rT} f\left(S_T\right)\right], \text{ with prob. } 1$$

- Moreover, by Central Limit Theorem (CLT)

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

E.g. Show estimator of $\mathbb{E}(e^{-rT} f(S_T))$ is consistent, and build 95% confidence interval for G_0

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

Build a 95% confidence interval for G_0 as well

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

Confidence interval:

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

9.5.1 European Call

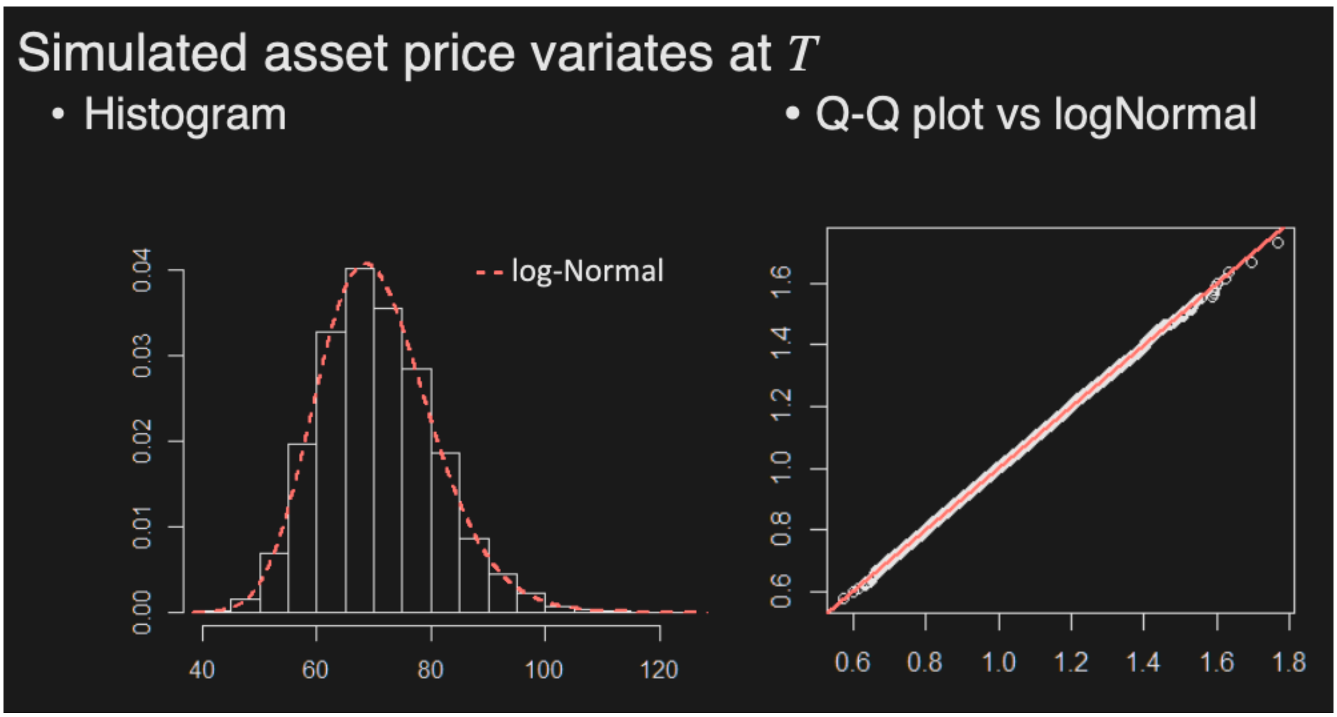
Estimate European call price w/ simulation

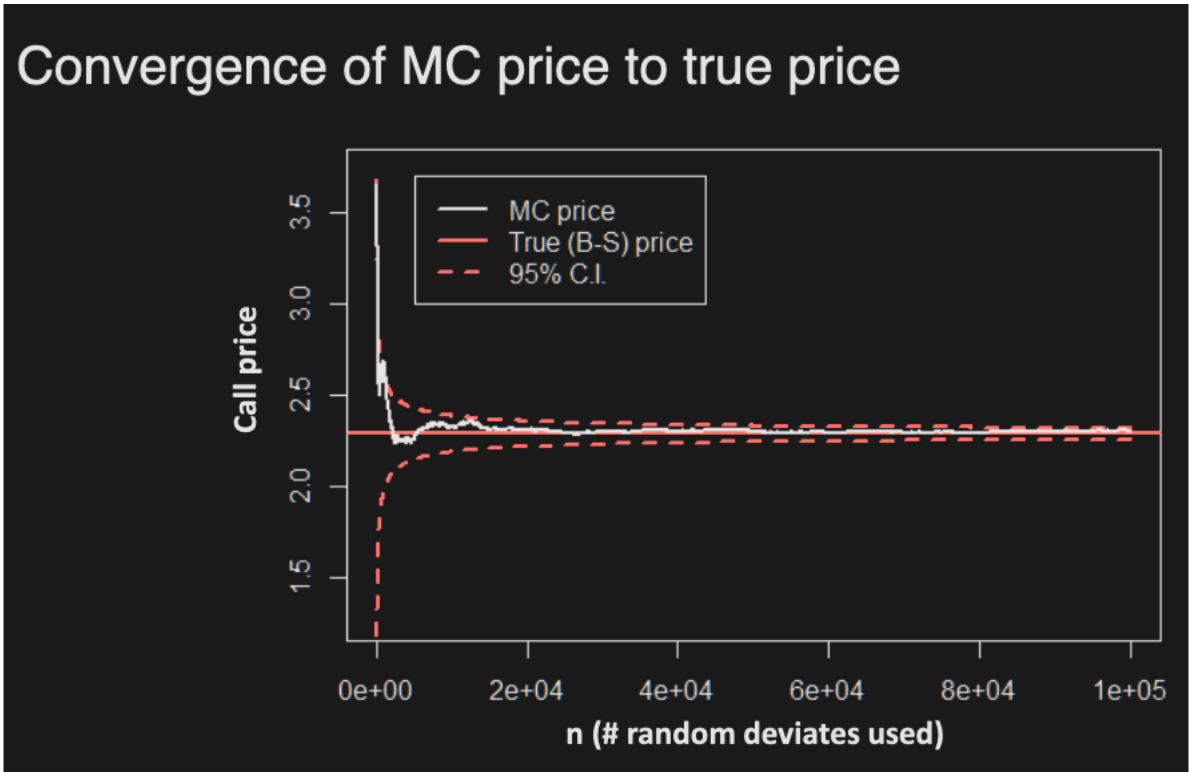
- Asset price dynamics: $dS_t = rS_t dt + \sigma S_t dW_t$
- Payoff function for strike K & maturity T: $f(S_T) = (S_T - K)_T$

Generate random asset price variates as:

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

where Z_i is standard Normal variate





9.5.2 Multiple assets

Payoff of some options depends on prices of multiple assets

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

$$\max\{S_1(T) - S_2(T), 0\} = (S_1(T) - S_2(T))_+$$

Monte Carlo option pricing requires simulating and averaging *multiple* asset prices/paths. We cannot simply simulate each asset separately since there could be cross-asset dependence.

9.6 Multivariate Brownian Motion

Define d-dimensional standard BM

$$\mathbf{W}(t) = \begin{bmatrix} W_1(t) \\ \vdots \\ W_d(t) \end{bmatrix} \text{ with correlation matrix } \rho = \begin{bmatrix} 1 & \cdots & \rho_{1d} \\ \vdots & \ddots & \vdots \\ \rho_{d1} & \cdots & 1 \end{bmatrix}$$

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!

9.6.1 Multivariate ABM

$\{\mathbf{X}(t)\}$ 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

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

Cholesky Factorization

A simple way to generate correlated Normal variates from independent ones

For a positive definite matrix Σ where $\mathbf{x}^T\Sigma\mathbf{x} > 0$, the Cholesky decomposition gives

$$\mathbf{x}^T\mathbf{L}\mathbf{L}^T\mathbf{x} > 0$$

It's essentially the square root matrix.

If $\mathbf{Z} \sim N_d(\mathbf{0}, \mathbf{I})$ and $\Sigma = \mathbf{L}\mathbf{L}^T$ is the Cholesky factorization of the covariance matrix Σ, then

$$\begin{aligned} \mathbb{V}(\mathbf{LZ}) &= \mathbf{L}\mathbb{V}(\mathbf{Z})\mathbf{L}^T = \mathbf{L}\mathbf{L}^T = \Sigma \\ &\implies \mathbf{LZ} \sim N_d(\mathbf{0}, \Sigma) \end{aligned}$$

Note that $\mathbf{L}$ is lower diagonal.

E.g.

$$\begin{bmatrix} W_1 \\ W_2 \\ W_3 \end{bmatrix} \sim N(\mathbf{0}, \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix}) \sim \begin{bmatrix} Z_1 \\ Z_1 + Z_2 \\ Z_1 + Z_2 + Z_3 \end{bmatrix}$$

where $Z_i \sim N(0,1)$

The covariance matrix can be decomposed as

$$\mathbf{L}\mathbf{L}^T = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

Note that $W_1 \sim N(0,1), W_2|W_1 = W_1 \sim N(0,1), W_3|W_2 = W_2 \sim N(0,1)$

10 W10: Pricing Exotic Derivatives

10.1 Path Dependent Options

Derivatives whose payoffs depend on (aspects of) the entire asset price path, instead of just the final price

10.1.1 Barrier Options

Options that come into existence/get knocked out depending on whether prices hit a barrier. Final payoff is equal to a call/put.

