0 Introduction

We model the movement of the price P_t as a random walk by assuming the profit at that period $v_t := P_t - P_{t-1}, t = 1, \dots$ are IID random variable with

$$E(v_1) = \mu, \text{Cov}(v_1) = \sigma^2$$

Definition 0.1 (Random Walk)**.** Let P_0 be a point, and $P_t = P_0 + v_1 + \dots + v_t$. Process $\{P_0, P_1, \dots\}$ is called a **random walk** and $\{v_1, v_2, \dots\}$ are its step sizes.
We have

$$E(P_t \mid P_0) = P_0 + t\mu, \quad \text{Cov}(P_t \mid P_0) = \sigma^2 t$$

Here, μ is called the drift. σ is called the volatility.
If v_i are normally distributed, we can the process a *normal* random walk.

Definition 0.2 (Gross Return)**.** The gross return over k periods is $G_t(k) = \frac{P_t}{P_{t-k}}$.

Definition 0.3 (Net Return)**.** Assuming no dividend, the net return over k holding periods is

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

Definition 0.4 (Log Return)**.** The log return over k periods is

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

Remark:

- when $|R_t(k)|$ is small, $r_t \approx R_t$.
- k-period log return is the sum of single-period log returns.

Definition 0.5 (Adjustment for Dividend)**.** If a dividend(interest) D_t is paid prior to time t, so the initial price of day t is

$$P_{t-1} - D_t$$

Then the gross return on day t is

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

Therefore, to maintain the returns' nice-looking equation, we need to let $\frac{P_t}{P_{t-1}-D_t} := \frac{P_t}{P'_{t-1}} := \frac{P_t}{P_{t-1} \times a}$. Apparently, $a = 1 - \frac{D_t}{P_{t-1}}$.

In the analysis of financial data over a long period of time, we should use adjusted price.

Definition 0.6 (Excess Return)**.** Excess return is the difference $r_t - r_t^*$ betwwen the asset's log return r_t and the log return of r_t^* on some reference asset(usually risk free).

Definition 0.7 (Returns of a Portfolio)**.** Suppose onehas a portfolio consisting of p different assets. Let w_i be weights such that $\sum_{i=1}^p w_i = 1$. Then, the value of the asset i is $w_i P_t$ when the total value of the portfolio is P_t.
Suppose R_{it} and r_{it} are the net return and log return of the asset i at time t. The value of portfolio provided by asset i at time t is $w_i P_{t-1}(1 + R_{it})$. So the total value of a portfolio is

$$P_t = \sum_{i=1}^p w_i P_{t-1}(1 + R_{it}) = (1 + \sum_{i=1}^p w_i R_{it}) P_{t-1}$$

Overall net return R_t and the log return r_t of the portfolio are respectively

$$R_t = \frac{P_t}{P_{t-1}} = \sum_{i=1}^p w_i R_{it}$$

and

$$r_t = \log(1 + \sum_{i=1}^p w_i R_{it}) \approx \sum_{i=1}^p w_i R_{it} \approx \sum_{i=1}^p w_i r_{it}$$

Theorem 0.1 (Random Walk Model for Log Return)**.** The random walk hypothesis states that the single-period log returns, $r_t = \log(P_t) - \log(P_{t-1})$ are independent. Thus, $\log(P_t) - log(P_0)$ is a random walk if r_t are IID.
Sometimes, we further assume

- $r_t \sim N(\mu, \sigma^2)$
- This leads to $\log(P_t) - \log(P_{t-k}) \sim N(k\mu, k\sigma^2)$.
- And also, $\frac{P_t}{P_{t-k}}$ will be lognormal.

We have $P_t = P_0 \exp(r_t + \dots + r_1)$. We call such aprocess whose *logarithm* a random walk a **geometric random walk**. If $r_1, \dots$ are IID $N(\mu, \sigma^2)$, then P_t is a log-normal geometric random walk with parameter (μ, σ^2). In the case of a risk-free asset, the rate of return is called interest rate. Ifthe interest rate isconstant R compounded once per unit period, then the value of the risk-free asset at time t is $P_t = P_0(1 + R)^t$.

1 Exploratory Data Analysis, Moments

- Normal Distribution: $N(\mu, \sigma^2)$

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

- t-distribution: $t(\nu)$

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

- Log-normal distribution:logN(μ, σ)

$$f(x; \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp(-\frac{(\log(x) - \mu)^2}{2\sigma^2}), \quad x > 0$$

Definition 1.1 (Moment, Central Moment)**.** Let X be a random variable. The kth **moment** of X is $E(X^k)$. The kth **central moment** is defined as

$$\mu_1 = E(X),\ \mu_k = E[(X - E(X))^k]$$

Remark:

- First moment is the mean μ.
- Second central moment is variance σ^2.
- If X is the return, the we call $\sigma(X)$ the risk.

Theorem 1.1 (Useful Results of Moments)**.**

- If $X \sim N(\mu, \sigma^2)$, then

$$E(X) = \mu, \mathrm{Cov}(X) = \sigma^2, \mu_3 = 0, \mu_4 = 3\sigma^4$$

- If $X \sim t(\nu)$, then

$$E(X) = 0, \mathrm{Cov}(X) = \frac{\nu}{\nu - 2}(\text{ for } \nu > 2), \mu_3 = 0, \mu_4 = \frac{3}{\nu - 4}\frac{\nu^2}{\nu - 2}(\text{ for } \nu > 4)$$

- If $X \sim \mathrm{logN}(\mu, \sigma^2)$, then

$$E(X^s) = \exp(s\mu + \frac{1}{2}s^2\sigma^2)$$

Definition 1.2 (Sharpe Ratio)**.** For random variable X, the Sharpe ratio is defined as

$$SR = \frac{E(X)}{\sigma(X)}$$

In finance, the Sharpe Ratio of return R is

$$SR(R) = \frac{E(R - r_f)}{\sigma(R)}$$

Definition 1.3 (Sample Mean and Variance, and Estimation of Population Mean and Variance)**.** *With sample* $Y_1, \ldots, Y_n$,

$$\hat{\mu} = \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

If $Y_1, \ldots, Y_n$ *are IID with mean* μ *and standard deviation* σ^2*. Then approximately*

- $\sqrt{n}(\hat{\mu} - \mu) \sim N(0, \sigma^2)$
- $\sqrt{n}(\hat{\sigma^2} - \sigma^2) \sim N(0, 2\sigma^4)$
- $\sqrt{n}(\hat{\sigma} - \sigma) \sim N(0, \frac{1}{2}\sigma^2)$

We can estimate the Sharpe Ratio by $\hat{SR} = \frac{\hat{\mu}}{\hat{\sigma}}$, with

$$\hat{SR} - SR \sim N(0, \frac{1}{n}(1 + \frac{1}{2}SR^2))$$

Definition 1.4 (Quantile)**.** *Suppose* $X \sim F(x)$*, for any* $0 < q < 1$*, the* q*th quantile of* X *is defined as*

$$Q_q(X) = \max\{x : P(X < x) \leq q\}$$

Definition 1.5 (Value At Risk)**.** If X is the investment , the $-Q_q(X)$ is called the $100q\%$ **value at risk**(VaR):

$$\mathrm{VaR}_q(X) = -Q_q(X) = -\max\{v : F(v) \leq q\}$$

It is easy to see that

- $\mathrm{VaR}_q(X + c) = \mathrm{VaR}_q(X) - c$
- If $X \leq Y$, the $\mathrm{VaR}_q(X) \geq \mathrm{VaR}_q(Y)$
- $\mathrm{VaR}_q(\lambda X) = \lambda\,\mathrm{VaR}_q(X)$ for any constant $\lambda > 0$

Definition 1.6 (Expected Shortfall)**.** The expected shortfall at level q is

$$\mathrm{ES}_q(X) = \frac{1}{q}\int_0^q \mathrm{VaR}_\alpha(X)\,\mathrm{d}\,\alpha$$

It is easy to see that

- $\mathrm{ES}_q(X + c) = \mathrm{ES}_q(X) - c$
- $X \leq Y \Rightarrow ES_q(X) \geq ES_q(Y)$
- $ES_q(\lambda X) = \lambda ES_q(X)$ for $\lambda > 0$.

