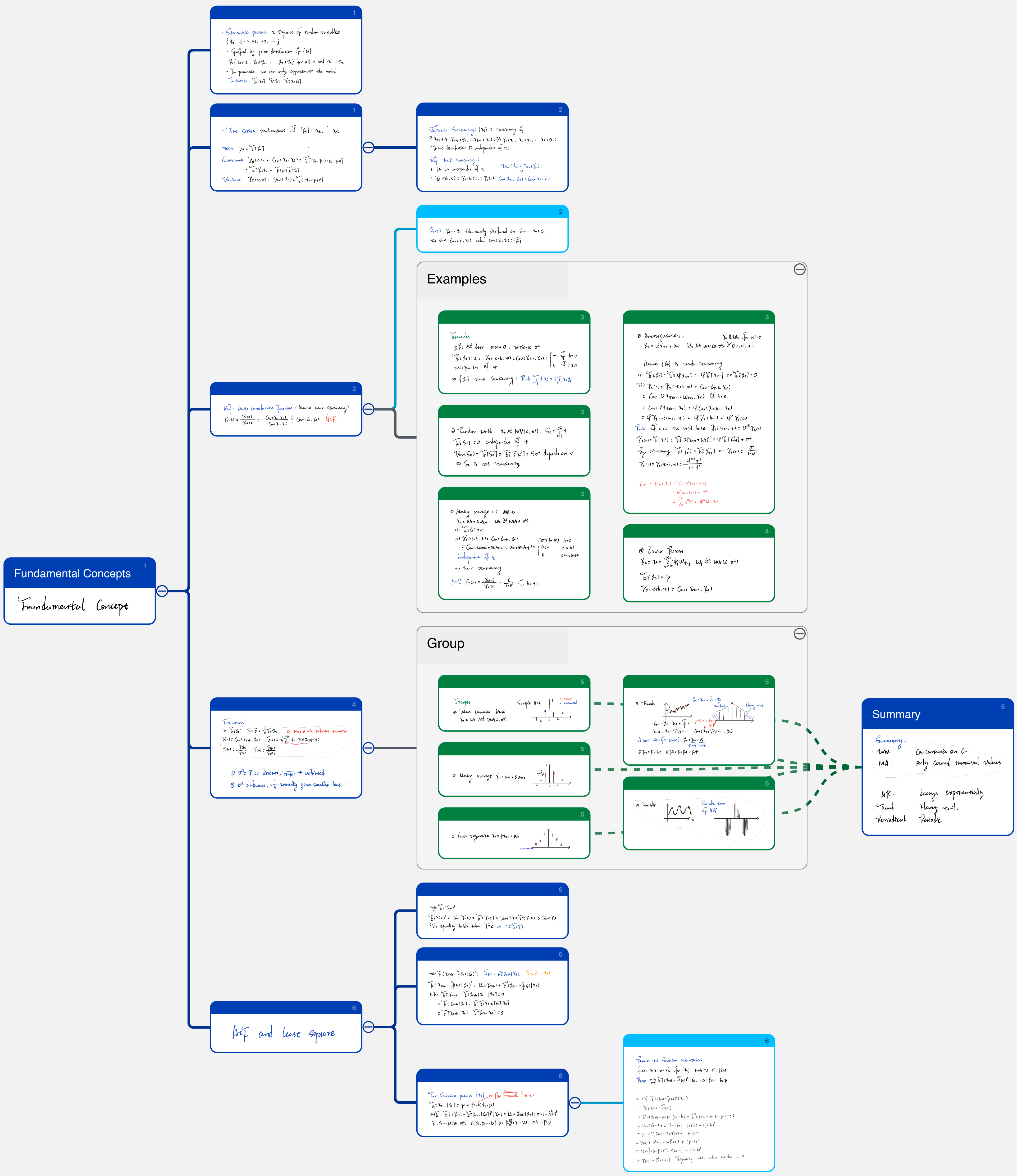