4 types of Barrier options:

Up-and-out (U&O):

- gets knocked out if prices moves above the barrier
- max must be below barrier for the option to be worth something

Down-and-out (D&O):

- gets knocked out if price moves below the barrier
- min must be above barrier for the option to be worth something

Up-and-in (U&I):

- comes into existence if price moves above the barrier
- max must be above barrier for the option to be worth something

Down-and-in (D&I):

- comes into existence if price moves below the barrier
- min must be below barrier for the option to be worth something

E.g. When $B < K$ (barrier < strike), $C_{U\&O} = 0$ b/c the payoff is not >0

Note that combining "out" and "in" options with the same B, K, T, etc. gives us a vanilla option. E.g. $P_{U\&I} + P_{U\&O} = P$

Pricing

Notation:

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

The prices are:

$$\begin{aligned}
C_{U\&O} &= e^{-rT}E\left[(S_T-K)_+ \mathbb{1}_{\{M_T<B\}}\right]\\
C_{D\&O} &= e^{-rT}E\left[(S_T-K)_+ \mathbb{1}_{\{m_T>B\}}\right]\\
C_{U\&I} &= e^{-rT}E\left[(S_T-K)_+ \mathbb{1}_{\{M_T>B\}}\right]\\
C_{D\&I} &= e^{-rT}E\left[(S_T-K)_+ \mathbb{1}_{\{m_T<B\}}\right]
\end{aligned}$$

10.2 Simulating GBM paths

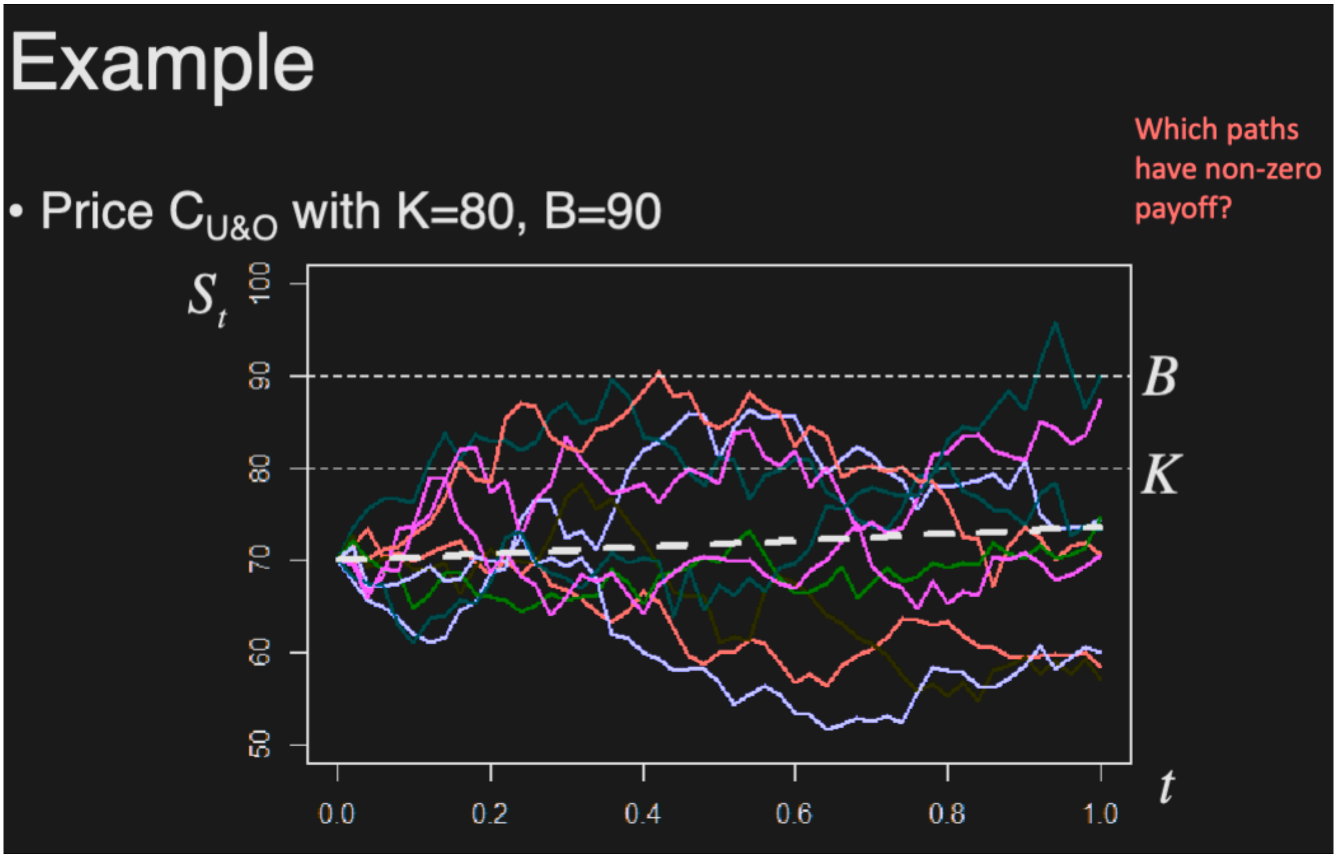
To price general path dependent options, we need to simulate asset price paths $\left\{S_t\right\}_{0\leq t\leq T}$

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

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

For GBM, $dS_t = rS_t dt + \sigma S_t dW_t$ has solution:

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



Only the neon purple option has a non-zero payoff, since it hasn't been knocked out (exceed B), and is above K

10.2.1 E.g. Quiz 5 question 1

Consider a digital call option that pays \$1 if the asset price at expiration (S_T) exceeds the strike price (K). Assume the asset price follows the usual Black-Scholes dynamics $dS_t = rS_t dt + \sigma S_t dW_t$, where $r > 0$ is the risk-free interest rate, $\sigma > 0$ is the volatility, and $\{W_t\}$ is standard Brownian Motion. Let $Z_i \sim$ iid $\mathrm{N}(0, 1)$, and write the digital call price estimate $\bar{X}_n$ as a function of the normal variates $\{Z_i\}_{i=1}^n$ and the parameters S_0, K, T, r, σ

From GBM, the asset price at expiration is given by

$$S_i(T) = S_0 \exp\left\{\left(r - \frac{1}{2}\sigma^2\right) T + \sigma\sqrt{T} Z_i\right\}$$

and the payoff by $I(S(T) \geq K) = \begin{cases} 1, & S(T) \geq K \\ 0, & \text{otherwise} \end{cases}$.

The MC estimator is given by the average discounted payoff

$$\begin{aligned}\bar{X}_n &= \frac{e^{-rT}}{n}\sum_{i=1}^{n} I\left(S_i(T) \geq K\right) \\ &= \frac{e^{-rT}}{n}\sum_{i=1}^{n} I\left(S_0 \exp\left\{\left(r - \frac{1}{2}\sigma^2\right)T + \sigma\sqrt{T}Z_i\right\} \geq K\right)\end{aligned}$$

10.3 Monte Carlo for Barrier Options

MC for barrier options based on simple discretization leads to biased prices!

All knock-out option prices will be overestimated, because the discretized minima/maxima will not be as extreme as the true ones (there may be some time point we did not simulate, during which the barrier could have been crossed, making the option worthless). Similarly, the knock-in option prices will be underestimated.

Bias can be reduced by increasing number of steps (m) in time discretization, but the computation would become increasingly expensive.

Trade-off between # paths (n) & # steps (m):

- n↑ $\iff$ Var↓ & m↑ $\iff$ Bias↓ (Bias-Variance trade-off)

10.3.1 E.g. Quiz 5 question 2

Assume you are using Monte Carlo simulation with path discretization for estimating the price of various path-dependent options under the Black-Scholes model (i.e., the price follows Geometric Brownian Motion). For each of the following option types, determine whether the estimate will under-estimate, over-estimate, or be unbiased.

- D&O Call: Overestimated
- D&O Put: Overestimated
- U&O Call: Overestimated
- U&I Call: Underestimated
- Asian Call with payoff $\left(\frac{1}{m}\sum_{j=1}^{m} S_{t_j} - K\right)_+$ depending on the average price at simulated times $t_1, \ldots, t_m$: unbiased

10.4 Reflection principle

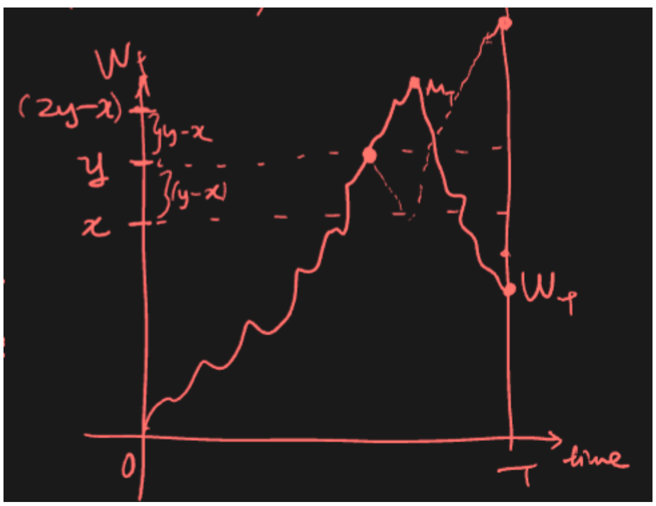
If the path of a Wiener process W_t reaches a value y at time t, then the subsequent path after time s has the same distribution as the reflection of the subsequent path about the value y.

$$
\begin{aligned}
&P\left(M_t > y, W_t \leq x\right) \\
=&P\left(\cancel{M_t > y}, \underbrace{W_t > y + (y - x)}_{\text{subset of } M_t > y}\right) \\
=&P\left(W_t > 2y - x\right)
\end{aligned}
$$

10.4.1 Max of standard BM ~ absolute normal

For standard BM $\{W_t\}$, the max by time T, $M_t = \max\{W_t\}_{0 \leq t \leq T}$ is distributed as a folded normal.