An alternative formula for expected shortfall is

$$ES_q(X) = -E(X \mid X < -\mathrm{VaR}_q(X)) = -\frac{1}{q}\int_{-\infty}^{-\mathrm{VaR}_q(X)} x f(x)\,\mathrm{d}$$

Definition 1.7 (Conherent Risk Measures)**.** A risk measure which satisfies the following four properties is coherent:

- Drift Invariance: $\rho(r + c) = \rho(r) - c$
- Homogeneity: $\rho(\lambda r) = \lambda\rho(r)$ for any $\lambda > 0$.
- Monotonicity: for a pair (r_1, r_2), if $r_1 \geq r_2$, then $\rho(r_1) \leq \rho(r_2)$.
- Subadditivity: $\rho(r_1 + r_2) \leq \rho(r_1) + \rho(r_2)$

Definition 1.8 (Skewness)**.** Skewness coefficient of X measures the degree of assymetry, and is measured as

$$\mathrm{SK}(X) = \frac{\mu_3}{\sigma^3} = \frac{\mu_3}{\mu_2^{\frac{3}{2}}}$$

- $\mathrm{SK}(X) = 0$ denotes symmetric distribution.
- $\mathrm{SK}(X) > 0$ denotes positive skewness, which indicates that the distributionhas a long right tail compared to left tail.
- $\mathrm{SK}(X) > 0$ denotes negative skewness.

Definition 1.9 (Kurtosis)**.** The kurtosis of a random variable Y is defined as

$$\mathrm{Kur}(Y) = \frac{\mu_4}{\sigma^4} = \frac{\mu_4}{\mu_2^2}$$

We also define Kurtosis excess:

$$\mathrm{Ex.\,Kur}(Y) = \mathrm{Kur}(Y) - 3$$

Kurtosis is a measure of whether the data are heavy-tailed or light-tailed relative to a normal distribution. distribution has heavy tail if kurtosis is greater than 3, while normal has kurtosis 3.

Theorem 1.2 (Kurtosis of common distributions).

- $N(\mu, \sigma^2)$ has kurtosis 3.
- $t(\nu)$ has kurtosis $3 + \frac{6}{\nu-4}$
- $\log\text{N}(\mu, \sigma^2)$ has kurtosis $e^{4\sigma^2} + 2e^{3\sigma^2} + 3e^{2\sigma^2} - 3$.

Theorem 1.3 (Estimation of Skewness and Kurtosis). Suppose $Y_1, \dots Y_n$ are samples from a distribution. Let sample mean and standard deviation be $\bar{Y}$ and s. Then the sample skewness and kurtosis are

$$\hat{SK} = \frac{1}{n}\sum_{i=1}^{n}(\frac{Y_i - \bar{Y}}{s})^3$$

and

$$\hat{Kur} = \frac{1}{n}\sum_{i=1}^{n}(\frac{Y_i - \bar{Y}}{s})^4$$

and

$$\hat{\mu}_k = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^k$$

Theorem 1.4 (Test of Skewness). Given Null hypothesis H_0 : X is symmetric, under H_0, we have

$$\sqrt{n}\hat{SK} \sim N(0, \sigma^2_{SK})$$

where $\sigma^2_{SK} \approx 9 + \frac{\hat{\mu}_6}{\hat{\mu}_2^3} - 6\frac{\hat{\mu}_4}{\hat{\mu}_2^2}$. If X has normal distribution, the $\sigma^2_{SK} = 6$.

Theorem 1.5 (Test of Kurtosis). Given Null hypothesis H_0' : X has kurtosis 3, under H_0', we have

$$\sqrt{n}(\hat{Kur} - 3) \sim N(0, \sigma^2_{kurt})$$

where $\sigma^2_{kurt} \approx 24_6\frac{\hat{\mu}_3^2}{\hat{\mu}_2^3} - 8\frac{\hat{\mu}_3\hat{\mu}_5}{\hat{\mu}_2^4}$. If X hasnormal distribution, $\sigma^2_{kurt} = 24$.

Theorem 1.6 (Test of Normality). Jarque-Bera test check whether data have the skewnessand kurtosis matching a normal distribution

$$JB = \frac{n-k+1}{6}(\hat{SK}^2 + \frac{1}{4}(\hat{Kurt} - 3)^2)$$

where n is number of observations, and k is number of regressionin a model, if the data is the residuals, or $k = 0$ if it is observed.
If data is from normal distribution, then

$$JB \sum \chi^2(2)$$

Definition 1.10 (Heavy Tailed Distribution). The distribution of r.v. X is said to have a heavy right tail if

$$\lim_{x\to\infty} e^{\lambda x}P(X > x) = \infty \text{ for all } \lambda > 0$$

if heavy left tail if

$$\lim_{x\to-\infty} e^{\lambda|x|}P(X < x) = \infty \text{ for all } \lambda > 0$$

If

$$\lim_{x\to\infty}\frac{P(X > x)}{P(Y > x)} \to \infty \text{ or } \lim_{x\to-\infty}\frac{P(X < x)}{P(Y < x)} \to \infty$$

we say X is heavier tail than Y.

Definition 1.11 (Fat Tailed Distribution). If $P(|X| > z) \propto z^{-\alpha}$for some $\alpha > 0$.

Normal distribution does not have heavy tail, but t has, and log normal has right heavy tail.
For MLE, refer to ST2132 note.

Definition 1.12 (AIC, BIC). Akaike's Information Criterion(AIC) is defined as

$$AIC = -2\log\{\mathcal{L}(\hat{\theta}_{ML}\} + 2p$$

and Bayesian Information Criterion(BIC) is defined as

$$BIC = -2\log\{\mathcal{L}(\hat{\theta}_{ML}\} + \log(n)p$$

where p equals the number of parameters in the model/distributi and n the sample size. We call the last term the complexity penality.
A distribution with smaller AIC/BIC is preferred.

2 Multivariate Statictical Models

Theorem 2.1 (Multivariate Normal and t). Multivariate normal $N(\mu, \mathbf{\Sigma})$ has pdf

$$f_{\mathbf{x}}(x_1, \dots, x_p) = \frac{1}{(2\pi)^{k/2}|\mathbf{\Sigma}|^{1/2}}\exp(-\frac{1}{2}(\mathbf{x} - \mu)^T\sigma^{-1}(\mathbf{x} - \mu))$$

where $|\sigma|$ is the determinant of $\sigma = (\sigma_{ij})_{i\in(1,n),j\in(1,n)}$.
The marginal distribution of $\mathbf{x}_k$ is $N(\mu_k, \sigma_{kk} = \sigma_k^2)$.
Multivariate t distribution $t(\nu, \mu, \sigma)$ has pdf

$$f_{\mathbf{x}}(x_1, \dots, x_p) = \frac{\Gamma[(\nu + p)/2]}{\Gamma(\nu/2)(\nu\pi)^{p/2}}|\sigma|^{-1/2}[1 + \frac{1}{\nu}(\mathbf{x} - \nu)^T\sigma^{-1}(\mathbf{x} - \mu)]^-$$

Theorem 2.2 (Hypothesis Testing on correlation). Given null hypothesis $H_0 : \rho_{X,Y} = 0$. Under H_0, we have

$$\sqrt{n-3}r_{X,Y} \sim N(0, 1)$$

Definition 2.1 (Portfolio). Let $R_i, i = 1, \dots, p$ be the return of p assets and R the vector of R_i. We clearly see the expected return of ith asset $r_i = E(R_i)$ and risk $\sigma_i = \sigma_{R_i}$. We have the covariance matrix of R to be $\Sigma := (\sigma_{ij})$.
A **portfolio** is a new asset consisting of existing assets

$$R_N = \sum_{i=1}^{p} R_p = w^T R$$

where w_i is the weight of ith asset, and satisfies $\sum_{i=1}^{p} w_i = 1$.
The expected return of portfolio $E(R_N) = w^T E(R)$.
Variance of the portfolio is $w^T\Sigma w$.
The Sharpe ratio is

$$\frac{E(R_N - r_F)}{\sigma_N}$$

Theorem 2.3 (Portfolio of One Risky Asset And One Risk-free Asset)**.** $R_N = wR_1 + (1-w)r_f$. where R_1 has expected return r_1 and risk σ_1 and r_f is the return of risk-free asset.
Under this case, we have

- $r_N = wr_1 + (1-w)r_f$
- $\mathrm{Var}(R_N) = w^2\sigma_1^2$
- Sharpe ratio is $\frac{r_1 - r_f}{\sigma_1}$
- The line (σ_N, r_N) is called the capital market line.

Theorem 2.4 (Global Minimum Variance Portfolio)**.** we have $w = \frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}^T\Sigma^{-1}\mathbf{1}}$.

