金融时间序列分析

- 2021秋季
- 孙便霞

> 如果想快速浏览重点，可以只看高亮部分的相关内容；如果想理解性记忆，建议按顺序看完。

Basic

- **Notation**
  - For a time series $\{X_t\}$, Denote $\mu_t = E(X_t)$ and $\gamma_{t,\,k} = Cov(X_t,\,X_{t+k})$. For simplicity, we assume $k \geqslant 0$ (the negative case is the same).
  - For a stationary series, γ_k is called auto-covariance function, and $\rho_k = \frac{\gamma_k}{\gamma_0}$ is called auto-correlation function (ACF).
  - Back-shift operator: B
    - $BX_t = X_{t-1},\; B^2 X_t = X_{t-2},\; \cdots,\; B^k X_t = X_{t-k}$

- **Stationarity**
  - Strictly stationary: $\forall t$ and $\forall k$, the joint distribution of $(X_{t_1},\,X_{t_2},\,\cdots,\,X_{t_k})$ is the same as that of $(X_{t+t_1},\,X_{t+t_2},\,\cdots,\,X_{t+t_k})$.
  - ==Weakly stationary== (covariance stationary): $\forall t$ and $\forall k$, $\mu_t \equiv \mu$ and $\gamma_{t,\,k} \equiv \gamma_k$, which means the mean is constant and the auto-covariance only depends on lag k.
    - When we talk about stationary in this class, we always refer to the weakly one.
  - Obviously, weakly stationary does not imply strictly stationary. But does strictly stationary imply weakly stationary? The answer is no. Suppose the joint distribution of $(X_{t_1},\,X_{t_2},\,\cdots,\,X_{t_k})$ is Cauchy distribution, whose moments do not exist, then the weakly stationary condition cannot be satisfied.
  - White noise: $\{a_t\} \overset{\text{i.i.d.}}{\sim} WN(0,\,\sigma^2)$
    - WN need not be normal distribution, but sometimes people assume a normal white noise.
    - White noise is stationary, even strictly stationary.

- **Estimation**
  - Sample auto-covariacne: $\hat{\gamma}_k = \dfrac{\sum\limits_{t=1}^{T-k}(X_t - \overline{X})(X_{t+k} - \overline{X})}{T-k-1}$
    - When $k = 0$, this is the sample variance that we are familiar with.
  - Sample ACF: $\frac{\hat{\gamma}_k}{\hat{\gamma}_0}$

- **Test for White Noise**
  - Ljung-Box test

$$Q_m = T(T+2)\sum_{k=1}^{m}\frac{\hat{\rho}_k^2}{T-k} \to \chi^2(m)$$

  - m is the max-lag that need to be chosen. Usually we choose $m = 10$.
  - H_0: The series tested is a white noise.

MA (Moving Average)

- **Definition**
  - Assume there is a white noise $\{a_t\} \sim WN(0, \sigma^2)$, then its moving average series

$$X_t = \mu + \theta_0 a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \cdots + \theta_q a_{t-q}$$

  is called an $MA(q)$ process.

  - Generally, θ_0 is set to be 1 and the '$+$' will be replaced by '$-$' for lag terms. So, in this class, an $MA(q)$ process will be written as

$$X_t = \mu + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q}$$

  - Back-shift form:

$$\begin{aligned}
X_t &= \mu + a_t - \theta_1 B a_t - \theta_2 B^2 a_t - \cdots - \theta_q B^q a_t \\
&= \mu + (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q) a_t \\
&= \mu + \Theta(B) a_t
\end{aligned}$$

- **Properties**
  - ==Always stationary== since

$$\begin{aligned}
E(X_t) &= \mu \\
\gamma_{t, k} &= Cov(X_t, X_{t-k}) \\
&= \begin{cases} (\theta_k + \theta_{k+1}\theta_1 + \cdots + \theta_q \theta_{q-k})\sigma^2, & k \leqslant q \\ 0, & k > q \end{cases} \\
&\equiv \gamma_k
\end{aligned}$$

  - Variance

$$Var(X_t) = \gamma_0 = (1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2)\sigma^2$$

  - ACF

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \begin{cases} \frac{\theta_k + \theta_{k+1}\theta_1 + \cdots + \theta_q \theta_{q-k}}{1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2}, & k \leqslant q \\ 0, & k > q \end{cases}$$

  - ==Finite memory==: Since ACF becomes 0 when $k > q$, an $MA(q)$ process only has q-period memory.
  - Forecast

- Use mean(s) to estimate unknown variable(s), e.g., we stand on t, i.e., we know all information before, including the current one. If we want to get the 1-step ahead forecast, we first write

$$X_{t+1} = \mu + a_{t+1} - \theta_1 a_t - \theta_2 a_{t-1} - \cdots - \theta_q a_{t-q+1}$$

Since we know all the information except a_{t+1}, we replace it by its mean 0 and get

$$\hat{X}_t(1) = \mu - \theta_1 a_t - \theta_2 a_{t-1} - \cdots - \theta_q a_{t-q+1}$$

- The error of forecast is just the difference between the true value and its estimation. Also consider the case above, the error is

$$e_t(1) = a_{t+1} - 0 = a_{t+1}$$

- The variance of forecast error is given by

$$Var(e_t(1)) = Var(a_{t+1}) = \sigma^2$$

- If we want to get the $(q+1)$-step ahead forecast, following the procedure above we will find that the forecast is μ (finite memory).

We have already know how to get forecast. But when you actually do it, you would run into a problem: how can we know the values of $a_t,\ a_{t-1},\ \cdots$ through the observations $X_t,\ X_{t-1},\ \cdots$?

- **Invertibility**
  - Definition
    - $\{X_t\}$ is called invertible if it can be written as

$$X_t = \pi_1 X_{t-1} + \pi_2 X_{t-2} + \cdots + a_t$$

with white noise $\{a_t\}$ and $\sum\limits_{i=1}^{\infty} |\pi_i| < \infty$.
    - Back-shift form:

$$X_t = \pi_1 B X_t + \pi_2 B^2 X_t + \cdots + a_t$$
$$(1 - \pi_1 B - \pi_2 B^2 - \cdots)X_t = a_t$$
$$\pi(B)X_t = a_t$$

  - If a time series is <mark>not invertible</mark>, then we cannot get $a_t,\ a_{t-1},\ \cdots$ from observations, which means it is <mark>impossible to make forecast</mark>.
  - Consider a zero-mean $MA(q)$ process $X_t = \Theta(B)a_t$. Multiply both sides by $\pi(B)$ and we have

$$\pi(B)X_t = \pi(B)\Theta(B)a_t$$

By the definition of invertibility, we need to let $\pi(B)\Theta(B) = 1$, i.e.,

$$(1 - \pi_1 B - \pi_2 B^2 - \cdots)(1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q) = 1$$

Since the constant term of the left side is obviously 1, we just need the coefficients of terms with B and its higher orders to be 0, i.e.,

$$-\theta_1 - \pi_1 = 0$$
$$-\theta_2 + \pi_1 \theta_1 - \pi_2 = 0$$
$$\cdots$$

Solving the equations we can get π_1, π_2, $\cdots$. By the restriction $\sum_{i=1}^{\infty} |\pi_i| < \infty$, we derive the ==condition of invertibility==: all of the roots of the corresponding polynomial of $\Theta(B)$ are outside the unit circle, i.e., all of the roots of

$$1 - \theta_1 x - \theta_2 x^2 - \cdots - \theta_q x^q = 0$$

have modulus bigger than 1.