We want to find the CDF $P(M_T \leq y)$



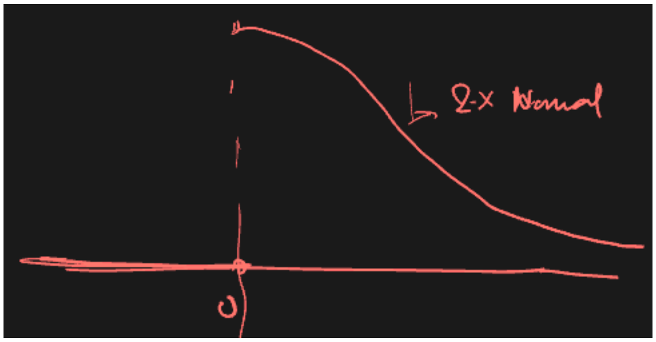
For $x \leq y$

$$
\begin{aligned}
P(W_T \leq x, M_T \leq y) &= P(W_T \leq x) - P(W_T \leq x, M_T \geq y) \quad \leftarrow P(A \cap B) = P(A) - P(A \cap B^C) \\
&= P(W_T \leq x) - P(W_T \geq 2y - x) \quad \leftarrow \text{reflection principle}
\end{aligned}
$$

When $x = y$, we have

$$
\begin{aligned}
P(W_T \leq y, M_T \leq y) &= P(M_T \leq y) \\
&= P(W_T \leq y) - P(W_T \geq 2y - y) \\
&= P(W_T \leq y) - P(W_T \geq y) \\
&= P(W_T \leq y) - P(W_T \leq -y) \\
&= P(-y \leq W_T \leq y) \\
&= P(|W_T| \leq y)
\end{aligned}
$$

Thus $M_T \sim |W_T|$ (consider first line vs last line above).



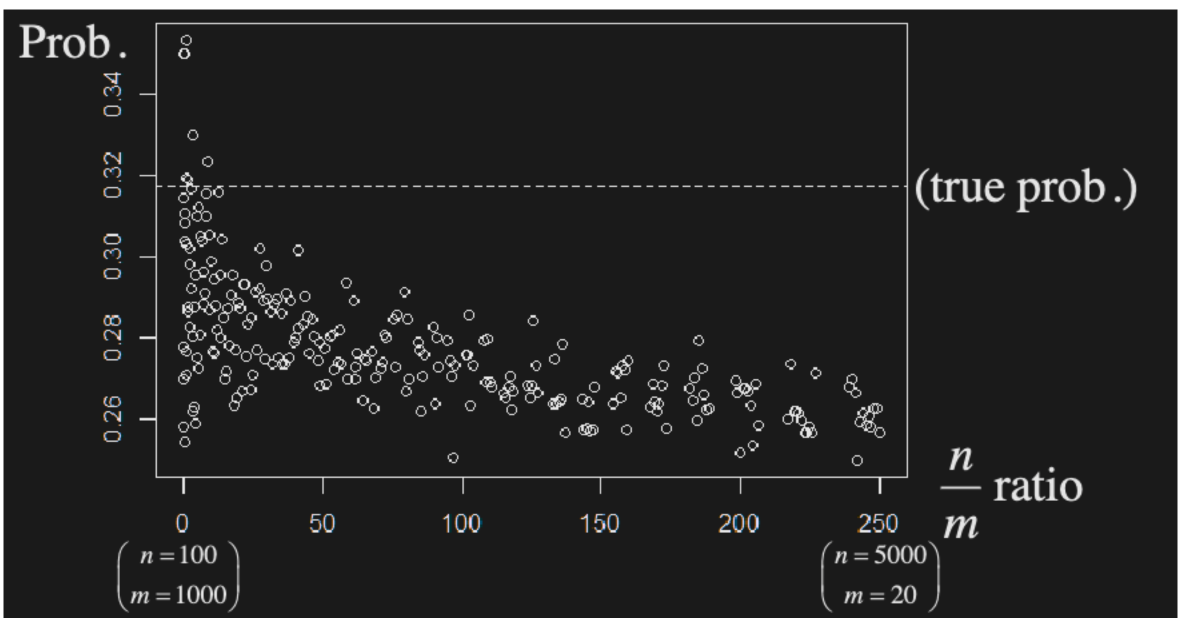
E.g. Find the probability that standard BM $\{W_t\}$ hits barrier $B = 1$ before time $T = 1$

Since $W_T \sim N(0, T)$, we have

$$\begin{aligned} P(M_1 \geq 1) &= P(|W_1| \geq 1) \\ &= 2P(Z \geq 1) = 2\Phi(-1) = .317862 \end{aligned}$$

10.5 Optimal n/m ratio

MC estimates of $P(\max \{W_t\}_{0 \leq t \leq 1} \geq 1)$ using path discretization w/ different n (paths), m (steps)



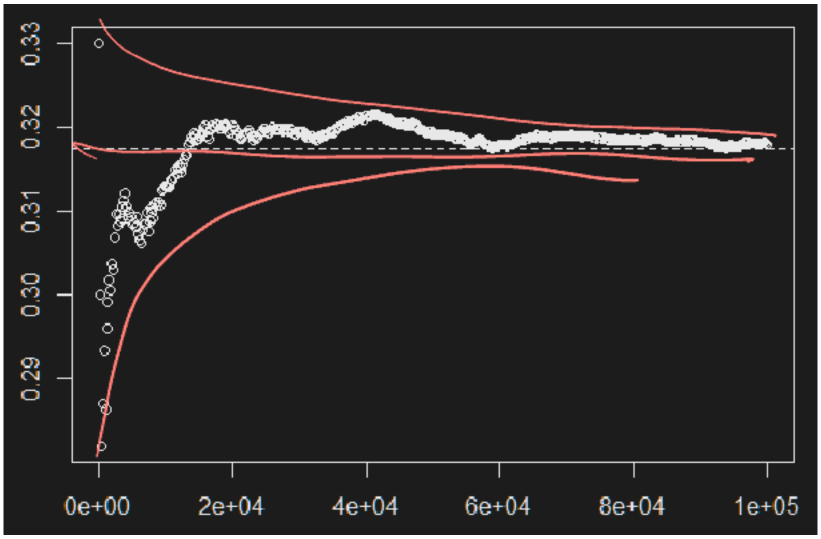
Best MSE lies in the middle:

- High variance when n/m is low
- High bias when n/m is high

E.g. Estimate probability that standard BM hits 1 before time 1, with MC but without bias.

Generate values of M_T directly by generating W_T and setting $M_T = |W_T|$. Then, estimate the probability by the proportion of $M_T's$ that are > 1

Below are MC estimates of $P(\max\{W_t\}_{0 \le t \le 1} \ge 1)$ using direct simulation of $\max\{W_t\}$ w/ $n = 100,000$



10.6 Extrema of Brownian Motion

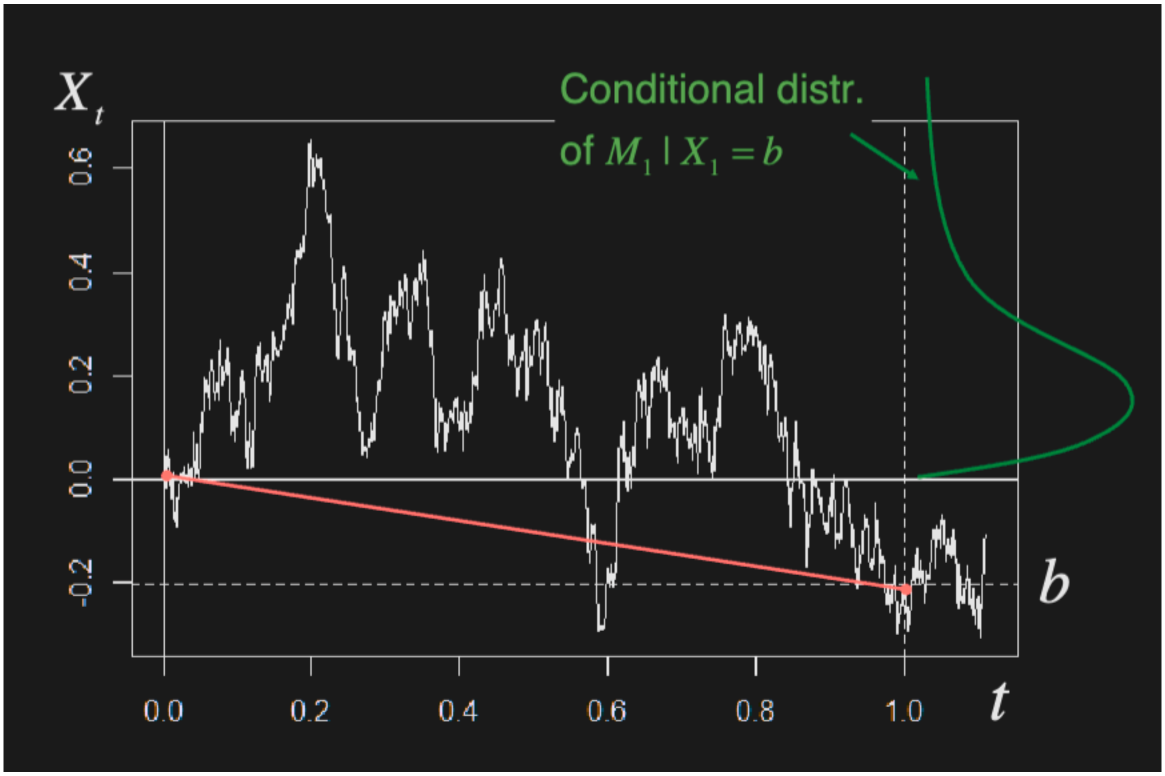
- For standard BM $\{W_t\}$, maximum M_{T} is distributed as $|W_T|$
- For arithmetic BM $\{X_t\}$, the distribution of the maximum is difficult to work with - reflection principle does not work b/c of drift
- However, one can easily simulate random deviates of maximum using Brownian bridge (Brownian motion with fixed end point)
  - Its construction allows for general treatment of extrema of various processes