Theorem 2.5 (Tangency Portfolio)**.** we have $w = \frac{\Sigma^{-1}(r - r_f\mathbf{1})}{\mathbf{1}^T\Sigma^{-1}(r - r_f\mathbf{1})}$

Theorem 2.6 (Efficient Frontier)**.** Suppose w_t with target returns r_t is efficient, then

$$w_t = \frac{c - br_t}{ac - b^2}\Sigma^{-1}\mathbf{1} + \frac{ar_t - b}{ac - b^2}\Sigma^{-1}r$$

where $a = 1^T\Sigma^{-1}1$, $b = 1^T\Sigma^{-1}r$ and $c = r^T\Sigma^{-1}r$.
Also, suppose w_A, w_B efficient, then $w_C = \alpha w_A + (1-\alpha)w_B$ is also efficient as long as $r_C > r_{Min.Var}$.

3 Copula

Definition 3.1 (Copula)**.** A copula is a special multivariate cumulative distribution function(CDF),

$$C(u_1, \ldots, u_p)$$

whose univariate marginal distribution are all $U(0,1)$.
As a result, copula has the following property:

- $(u_1, \ldots, u_p) \in [0,1]^p$.
- $C(u_1, \ldots, u_p)$ is increasing on $[0,1]^p$.
- $C(u_1, \ldots, u_p)$ has margin CDF as

  $$C_k(u) = C(1, \ldots, 1, u, 1, \ldots, 1) = u$$

  for all $u \in [0,1]$.

The density of a copula $c(u_1, \ldots, u_p)$ is given by

$$c(u_1, \ldots, u_p) = \frac{\partial^p C(u_1, \ldots, u_p)}{\partial u_1 \ldots \partial u_p}$$

Theorem 3.1 (Sklar's theorem)**.** Any continuous random vector $X = (\mathbf{x}_1, \ldots, \mathbf{x}_p)^T$ has a copula.

- If $\mathbf{x}$ is continuous univariate with cumulative distribution $F_{\mathbf{x}}(x)$, then

  $$F_{\mathbf{x}}(x) \sim U[0,1]$$

- For any random vector $X = (\mathbf{x}_1, \ldots, \mathbf{x}_p)^T$ with joint CDF $F(x_1, \ldots, x_p)$ and marginal CDF $F_1(x_1), \ldots, F_p(x_p)$, let $\mathbf{u}_i = F_i(\mathbf{x}_i)$, and $\mathbf{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_p)^T$, then

  $$\begin{aligned} C^*(u_1, \ldots, u_p) :&= P(\mathbf{u}_1 \le u_1, \ldots, \mathbf{u}_p \le u_p) \\ &= P(\mathbf{x}_1 \le F_1^{-1}(u_1), \ldots, \mathbf{x}_p \le F_p^{-1}(u_p)) \\ &= F(F_1^{-1}(u_1), \ldots, F_p^{-1}(u_p)) \end{aligned}$$

  is a copula.
  Under such definition, we have $F(x_1, \ldots, x_p) = C^*(F_1(x_1),$

- For the copula above, its density is

  $$c^*(u_1, \ldots, u_p) = \frac{f(F_1^{-1}(u_1), \ldots, F_p^{-1}(u_p))}{f_1(F_1^{-1}(u_1)) \times \cdots \times f_p(F_p^{-1}(u_p))}$$

- For any marginal distribution $F_1(u_1), \ldots, F_p(u_p)$, we can introduce a copula $C_{new}(u_1, \ldots, u_p)$ to combine them, and generate a joint distribution

  $$F_{new}(y_1, \ldots, y_p) = C_{new}\{F_{\mathbf{y}_1}(y_1), \ldots, F_{\mathbf{y}_p}(y_p)\}$$

  and density

  $$f_{new}(y_1, \ldots, y_p) = c_{new}\{F_{\mathbf{y}_1}(y_1), \ldots, F_{\mathbf{y}_p}(y_p)\}f_{\mathbf{y}_1}(y_1) \times \cdots$$

- X and Y are independent if and only if their copular is

  $$C(u,v) = uv, 0 \le u, v \le 1$$

  or

  $$c(u,v) = 1, 0 \le u, v \le 1$$

- Independence copula: $C(u_1, \ldots, u_p) = u_1 \cdots u_p$
- Co-monotonicity copula: $M(u_1, \ldots, u_p) = \min(u_1, \ldots, u_p)$
- Counter-monotonicity copula: $W(u_1, \ldots, u_p) = \max\{1 - p + \sum_{i=1}^p u_i, 0\}$.

For any copula $C(u_1, \ldots, u_p)$, we have

$$W \le C \le M$$

Definition 3.2 (Gaussian Copula)**.** For a given covariance matrix $\Sigma \in \mathbb{R}^{p \times p}$, the Gaussian copula with parameter matrix Σ and expectation μ can be written as

$$C_\Sigma^{\text{gauss}}(u) = \Phi_\Sigma(\Phi_1^{-1}(u_1), \ldots, \Phi_p^{-1}(u_p))$$

where Φ_k^{-1} is the inverse cumulative distribution of $N(0, \sigma_{kk})$ and Φ_Σ is the joint cumulative distribution function of a multivariante normal distribution with mean vector 0 and covariance matrix equal to the correlation matrix $\Sigma = R$.
For bivariate case, we have

$$C_\rho^{\text{gauss}}(u_1, u_2) = \Phi_\rho(\Phi_1^{-1}(u_1), \Phi_2^{-1}(u_2))$$

or equivalently,

$$\begin{aligned} &C(u_1, u_2; \rho) \\ &= \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\{-\frac{s_1^2 - 2\rho s_1 s_2 + s_2^2}{2(1-\rho^2)}\} \, \mathrm{d}\, s_1 \, \mathrm{d} \end{aligned}$$

where Φ^{-1} is inverse of CDF of standard normal.
Note this copula is jointly normal, even if the distribution is not normal.

Definition 3.3 (t copula)**.**

$$C_{\nu,\Sigma}(u_1,\dots,u_p) = \int_{-\infty}^{t_\nu^{-1}(u_1)} \cdots \int_{-\infty}^{t_\nu^{-1}(u_p)} \frac{\Gamma[(\nu+p)/2]}{\Gamma(\nu/2)(\nu\pi)^{p/2}} |\Sigma|^{-1/2} [1 + \frac{1}{\nu} X^T \Sigma^{-1} X]^{-(\nu+p)/2} \, \mathrm{d}\, x_1 \cdots \mathrm{d}\, x_p$$

where t_ν^{-1} is the quantile function of a standard univariate t_v distribution, and $X = (x_1,\dots,x_p)^T$.

Theorem 3.2. The copula generated by $X = (x_1,\dots,x_p)$ is the same as that generated by $\bar{X} := (\frac{x_1-\mu_1}{\sigma_1},\dots,\frac{x_p-\mu_p}{\sigma_p})$

Definition 3.4 (Archimedean Copula)**.** $C(u_1,\dots,u_p) = \psi(\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_p))$ where ψ is called *generater.*
The table for famous generaters is omitted here.

Theorem 3.3 (Mixture of copula is still a copula)**.** If C and D are two copulas, then for any $w \in [0,1]$, $wC + (1-w)D$ is still a copula.

Let $I_i = \mathbf{1}(y_i \le F_i^{-1}(u)), i = 1,2$. We can consider I_i to be the indicator for default.
Then $E(I_1 I_2) = F(F_1^{-1}(u), F_2^{-1}(u)) = C(u,u)$ is a copula.
Furthermore, we have

$$Cor(I_1, I_2) = \frac{C(u,u) - u^2}{u - u^2}$$

So the lower tail correlation of y_1 and $y-2$ is

$$\lim_{u\to 0} Cor(I_1, I_2) = \lim_{u \to 0} \frac{C(u,u)}{u}$$

Definition 3.5 (Lower Tail Dependence, Upper Tail Dependence)**.** We define lower tail dependence λ_L

$$\lambda_L := \lim_{q\to 0^+} P(\mathbf{y}_2 \le F_{\mathbf{y}_2}^{-1}(q) \mid \mathbf{y}_1 \le F_{\mathbf{y}_1}^{-1}(q)) = \lim_{q \to 0^+} \frac{C_Y(q,q)}{q}, \text{ where } Y = (\mathbf{y}_1, \mathbf{y}_2)^T$$

and upper tail dependence λ_U

$$\lambda_U := \lim_{q\to 1^-} P(\mathbf{y}_2 \ge F_{\mathbf{y}_2}^{-1}(q) \mid \mathbf{y}_1 \ge F_{\mathbf{y}_1}^{-1}(q)) = \lim_{q \to 1^-} \frac{1 - 2q + C_Y(q,q)}{1-q}$$