- Examples
  - Consider an $MA(1)$ process, $\Theta(B) = 1 - \theta_1 B$, which means the corresponding polynomial is $1 - \theta_1 x$. To ensure the invertibility, the modulus of the root, $\left|\frac{1}{\theta_1}\right|$, should be bigger than 1, i.e., ==$|\theta_1| < 1$==.
  - Consider an $MA(2)$ process, $\Theta(B) = 1 - \theta_1 B - \theta_2 B^2$, which means the corresponding polynomial is $1 - \theta_1 x - \theta_2 x^2$. To ensure the invertibility, the modulus of the roots, $\left|\frac{\theta_1 \pm \sqrt{\theta_1^2 + 4\theta_2}}{-2\theta_2}\right|$, should be bigger than 1. However, the expression is too complicated to solve the conditions for θ_1 and θ_2. We can use Vieta Theorem to get the sum and the product of roots. Suppose we have 2 roots x_1 and x_2, then

$$x_1 + x_2 = -\frac{-\theta_1}{-\theta_2}$$
$$x_1 x_2 = \frac{1}{-\theta_2}$$

which means
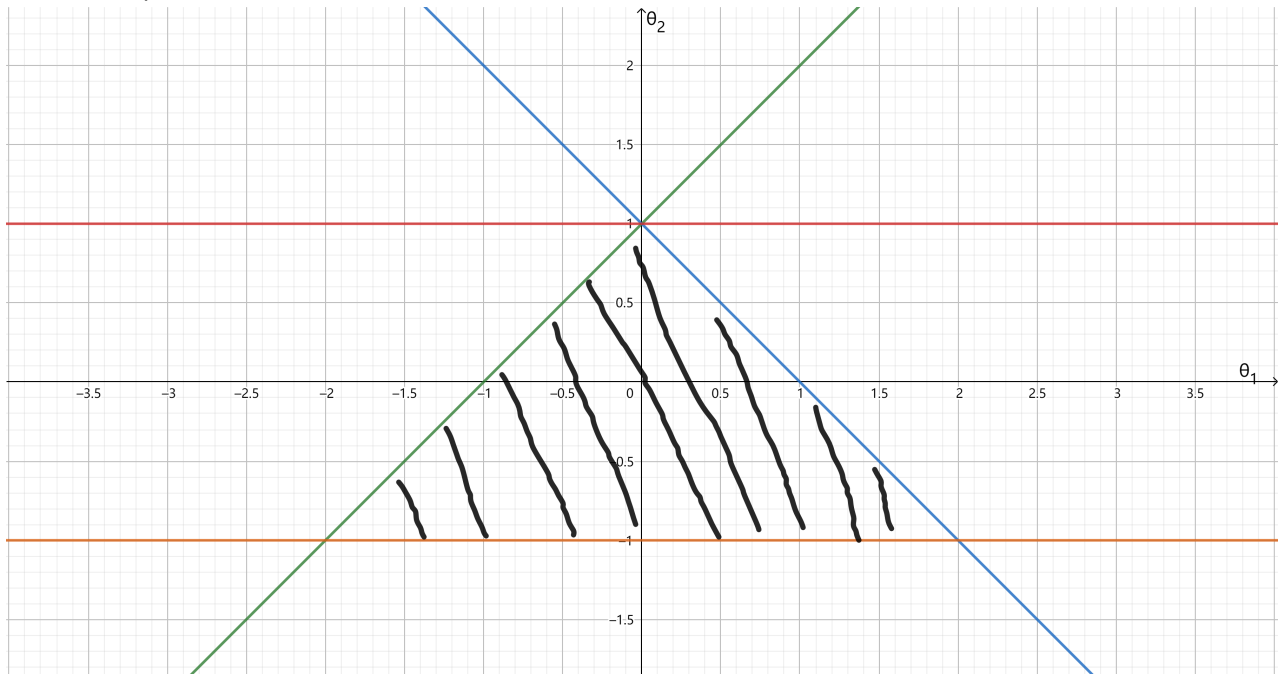
$$\theta_2 = -\frac{1}{x_1 x_2}$$
$$\theta_1 = \frac{x_1 + x_2}{x_1 x_2}$$

where $|x_1| > 1$ and $|x_2| > 1$.

Since $|x_1 x_2| > 1$, we need $|\theta_2| < 1$; Also, We can easily see that $(1 - \frac{1}{x_1})(1 - \frac{1}{x_2}) > 0$,

i.e., $1 - \frac{1}{x_1} - \frac{1}{x_2} + \frac{1}{x_1 x_2} > 0$, which means $\theta_1 + \theta_2 = \frac{x_1 + x_2 - 1}{x_1 x_2} < 1$; Similarly, $(1 + \frac{1}{x_1})(1 + \frac{1}{x_2}) > 0$ yields $\theta_2 - \theta_1 < 1$.

Check the sufficiency of these conditions and we have the sufficient and necessary conditions of invertibiity for $MA(2)$: $|\theta_2| < 1$ and $\theta_2 \pm \theta_1 < 1$. The conditions are shown below (the dash area).

We can see from the graph that $|\theta_1|$ does not need to be less than 1.

- Similarly we can derive necessary conditions of invertibility for an $MA(q)$ process: $|\theta_p| < 1$ and $\theta_1 + \theta_2 + \cdots + \theta_p < 1$.

- **Model Identification**
  - Approximately determine the lag order by looking at the sample ACF.
    - If the sample ACF cuts off at lag q, i.e., $ACF(n) = 0$ (lie between the confidence interval) $\forall n > q$ and $ACF(q) \neq 0$ (lie outside the confidence interval), then the time series can be identified as an $MA(q)$ process.
  - MLE (Maxilimum Likelihood Estimation) to estimate the parameters, i.e., θ_1, θ_2, $\cdots$.
  - IC (Information Criterion), a more precise way to determine the lag order:
    1. AIC;
    2. BIC (SC).
    - Based on log-likelihood and penalty of long lag (prevent over-fitting).
    - Bigger is better.
  - Model Checking
    - Check whether the residuals are white noise (do a Ljung-Box test for the residuals).