Consider ABM: $dX_t = \mu dt + \sigma dW_t$

Conditional on $X_T = b$, the maximum $(M_T|X_T) = \max_t \{X_t \mid X_T\}$ of the Brownian bridge process has a *Rayleigh distribution*:

$$P\left(M_T \le m \mid X_T = b\right) = 1 - \exp\left\{-2\frac{m(m-b)}{\sigma^2 T}\right\}, \forall m \ge 0 \vee b$$

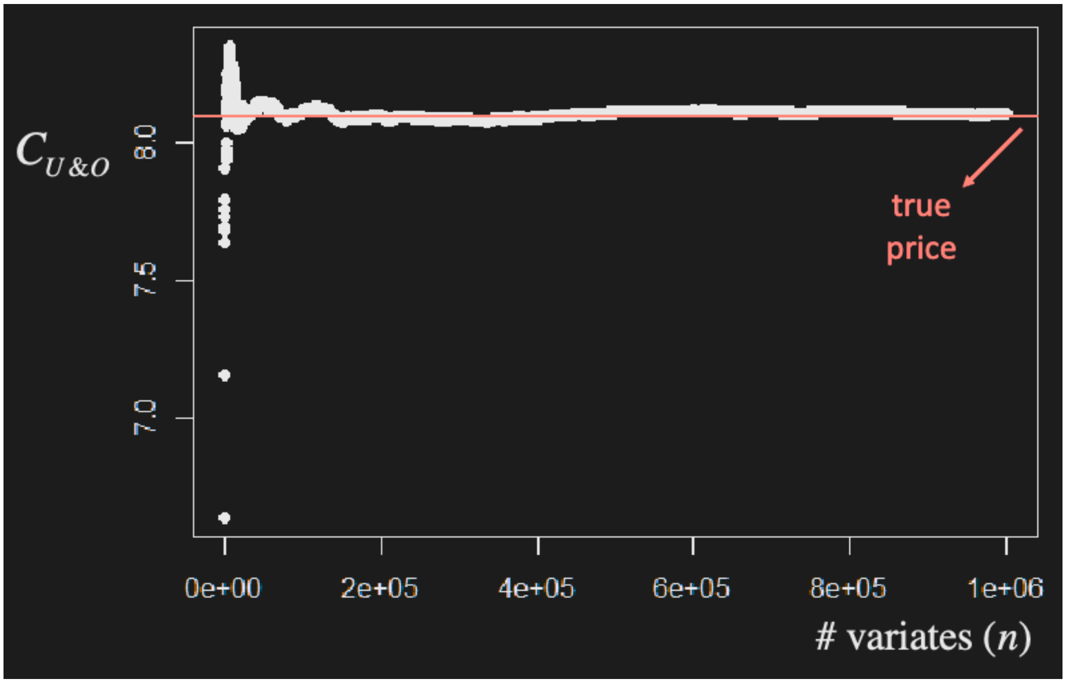
Note that distribution of conditional maximum is *independent* of the drift, given $X_T = b$



10.6.1 Simulating maxima of ABM

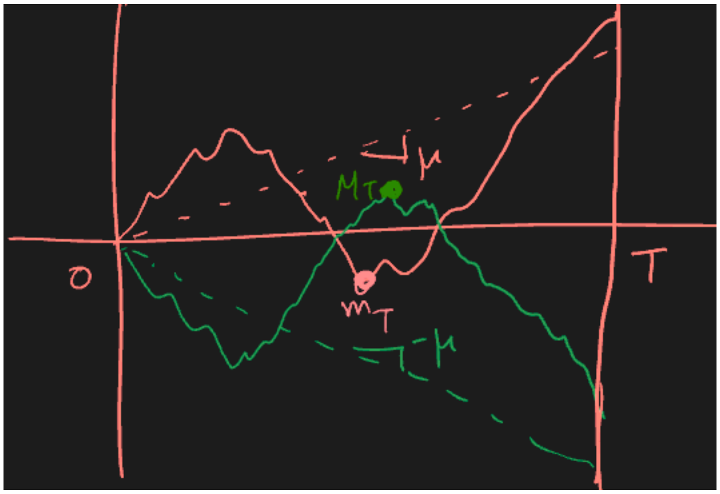
1. Generate $X_T \sim N\left(\mu T, \sigma^2 T\right)$
2. Generate $U \sim \text{Uniform}(0, 1)$
3. Calculate $M_T \mid X_T = \frac{X_T + \sqrt{X_T^2 - 2\sigma^2 T \log(U)}}{2}$

For maxima of GBM, exponentiate ABM result



10.6.2 Simulating minima of ABM

By symmetry, min of ABM with $\mu \equiv$ max of ABM with $-\mu$



10.7 Time Discretization

Path dependent options generally require simulation of entire discretized path. Exceptions are options depending on maximum (e.g. barrier, look-back).

If prices do not follow GBM, it is generally not possible to simulate from exact distribution of asset prices, so we need to approximate sample path distribution over discrete times

10.7.1 Euler Discretization

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

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

There is no general explicit solution for S_t, i.e. distribution of S_t is unknown (in closed form)

To approximate the behaviour of S_t:

- Discretize time $t_i = i(T/m) = i\Delta t, i = 0, \ldots, m$
- Simulate (approx.) path recursively, using $Z_i \overset{iid}{\sim} N(0,1), i = 1, \ldots, m$

$$S\left(t_i\right) = S\left(t_{i-1}\right) + \mu\left(S\left(t_{i-1}\right), t_{i-1}\right) \Delta t + \sigma\left(S\left(t_{i-1}\right), t_{i-1}\right) \sqrt{\Delta t} Z_i$$

To approximate distribution of $S(T)$, generate multiple (#n) discretized paths

11 W11: Simulation - Variance Reduction Techniques

11.1 Antithetic Variables

For each normal variate Z_i, consider its negative $-Z_i$. Note that they are dependent. For Uniform(0, 1), use U_i and $1 - U_i$

Calculate the discounted payoff under both:

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

Estimate price as the *mean* of the RVs

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

The idea is to balance payoffs of paths with opposite returns.

11.1.1 Pros and cons

This technique is simple, but not always useful. It only helps if the original and antithetic variates are *negatively related*.

We can prove this by comparing its variance $V(\bar{Y}_{AV})$ to the variance of the naive mean $V\left(\frac{1}{2n}\sum_{i=1}^{2n} Y_i\right)$. Under what condition does the variance get reduced?

11.1.2 Variance reduction proof

Variance of naive mean:

$$V\left(\frac{1}{2n}\sum_{i=1}^{2n} Y_i\right) = \left(\frac{1}{2n}\right)^2 (2nV(Y_i)) = \frac{1}{2n}V(Y_i)$$

Variance of antithetic mean:

$$\begin{aligned}
V(\bar{Y}_{AV}) &= V\left(\frac{1}{2n}\left(\sum_{i=1}^{n} Y_i + \sum_{i=1}^{n} \tilde{Y}_i\right)\right) \\
&= \left(\frac{1}{2n}\right)^2 V\left(\sum_{i=1}^{n} Y_i + \sum_{i=1}^{n} \tilde{Y}_i\right) \\
&= \left(\frac{1}{2n}\right)^2 \left[nV(Y_i) + nV(\tilde{Y}_i) + 2Cov\left(\sum_{i=1}^{n} Y_i, \sum_{j=1}^{n} \tilde{Y}_j\right)\right] \\
&= \left(\frac{1}{2n}\right)^2 \left[2nV(Y_i) + 2\sum_{i=1}^{n}\sum_{j=1}^{n} Cov(Y_i, \tilde{Y}_j)\right] \\
&= \left(\frac{1}{2n}\right)^2 \left[2nV(Y_i) + 2\sum_{i=1}^{n} Cov(Y_i, \tilde{Y}_i)\right] \\
&= \left(\frac{1}{2n}\right)^2 \left[2nV(Y_i) + 2nCov(Y_i, \tilde{Y}_i)\right] \\
&= \frac{1}{2n}(V(Y_i) + Cov(Y_i, \tilde{Y}_i))
\end{aligned}$$

For $V\left(\bar{Y}_{AV}\right) < V\left(\frac{1}{2n}\sum_{i=1}^{2n} Y_i\right)$ to hold, we must have

$$Cov(Y_i, \tilde{Y}_i) \leq 0$$

Even function => worst case scenario (-Z gives the same value)

11.1.3 Asymptotic distribution of estimator

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

By CLT, we have

$$\bar{Y}_{AV} \overset{approx}{\sim} N\left(E\left(\frac{Y_i+\tilde{Y}_i}{2}\right), \frac{1}{n}V\left(\frac{Y_i+\tilde{Y}_i}{2}\right)\right)$$

where the mean is

$$\frac{1}{2}(E(Y_i) + E(\tilde{Y}_i)) = \frac{1}{2}2E(Y) = E(Y) = E(f(Z))$$

and the variance is

$$
\begin{aligned}
&\frac{1}{n}\cdot\frac{1}{4}(V(Y_i)+V(\tilde{Y}_i)+2Cov(Y_i,\tilde{Y}_i))\\
=&\frac{1}{4n}(2V(Y_i)+2Cov(Y_i,\tilde{Y}_i))\\
=&\frac{1}{2n}(V(Y_i)+Cov(Y_i,\tilde{Y}_i))
\end{aligned}
$$