The tail dependence of copula table is included in the lecture note and is omitted here. Consider only bivariate case $Z = (\mathbf{x}, \mathbf{y})$, whose distribution, both marginal and joint) is unknown and needs to estimate.
For observations $(x_i, y_i), i = 1,\dots,n$, the empricial CDF of Z is

$$\hat{F}(x,y) = \frac{1}{n+1} \sum_{i=1}^n I(x_i \le x, y_i \le y)$$

and the marginal empirical CDF is

$$\hat{F}_{\mathbf{x}}(x) = \frac{1}{n+1} \sum_{i=1}^n I(x_i \le x) = \frac{\#\{x_i : x_i \le x\}}{n+1}$$

and similarly

$$\hat{F}_{\mathbf{y}}(y) = \frac{1}{n+1} \sum_{i=1}^n I(y_i \le y) = \frac{\#\{y_i : y_i \le y\}}{n+1}$$

We define $u_j = \hat{F}_{\mathbf{x}}(x_j) = \frac{\text{rank of } x_j \text{ in } x_1,\dots,x_n}{n+1}$ and $v_j = \hat{F}_{\mathbf{y}}(y_j) = \frac{\text{rank of } y_j \text{ in } y_1,\dots,y_n}{n+1}$, and they are observations from $\mathbf{u}_{\mathbf{x}} := F_{\mathbf{x}}(\mathbf{x})$ and $\mathbf{u}_{\mathbf{y}}$. It is easy to see they can form a corpula.
We define the **empirical copula** to be

$$\hat{C}(u_1, u_2) = \frac{1}{n} \sum_{i=1}^n I(u_i \le u_1, v_i \le u_2)$$

To model a joint distribution of $(\mathbf{y}_1,\dots,\mathbf{y}_p)$, we only need to model the marginal distribution of each variable $F_k(y_k \mid \theta_k), k = 1,\dots,p$, as well as the copula density $c(u_1,\dots,u_p \mid \theta_C)$.
For example, we can have $F(x,y) = C_t(t_{\nu_1}(\frac{x-\mu_1}{s_1}), t_{\nu_2}(\frac{x-\mu_2}{s_2}) \mid \rho, \nu)$.
More generally, the joint density is $c(F_1(y_1 \mid \theta_1),\dots,F_p(y_p \mid \theta_p) \mid \theta_C) f_1(y_1 \mid \theta_1) \cdots f_p(y_p \mid \theta_p)$. For samples, $Y_1 = (y_{11},\dots,y_{1p}$ $(y_{n1},\dots,y_{np})^T$, and the log-likelihood is

$$\log\{L\} = \sum_{i=1}^n \log[c(F_1(y_1 \mid \theta_1),\dots,F_p(y_p \mid \theta_p) \mid \theta_C) f_1(y_1 \mid \theta_1) \cdots$$

The maximum values for the parameters are the MLE of the parameters $(\theta_1,\dots,\theta_p;\theta_C)$.
Note that $F_k(y_{ik} \mid \theta_k) \approx u_{ik}$, so we can estimate θ_C directly by maximizing

$$\log\{L\} = \sum_{i=1}^n \log\{c(u_{ik},\dots,u_{ik} \mid \theta_C)\}$$

If we need to estimate the distribution function, we then only need to estimate each marginal density separately, by maximizing

$$\sum_{i=1}^n \log\{f_k(y_{ik} \mid \theta_k)\}$$

We can also use AIC or BIC to select among different copulas. We prefere the copula with smallest AIC or BIC. Suppose $G(u), F(x_1,\dots,x_p)$ are strictly continuous increasing CDFs, and $F_i(x_i), i = 1,\dots,p$ are the marginal CDF of $F(x_1,\dots,x_p)$.
Suppose $\mathbf{z} \sim G(x)$, and let $\mathbf{u} = G(\mathbf{z})$, then

$$\mathbf{u} \sim U(0,1)$$

If $\mathbf{u} \sim U(0,1)$, and let $\mathbf{z} = G^{-1}(u)$, then

$$\mathbf{z} \sim G(z)$$

Let $X = (\mathbf{x}_1,\dots,\mathbf{x}_p) \sim F(x_1,\dots,x_p)$, then

- Marginal $\mathbf{x}_i \sim F_i(x_i) = F(\infty,\dots,x_i,\dots,\infty)$.
- Let $U = (F_1^{-1}(\mathbf{x}_1),\dots,F_p^{-1}(\mathbf{x}_p))$, then

$$U \sim F(F_1^{-1}(u_1),\dots,F_p^{-1}(u_p))$$

Let $U = (\mathbf{u}_1,\dots,\mathbf{u}_p) \sim C(u_1,\dots,u_p)$, and let $X = (F_1^{-1}(\mathbf{u}_1),\dots$
For a general distribution $\mathbf{x} \sim F(x)$, we can generate random numbers from it

$$\mathbf{x}_i = F^{-1}(\mathbf{u}_i)$$

where $\mathbf{u}_i \sim U(0,1)$.
As a result,

$$E(\mathbf{x}^k) = \int x^k \,\mathrm{d}\,F(x) = \lim_{N\to\infty} \frac{\sum_{i=1}^N \mathbf{x}_i^k}{N}$$

and

$$\mathrm{VaR}_q(X) = -Q_q(\mathbf{x}) = -\lim_{N\to\infty} \mathit{quantile}(\{x_i, i=1,\dots,N\}, \mathit{probs}=q)$$

and ES

$$\mathrm{ES}_q(\mathbf{x}) = \frac{1}{q}\int_{-\infty}^{Q_q(\mathbf{x})} x \,\mathrm{d}\,F(x) = -\lim_{N\to\infty} \frac{\sum_{\mathbf{x}_i;\mathbf{x}_i<Q_q(\mathbf{x})} \mathbf{x}_i}{\#\{\mathbf{x}_i : \mathbf{x}_i < Q_q(\mathbf{x})\}}$$

4 CAPM

Suppose N assets denoted by $i = 1,\dots,N$, with returns in period t, $R_{it}, t = 1,\dots,T$

$$R_{it} = \alpha_i + \beta_i R_{mt} + \epsilon_{it}, i = 1,\dots,N; t = 1,\dots,T$$

where

- α_i, β_i constant over time,
- R_{mt} is return on **diversified** market index
- ϵ_{it} random error term unrelated to R_{mt}.

We also assume

- $\mathrm{Cov}(R_{mt}, \epsilon_{is}) = 0$ for all i, t, s.
- $\mathrm{Cov}(\epsilon_{is}, \epsilon_{jt})$ for all $i \neq j, t$ and s
- $\epsilon_{it} \sim IIDN(0, \sigma^2_{\epsilon,i})$
- $R_{m,t} \sim IIDN(\mu_m, \sigma^2 m)$

We can easily see that, by computing the numerator

$$\beta_i = \frac{\mathrm{Cov}(R_{it}, R_{mt})}{\mathrm{Cov}(R_{mt})} = \frac{\sigma_{iM}}{\sigma^2_m}$$

which is the contribution of asset i to the volatility of the market index.
Also, we notice that returns is correlated only through their exposure to common market-wide news.

Theorem 4.1 (Statistical Properties of SI Model)**.** *For SI Model $R_{it} = \alpha_i + \beta_i R_{mt} + \epsilon_{it}$,*

- $\mu_i = \alpha_i + \beta_i + \mu_m$
- $\sigma^2_i = \beta^2_i \sigma^2_m + \sigma^2_{\epsilon,i}$
- $\sigma_{ij} = \mathrm{Cov}(R_{it}, R_{jt}) = \sigma^2_m \beta_i \beta_j$.
- $R_{it} \sim N(\mu_i, \sigma^2_i)$.

Theorem 4.2 (Decomposition of Total Variance)**.**

$$\sigma^2_i = \beta^2_i \sigma^2_m + \sigma^2_{\epsilon,i}$$

and we have

$$1 = \frac{\beta^2_i \sigma^2_m}{\beta^2_i} + \frac{\sigma^2_{\epsilon,i}}{\sigma^2_i} = R^2_i + (1 - R^2_i)$$

where R^2_i is proportion of market variance, and $1-R^2_i$ proportion of non-market variance.
Similarly, covariance matrix can be decomposed

$$\Sigma = \sigma^2_m \beta^T \beta + Diag[\sigma^2_{\epsilon,1} \dots \sigma^2_{\epsilon,n}]$$

The capital asset pricing model is

$$R_i = r_f + \beta_i(R_m - r_f) + \epsilon_i$$

where

- R_m is return of the market
- $E(R_m) - r_f$ is known as market premium.
- $E(R_i) - r_f$ is known as risk premium.
  CAPM can be written as $E(R_i) - r_f = \beta_i(E(R_m) - r_f)$
- β_i is the sensitivity of expected excess asset returns to the expected excess market returns, or

$$\beta_i = \frac{\mathrm{Cov}(R_i, R_m)}{\mathrm{Cov}(R_m)}$$

- We can also write β_i as $\beta_i = \rho_{R_i,R_m}\frac{\sigma_{R_i}}{\sigma_{R_m}}$, where ρ is the correlation coefficient.
- The expected market return is usually estimated by measuring the log-return of the historical returns.