AR (Auto-Regression)

- **Definition**
  - Consider a time series $\{X_t\}$ and white noise $\{a_t\} \sim WN(0, \sigma^2)$, an $AR(p)$ model is written as

  $$X_t = \phi_0 + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + a_t$$

  - Back-shift form:

  $$X_t = \phi_0 + \phi_1 B X_t + \phi_2 B^2 X_t + \cdots + \phi_p B^p X_t + a_t$$
  $$(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)X_t = \phi_0 + a_t$$
  $$\Phi(B)X_t = \phi_0 + a_t$$

  > After introducing the AR model, we can see that the condition of invertibility can be expressed as: a time series can be written as an $AR(\infty)$ process with coefficient series absolutely convergence.
  >
  > Similarly, the condition of stationarity can be expressed as: a time series can be written as an $MA(\infty)$ process with coefficient series absolutely convergence.

- **Properties**
  - Obviously, an $AR(p)$ process is always invertible.
  - Similar to the condition of invertibility of MA processes, the condition of stationarity of an $AR(p)$ process is: all of the roots of the corresponding polynomial of $\Phi(B)$ are outside the unit circle, i.e., all of the roots of

  $$1 - \phi_1 x - \phi_2 x^2 - \cdots - \phi_p x^p = 0$$

  have modulus bigger than 1.
    - $AR(1)$: $|\phi_1| < 1$
    - $AR(2)$: $|\phi_2| < 1$ and $\phi_2 \pm \phi_1 < 1$
    - $AR(p)$ (necessary but not sufficient): $|\phi_p| < 1$ and $\phi_1 + \phi_2 + \cdots + \phi_p < 1$
    > Why do we always care about the stationarity? If the series is stationary, then the mean is time invariant, which means a stationary series is a mean-reverting process (can get accurate forecast).

  - Mean of a stationary $AR(p)$ process:

  $$E(X_t) = \phi_0 + \phi_1 E(X_t) + \phi_2 E(X_t) + \cdots + \phi_p E(X_t)$$
  $$E(X_t) = \frac{\phi_0}{1 - \phi_1 - \phi_2 - \cdots - \phi_p}$$

- <mark>Variance and ACF for a stationary $AR(p)$ process</mark>:
  - Consider $p = 2$, after taking variance on both sides, we have

$$Var(X_t) = \phi_1^2 Var(X_{t-1}) + \phi_2^2 Var(X_{t-2}) + \phi_1\phi_2 Cov(X_{t-1}, X_{t-2}) + \sigma^2$$
$$\gamma_0 = \phi_1^2\gamma_0 + \phi_2^2\gamma_0 + \phi_1\phi_2\rho_1\gamma_0 + \sigma_2$$
$$\gamma_0 = \frac{\sigma^2}{1 - \phi_1^2 - \phi_2^2 - \phi_1\phi_2\rho_1}$$

  Since ρ_1 is unknown now, we turn to calculate ACF.
  - To calculate ACF, we first demean, i.e., let $Y_t = X_t - E(X_t)$, then we have a zero-mean $AR(2)$ process

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + a_t$$

  - Then, we multiply both sides by Y_{t-1} and take expectation on both sides:

$$E(Y_t Y_{t-1}) = \phi_1 E(Y_{t-1}^2) + \phi_2 E(Y_{t-2} Y_{t-1})$$

  Since the mean is zero, by the definition of covariance, the equation actually is $\gamma_1 = \phi_1\gamma_0 + \phi_2\gamma_1$
  - Finally, divide both sides by γ_0 and we get

$$\rho_1 = \phi_1 + \phi_2\rho_1$$
$$\rho_1 = \frac{\phi_1}{1 - \phi_2}$$

  which means now γ_0 can be known.
  - For $ACF(k)$ with $k \geqslant p$, we can use the same method, e.g., for $ACF(2)$ we have $\rho_2 = \phi_1\rho_1 + \phi_2$.
  - Generally, the ACF of an stationary $AR(p)$ process <mark>decays exponentially</mark>.

> Recall that we can determine the lag order of MA processes by looking at the sample ACF plot. However, the sample ACF of stationary AR processes <mark>tails off</mark>, which means we need an alternative way to help us determine the lag order.

- **PACF (Partial Auto-Correlation Function)**
  - Definition (omitted)
  - Calculation
    - To get $PACF(k)$, we write an $AR(k)$ model:

$$X_t = \phi_{0,k} + \phi_{1,k}X_{t-1} + \phi_{2,k}X_{t-2} + \cdots + \phi_{k,k}X_{t-k} + a_t$$

and $PACF(k)$ is just $\phi_{k,k}$.

- For an $AR(p)$ process, the corresponding equation of $PACF(p)$ is just the process itself, i.e., $\phi_{p,\,p} = \phi_p$, which means $PACF(p) = \phi_p$.
- Since the true model is $AR(p)$, $\forall k > p$, $\phi_{k,\,k}$ should be 0, which means $PACF(k) = 0$.
- Similar to the way of calculating ACF, we should convert the $AR(k)$ model to a system of equations like this:

$$
\begin{cases}
\rho_1 &= \phi_{1,\,k} + \phi_{2,\,k}\rho_1 + \cdots + \phi_{k,\,k}\rho_{k-1} \\
\rho_2 &= \phi_{1,\,k}\rho_2 + \phi_{2,\,k} + \cdots + \phi_{k,\,k}\rho_{k-2} \\
&\cdots \\
\rho_k &= \phi_{1,\,k}\rho_{k-1} + \phi_{2,\,k}\rho_{k-2} + \cdots + \phi_{k,\,k}
\end{cases}
$$

or in matrix form:

$$
\begin{bmatrix}
\rho_1 \\ \rho_2 \\ \vdots \\ \rho_k
\end{bmatrix}
=
\begin{bmatrix}
1 & \rho_1 & \cdots & \rho_{k-1} \\
\rho_2 & 1 & \cdots & \rho_{k-2} \\
\vdots & \vdots & \ddots & \vdots \\
\rho_{k-1} & \rho_{k-2} & \cdots & 1
\end{bmatrix}
\begin{bmatrix}
\phi_{1,\,k} \\ \phi_{2,\,k} \\ \vdots \\ \phi_{k,\,k}
\end{bmatrix}
$$

which is called Yule-Walker equation.

Using the sample ACF, we can solve for $\phi_{k,\,k}$, i.e., $PACF(k)$.

- Note that when $k = 1$, the Yule-Walker equation becomes $\rho_1 = \phi_{1,\,1}$, which means for any $AR(p)$ process, its $PACF(1) = ACF(1)$.