11.1.4 Example

Antithetic variable pricing of a 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

11.2 Stratification

Split the RV domain into equiprobable strata, and draw equal number of variates from within each one

Consider m equiprobable Normal strata $\{A_j\}$

$$P\left(Z\in A_j\right)=\frac{1}{m}\text{ for } j=1,\dots,m,\text{and } Z\sim N(0,1)$$

Stratified estimator of $Y=f(Z)$ is given by

$$
\begin{aligned}
\bar{Y}_{Str}&=\frac{1}{m}\sum_{j=1}^{m}\bar{Y}^{(j)}\\
&=\frac{1}{m}\sum_{j=1}^{m}\left(\frac{1}{n}\sum_{i=1}^{n}f\left(Z_i^{(j)}\right)\right)
\end{aligned}
$$

where $\bar{Y}^{(j)}$ is the estimator within each stratum j, and $Z_i^{(j)}\overset{iid}{\sim}N\left(0,1\mid Z_i^{(j)}\in A_j\right),j=1,\dots,m$

11.2.1 Mean of estimator

Verify that $\bar{Y}_{Str}$ is an unbiased estimator of $E(f(Z))$

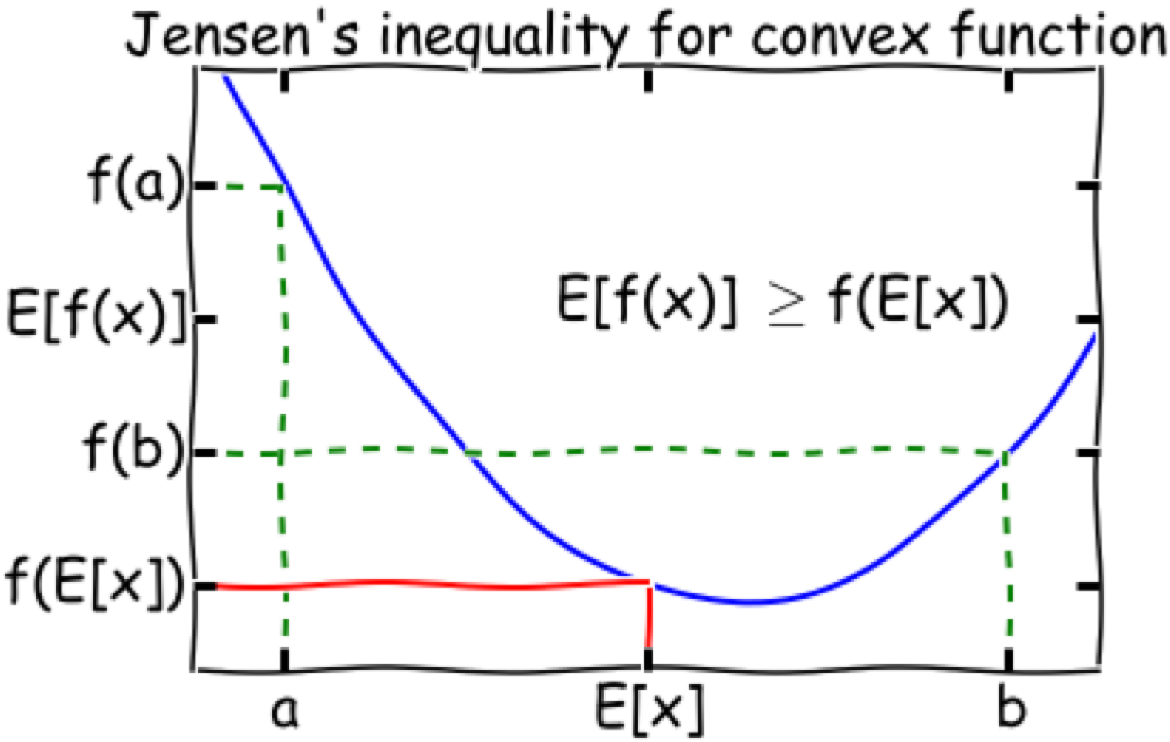
$$\begin{aligned}
E(\bar{Y}_{Str}) &= E\left(\frac{1}{m}\sum_{j=1}^{m}\bar{Y}^{(j)}\right) \\
&= \frac{1}{m}\sum_{j=1}^{m}E(\bar{Y}^{(j)}) \\
&= \frac{1}{m}\sum_{j=1}^{m}E\left(\frac{1}{n}\sum_{i=1}^{n}Y_j^{(j)}\right) \\
&= \frac{1}{m}\sum_{j=1}^{m}\frac{1}{n}\sum_{i=1}^{n}\underbrace{E(Y_i^{(j)})}_{E(Y|Z\in A_j)} \\
&= \underbrace{\frac{1}{m}}_{P(Z\in A_j)}\sum_{j=1}^{m}\frac{1}{n}nE(f(Z)|Z\in A_j) \\
&= \sum_{j=1}^{m}E(f(Z)|Z\in A_j)P(Z\in A_j) \\
&= E(f(Z)) \quad \text{by LOTP} \\
&= E(Y)
\end{aligned}$$

11.2.2 Variance reduction proof

Show that $V[\bar{Y}_{Str}] < V[\bar{Y}]$, where $\bar{Y} = \frac{1}{nm}\sum_{i=1}^{nm} f(Z_i)$

$$\begin{aligned}
V(\bar{Y}) &= V\left(\frac{1}{nm}\sum_{i=1}^{nm}Y_i\right) = \frac{1}{nm}V(Y_i) = \frac{1}{nm}\left(E(Y_i^2)-(E(Y_i))^2)\right) = \frac{1}{nm}(E(f^2(Z))-\mu^2) \\
V(\bar{Y}_{Str}) &= V\left(\frac{1}{m}\sum_{j=1}^{m}\bar{Y}^{(j)}\right) = \frac{1}{m^2}\sum_{j=1}^{m}V(\bar{Y}^{(j)}) = \frac{1}{m^2}\sum_{j=1}^{m}V\left(\frac{1}{n}\sum_{i=1}^{n}Y_i^{(j)}\right) = \frac{1}{m^2n^2}\sum_{j=1}^{m}\sum_{i=1}^{n}V(Y_i^{(j)}) \\
&= \frac{1}{m^2n^2}\sum_{j=1}^{m}nV(Y^{(j)}) = \frac{1}{m^2n}\sum_{j=1}^{m}V(Y|Z\in A_j) = \frac{1}{m^2n}\sum_{j=1}^{m}(E(f^2(Z)|Z\in A_j)-(E(f(Z))|Z\in A_j)^2) \\
&= \frac{1}{mn}\left\{\underbrace{\sum_{j=1}^{m}E(f^2(Z)|Z\in A_j)\underbrace{\frac{1}{m}}_{P(Z\in A_j)}}_{E(f^2(Z))} - \frac{1}{m}\sum_{j=1}^{m}(\underbrace{E(f(Z)|Z\in A_j)}_{\mu_j})^2\right\} \\
&= \frac{1}{mn}\left\{E(f^2(Z)) - \frac{1}{m}\sum_{j=1}^{m}\mu_j^2\right\}
\end{aligned}$$

So we have $V[\bar{Y}_{Str}] < V[\bar{Y}]$ since $\frac{1}{m}\sum_{j=1}^{m}\mu_j^2 \geq \mu^2$ by Jensen's inequality



So for $f(x) = x^2$, a convex function, we have

$$\underbrace{E(x^2)}_{=\frac{1}{m}\sum_{j=1}^{m}\mu_j^2} \geq \underbrace{(E(x))^2}_{=\mu^2=\left(\frac{1}{m}\sum_{j=1}^{m}\mu_j\right)^2}$$

11.2.3 Pros and cons

This method ensures equal representation of each stratum in the RV's domain. It *always* reduces variance.

It works best when target RV (payoff) *changes* over its domain, i.e. is highly variable (as opposed to a flat payoff).

It is computationally difficult for multidimensional RV's. Getting the conditional distribution within each stratum can be difficult, and the CDF is often unknown.

11.2.4 Example

Stratified pricing of a 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

11.3 Control Variates

Estimate $E[Y] = E[f(Z)]$ using MC: generate iid Z_i and use

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

where $f(\cdot)$ is option's discounted payoff

Assume there is another option with payoff $g(\cdot)$ whose price $E[X] = E[g(Z)]$ is known. The idea is to use MC with the same variates to estimate both $E[Y]$ and $E[X]$, but adjust the estimate $\bar{Y}$ to take into account the error of estimate $\bar{X}$. E.g. if $\bar{X}$ underestimates $E[X]$, then increase $\bar{Y}$.

Adjust $\bar{Y}$ for estimation error $\bar{X} - E[X]$ linearly, as

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

where the coefficient b controls adjustment.