To price using CAPM, notice

$$E(R_T) = \frac{E(P_T) - P_0}{P_0}$$

Here, P_0 is fixed but P_T is a random variable.
By CAPM,

$$E(R_T) = r_f + \frac{\mathrm{Cov}(R_T, R_m)}{\mathrm{Cov}(R_m)}(E(R_m) - r_f)$$

So,

$$P_0 = \frac{1}{1+r_f}[E(P_T) - \frac{\mathrm{Cov}(P_T, R_m)}{\mathrm{Cov}(R_m)}(E(R_m) - r_f)]$$

where P_T is the expected price of asset. Suppose p risky asset R_i with expected return r_i and risk σ_i, $i = 1,\dots,p$. Let r_f be risk free return.
The optimal market portfolio

$$R_m = \tilde{w}_i R_1 + \dots + \tilde{w}_p R_p$$

is the tangency portfolio, or at the CML, with expected return ,risk and covariance

$$r_m := E(R_m), \sigma_m^2 = \mathrm{Cov}(R_m), \sigma_{i,m} = \mathrm{Cov}(R_i, R_m)$$

Consider a portfolio,

$$R_N = \sigma_{i=1}^{p} w_i R_i$$

with expected return

$$r_N = \sum_{i=1}^{p} w_i E(R_i)$$

Obviously,

$$\sigma_N^2 = \sum_{i=1}^{p} w_i^2 \sigma_i^2 + \sum_{i} \sum_{j \neq i} w_i w_j \mathrm{Cov}(R_i, R_j)$$

We define

$$C = \sigma_N + \lambda[r_N - \sum_{i=1}^{p} w_i E(R_i) - \underbrace{(1 - \sum_{i=1}^{p} w_i) r_f}_{} = 0]$$

We need to minimize C. Using Lagrange multiplier, we have

$$\sigma_N = \lambda(r_N - r_f)$$

where $\lambda = \frac{r_m - r_f}{\sigma_m}$. Then we use Lagrange multiplier equation to get the equation on $E(R_i)$, which gives CAPM. Suppose asset return R_i is driven by K common factors, $f_1, \ldots, f_K$ and idiosyncratic noise u_i. A multi-factor regression model is

$$R_i = \alpha_i + \sum_{j=1}^{K} \beta_{ij} f_j + u_i$$

for $i = 1, \ldots, n$, a total of n assets.
Here, α_i is regression intercept for return of asset i.
f_j are common factors driving **all** asset returns, with $\mathrm{Cov}(f_i) = \lambda_i$, and $\mathrm{Cov}(f_i, f_j) = 0$ for $i \neq j$.
β_{ij} gives how sensitive the return of asset i is, with respect to jth factor, which is called the factor loading of asset i on factor f_j.
u_i is idiosyncratic component in ith return, with $E(u_i) = 0$ and $\mathrm{Cov}(u_i) = \sigma_{u_i}^2$, and $\mathrm{Cov}(u_i, u_j) = 0$ for $i \neq j$.
u_i is independent of $f_1, \ldots, f_K$.

Theorem 4.3. If market is efficient, and factors are indeed complete, the the model is

$$R_i - r_f = \sum_{j=1}^{K} \beta_{ij} f_j' + u_i$$

where f_j' needs to remove risk-free return.

Theorem 4.4 (Expectation and Variance of R_i)**.**

$$E(R_i) - r_f = \sum_{j=1}^{K} \beta_{ij} E(f_j')$$

$$\mathrm{Cov}(R_i) = \sum_{j=1}^{K} \beta_{ij}^2 \mathrm{Cov}(f_j) + \sigma_{u_i}^2$$

$$\mathrm{Cov}(R_i, R_j) = \sum_{k=1}^{K} \beta_{ik} \beta_{jk} \mathrm{Cov}(f_k)$$

Definition 4.1 (Fama-French Model)**.**

$$R_{i,t} = r_{f,t} + \beta_{i,m}(R_{t,m} - r_{f,t}) + \beta_{i,s} SMB_t + \beta_{i,v} HML_t + \alpha_i + \epsilon_{i,t}$$

where

- $R_{i,t}$ the stock return during period t.
- $r_{f,t}$ the return on risk-free asset.
- α_i zero, under certain assumptions.
- β is beta
- $\epsilon_{i,t}$ regression error.
- SMB is the return of small capital firms minus that of big.
- HML is return of high (book value)/(market value) ratio minus that of low ratio

Definition 4.2 (Arbitrage Pricing Model)**.**

$$E(r_j) = r_f + \sum_{k=1}^{K} \beta_{jk} RP_k$$

where RP_k is the risk premium of the factor, ususally $RP_k = E(f_k - r_f)$, or $= E(f_k)$.

Given a p dimensional random variable $X = (x_1, \ldots, x_p)^T$, with $E(X) = \mu$ and $\Sigma = \mathrm{Cov}(X)$.
Consider

$$\Sigma = W \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \vdots \\ 0 & 0 & \cdots & \lambda_p \end{pmatrix} W^T$$

which is the eigenvalue-eigenvector decomposition, where $\lambda_1 \geq \ldots \geq \lambda_p \geq 0$, and $W = (\mathbf{w}_1, \ldots, \mathbf{w}_p)$ is an orthogonal matrix, i.e. $W_W^T = W^T W = I_p$.
The orthogonal means $\mathbf{w}_i \mathbf{w}_j = 1$ if and only if $i = j$, and 0 otherwise.
Also, we have $\Sigma \mathbf{w}_i = \lambda_i \mathbf{w}_i$.
We can write $\Sigma = \lambda_1 \mathbf{w}_1 \mathbf{w}_1^T + \ldots + \lambda_p \mathbf{w}_p \mathbf{w}_p^T$, and

$$\mathbf{w}_i^T \Sigma \mathbf{w}_i = \lambda_i$$

We can $\mathbf{z}_i = \mathbf{w}_i^T X$, the ith principal component, where

$$\mathrm{Cov}(\mathbf{z}_i) = \lambda_i$$

and

$$\mathrm{Cov}(\mathbf{z}_i, \mathbf{z}_j) = 0 \text{ if } i \neq j$$

The first principal component of X is the linear combination $\mathbf{z}_1 = \mathbf{w}_1^T X$ that maximizes $\mathrm{Cov}(\mathbf{z}_1)$ subject to the constraint

$\mathbf{w}_1^T\mathbf{w}_1 = 1$.
Similarly, $\mathbf{z}_2$ maximizes $\text{Cov}(\mathbf{z}_2)$ subject to $\mathbf{w}_2^T\mathbf{w}_2 = 1$ and $\text{Cov}(\mathbf{z}_2, \mathbf{z}_1) = 0$.
This applies to all other $\mathbf{z}_i$'s.
We can write $(\mathbf{z}_1, \ldots, \mathbf{z}_p)^T = (\mathbf{w}_1, \ldots, \mathbf{w}_p)^T X$.
For total variance $\sum_{i=1}^p \text{Cov}(x_i)$, we have

$$\sum_{i=1}^p \text{Cov}(x_i) = tr(\Sigma) = tr(W \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0\vdots \\ 0 & 0 & \cdots & \lambda_p \end{pmatrix} W^T)$$
$$= tr(\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0\vdots \\ 0 & 0 & \cdots & \lambda_p \end{pmatrix})$$
$$= \sigma_{i=1}^p \lambda_i$$

And $\frac{\lambda_i}{\sum_{j=1}^p \lambda_j}$ is the proportion of total variance in X explained by the ith principal component.
Suppose we define $z = w^T X$, and use it to predict each variables in X by model

$$x_i = \alpha_i + \beta_i \mathbf{z} + \epsilon_i$$

where $E(\epsilon_i) = 0$ and $\text{Cov}(z, \epsilon_i) = 0$. The prediction error for this stock i is defined as

$$\epsilon_i = x_i - \alpha_i - \beta_i z$$

We hope to minimize

$$\min_w E(\sum_{i=1}^p \epsilon_i^2)$$

then the solution is $w = \mathbf{w}_1$ and common factor is $\mathbf{z} = \mathbf{w}_1^T X$. where $\mathbf{w}_1$ is the first principal component.