- **IRF (Inpulse Response Function)**
    - Definition

$$
IRF(k) = \frac{\partial X_t}{\partial a_{t-k}}
$$

    - $IRF(k)$ describes the effect on X_t from lag k.
    - It is easier to get IRF under MA representation.

- **Model Identification**
    - From the properties of PACF we know that, the PACF of a stationary $AR(p)$ process cuts off at lag p, which means we can approximately choose the lag order by looking at the PACF plot.
    - Use sample ACF in Yule-Walker equation to estimate the coefficients.
    - Can also use IC as a more precise way to determine the lag order.

ARMA

- **Definition**
  - A combination of MA and AR:

$$\Phi(B)X_t = \phi_0 + \Theta(B)a_t$$

  - $ARMA(p,\ q)$ is a model with p-order $\Phi(B)$ and q-order $\Theta(B)$.
- **Properties**
  - Condition of <mark>stationarity</mark>: the same as $AR(p)$.
  - Condition of <mark>invertibility</mark>: the same as $MA(q)$.
  - An $ARMA(p,\ q)$ process can be expressed as an $MA(\infty)$ or $AR(\infty)$ process.
    - It is easier to get forecast under AR representation. If the process is not invertible, i.e., the coefficient series of $AR(\infty)$ does not converge, we can not make any forecast.
    - It is easier to calculate the variance of forecast error under MA representation. If the process is non-stationary, i.e., the coefficient series of $MA(\infty)$ does not converge, we get infinity when we calculate the error variance.
    - Under AR representation, we can easily find that <mark>ACF and PACF both tail off</mark>.

How to determine the lag orders?

- **Model Identification**
  - Use EACF (Extended ACF) to determine the lag orders.
  - IC can be used.
  - ARMA processes can be approximated by AR processes with higher lag order.
  - AR models are easier to estimate and forecast than MA and ARMA models.

Quiz 1

Consider an $AR(2)$ model $X_t = 0.2 + X_{t-1} - 0.6X_{t-2} + a_t$ where $\{a_t\}$ is white noise with variance 0.25.

1. Is this $AR(2)$ process stationary? Why?

Yes. Since $|-0.6| < 1$ and $-0.6 \pm 0.2 < 1$.

2. Please compute $E(X_t)$ and $Var(X_t)$.

Since the series is stationary, we have

$$E(X_t) = 0.2 + E(X_t) - 0.6E(X_t) \implies E(X_t) = \frac{1}{3} \approx 0.33$$

$$Var(X_t) = Var(X_t) + 0.6^2 Var(X_t) - 2 \times 0.6\rho_1 Var(X_t) + 0.25 \implies Var(X_t) = \frac{0.25}{1.2\rho_1 - 0.36}$$

Let $Y_t = X_t - E(X_t)$, then $Y_t = Y_{t-1} - 0.6Y_{t-2} + a_t$. Multiply both sides by Y_{t-1}, take expectaion and divide both sides by γ_0, we have

$$\rho_1 = 1 - 0.6\rho_1 \implies \rho_1 = 0.625$$

which means $Var(X_t) = \frac{0.25}{1.2\times0.625-0.36} \approx 0.64$.

3. Please compute the lag-1, -2 and -3 ACF and PACF of $\{X_t\}$.

We have already known $\rho_1 = 0.625$. Using the same method above, we have $\rho_2 = \rho_1 - 0.6 = 0.025$ and $\rho_3 = \rho_2 - 0.6\rho_1 = -0.35$.

$PACF(1) = \rho_1 = 0.625$, $PACF(2) = -0.6$ and $PACF(3) = 0$.

4. Please compute the lag-1, -2 and -3 IRF of $\{X_t\}$.

$$IRF(1) = \frac{\partial X_t}{\partial a_{t-1}} = \frac{\partial X_{t-1}}{\partial a_{t-1}} - 0.6\frac{\partial X_{t-2}}{\partial a_{t-1}} = \frac{\partial X_t}{\partial a_t} = 1$$

$$IRF(2) = \frac{\partial X_t}{\partial a_{t-2}} = \frac{\partial X_{t-1}}{\partial a_{t-2}} - 0.6\frac{\partial X_{t-2}}{\partial a_{t-2}} = IRF(1) - 0.6\frac{\partial X_t}{\partial a_t} = 1 - 0.6 = 0.4$$

$$IRF(3) = \frac{\partial X_t}{\partial a_{t-3}} = \frac{\partial X_{t-1}}{\partial a_{t-3}} - 0.6\frac{\partial X_{t-2}}{\partial a_{t-3}} = IRF(2) - 0.6IRF(1) = 0.4 - 0.6 = -0.2$$

5. Suppose $X_{100} = X_{99} = 0.2$, please write down the 1-, 2- and 3-step-ahead forecasts of the series at the forecast origin $t = 100$. What are the corresponding variances of the forecast errors?

$$\hat{X}_{100}(1) = 0.2 + X_{100} - 0.6X_{99} + 0 = 0.2 + 0.2 - 0.6 \times 0.2 = 0.28$$
$$e_{100}(1) = a_{101} \implies Var(e_{100}(1)) = 0.25$$

$$\hat{X}_{100}(2) = 0.2 + \hat{X}_{100}(1) - 0.6X_{100} + 0 = 0.2 + 0.28 - 0.6 \times 0.2 = 0.36$$
$$e_{100}(2) = e_{100}(1) + a_{102} = a_{101} + a_{102} \implies Var(e_{100}(2)) = 0.25 \times 2 = 0.5$$

$$\hat{X}_{100}(3) = 0.2 + \hat{X}_{100}(2) - 0.6\hat{X}_{100}(1) + 0 = 0.2 + 0.36 - 0.6 \times 0.28 = 0.392$$

$$e_{100}(3) = e_{100}(2) - 0.6e_{100}(1) + a_{103} = (a_{101} + a_{102}) - 0.6a_{101} + a_{103} = 0.4a_{101} + a_{102} + a_{103}$$

$$\implies Var(e_{100}(3)) = 0.25 \times (0.4^2 + 1 + 1) = 0.54$$

Non-stationary Processes

- **Examples**
  - ==Trend==
    - If a time series $\{X_t\}$ can be written as

      $$X_t = \mu(t) + Y_t$$

      where $\mu(t)$ is a deterministic function of t and $\{Y_t\}$ is a stationary process with zero mean, then we call $\{X_t\}$ has deterministic trend.
    - We can de-trend to make the series stationary, i.e., modeling $\{Y_t\}$ instead of $\{X_t\}$.
  - ==Random walk== (with drift)
    - $X_t = \mu + X_{t-1} + a_t$ with constant μ and white noise $\{a_t\}$.
    - The process above is actually an $AR(1)$ process with $\phi_1 = 1$, which reject the condition of stationarity.
    - To get a stationary process, we simply take difference, i.e., $\{X_t - X_{t-1}\}$.
  - ==Seasonal effect==
    - $X_t = \mu + X_{t-s} + a_t$ with constant μ and white noise $\{a_t\}$, s is the seasonal frequency, e.g., 4 for quarterly effect and 12 for monthly effect.
    - The process above is actually an $AR(s)$ process with $\phi_s = 1$, which reject the condition of stationarity.
    - To get a stationary process, we simply take difference, i.e., $\{X_t - X_{t-s}\}$.