11.3.1 Mean of estimator (unbiased proof)

Show that $\bar{Y}(b)$ is unbiased for any b (provided $\bar{Y}, \bar{X}$ are unbiased)

$$
\begin{aligned}
E(\bar{Y}(b)) &= E(\bar{Y} - b(\bar{X} - E(X))) \\
&= E(\bar{Y}) - b(E(\bar{X}) - E(X)) \\
&= E(Y) - b(\underbrace{E(X) - E(X)}_{\text{0 adjustment}}) \\
&= E(Y)
\end{aligned}
$$

11.3.2 Variance of estimator

$$
\begin{aligned}
\mathbb{V}[\bar{Y}(b)] &= \mathbb{V}[\bar{Y} - b(\bar{X} - \mathbb{E}(X))] = \mathbb{V}[\bar{Y} - b\bar{X}] \\
&= \mathbb{V}(\bar{Y}) + b^2\mathbb{V}(\bar{X}) - 2b\operatorname{Cov}(\bar{Y}, \bar{X}) \\
&= \frac{1}{n}\mathbb{V}(Y) + b^2\frac{1}{n}\mathbb{V}(X) - 2b\operatorname{Cov}(\frac{1}{n}\sum_{i=1}^{n}\underbrace{f(X_i)}_{Y_i}, \frac{1}{n}\sum_{i=1}^{n}\underbrace{g(Z_i)}_{X_i})) \\
&= \frac{1}{n}\mathbb{V}(Y) + b^2\frac{1}{n}\mathbb{V}(X) - 2b\frac{1}{n^2}n\operatorname{Cov}(f(Z)), g(Z)) \\
&= \frac{1}{n}\left[\sigma_Y^2 + b^2\sigma_X^2 - 2b\sigma_{XY}\right]
\end{aligned}
$$

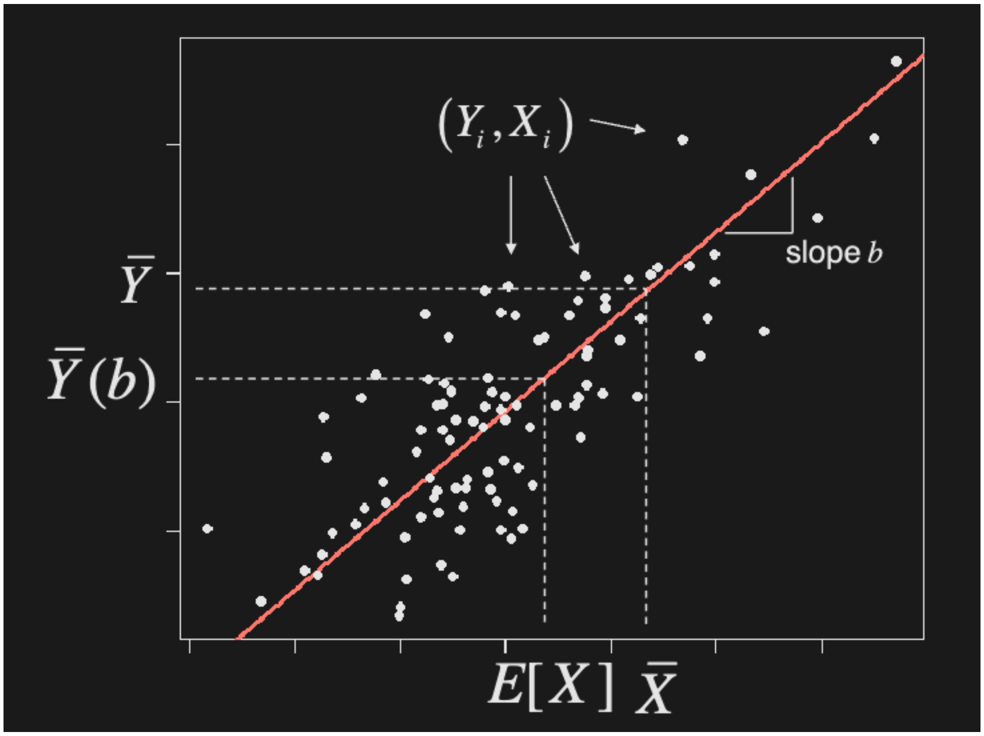
11.3.3 Optimal value of adjustment coefficient

Show that the optimal value of b is $b^* = \frac{Cov(X,Y)}{Var(X)}$. This is the regression slope coefficient.

$$
\begin{aligned}
\frac{\partial}{\partial b}\mathbb{V}(\bar{Y}(b)) &= 0 \\
\frac{\partial}{\partial b}\left(\frac{1}{n}\left[\sigma_Y^2 + b^2\sigma_X^2 - 2b\sigma_{xXY}\right]\right) &= 0 \\
b\sigma_X^2 - \sigma_{XY} &= 0 \\
b = \frac{\sigma_{XY}}{\sigma_X^2} = \frac{\rho_{XY}\sigma_X\sigma_Y}{\sigma_X^2} &= \rho_{XY}\frac{\sigma_Y}{\sigma_X}
\end{aligned}
$$

In practice, $\operatorname{Cov}[X, Y], \operatorname{Var}[X]$ are unknown, so we estimate b^* using MC sample

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



11.3.4 Optimal variance

Show that the optimal variance is $V(\bar{Y}(b^*)) = V(\bar{Y})(1 - \rho_{XY}^2)$

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

In practice, we need to use sample estimates of $\mathrm{Var}[\bar{Y}], \rho_{XY}$

11.3.5 Correlation of control

Good control variates have *high absolute correlation with option payoff* (high ρ_{XY})

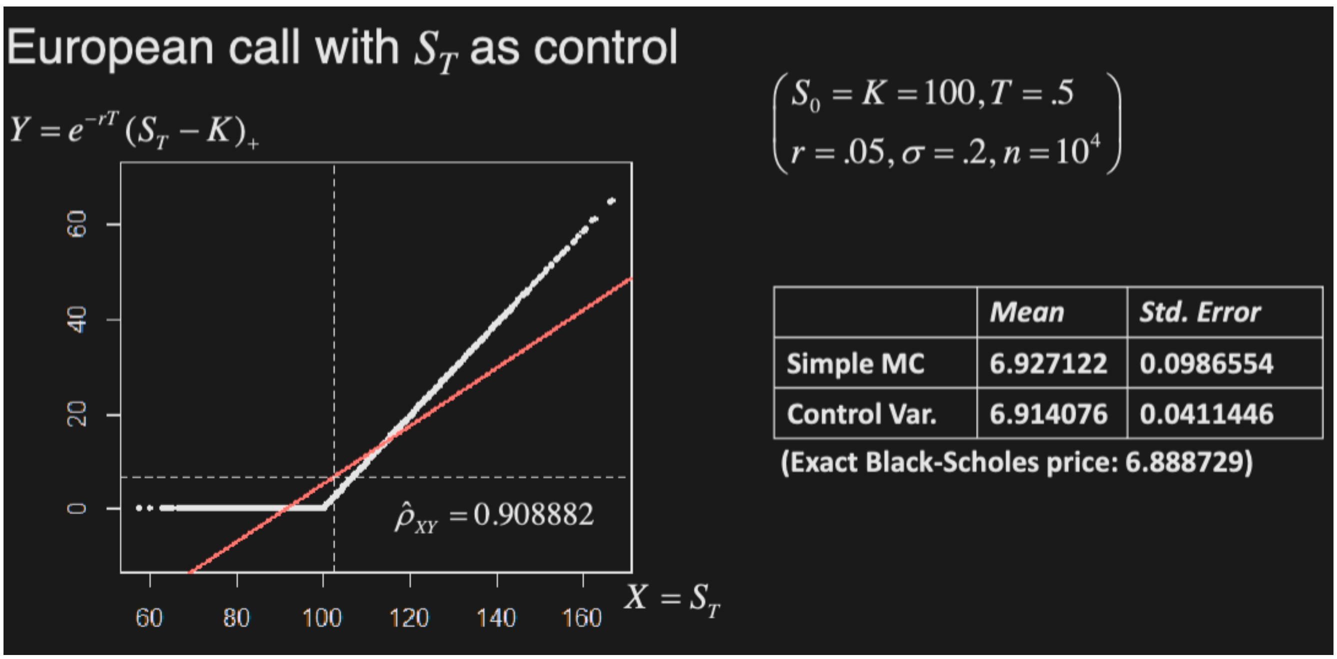
- In-the-money call: $\rho_{XY} \approx 1$
  - $Cov(S_T, S_T - K) > 0$

- Out-of-the-money call: $\rho_{XY} \approx 0$
  - $Cov(S_T, 0) = 0$
- In-the-money put: $\rho_{XY} \approx -1$
- Out-of-the-money put: $\rho_{XY} \approx 0$

11.3.6 Example

Price European option using final asset price (S_t) as control, assuming GBM with r, σ

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



11.4 Importance Sampling

We can reduce variance by changing the distribution (probability measure) from which paths are generated to give more weight to important outcomes, thereby increasing sample efficiency. The performance of this method relies *heavily* on the equivalent measure being used.

E.g. for European call, we put more weight to paths with positive payoff (i.e. paths for which we exercise)

Let $\phi(z)$ be pdf of Z, we want to estimate

$$\alpha = E_\phi[f(Z)] = \int_z f(z)\phi(z)dz$$

Using simple MC, generate sample $Z_i \sim^{iid} \phi, i = 1, \ldots, n$. The estimate is thus

$$\hat{\alpha} = \sum_{i=1}^{n} f\left(Z_i\right)/n$$

If we have sample $Z_i' \overset{iid}{\sim} \psi, i = 1, \ldots, n$ from a new pdf ψ, we can still estimate α as follows

$$\alpha = \int_z f(z)\phi(z)dz = \underbrace{\int_z f(z)\frac{\phi(z)}{\psi(z)}\psi(z)dz = E_\psi\left[f\left(Z'\right)\frac{\phi\left(Z'\right)}{\psi\left(Z'\right)}\right]}_{\int x f_X(x)\,dx = E(X)}$$

$$\Rightarrow \hat{\alpha}' = \frac{1}{n}\sum_{i=1}^{n} f\left(Z_i'\right)\frac{\phi\left(Z_i'\right)}{\psi\left(Z_i'\right)}$$

11.4.1 Mean of estimator

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

Note that this estimate is unbiased (provided simple MC estimate $\hat{\alpha}$ is unbiased).