5 Financial Time Series

Definition 5.1 (Time Series, Realization)**.** A **time series** $\{Z_t\}$ is a sequence of random variables indexed by time t

$$\{\cdots, Z_1, \cdots, Z_{t-1}, Z_t, Z_{t+1}, \cdots\}$$

A realization of a stochastic process is the sequence of observed data

$$\{\cdots, Z_1 = z_1, \cdots, Z_{t-1} = z_{t-1}, Z_t = z_t, Z_{t+1} = z_{t+1}, \cdots\}$$

We use $\{z_t\}$ for both theoretical time series and its observations.

Definition 5.2 (Mean, Variance and Covariance Function)**.** Let $\{z_t\}$ be a time series with $E(z_t^2) < \infty$.
The mean function is $\mu_t = E(z_t)$.
The variance function is $\sigma_t^2 = \text{Cov}(z_t) = E[(z_t - \mu_t)^2]$
The covariance function is $\gamma(r, s) = \text{Cov}(z_r, z_s) = E[(z_r - \mu_r)(z_s - \mu_s)]$

Definition 5.3 (Strictly Stationary)**.** $\{z_t\}$ is strictly stationary if for any given finite integer k, and for any set of subscript $t_1, \ldots, t_k$, the joint distribution of $z_{t_1}, \ldots, z_{t_k}$ depends only on $t_1 - t_2, \ldots, t_{k-1} - t_k$, but not directly on $t_1, \ldots, t_k$.

Some of the consequences are

- If $\{Z_t\}$ is an IID sequence, then it is strictly stationary.
- Let $\{Z_t\}$ be an iid sequence and let X independent of $\{Z_t\}$. Let $Y_t = Z_t + X$, and then $\{Y_t\}$ is strictly stationary.
- For any function g, $\{g(z_t)\}$ is also strictly stationary.

Definition 5.4 (Weakly Stationary)**.** $\{z_t\}$ is weakly stationary if

- μ_t does not depend on t, denoted by μ
- $\gamma(r, s) = \gamma(|r - s|)$.

By the definition, we have $\text{Cov}(z_t) = 0$.

Theorem 5.1. If $\{Z_t\}$ is strictly stationary and $\text{Cov}(Z_t) < \infty$, then $\{Z_t\}$ is also weakly stationary. However, a weakly stataionary time series is usually not strictly stationary.

Definition 5.5 (Autocovariance Function)**.** If $\{z_t\}$ is stationary, then

- $\gamma(h) = \text{Cov}(z_{t+h}, z_t)$ is called autocovariance function(ACVF) of $\{z_t\}$.
- $\rho_z(h) := \frac{\gamma(h)}{\gamma(0)} = \text{Corr}(z_{t+h}, z_t)$ is called **autocorrelation** function(ACF) at lag h of $\{z_t\}$.

Basic properties of ACVF include

- $\gamma(0) \geq 0$
- $|\gamma(h)| \leq \gamma(0)$ for all h
- $\gamma(h) = \gamma(-h)$

Definition 5.6 (White Noise Sequence)**.** If $\{z_t\}$ is a stationary with $E(z_t) = 0$, $\gamma(0) = \sigma^2$ and

$$\gamma(h) = 0, \text{ for any } h \neq 0$$

then we call $\{z_t\}$ white noise sequence, denoted by $z_t WN(0, \sigma_z^2)$.

Suppose $\{y_t\}$ stationary, and $\{\epsilon_t\}$ a $WN(0, \sigma^2)$. If they are independent, i.e. $E(y_t\epsilon_{t+k}) = E(y_t)E(\epsilon_{t+k})$ for all t, k, then

$$z_t = (a + by_t^2)^{1/2}, a \geq 0, b \geq 0$$

is also a WN.
Suppose $\{\epsilon_t\}$ is a sequence of IID $N(0, \sigma^2)$, thus WN. Let

$$z_t = (a + bz_{t-1}^2)^{1/2}\epsilon_t, a \geq 0, b \geq 0$$

then $\{z_t\}$ is also a WN. THis is the Arch model.
We can build stationary sequences from White Noise. For example, $y_t = z_t + \theta z_{t-1}$, where $\{z_t\}$ is a white noise.

Theorem 5.2 (Wold's Decomposition Theorem)**.** Any weakly stationary time series $\{Y_t\}$ can be represented in the form

$$\begin{aligned} Y_t &= \mu + \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + \cdots \\ &= \mu + \sigma_{k=0}^{\infty} \phi_k \epsilon_{t-k} \end{aligned}$$

where

$$\phi_0 = 1, \sigma_{k=1}^{\infty} \phi_k^2 < \infty$$

and

$$\{\epsilon_t\} \text{ is } WN(0, \sigma^2)$$

Such Y_t has the following properties:

- $E(Y_t) = \mu$.
- $\gamma(0) = \mathrm{Cov}(Y_t) = \sigma^2 \sum_{k=0}^{\infty} \phi_k^2$
- When $h > 1$, $\gamma(h) = \sigma^2 \sum_{j=0}^{\infty} \phi_j \phi_{h+j}$.

In finance data, the stationary can usually be obtained by considering the first-difference

$$z_t = x_t - x_{t-1}$$

If z_t is stationary, then x_t is integrated of order 1.
We can also consider kth-difference. Suppose we have observation $z_1, \ldots, z_n$. Suppose that $\{z_t\}$ is stationary, we then have to estimate μ_z, $\gamma(h)$ and $\rho_z(h)$, with $h = 0, 1, \ldots$.
We have

- Sample Mean $\bar{z} = \frac{1}{n} \sum_{t=1}^{n} z_t$
- Sample Variance $\hat{\sigma}_z^2 = \frac{1}{n} \sum_{t=1}^{n} (z_t - \bar{z})^2 = \hat{\gamma}(0)$
- Sample autocovariance function at lag h (SACVF):

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-h} (z_t - \bar{z})(z_{t+h} - \bar{z})$$

- Sample autocorrelation coefficient function at lag h, (SACF):

$$r\hat{h}o(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$$

Theorem 5.3 (Testing $\rho(h) = 0$)**.** Suppose $z_1, \ldots, z_n$ is realization of stationary time series.
The SACF at lag h is

$$r_h = \frac{\sum_{t=1}^{n-h} (z_t - \bar{z})(z_{t+h} - \bar{z})}{\sum_{t=1}^{n} (z_t - \bar{z})^2}$$

where

$$\bar{z} = \frac{1}{n} \sum_{i=1}^{n} z_t$$

Under $H_0' : \rho(h) = 0$ for all $h > 0$,

$$s_{r_h} = (\frac{1}{n})^{\frac{1}{2}}$$

If $|r_h| < 2 s_{r_h}$, then accept H_0. Otherwise, reject H_0.

Theorem 5.4 (Test of White Noise)**.** The Ljung-Box test has test statistics

$$Q(h) = n(n+2) \sum_{k=1}^{h} \frac{r(k)^2}{n-k}$$

where n is sample size, $r(k)$ is sample autocorrelation at lag k, and h the number of lags being tested.
Under $H_0 : \rho(k) = 0, k = 1, \ldots, h$,

$$Q(h) \sim \xi^2(h)$$

For significance level α, the critical region for rejection of hypothesis of randomness is rejected if the statistic

$$Q^*(h) > \xi_{1-\alpha}^2(h)$$

where $\xi_{1-\alpha}^2(h)$ is the $(1-\alpha)$-quantile of the chi-square distribution with h degrees of freedom.

$$p\text{-value} = P(\xi^2(h) > Q^*(h))$$

We accept H_0 if p-value$> \alpha$.

For any integer $q \geq 1$, the moving average model of order q, $MA(q)$ is

$$x_t = \mu + a_t + \theta_1 a_{t-1} + \cdots + \theta_q a_{t-q}$$

More generally, let

$$x_t = \mu + \sum_{j=0}^{\infty} \psi_j a_{t-j}$$

where $\psi_0 = 1$, and $\{a_t\} \sim WN(0, \sigma^2)$ and

$$\sum_{j=0}^{\infty} |\psi_j| < \infty$$

$\{x_t\}$ is called a stationary general linear process or a $MA(\infty)$ process.
For $MA(q)$ we have

- $E(x_t) = \mu$
- $\mathrm{Cov}(x_t) = (1 + \theta_1^2 + \cdots + \theta_q^2)\sigma^2$
- If $|h| \leq q$,

$$\mathrm{Cov}(x_t, x_{t-h}) = \sigma^2 \sum_{j=0}^{q-|h|} \theta_j \theta_{j+|h|}$$

- $\mathrm{Cov}(x_t, x_{t-h}) = 0$ if $|h| > q$

An autoregressive model of order p, AR(p) is

$$x_t = \phi_0 + \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + a_t$$

where $a_t \sim WN(0, \sigma^2)$.
It follows

$$E(x_t) = \phi_0 + \phi_1 E(x_{t-1}) + \cdots + \phi_p E(x_{t-p})$$

If it is stationary, let $\mu = E(x_t)$. We have

$$\mu = \frac{\phi_0}{1 - \phi_1 - \cdots - \phi_p}$$

So AR(p) can be written as

$$x_t = (1 - \phi_1 - \cdots - \phi_p)\mu + \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + a_t$$

i.e.

$$y_t = \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + a_t$$

where $y_i = x_i - \mu$. Therefore, in theoretical analysis, we only need to consider $\mu = 0$.