In the last 2 examples, we can see that when testing their stationarity, we find a root $x = 1$, which is called ==unit root==.

- **Unit Root Test**
  - DF (Dickey-Fuller) test
    - Originally test for

      $$X_t = \phi X_{t-1} + a_t$$

    - H_0: Unit root exists.
    - Test with intercept

      $$X_t = \mu + \phi X_{t-1} + a_t$$

    - Test with intercept and linear trend

      $$X_t = \mu + \beta t + \phi X_{t-1} + a_t$$

  - ==ADF== (Augmented DF) test
    - Originally test for

$$X_t = \phi X_{t-1} + u_t$$

where $u_t = (1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_p B^p)^{-1} a_t$.

- H_0: Unit root exists. If H_0 holds, we have

$$X_t = \phi X_{t-1} + \varphi_1 \Delta X_{t-1} + \varphi_2 \Delta X_{t-2} + \cdots + \varphi_p \Delta X_{t-p} + a_t$$

- Similarly to DF test, we can also add intercept and trend terms.

- PP test
  - Consider the auto-correlation and heteroscedasticity of the error term.
  - H_0: Unit root exists.

> Empirically, people find that they need strong evidence to pass these tests, i.e., most of the financial time series data are hard to pass the tests, which means they all have unit root. But this may not be true. So, an alternative test will be introduced.

- KPSS test
  - H_0: Unit root does not exist.

- **ARIMA**
  - If a series becomes an $ARMA(p,\ q)$ process after several times of differencing, we call it an $ARIMA(p,\ d,\ q)$ process, where d is the times of differencing.
  - An $ARMA(p,\ q)$ process can be represented by $ARIMA(p,\ 0,\ q)$.
  - Seasonal ARIMA
    - For a time series with seasonal effect, we always take difference for the seasonal frequency. However, after differencing, the series may still be non-stationary. Thus, we still need to do the simple 1-lag difference for several times.
    - For example,

$$(1 - B)(1 - B^s)X_t = (1 - \theta B)(1 - \Theta B^s)a_t$$

is an $ARIMA(0,\ 1,\ 1) \times (0,\ 1,\ 1)_s$ model.

    - The first $(0,\ 1,\ 1)$ is the simple 1-lag difference part. Since we take the simple difference once, i.e., we multiply X_t by $(1 - B)$, d should be 1. Since the model has $MA(1)$ form in the 1-lag part after differencing, i.e., we multiply a_t by $(1 - \theta B)$, q should be 1.
    - The second $(0,\ 1,\ 1)$ with a subscript s is the seasonal difference part. Since we take seasonal difference once, i.e., we multiply X_t by $(1 - B^s)$, d should be 1. Since the model has $MA(1)$ form in the seasonal part after differencing, i.e., we multiply a_t by $(1 - \Theta B^s)$, q should be 1.
    - Take another example,

$$(1 - \phi B)(1 - B^s)^2 X_t = (1 - \theta_1 B - \theta_2 B^2)(1 - \Theta B^s)$$

is an $ARIMA(1,\ 0,\ 2) \times (0,\ 2,\ 1)_s$ model. Check yourself.

Multivariate Time Series

- **Stationarity**
  - Consider bivariate cases:

$$\boldsymbol{X_t} = \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix}$$

  where $\{x_{1t}\}$ and $\{x_{2t}\}$ are 2 time series.
  - If both

$$E(\boldsymbol{X_t}) = \begin{bmatrix} E(x_{1t}) \\ E(x_{2t}) \end{bmatrix} = \boldsymbol{\mu}$$

  and

$$Cov(\boldsymbol{X_t}, \ \boldsymbol{X_{t-k}}) = \begin{bmatrix} Cov(x_{1t}, \ x_{1,\, t-k}) & Cov(x_{1t}, \ x_{2,\, t-k}) \\ Cov(x_{2t}, \ x_{1,\, t-k}) & Cov(x_{2t}, \ x_{2,\, t-k}) \end{bmatrix} = \boldsymbol{\Gamma_k}$$

  are time invariant, then $\boldsymbol{X_t}$ is stationary.
  - Note that $\boldsymbol{\Gamma_k}$ may not be symmetric if $k \neq 0$ since $Cov(x_{1t}, \ x_{2,\, t-k})$ and $Cov(x_{2t}, \ x_{1,\, t-k})$ can be different.
  - If $\boldsymbol{X_t}$ is stationary, then x_{1t} and x_{2t} are both stationary. Conversely, ==if x_{1t} and x_{2t} are both stationary, it does not mean $\boldsymbol{X_t}$ is stationary.==

- **VAR (Vector AR)**
  - From scalar to vector:

$$\boldsymbol{X_t} = \boldsymbol{\Phi_0} + \boldsymbol{\Phi_1} \boldsymbol{X_{t-1}} + \boldsymbol{\Phi_2} \boldsymbol{X_{t-2}} + \cdots + \boldsymbol{\Phi_p} \boldsymbol{X_{t-p}} + \boldsymbol{a_t}$$

  where $\boldsymbol{a_t}$ is a sequence of i.i.d. multivariate normal random vectors with mean zero and covariance matrix $\boldsymbol{\Sigma}$.
  - Consider a bivariate $VAR(1)$:

$$\begin{bmatrix} x_{1t} \\ x_2t \end{bmatrix} = \begin{bmatrix} \phi_{10} \\ \phi_{20} \end{bmatrix} + \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} x_{1,\, t-1} \\ x_{2,\, t-1} \end{bmatrix} + \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}$$

  which can be rewritten as

$$x_{1t} = \phi_{10} + \phi_{11} x_{1,\, t-1} + \phi_{12} x_{2,\, t-1} + a_{1t}$$
$$x_{2t} = \phi_{20} + \phi_{21} x_{1,\, t-1} + \phi_{22} x_{2,\, t-1} + a_{2t}$$

  - ==Granger Causality==

- In the case above, if $\phi_{12} \neq 0$, i.e., x_{1t} does not depend on $x_{2,\,t-1}$, we say x_2 Granger cause x_1.
- Test the Granger causality is to test whether the cross-coefficient, e.g., ϕ_{12} or ϕ_{21}, is different from 0.
- H_0: x_1 (x_2) does not Granger cause x_2 (x_1).
- Since the test is about the cross-coefficient, Granger causality test depends on the lag order of the model, which means we should select order first.
- Granger causality is just a statistical concept, the true causality can not be tested.
  - Model Identification
    - Use IC or a stepwise χ^2 test to select order.
    - Use OLS (Ordinary Least-Squares) method to estimate parameters.
    - Remember to check the residuals after modeling.