11.4.2 Variance of estimator

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

11.4.3 Variance reduction proof & condition

Show that $\text{Var}_\psi\left[\hat{\alpha}'\right] \leq \text{Var}_\phi[\hat{\alpha}] \iff E_\phi\left[f^2(Z)\frac{\phi(Z)}{\psi(Z)}\right] \leq E_\phi\left[f^2(Z)\right]$

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

The LHS is equivalent to

$$\begin{aligned}
\int \underbrace{f^2(z')\frac{\phi^2(z')}{\psi^2(z')}}\psi(z')\,dz' &= \int f^2(z')\underbrace{\frac{\phi(z')}{\psi(z')}\phi(z')}\,dz' \\
&= \mathbb{E}_\phi\left[f^2(z)\frac{\phi(z)}{\psi(z)}\right]
\end{aligned}$$

11.4.4 Optimal variance condition

Show that for positive f, $Var_\psi[\hat{\alpha}'] = 0$ if $\psi(z) \propto f(z)\phi(z)$

i.e. importance sampling works best when new pdf ψ resembles $f \times \phi$ (payoff $\times$ original pdf)

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

11.4.5 Multiple random variates

Importance sampling can be extended to multiple random variates per path

For example, for a path-dependent option with payoff $f(Z_1, \dots, Z_m)$, which is a function of m variates forming discretized path, the mean of the estimate is

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

If in addition, $Z_j \overset{iid}{\sim} \phi_j$, $Z_j' \overset{iid}{\sim} \psi_j$, then

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

11.4.6 Example

Consider a deep out-of-the-money European call with $S_0 = 50, K = 65$

With simple MC, generate final prices as

$$S_T = S_0 e^Z, \text{ where } Z \sim \phi = N\left(\left(r - \frac{\sigma^2}{2}\right)T, \sigma^2 T\right)$$

Which of the following is a better candidate for ψ ?

$$Z' \sim \psi = N\left(\log\left(\frac{90}{50}\right) - \frac{\sigma^2}{2}T, \sigma^2 T\right) \text{ or } N\left(\log\left(\frac{30}{50}\right) - \frac{\sigma^2}{2}T, \sigma^2 T\right)$$

The former, since it is ITM. We want to simulate from distributions with higher means (closer to $K = 65$).

12 W12: Optimization in Finance

Most real world problems involve making decisions, often under uncertainty. Making good/optimal decisions typically involves some optimization.

In finance, we must typically decide how to invest over time and across assets.

E.g. mean-variance analysis or Kelly criterion

12.1 Types of Optimization Problems

- Straightforward (closed form or polynomial complexity):
  - Linear, quadratic, convex
  - Equality/linear/convex constraints
- Difficult:
  - Discrete optimization (discrete variable)
    - E.g. indivisible assets, transaction costs
  - Dynamic optimization (previous decisions affect future ones)
    - Investing overtime
  - Stochastic optimization (uncertainty)

12.2 Discrete & Dynamic Optimization

Assume you can perfectly foresee the price of a stock. You want to make optimal use of such knowledge, assuming

- you can only trade integer units of the asset
- every transaction costs you a fixed amount
- you cannot short sell the asset

This is a **discrete, dynamic optimization problem**. Although there is no randomness (we have perfect knowledge), the problem is not trivial.

We could consider all possible strategies, but that would be expensive - the search space size is 2^n.

We can use dynamic programming (backward induction) instead:

- At any time t, there are 2 states: owning or not owning the asset

- The optimal value of each state at t = the best option out of transitioning to another state + the optimal value of that state at t+1
- Start from the end, and consider optimal value going backwards to discover the best strategy

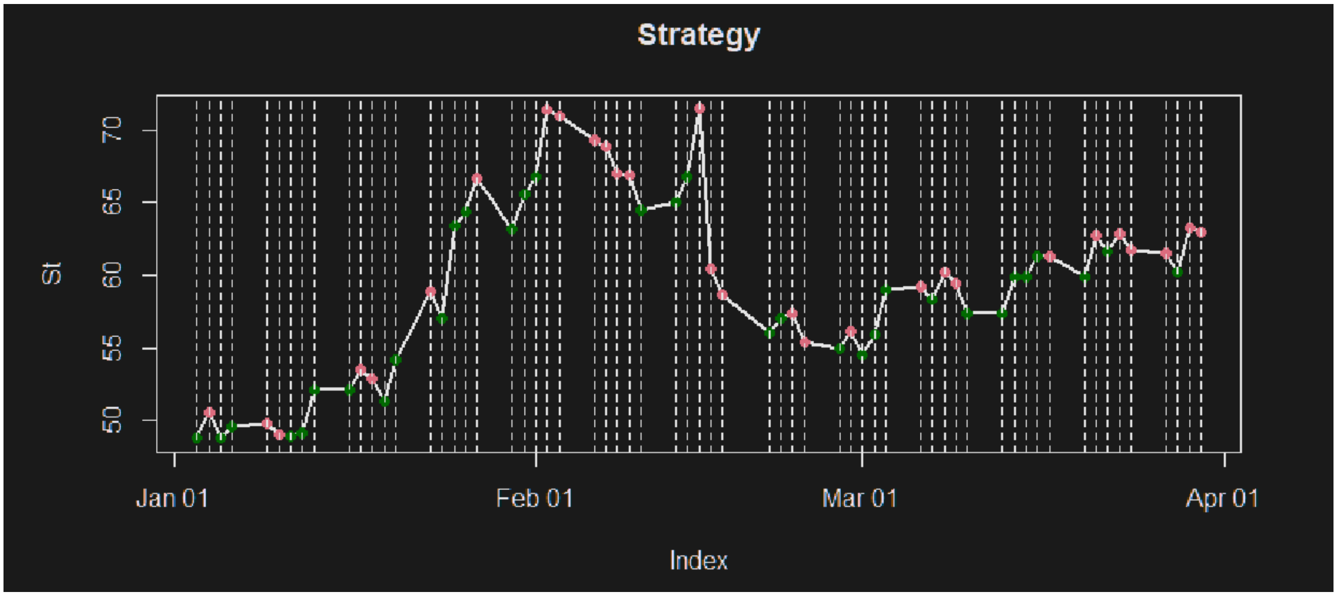
E.g. Find evolution of value, assuming no position at $t = 0$ and $t > n$

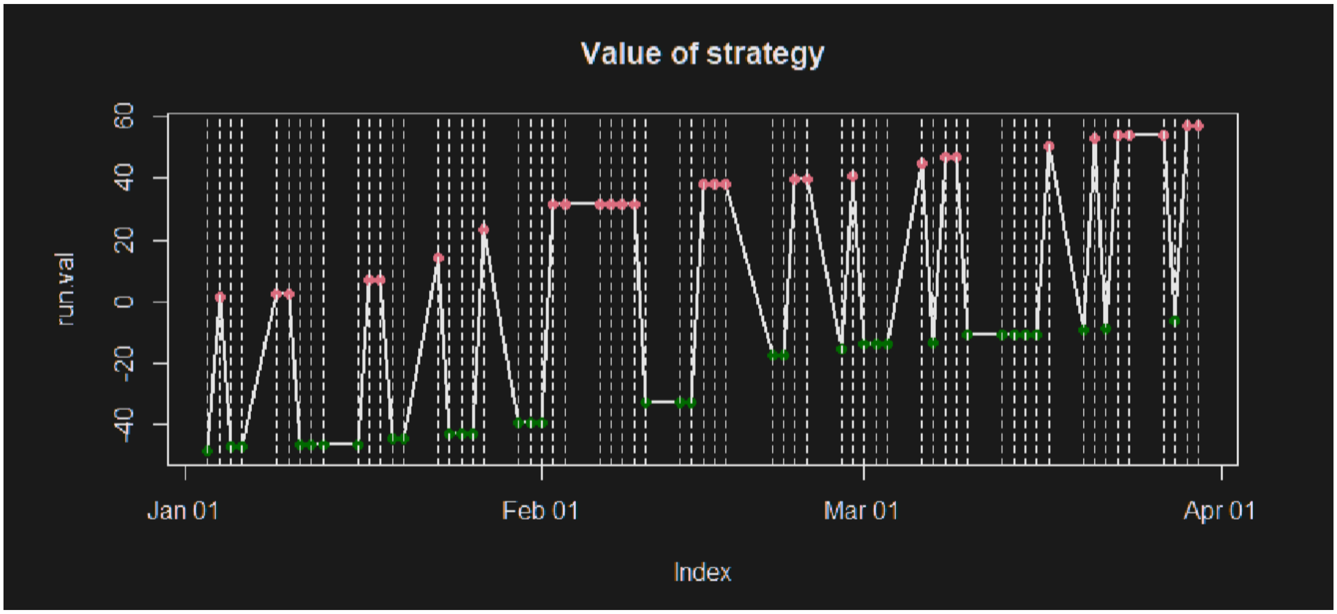
Let $S(t)$ = asset price at t, tc = transaction cost, $V_{np}(t)$ = opt. value for no position at t, $V_{lp}(t)$ = opt. value for long position at t

state \ time	t=1	t=2	t=3=n	> n
no position	$V_{np}(1) = \max\{0 + V_{np}(2), -S(1) - tc + V_{lp}(2)\}$	$V_{np}(2) = \max\{0 + V_{np}(3), -S(2) - tc + V_{lp}(3)\}$	$V_{np}(3) = 0$	0

state \ time	t=1	t=2	t=3=n	> n
long position	/	$V_{lp}(2) = \max\{0 + V_{lp}(3), S(2) - tc + V_{np}(3)\}$	$V_{lp}(3) = S(3) - tc$	/

E.g.