Consider AR(1), where $x_t \phi_1 x_{t-1} + a_t$. If $|\phi_1| < 1$, we have

$$x_t = \sum_{k=0}^{\infty} \phi_1^k a_{t-k}$$

which is a general linear process with $\sum_{m=0}^{\infty} |\phi_1^m| < \infty$. Therefore,

- If $|\phi_1| < 1$, AR(1) is stationary.
- If $|\phi_1| > 1$, the time series will diverge to ∞.
- If $|\phi_1| = 1$, the times series will stay in a reasonable region, and $x_t = x_{t-1} + a_t$. This is not stationary, but a random walk.

Definition 5.7 (Backshift Operator)**.** For any time series y_t, define

$$Ly_t = y_{t-1}, L^k y_t = y_{t-k}, k \geq 1$$

A general AR(p) model can be written as

$$\phi_p(L) y_t = a_t$$

where $\phi_p(L) = 1 - \phi_1 L - \cdots - \phi_p L^p$.

A general $MA(q)$ model can be written as

$$y_t = \theta_q(L) a_t$$

where $\theta_q(L) = 1 + \theta_1 L + \cdots + \theta_q L_q$.
Note that $(1 - L)X_t = X_t - X_{t-1}$, which indicates that 1_L is a difference operator. We can also consider higher order difference $(1 - L)^d$.
We have proved that for AR(1), if $|\phi_1| < 1$, then x_t is stationary. Equivalently, if the root of

$$1 - \phi_1 L = 0$$

is outside the unit circle $|z|^2 = 1$,i,e, $|L| = |1/\phi_1 > 1$, then AR(1) modek is stationary.
For AR(p) model, if all roots of $\phi_p(L) = 0$ are outside of unit circle, then the time series is stationary.
Generally, if AR(p)

$$y_t = \phi_0 + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + a_t$$

is stationary, then it can be written as

$$y_t = \tilde{\mu} + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots$$

where $\sum_{k=1}^{\infty} |\psi_k| < \infty$. Thus,

- $\mathrm{Cov}(y_t, a_t) = \sigma^2$ and $\mathrm{Cov}(y_t, a_{t+h}) = 0$ for any $h > 0$
- $\mathrm{Cov}(y_t, y_{y-h}) = \phi_1 \mathrm{Cov}(y_{t-1}, y_{t-h}) + \cdots + \phi_p \mathrm{Cov}(y_{t-p}, y_{t-h}$ when $h > 0$
- $\mathrm{Cov}(y_t, y_t) = \phi_1 \mathrm{Cov}(y_{t-1}, y_t) + \cdots + \phi_p \mathrm{Cov}(y_{t-p}, y_t) + \sigma^2$

Definition 5.8 (ARMA model, ARIMA model)**.** *The* ARMA(p, *model is*

$$X_t = \phi_0 + \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + a_t + \psi_1 a_{t-1} + \cdots + \psi_q a_{t-q}$$

Again, let $\mu = E(X_t)$*, we have*

$$X_t - \mu = \phi_0 + \phi_1(X_{t-1} - \mu) + \cdots + \phi_p(X_{t-p} - \mu) + a_t + \psi_1 a_{t-1} + \cdots + \psi$$

Sometimes, we consider $Z_t = (1-L)^d X_t$*. If* Z_t *follows* ARMA(p, *model, then we call* X_t *follows* ARIMA(p, d, q) *model.*

The ARMA(p, q) model has parameter

$$\theta = (\mu, \phi_1, \ldots, \phi_p, \psi_1, \ldots, \psi_q, \sigma^2)$$

to be estimated.
If p, q konwn, we can use maximum log-likelihood. If p, q unknown, we can select them by AIC and BIC, where $D = p + q + 1$.
For models, we also check the residuals(innovation): if the model is appropriate, the innovation should be white noise, which means there is no information left in the innovations for linear prediction.

Definition 5.9 (Residuals)**.** The residuals are

$$\hat{a}_t = x_t - \{\hat{\mu} + \hat{\phi}_1(x_{t-1} - \hat{\mu} + \cdots + \hat{\phi}_p(x_{t-p} - \hat{\mu}) + \hat{\psi}_1 \hat{a}_{t-1} + \cdots + \hat{\psi}_q \hat{a}_{t-q}$$

And we can use Ljung-Box test.

Let $\{y_t\}$ be a stationary process, e.g. ARMA(p, q) process with Wold representation

$$y_t = \mu + a_t + \phi_1 a_{t-1} + \phi_2 a_{t-2} + \cdots$$

For **prediction without any information**, the prediction of y_{t+1} is

$$E(y_{t+1}) = \mu + E(a_{t+1}) + \phi_1 E(a_t) + \cdots = \mu$$

For AR(p) model,

$$E(y_{t+1}) = \phi_0 + \phi_1 E(y_t) + \cdots + \phi_p E(y_{t-p+1}) + E(a_{t+1}) = \mu$$

For **prediction with information up to time** t, let $I_t = \{y_t, y_{t-1}, \ldots\}$ be information up to time t. Our purpose is to predict $y_{t+h}, h \geq 1$ based on I.
A linear predictor is

$$\begin{aligned} & E(y_{T+h|t} \mid a_T, a_{T-1}, \ldots) \\ =& \mu + E(a_{T+h}) + \phi_1 E(a_{T+h-1}) + \cdots + \phi_{h-1} E(a_{T+1}) + \phi_h E(a_T \\ =& \mu + \phi_h a_T + \phi_{h+1} a_{T-1} + \cdots \end{aligned}$$

Definition 5.10 (Forecast Error)**.** Define $y_{t+h|t}$ as forecast of y_{t+h} based on I_t. The forecast error is

$$e_{t+h|t} = y_{t+h} - y_{t+h|t}$$

The mean squared error(MSE) of forecast is

$$MSE(a_{t+h|t}) = E[(y_{t+h} - y_{t+h|t})^2]$$

The forecast error of the best linear predictor is

$$e_{t+h|t} = y_{t+h} - y_{t+h|t} = a_{t+h} + \phi_1 a_{t+h-1} + \cdots \phi_{h-1} a_{t+1}$$

The MSE of the forecast error is

$$MSE(a_{t+h|t}) = \sigma^2(1 + \phi_1^2 + \cdots + \phi_{h-1}^2)$$

If $\{a_t\}$ gaussian then

$$y_{t+h|I_t} \sim N(y_{t+h}, \sigma^2(1 + \phi_1^2 + \cdots + \phi_{h-1}^2))$$

The 95% confidence interval for h-step ahead prediction has the form

$$y_{t+h|t} \pm 1,96\sqrt{\sigma^2(1 + \phi_1^2 + \cdots + \phi_{h-1}^2)}$$

And standard error of prediction

$$s.e. = \sqrt{\sigma^2(1 + \phi_1^2 + \cdots + \phi_{h-1}^2)}$$

6 Conditional Heteroscedastic Model

For time series $\{y_t\}$, denote information up to time t by $I_t = \sigma\{y_t, \cdots\}$.
The conditional mean $\mu_t = E(y_t \mid I_{t-1})$ can be used to predict y_t based on past information I_{t-1}.
Conditional variance, or **volatility** is $\sigma_t^2 = \mathrm{Cov}(yz_t \mid I_{t-1}) = E[(y_t - \mu_t)^2 \mid I_{t-1}]$.
We can calculate volatility from high-frequency data, from implied volatility, or from weighted average.