> There are also VMA (Vector MA) and VARMA (Vector ARMA) similar to VAR. However, they may have the identifiability problem.

- **Identifiability problem for VMA and VARMA**
  - Case I:
    A $VMA(1)$ model

    $$\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} - \begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} a_{1,\,t-1} \\ a_{2,\,t-1} \end{bmatrix}$$

    is identical to the $VAR(1)$ model

    $$\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} - \begin{bmatrix} 0 & -2 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} x_{1,\,t-1} \\ x_{2,\,t-1} \end{bmatrix} = \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}$$

  - Case II:
    A $VARMA(1,\ 1)$ model

    $$\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} - \begin{bmatrix} 0.8 & -2+\eta \\ 0 & \omega \end{bmatrix}\begin{bmatrix} x_{1,\,t-1} \\ x_{2,\,t-1} \end{bmatrix} = \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} - \begin{bmatrix} -0.5 & \eta \\ 0 & \omega \end{bmatrix}\begin{bmatrix} a_{1,\,t-1} \\ a_{2,\,t-1} \end{bmatrix}$$

    is identical to the $VARMA(1,\ 1)$ model

    $$\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} - \begin{bmatrix} 0.8 & -2 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} x_{1,\,t-1} \\ x_{2,\,t-1} \end{bmatrix} = \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} - \begin{bmatrix} -0.5 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} a_{1,\,t-1} \\ a_{2,\,t-1} \end{bmatrix}$$

  - Case I is harmless since we can use either VMA or VAR model while case II indicates that a VARMA model may not be uniquely identified based on the sample data.

- **IRF**

- Similar to IRF in AR models, we can obtain IRF of a VAR model by converting the model to VMA representation, i.e., consider a stationary $VAR(p)$ model

$$(I - \Phi_1 B - \Phi_2 B^2 - \cdots - \Phi_p B^p)X_t = \Phi_0 + a_t$$

  which can be rewritten as

$$X_t = \mu + a_t + \Psi_1 a_{t-1} + \Psi_2 a_{t-2} + \cdots$$

- In a bivariate case, we have

$$\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \begin{bmatrix} \psi_{11}^{(0)} & \psi_{12}^{(0)} \\ \psi_{21}^{(0)} & \psi_{22}^{(0)} \end{bmatrix} \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} + \begin{bmatrix} \psi_{11}^{(1)} & \psi_{12}^{(1)} \\ \psi_{21}^{(1)} & \psi_{22}^{(1)} \end{bmatrix} \begin{bmatrix} a_{1, t-1} \\ a_{2, t-1} \end{bmatrix} + \cdots$$

  The response of x_2 to the unit impulse on x_1 is $\psi_{21}^{(0)}$, $\psi_{21}^{(1)}$, $\cdots$. The accumulated response is just the sum of them.

  IRF implicitly assumes that when one component of a_t changes, others should keep unchanged, which means the components should be uncorrelated with each other. However, the error terms in VAR are always correlated with each other.

- Now we try to convert a_t to a orthogonal vector whose components are uncorrelated with each other. Note that the covariance matrix of a_t, $\Sigma = E(a_t a_t^\top)$, is positive definite, which means there exists a lower triangular matrix L with with unit diagonal elements and a diagonal matrix G s.t. $\Sigma = LGL^\top$ (Cholesky decomposition). Therefore, we have

$$E(a_t a_t^\top) = LGL^\top$$
$$E((La_t)(La_t)^{-1}) = G$$

  which means La_t is a orthogonal vector. We denote it by b_t.
- After transforming, the VMA form of the VAR model can be rewritten as

$$X_t = \mu + LL^{-1}a_t + \Psi_1 LL^{-1}a_{t-1} + \Psi_2 LL^{-1}a_{t-2} + \cdots$$
$$= \mu + Lb_t + (\Psi_1 L)b_{t-1} + (\Psi_2 L)b_{t-2} + \cdots$$

  However, the method above depends on the ordering of the component of X_t, i.e., we may have different IRFs setting $X_t = \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix}$ and $X_t = \begin{bmatrix} x_{2t} \\ x_{1t} \end{bmatrix}$.

- Generalized impulse response analysis
  - Free of the ordering.

- **Cointegration**

- In a bivariate VAR model, suppose 2 components x_{1t} and x_{2t} are unit-root nonstationary, but a linear combination of them, say, $\alpha x_{1t} + \beta x_{2t}$ is stationary, then x_{1t} and x_{2t} are co-integrated.
- ==Test for cointegration==
  - Consider a $VAR(p)$ model with k dimension. The ==precondition== of cointegration is that each component of $\boldsymbol{X}_t$ is an $I(1)$ series, i.e., it will be stationary after differencing once.

  > Actually, the precondition allow components to be $I(2)$ or even higher order. But we only consider $I(1)$ in this class.

  - Let $\boldsymbol{Y}_t = \boldsymbol{X}_t - \boldsymbol{X}_{t-1}$, then the model can be rewritten as VECM (Vector Error-Correction Model) with $p-1$ order:

$$Y_t = \Pi X_{t-1} + \sum_{i=1}^{p-1} \Phi_i^* Y_{t-i} + a_t$$

  where

$$\Phi_{p-1}^* = -\Phi_p$$
$$\Phi_{p-2}^* = -\Phi_{p-1} - \Phi_p$$
$$\cdots$$
$$\Phi_1^* = -\Phi_2 - \cdots - \Phi_p$$
$$\Pi = \Phi_1 + \cdots + \Phi_p - I$$

- Denote $m = \text{rank}(\boldsymbol{\Pi})$. Since the precondition requires all k components are $I(1)$ series, we must have $m < k$. Also, if $m = 0$, there is no cointegration. Therefore, ==cointegration means $0 < m < k$==.
- There are ==$k - m$ unit roots== in $\boldsymbol{X}_t$ and ==m linear combinations of $\boldsymbol{X}_t$ that are stationary==.
- $\boldsymbol{\Pi}$ can be decomposition as $\boldsymbol{\alpha\beta}^\mathsf{T}$ where $\boldsymbol{\alpha}$ is a $k \times m$ matrix and $\boldsymbol{\beta}^\mathsf{T}$ is a $m \times k$ matrix called the co-integrating vector. $\boldsymbol{Z}_t = \boldsymbol{\beta}^\mathsf{T} \boldsymbol{X}_t$ is stationary.
- After testing cointegration, the VECM can be rewritten as

$$Y_t = \alpha\beta^\mathsf{T} X_{t-1} + \sum_{i=1}^{p-1} \Phi_i^* Y_{t-i} + a_t$$

  where $\boldsymbol{\alpha}$ is called ==adjustment coefficient==, $\boldsymbol{\beta}$ is called ==long-run coefficient== and $\boldsymbol{\Phi}_i^*$ is called ==short-run coefficient==.