12.3 Stochastic Optimization

Consider a similar problem without exact price knowledge (e.g. prices follow binomial tree with some probabilities)

We want to find the best trading strategy (which maxes the expected P&L)

Define the following:

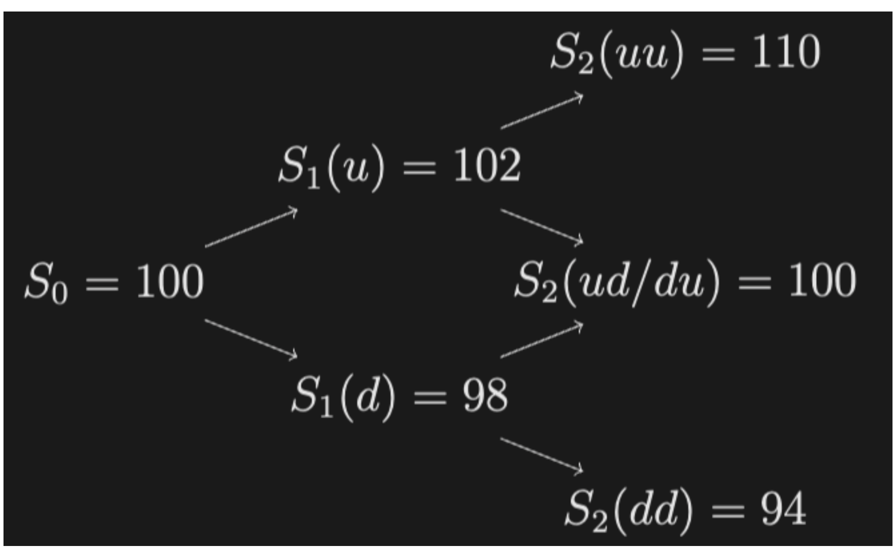
- X_t is the state RV (price and position, e.g. long/short)
- a_t is action (change in state e.g. buy/sell)
- f_t is a reward function (e.g. cashflow)

We want to maximize expected reward over stochastic actions

Letting $V(t, X_t)$ be the optimal value function, we have

$$
\begin{aligned}
V(t, X_t) &= \max_{a_{t\to T}} \left\{ \mathbb{E}\left[\sum_{s=t}^{T} f(s, X_s, a_s) \mid X_t \right] \right\} \\
&= \max_{a_{t\to T}} \left\{ f(t, X_t, a_t) + \mathbb{E}\left[\sum_{s=t+1}^{T} f(s, X_s, a_s) \mid X_t \right] \right\} \\
&= \max_{a_t} \left\{ f(t, X_t, a_t) + \max_{a_{(t+1)\to T}} \left\{ \mathbb{E}\left[\sum_{s=t+1}^{T} f(s, X_s, a_s) \mid X_t \right] \right\} \right\} \\
&= \max_{a_t} \left\{ f(t, X_t, a_t) + \mathbb{E}\left[\max_{a_{(t+1)\to T}} \left\{ \mathbb{E}\left[\sum_{s=t+1}^{T} f(s, X_s, a_s) \mid X_{t+1}^{(a_t)} \right] \right\} \mid X_t \right] \right\} \\
&= \max_{a_t} \left\{ f(t, X_t, a_t) + \mathbb{E}\left[V\left(t+1, X_{t+1}^{(a_t)}\right) \mid X_t \right] \right\}
\end{aligned}
$$

E.g. Consider the following Binomial tree, with up/down probability of 1/2:



Find the optimal strategy that maximizes the expected P/L assuming you can long and short the asset, and there is a transaction cost of \$0.1 /share. Note that there are three possible states now (long, neutral, short), and $3^3 = 27$ possible strategies. Find the optimal strategy and its value using dynamic programming, and optionally verify it with an exhaustive search.

Solution: It is not difficult to reason that the best strategy is to go long at u and short at d, since the up paths have positive expected P/L and the down paths have negative expected P/L (greater than the transaction costs), and this strategy minimizes the expected costs (you only long/short when you need). You can actually verify this by calculating the expected P/L of all 27 strategies by brute force (e.g., in R) to get:

Strategy	Expected P//L
s,s,s	-1-2tc
s,s,n	-1.5-2tc
s,s,l	-2-3tc
s,n,s	0.5-2tc
s,n,n	0-2tc
s,n,l	-0.5-3tc
s,l,s	2-3tc
s,l,n	1.5-3tc
s,l,l	1-4tc
n,s,s	-1-2tc
n,s,n	-1.5-1tc
n,s,l	-2-2tc
n,n,s	0.5-1tc
n,n,n	0
n,n,l	-0.5-1tc
n,l,s	2-2tc
n,l,n	1.5-1tc
n,l,l	1-2tc
l,s,s	-1-4tc
l,s,n	-1.5-3tc
l,s,l	-2-3tc
l,n,s	0.5-3tc
l,n,n	0-2tc
l,n,l	-0.5-2tc
l,l,s	2-3tc
l,l,n	1.5-2tc
l,l,l	1-2tc

But you can drastically reduce the required calculations using backward induction/dynamic programming. Let $V(t, (S_t, p))$ denote the optimal value at time t for price S_t and "position" $p \in \{s, n, l\}$.

At time $t = 2$ we have:

$$\begin{aligned} V(2,(110,l)) &= 110 - tc \\ V(2,(110,n)) &= 0 \\ V(2,(110,s)) &= -110 - tc \\ \\ V(2,(100,l)) &= 100 - tc \\ V(2,(100,n)) &= 0 \\ V(2,(100,s)) &= -100 - tc \\ \\ V(2,(94,l)) &= 94 - tc \\ V(2,(94,n)) &= 0 \\ V(2,(94,s)) &= -94 - tc \end{aligned}$$

At time $t = 1$ and $S_1 = 102$, we have

$$\begin{aligned} V(1,(102,l)) &= \max \left\{ \begin{array}{l} 0 + \frac{1}{2}[V(2,(110,l)) + V(2,(100,l))] = 105 - tc, \\ 102 - tc + \frac{1}{2}[V(2,(110,n)) + V(2,(100,n))] = 102 - tc, \\ 2(102 - tc) + \frac{1}{2}[V(2,(110,s)) + V(2,(100,s))] = 99 - 3tc \end{array} \right\} \\ &= 105 - tc \\ V(1,(102,n)) &= \max \left\{ \begin{array}{l} -102 - tc + \frac{1}{2}[V(2,(110,l)) + V(2,(100,l))] = 3 - 2tc, \\ 0 + \frac{1}{2}[V(2,(110,n)) + V(2,(100,n))] = 0 \\ 102 - tc + \frac{1}{2}[V(2,(110,s)) + V(2,(100,s))] = -3 - 2tc \end{array} \right\} \\ &= 3 - 2tc \\ V(1,(102,s)) &= \max \left\{ \begin{array}{l} -2(102 + tc) + \frac{1}{2}[V(2,(110,l)) + V(2,(100,l))] = -99 - 3tc, \\ -102 - tc + \frac{1}{2}[V(2,(110,n)) + V(2,(100,n))] = -102 - tc, \\ 0 + \frac{1}{2}[V(2,(110,s)) + V(2,(100,s))] = -105 - tc \end{array} \right\} \\ &= -99 - 3tc \end{aligned}$$

At time $t = 1$ and $S_1 = 98$, we have

$$V(1,(98,l)) = \max\left\{\begin{array}{l} 0 + \frac{1}{2}[V(2,(100,l)) + V(2,(94,l))] = 97 - tc \\ 98 - tc + \frac{1}{2}[V(2,(100,n)) + V(2,(94,n))] = 98 - tc \\ 2(98 - tc) + \frac{1}{2}[V(2,(100,s)) + V(2,(94,s))] = 99 - 3tc \end{array}\right\}$$

$$= 99 - 3tc$$

$$V(1,(98,n)) = \max\left\{\begin{array}{l} -98 - tc + \frac{1}{2}[V(2,(100,l)) + V(2,(94,l))] = -1 - 2tc, \\ 0 + \frac{1}{2}[V(2,(100,n)) + V(2,(94,n))] = 0 \\ 98 - tc + \frac{1}{2}[V(2,(100,s)) + V(2,(94,s))] = 1 - 2tc \end{array}\right\}$$

$$= 1 - 2tc$$

$$V(1,(98,s)) = \max\left\{\begin{array}{l} -2(98 + tc) + \frac{1}{2}[V(2,(100,l)) + V(2,(94,l))] = -99 - 3tc, \\ -98 - tc + \frac{1}{2}[V(2,(100,n)) + V(2,(94,n))] = -98 - tc, \\ 0 + \frac{1}{2}[V(2,(100,s)) + V(2,(94,s))] = -97 - tc \end{array}\right\}$$

$$= -97 - tc$$

At time $t = 0$ and state n (the only relevant one at the start) we have

$$V(0,(100,n)) = \max\left\{\begin{array}{l} -100 - tc + \frac{1}{2}[V(1,(102,l)) + V(1,(98,l))] \\ 0 + \frac{1}{2}[V(1,(102,n)) + V(1,(98,n))] \\ +100 - tc + \frac{1}{2}[V(1,(102,s)) + V(1,(98,s))] \end{array}\right\}$$

$$= \max\left\{\begin{array}{l} -100 - tc + \frac{1}{2}[(105 - tc) + (99 - 3tc)] = 2 - 3tc, \\ 0 + \frac{1}{2}[(3 - 2tc) + (1 - 2tc)] = 2 - 2tc, \\ +100 - tc + \frac{1}{2}[(-99 - 3tc) + (-97 - tc)] = 2 - 3tc \end{array}\right\}$$

$$= 2 - 2tc$$