Theorem 6.1 (Weighted Average)**.** Suppose the return u_i has mean 0, or we consider $u_i - E(u_i)$.
We estimate the variance by

$$\sigma_t^2 = \sum_{i=1}^{m} w_i u_{t-i}^2 \text{ for prediction}$$

or

$$\sigma_t^2 = \sum_{i=0}^{m-1} w_i u_{t-i}^2 \text{ for fitting}$$

where m can be infinity providing $\sum_{w_i} = 1$.
A special case is

$$w_k = (1-\lambda)\lambda^{k-1}, k = 1, \dots$$

and

$$\sigma_t^2 = (1-\lambda)\sum_{i=1}^{\infty} \lambda^{i-1} u_{t-i}^2$$

This leads to

$$\sigma_t^2 = (1-\lambda)u_{t-1}^2 + \lambda\sigma_{t-1}^2$$

Theorem 6.2 (GARCH Model)**.** GARCH model has the following recurrence relation

$$\sigma_t^2 = \omega + \alpha u_{t-1}^2 + \beta\sigma_{t-1}^2$$

with $\omega \geq 0, \alpha \geq 0, \beta \geq 0$, but $\alpha + \beta \leq 1$.

Theorem 6.3 (Basic Structure of Conditional Mean and Conditional Variance)**.** For return r_t, denote conditional mean by u_t

$$r_t = \mu_t + u_t$$

and usually we model μ_t by ARMA(p, q),

$$\mu_t = \phi_0 + \sum_{i=1}^{p} \phi_i r_{t-i} + \sum_{i=1}^{q} \psi_i u_{t-i}$$

Volatility models are concerned with modelling of

$$\sigma_t^2 = \mathrm{Cov}(r_t, \mid r_{t-1}, \dots) = \mathrm{Cov}(u_t \mid r_{t-1}, r_{t-2}, \dots)$$

the conditional variance of the return.
We can use univariate volatility models such that

- ARCH
- GARCH
- GARCH-M
- $\cdots$

Theorem 6.4 (ARCH Model)**.**

$$\begin{aligned} r_t &= \mu_t + u_t \\ u_t &= \sigma_t \epsilon_t \\ \sigma_t^2 &= \omega + \alpha_1 u_{t-1}^2 + \cdots + \alpha_m u_{t-m}^2 \end{aligned}$$

where $\{\epsilon_t\}$ is a sequence of iid random variables, with

$$E(\epsilon_t) = 0, \mathrm{Cov}(\epsilon_t) = 1$$

and $\omega > 0$, and $\alpha_i \geq 0$ for $i > 0$.

$$\epsilon_t \perp \{\sigma_s, s \leq t\}$$

We call $u_t = r_t - \mu_t$ **residuals** and ϵ_t **standardized residuals**.
Commonly used distribution for ϵ_t can be standardized normal, or standardized student-t.
Let's take ARCH(1) as example:

$$r_t = u_t + \mu_t, \;\; u_t = \sigma_t\epsilon_t, \;\; \sigma_t^2 = \omega + \alpha_1 u_{t-1}^2$$

where $\omega > 0$and $\alpha_1 \geq 0$.
We have

- $E(u_t) = 0$
- u_t is called residuals, ϵ_t standardized residuals.
- Model can be written as

  $$u_t^2 = \omega + \alpha_1 u_{t-1}^2 + \eta_t$$

  where $\eta_t = \sigma_t^2(\epsilon_t^2 - 1)$.

- If $\alpha_1 < 1$, $E(u_t^2)$ is constant, and u_t also stationary.
- Given $\mathrm{Cov}(u_t)$ is constant for all t, then

$$\mathrm{Cov}(u_t) = \frac{\omega}{(1-\alpha_1)} \text{ if } 0 < \alpha_1 < 1$$

- Under normality of ϵ_t, and given the moments are constants for all t, we have

$$\mu_4 = \frac{3\omega^2(1+\alpha_1)}{(1-\alpha_1)(1-3\alpha_1^2)}$$

provided $0 < \alpha_1^2 < 1/3$. However, the 4th moment does not exist if $\alpha_1^2 > 1/3$.

- THe kurtosis is

$$\frac{\mu_4}{\mu_2^2} = 3\frac{(1-\alpha_1^2)}{(1-3\alpha_1^2)} > 3$$

which exmplains the heavy-tailedness of data.

We can use MLE to estimate $\omega, \alpha_i, i = 1, \dots, m$, and by AIC, BIC, we can choose a suitable m.
The prediction of σ_{t+h}^2 can be done using the recursion relationship.

A general $\mathrm{ARMA}(p,q) + \mathrm{GARCH}(m,s)$ model is

$$r_t = \mu_t + u_t, u_t = \sigma_t\epsilon_t$$

$$\sigma^2 = \omega + \sum_{i=1}^{m}\alpha_i u_{t-i}^2 + \sum_{j=1}^{s}\beta_j\sigma_{t-j}^2$$

where μ_t is $\mathrm{ARMA}(p,q)$ modek

$$\mu_t = \phi_0 + \sum_{i=1}^{p}\phi_i r_{t-i} + \sum_{i=1}^{q}\psi_i u_{t-i}$$

where ϵ_t IID with $E(\epsilon_t) = 0$ and $E(\epsilon_t^2) = 1$, $\omega > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$, and

$$\sum_{i=1}^{\max(m;s)}(\alpha_i + \beta_i) < 1$$

Let $\eta_t = u_t^2 - \sigma_t^2$, then $\{\eta_t\}$ is uncorrelated series. GARCH model becomes

$$u_t^2 = \omega + \sum_{i=1}^{\max(m;s)}(\alpha_i + \beta_i)u_{t-i}^2 + \eta_t - \sum_{j=1}^{s}\beta_j\eta_{t-j}$$

The probabilities properties for $\mathrm{GARCH}(1,1)$ include

$$\sigma_t^2 = \omega + \alpha_1 u_{t-1}^2 + \beta_1\sigma_{t-1}^2$$

- Weakly stationary: $0 \geq \alpha_1$, $\beta_1 \leq 1$, $\alpha_1 + \beta_1 < 1$
- Volatility clustering
- Unconditional variance

$$E(u_t^2) = \frac{\omega}{1-\alpha_1-\beta_1}$$

- Heavy tails: if $\epsilon_t \sim N(0,1)$, and $1 - 2\alpha_1^2 - (\alpha_1+\beta_1)^2 > 0$, then

$$\frac{E(u_t^4)}{[E(u_t^2)]^2} = \frac{3[1-(\alpha_1+\beta_1)^2]}{1-(\alpha_1+\beta_1)^2-2\alpha_1^2} > 3$$

Garch In-Mean(GARCH-M) Model is

$$r_t = \mu_t + c\sigma_t + u_t$$

$$u_t = \sigma_t\epsilon_t$$

$$\sigma_t^2 = \omega + \alpha_1 u_{t-1}^2 + \beta_1\sigma_{t-1}^2$$

where c is referred to as risk premium, which is expected to be positive. In some financial time series, large negative returns appear to increase volativility more than positive returns of the same magnitude do.
$APARCH(p,q)$ model for conditional standard deviation is

$$\sigma_t^\delta = \omega + \sum_{i=1}^{p}\alpha_i(|u_{t-1}| - \gamma_i u_{t-1})^\delta + \sum_{j=1}^{q}\beta_j\sigma_{t-j}^\delta$$

where $\delta > 0$ and $-1 < \gamma_j < 1$. Note that $\delta = 2$ and $\gamma_i = 0$ gives standard GARCH model.
Given past information $r_1, \dots, r_{t-1}$, for a small $\alpha > 0$, we define the value at risk as

$$\mathrm{VaR}_t(\alpha) = -\max\{v : P(r_t \leq v \mid r_{t-1}, \dots) \leq \alpha\}$$

If we assume $r_t = \mu_t + \sigma_t\epsilon_t$, and $\epsilon_t \perp \sigma_t$, then

$$\mathrm{VaR}_t(\alpha) = -\mu_t - \sigma_t z_\alpha = -\mu_t + \sigma_t \mathrm{VaR}_\alpha(\epsilon)$$

where z_α is the α-th quantile of ϵ_t.
If $\epsilon_t \sim N(0,1)$, then

$$z_\alpha = \Phi^{-1}(\alpha)$$

If t distribution with freedom ν, then $z_\alpha t_\nu^{-1}(\alpha)/\sqrt{\frac{\nu}{\nu-2}}$.
The VaR_t with level alpha means $P(u_t > -\mathrm{VaR}_t(\alpha)) = 100(1-\alpha)\%$.
To validate the calculation of $\mathrm{VaR}_t(\alpha)$, the estimated probability

$$\hat{p} = \frac{\#\{u_t > -\mathrm{VaR}_t(\alpha), t = 1, \dots, T\}}{T}$$

should be approximately $100(1-\alpha)\%$.
If $\hat{p} > 100(1-\alpha)\%$, then the model is too conservative, and otherwise, model is too aggressive.
We first predict σ_{T+h} then, together with the distribution, calculate VaR at day $T+h$ at level α:

$$-\mu_{T+h|T} - \sigma_{T+h|T}Q(\alpha)$$