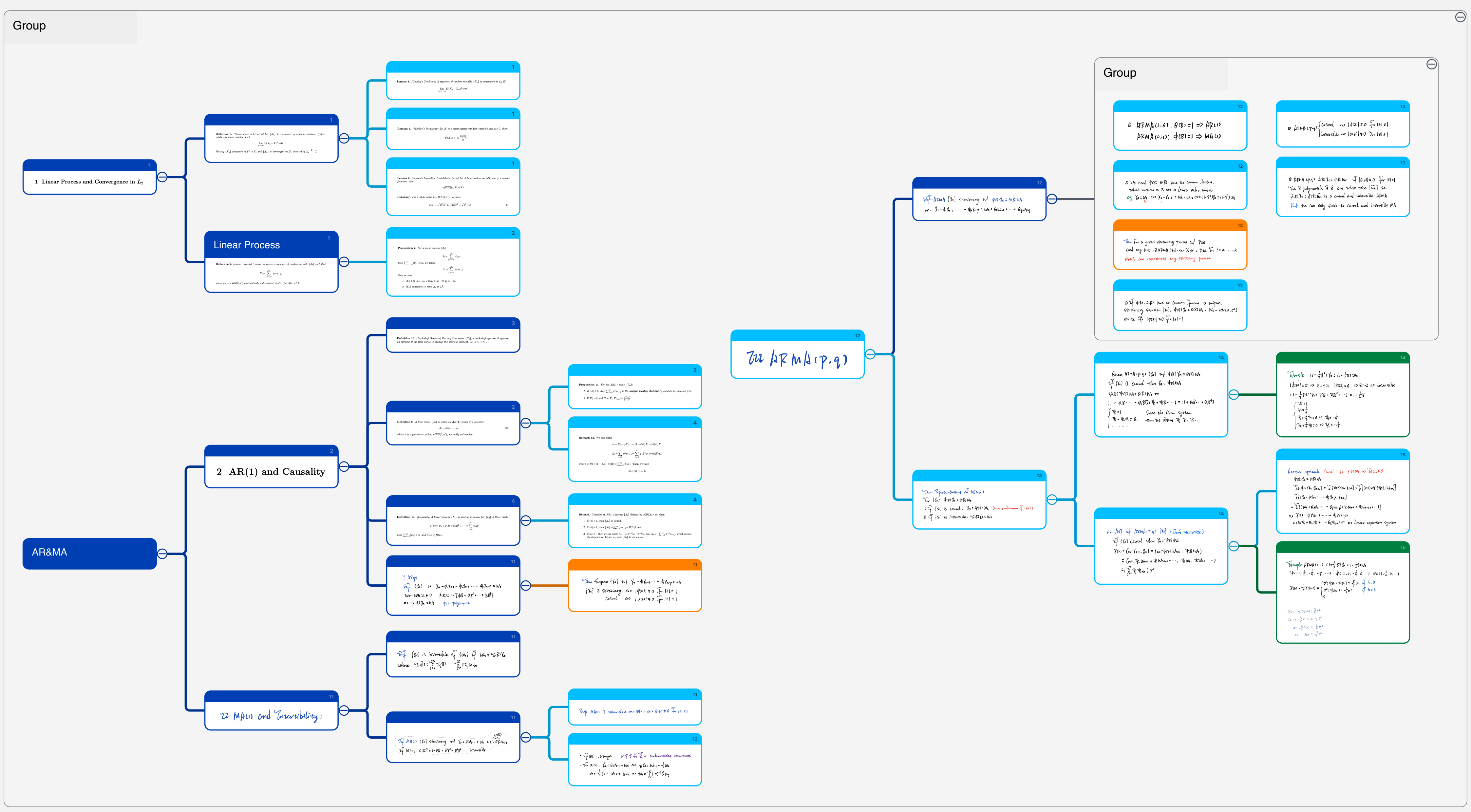
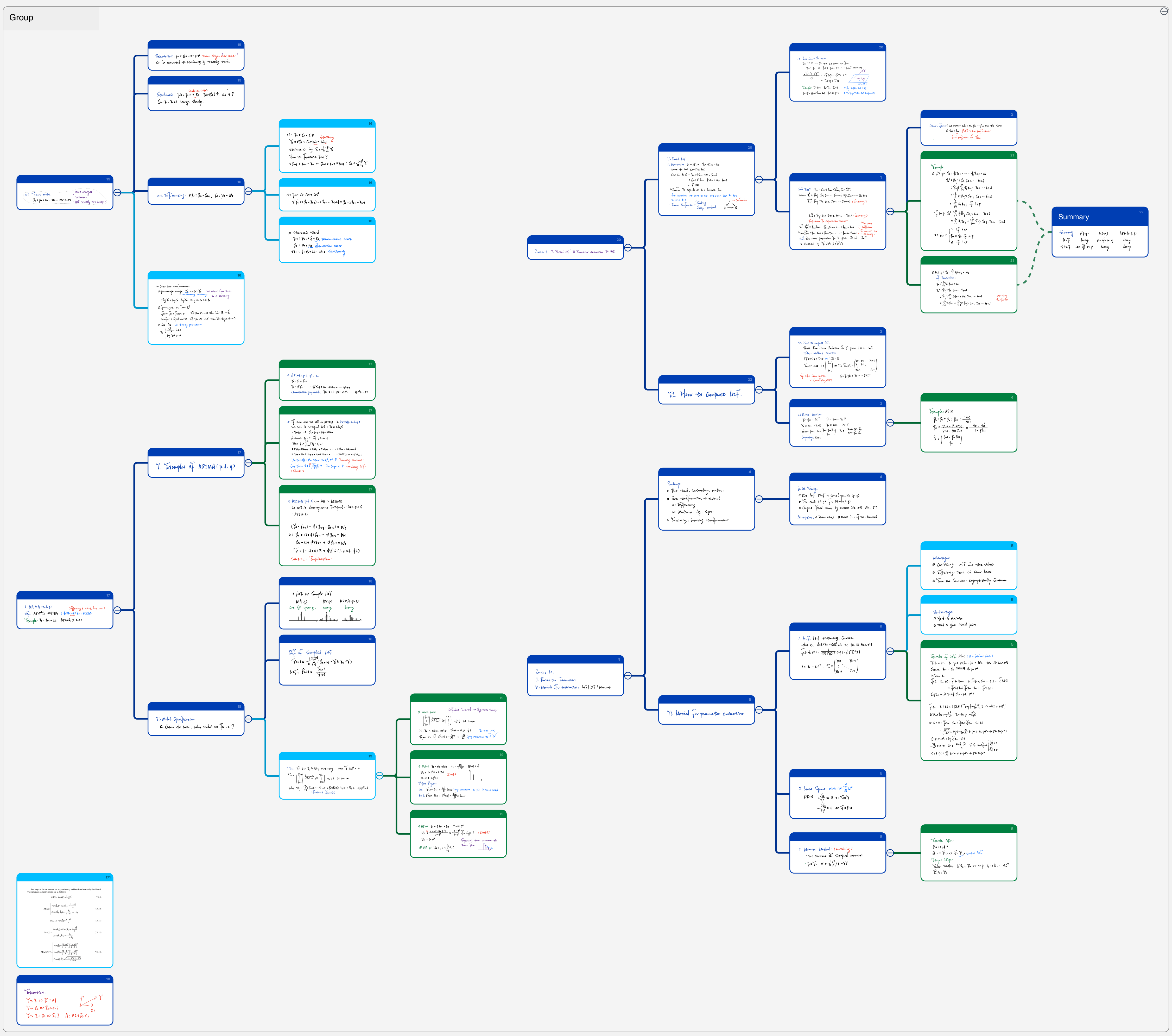
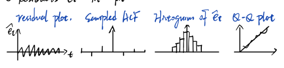