- Example
  - Consider a $VAR(1)$ model with 2 $I(1)$ components

$$\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \begin{bmatrix} 0.4 & 1.5 \\ -0.2 & 1.5 \end{bmatrix} \begin{bmatrix} x_{1,\,t-1} \\ x_{2,\,t-1} \end{bmatrix} + \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}$$

- Since $\mathbf{\Pi} = \mathbf{\Phi}_1 - \mathbf{I} = \begin{bmatrix} -0.6 & 1.5 \\ -0.2 & 0.5 \end{bmatrix}$ has rank 1, there exist an unit root and a linear combination of components that is stationary.

- To find what the linear combination is, the simplest way is to set the coefficient of x_{1t} as 1, i.e., we decompose $\mathbf{\Pi}$ like this:

$$\begin{bmatrix} -0.6 & 1.5 \\ -0.2 & 0.5 \end{bmatrix} = \begin{bmatrix} ? \\ ? \end{bmatrix} \begin{bmatrix} 1 & ? \end{bmatrix} = \begin{bmatrix} 0.6 \\ 0.2 \end{bmatrix} \begin{bmatrix} 1 & -2.5 \end{bmatrix}$$

Therefore, $x_{1t} - 2.5x_{2t}$ is stationary.

- The VECM of this model can be written as

$$\begin{bmatrix} \Delta x_{1t} \\ \Delta x_{2t} \end{bmatrix} = \begin{bmatrix} 0.6 \\ 0.2 \end{bmatrix} \begin{bmatrix} 1 & -2.5 \end{bmatrix} \begin{bmatrix} x_{1,\,t-1} \\ x_{2,\,t-1} \end{bmatrix} + \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}$$

○ Application

- Pair Trading: Suppose the log-price series of 2 assets are co-integrated, we can trade the linear combination to make profit since it is mean-reverting.

Volatility Modeling

- **Calculation of Volatility**
  - <mark>True volatility cannot be observed directly</mark>. People can only use different measures to define it, e.g.,
    1. Daily range (high minus low);
    2. Realized volatility (using intraday data, daily data, or data with other frequency);
    3. Implied volatility (based on Black-Scholes formula);
    4. Econometric modeling (assuming volatility is the function of available information) which is easy to estimate and forcast.
  - If we calculate realized volatility σ through daily return, then the annualized volatility is $\sigma_{year} = \sqrt{N}\sigma$ where N is the trading days in one year, e.g., $N = 252$ for US and $N = 250$ for China. Similarly, if we calculate realized volatility σ through monthly return, then $\sigma_{year} = \sqrt{12}\sigma$.

- **Properties of Financial Volatility**
  - <mark>Volatility clustering</mark>: High volatility would cluster in a period of time, so would low volatility.
  - Volatility evolves continuously with rare jumps.
  - Volatility does not diverge to infinity, i.e., it varies within some fixed range.
  - <mark>Leverage effect (also called volatility asymmetry)</mark>: Negative unexpected return causes higher volatility.
  - <mark>Long memory</mark>.

- **ARCH (Auto-Regressive Conditional Heteroscedastic)**
  - After modeling the return series, e.g., using an $ARMA(p,\ q)$ model or some other models, we get a white noise residual series $\{a_t\}$, i.e., without linear correlation. However, the residual series may have non-linear correlation.
  - <mark>Definition</mark>:
    An $ARCH(m)$ model is

    $$a_t = \sigma_t \epsilon_t$$
    $$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \alpha_2 a_{t-2}^2 + \cdots + \alpha_m a_{t-m}^2$$

    where $\{\epsilon_t\} \overset{i.i.d.}{\sim} WN(0,\ 1)$, $\alpha_0 > 0$ and $\alpha_i \geqslant 0\ (i = 1,\ 2,\ \cdots,\ m)$.
  - <mark>Properties</mark>
    - At $t - 1$, σ_t is known since it is based on the past information.
    - The distribution of $\{\epsilon_t\}$ can be standard normal, standart t, or standard GED (Generalized Error Distribution).
    - Let $\eta_t = a_t^2 - \sigma_t^2$, we can easily find that $\{\eta_t\}$ is also a white noise. Then, we can rewrite the above model as

      $$a_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \alpha_2 a_{t-2}^2 + \cdots + \alpha_m a_{t-m}^2 + \eta_t$$

which is an $AR(m)$ model for $\{a_t^2\}$.

- Consider $m = 1$ and $0 < \alpha_1 < 1$ to ensure the stationarity of $\{a_t^2\}$,

$$E(a_t) = 0$$
$$Var(a_t) = E(a_t^2) = \frac{\alpha_0}{1 - \alpha_1}$$

Under normality, the 4-th moment of a_t is given by

$$
\begin{aligned}
E(a_t^4) &= E\left(E_{t-1}\left(a_t^4\right)\right) \\
&= E\left(3\left(E_{t-1}\left(a_t^2\right)^2\right)\right) \\
&= E\left(3\left(E_{t-1}\left(\sigma_t^2\right)\right)^2\right) \\
&= E\left[3\left(\alpha_0 + \alpha_1 a_{t-1}^2\right)^2\right] \\
&= 3\left(\alpha_0^2 + 2\alpha_0\alpha_1 E\left(a_{t-1}^2\right) + \alpha_1^2 E\left(a_{t-1}^4\right)\right) \\
&= \frac{3\alpha_0^2(1 + \alpha_1)}{(1 - \alpha_1)(1 - 3\alpha_1^2)}
\end{aligned}
$$

where $\alpha_1^2 < \frac{1}{3}$ to ensure the stationarity of $\{a_t^4\}$.

The kurtosis of a_t is given by

$$\frac{E(a_t^4)}{(Var(a_t))^2} = \frac{\frac{3\alpha_0^2(1+\alpha_1)}{(1-\alpha_1)(1-3\alpha_1^2)}}{\left(\frac{\alpha_0}{1-\alpha_1}\right)^2} = 3\frac{1 - \alpha_1^2}{1 - 3\alpha_1^2} > 3$$

which implies heavy tails.

Remember to distinguish $E(\cdot)$ and $E_{t-1}(\cdot)$. The former one is unconditional expectation which describes the long-run mean, while the latter one is conditional expectaion which describes the mean of next time state based on history.