Fundamental Concepts
Fundamental Concept
Examples
Group
Summary
ACF and least square

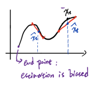

Group
1 Linear Process and Convergence in L_2
Linear Process
AR&MA
2 AR(1) and Causality
Group


Group
Summary

- **Trends**
  - 7 — §. Trends $X_t = \mu_t + e_t$ μ_t: trend e_t: noise, Stationary, $E(e_t) = 0$
    - 7 — ① $\mu_t \equiv \mu$ Constant.
      $\hat{\mu} = \frac{1}{n}\sum_{t=1}^{n} X_t := \bar{X}$
      $Var(\hat{\mu}) = Var(\bar{X}) = Var(\frac{1}{n}\sum_i X_i) = \frac{1}{n^2}\sum_{i,j} Cov(X_i, X_j)$
      $= \frac{1}{n^2}[\sum_i Var(X_i) + 2\sum_{i=1}^{n-1}\sum_{h=1}^{n-i} Cov(X_i, X_{i+h})]$
      $= \frac{1}{n^2}[n\gamma(0) + 2\sum_{i=1}^{n-1}(n-i)\gamma(i)]$
      $= \frac{\gamma(0)}{n}(1 + 2\sum_{i=1}^{n-1}(1-\frac{i}{n})\rho(i))$
      - 7 — Comparison: For $Y_1 \dots Y_n$ iid (μ, σ^2)
        $Var(\bar{Y}) = \frac{\sigma^2}{n} = \frac{\gamma(0)}{n} \le Var(\bar{X})$
    - 7 — ② $\mu_t = \beta_0 + \beta_1 t$
      In linear regression $Y = X\beta + \varepsilon$
      $\min_\beta \|Y - X\beta\|_2^2 \Rightarrow \hat{\beta} = (X^TX)^{-1}X^TY$
      For $\mu_t = \beta_0 + \beta_1 t$ $\hat{\beta}_1 = \frac{\sum_t (y_t - \bar{y})(t - \bar{t})}{\sum_t (t - \bar{t})^2}$ $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{t}$, $\bar{t} = \frac{n+1}{2}$
      - 7 — Interpretation: $R^2 = \frac{Var(\hat{Y})}{Var(Y)} = Corr^2(\hat{Y}, Y) = \cos^2(\theta)$?
    - 7 — ③ Residuals $\hat{e}_t = X_t - \hat{\mu}_t$
      residual plot. Sampled ACF Histogram of $\hat{e}_t$ Q-Q plot

  - **Fitting**
    - 8 — • Parametric fitting $\Rightarrow \mu_t = \beta_0 + \beta_1 t$
    - 8 — No assumption $\Rightarrow$ Smoothing
      • Smoothing: $\hat{\mu}_t = \frac{1}{2q+1}\sum_{h=-q}^{q} X_{t+h}$
      If μ_t is linear in $[t-q, t+q]$
      then $\hat{\mu}_t = \mu_t + \frac{1}{2q+1}\sum_{t=-q}^{q} e_t$
      Small $q \to$ high bias
      Large $q \to$ accurate estimation, but violates local linearity $\to$ Large bias.
      Smoothing: Special case of linear filtering. $\hat{\mu}_t = \sum_{h=-q}^{q} a_h X_{t+h}$
      • Issue: ① End point ② tunning parameter q

    - 8 — • Isotonic regression. $\min_{a_1 \dots a_n} \sum_{t=1}^{n} (X_t - a_t)^2$ st $a_1, \dots a_n$ monotonic
      merits: ① Solved in $O(n\log n)$
      ② no tunning parameter
      ③ no end-point issue
      ④ non-parametric
      Drawback: Assuming monotonicity
      Dif btw para & nonpara
      para: # para $\ll n$
      nonpara: # para $\sim n$

Scribing notes 3~4 people in a group

CH 1. Introduction

{ Time domain approach
{ Frequency domain approach

1.1. The nature of time series data

Features: ① trend ② seasonality ③ sporadic pattern

Return: $\text{Return}_t = \log P_t - \log P_{t-1} = \log\left(\frac{P_t}{P_{t-1}} - 1 + 1\right) \approx \frac{P_t}{P_{t-1}} - 1.$

P_t: Price at time t

Symbols: WN: Gaussian, normal

Lecture 2 Fundamental Concept

- Stochastic process: a sequence of random variables
  $\{X_t : t = 0, \pm 1, \pm 2, \cdots\}$
  - Specified by joint distribution of $\{X_t\}$
    $\Pr\{X_1 \le x_1, X_2 \le x_2, \cdots, X_t \le x_t\}$ for all t and $x_1 \cdots x_t$
  - In practice, we can only approximate the model
    Interests: $E[X_t]$, $E[X_s]$, $E[X_t X_s]$
- Time series: realizations of $\{X_t\}$: $x_{t_1}, \cdots x_{t_n}$

mean: $\mu_t = \mathbb{E}[X_t]$

Covariance $\gamma_X(t,s) = \mathrm{Cov}(X_t, X_s) = \mathbb{E}[(X_t-\mu_t)(X_s-\mu_s)]$

$$= \mathbb{E}[X_t X_s] - \mathbb{E}[X_t]\,\mathbb{E}[X_s]$$

Variance $\gamma_X(t,t) = \mathrm{Var}(X_t) = \mathbb{E}[(X_t-\mu_t)^2]$

Definition (Stationary) $\{X_t\}$ is stationary if

$$P(X_{t+1} \le x_1, X_{t+2} \le x_2, \dots X_{t+n} \le x_n) = P(X_1 \le x_1, X_2 \le x_2, \dots X_n \le x_n)$$