- Forecast
    - Still consider the $ARCH(1)$ case. The 1-step ahead forecast at origin $t = h$ is $\sigma_h(1) = \alpha_0 + \alpha_1 a_h^2$.
    - For multi-step ahead forecasts, we rewrite the model as

$$\sigma_{t+1}^2 = \alpha_0 + \alpha_1 \sigma_t^2 \epsilon_t^2$$

which means we can get forecasts step-by-step:

$$\sigma_h^2(l) = \alpha_0 + \alpha_1 \sigma_h^2(l - 1) \quad l > 1$$

- **GARCH (Generalized ARCH)**
  - Definition:

    A $GARCH(m, s)$ model is

    $$a_t = \sigma_t \epsilon_t$$

    $$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{m} \alpha_i a_{t-1}^2 + \sum_{j=1}^{s} \beta_j \sigma_{t-j}^2$$

    where $\{\epsilon_t\} \overset{\text{i.i.d.}}{\sim} WN(0, 1)$, $\alpha_0 > 0$ and $\alpha_i \geqslant 0$ and $\beta_j \geqslant 0$. m is called ARCH order and s is called GARCH order.

  - **Properties**
    - In properties of ARCH we have already know that $\eta_t = a_t^2 - \sigma_t^2$ is a white noise, then the $GARCH(m, s)$ model can be rewritten as

      $$a_t^2 = \alpha_0 + \sum_{i=1}^{\max(m, s)} (\alpha_i + \beta_i) a_{t-1}^2 + \eta_t - \sum_{j=1}^{s} \beta_j \eta_{t-j}$$

      which is an $ARMA(\max(m, s), s)$ model for $\{a_t^2\}$. Note that if $m > s$, then $\beta_i = 0$ $(i = s+1, s+2, \cdots, m)$; if $m < s$, then $\alpha_i = 0$ $(i = m+1, m+2, \cdots, s)$.

      Recall that $\sum_{i=1}^{\max(m, s)} (\alpha_i + \beta_i) < 1$ is a necessary condition of stationarity of this model. With the nonnegative restrictions, this condition becomes sufficient.

  - **Forecast**
    - Consider a $GARCH(1, 1)$ model. The 1-step ahead forecast at origin $t = h$ is $\sigma_h(1) = \alpha_0 + \alpha_1 a_h^2 + \beta_1 \sigma_h^2$.
    - For multi-step ahead forecasts, we rewrite the model as

      $$\sigma_{t+1}^2 = \alpha_0 + (\alpha_1 + \beta_1)\sigma_t^2 + \alpha_1 \sigma_t^2 (\epsilon_t^2 - 1)$$

      which means we can get forecasts step-by-step:

      $$\sigma_h^2(l) = \alpha_0 + (\alpha_1 + \beta_1)\sigma_h^2(l - 1) \quad l > 1$$

  - **GARCH family**
    - IGARCH (Integrated GARCH)
      - If there is an unit root in $\{a_t^2\}$, then we call the model IGARCH.
      - Consider $IGARCH(1, 1)$, the existence of unit root means the model would be like

        $$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + (1 - \alpha_1)\sigma_{t-1}^2$$

      - Forecast
        - The 1-step ahead forecast is $\sigma_h^2(1) = \alpha_0 + \alpha_1 a_h^2 + (1 - \alpha_1)\sigma_h^2$

- The multi-step ahead forecasts are $\sigma_h^2(l) = \sigma_h^2(1) + (l-1)\alpha_0$, which means the effect of $\sigma_h^2(1)$ is persistent.
  - The special case with $\alpha_0 = 0$ is used in RiskMetrics to calculate VaR:

$$\sigma_t^2 = 0.06a_{t-1}^2 + 0.94\sigma_{t-1}^2$$

  which is an <mark>EWMA</mark> (Exponential Weighted Moving Average) method.
- GARCH-M (GARCH in the Mean)
  - To describe the phenomenon that high risk demands high return.
  - It does not change the structure of the volatility model. Instead, it only changes the mean model like

$$r_t = \mu + c\sigma_t^2 + a_t$$

  where c is called the risk premium parameter. $c\sigma_t^2$ can be changed to be $c\sigma_t$ or $c\ln(\sigma_t^2)$.

- EGARCH (Exponential GARCH)
  - To describe the leverage effect.
  - An $EGARCH(m,\ s)$ model is

$$a_t = \epsilon_t$$
$$\ln(\sigma_t^2) = \alpha_0 + \frac{1 + \beta_1 B + \beta_2 B^2 + \cdots + \beta_{s-1}B^{s-1}}{1 - \alpha_1 B - \alpha_2 B^2 - \cdots - \alpha_m B^m}g(\epsilon_{t-1})$$

  where $g(\epsilon_t) = \theta\epsilon_t + \gamma\left[|\epsilon_t| - E(|\epsilon_t|)\right]$ is weighted innovation with zero mean.
  - Since

$$g(\epsilon_t) = \begin{cases} (\theta + \gamma)\epsilon_t - \gamma E(|\epsilon_t|), & \epsilon \geqslant 0 \\ (\theta - \gamma)\epsilon_t - \gamma E(|\epsilon_t|), & \epsilon < 0 \end{cases}$$

  the model shows asymmetric response of unexpected return.
  - The model is non-linear if $\theta \neq 0$. Thus, θ is called leverage parameter. Since empirically negative unexpected return causes higher volatility, θ should be negative. θ shows the leverage effect and γ denotes the magnitude effect.
  - Alternative form:

$$\ln(\sigma_t^2) = \alpha_0 + \sum_{i=1}^{s}\alpha_i\frac{|a_{t-i}| + \gamma_i a_{t-i}}{\sigma_{t-i}} + \sum_{j=1}^{m}\beta_j\ln(\sigma_{t-j}^2)$$

  where γ_i is the leverage parameter.
- TGARCH (Threshold GARCH)
  - Similar to EGARCH but easier to estimate.
  - A $TGARCH(m,\ s)$ model is

$$a_t = \sigma_t \epsilon_t$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{s}(\alpha_i + \gamma_i N_{t-i})a_{t-i}^2 + \sum_{j=1}^{m}\beta_j \sigma_{t-j}^2$$

where N_{t-i} is an indicator for negative a_{t-i}, i.e., $N_{t-i} = \begin{cases} 1, & a_{t-i} < 0 \\ 0, & otherwise \end{cases}$ and γ_i is

the leverage parameter which is expected to be positive.
- APGARCH (Asymmetric Power GARCH)
  - Also consider the leverage effect.
  - An $APGARCH(m, \ s)$ model is like the alternative form of TGARCH:

$$\sigma_t^\delta = \omega + \sum_{i=1}^{s}\alpha_i(|a_{t-i}| + \gamma_i a_{t-i})^\delta + \sum_{j=1}^{m}\beta_j \sigma_{t-j}^\delta$$

where $\delta \geqslant 0$ is the power parameter.
  - Actually, TGARCH model is a special case of APGARCH models.

- **VaR**

  - **Definition**