(Joint distribution is independent of t)

Def (Weak stationary)

1. μ_t is independent of t $\qquad \mathrm{Var}(X_t) = \mathrm{Var}(X_0)$

$$\Uparrow$$

2. $\gamma_X(t+h, t) = \gamma_X(h, 0) := \gamma_X(h)$ $\qquad \mathrm{Cov}(X_{t+h}, X_t) = \mathrm{Cov}(X_t, X_0)$

Def. Auto Correlation Function (Assume weak stationary)

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = \frac{\mathrm{Cov}(X_h, X_0)}{\mathrm{Cov}(X_0, X_0)} \overset{?}{=} \mathrm{Corr}(X_h, X_0) \qquad \text{ACF}$$

Prop 1. $X_1 \cdots X_n$ identically distributed and $X_1 + \cdots + X_n = C$, with the same $\mathrm{Cov}(X_i, X_j)$ then $\mathrm{Corr}(X_1, X_2) = \frac{-1}{n-1}$

Proof:

$$0 = \mathrm{Var}(X_1 + \cdots + X_n) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + \sum_{i \ne j} \mathrm{Cov}(X_i, X_j)$$

$$= n\,\mathrm{Var}(X_1) + n(n-1)\,\mathrm{Cov}(X_1, X_2)$$

$$\Rightarrow \mathrm{Corr}(X_1, X_2) = \frac{\mathrm{Cov}(X_1, X_2)}{\mathrm{Var}(X_1)} = \frac{-1}{n-1}$$

Examples:

① $X_i \overset{iid}{\sim}$ dstn, mean 0, variance σ^2

$E[X_t]=0$, $\gamma_X(t+h,t)=Cov(X_{t+h},X_t)=\begin{cases}\sigma^2 & \text{if } h=0\\ 0 & \text{if } h\neq 0\end{cases}$

independence of t

$\Rightarrow \{X_t\}$ weak stationary. Rmk $E\sum_{i\neq j} X_iX_j=\sum_{i\neq j} E X_iX_j$.

② Random walk: $X_i \overset{iid}{\sim} WN(0,\sigma^2)$, $S_t=\sum_{i=1}^{t} X_i$

$E[S_t]=0$ independence of t

$Var(S_t)=E[S_t^2]=E[\sum X_i^2]=t\sigma^2$ depends on t

$\Rightarrow S_t$ is not stationary

③ Moving average (1). MA(1)

$X_t=W_t+\theta W_{t-1}$. $W_t \overset{iid}{\sim} WN(0,\sigma^2)$

(i) $E[X_t]=0$

(ii) $\gamma_X(t+h,t)=Cov(X_{t+h},X_t)$

$=Cov(W_{t+h}+\theta W_{t+h-1}, W_t+\theta W_{t-1})=\begin{cases}\sigma^2(1+\theta^2) & h=0\\ \theta\sigma^2 & h=\pm 1\\ 0 & \text{otherwise}\end{cases}$

independence of t

$\Rightarrow$ weak stationary

ACF. $\rho_X(h)=\frac{\gamma_X(h)}{\gamma_X(0)}=\frac{\theta}{1+\theta^2}$ if $h=\pm 1$

④ Autoregressive (1)

$X_0 \perp\!\!\!\perp W_t$ for all t

$X_t=\varphi X_{t-1}+W_t$ $\quad W_t \overset{iid}{\sim} WN(0,\sigma^2)$, $0<|\varphi|<1$.

Assume $\{X_t\}$ is weak stationary

(i) $E[X_t] = E[\varphi X_{t-1}] = \varphi E[X_{t-1}] \Rightarrow E[X_t] = 0$

(ii) $\gamma_X(h) = \gamma_X(t+h, t) = Cov(X_{t+h}, X_t)$

$= Cov(\varphi X_{t+h-1} + W_{t+h}, X_t)$ if $h > 0$:

$= Cov(\varphi X_{t+h-1}, X_t) = \varphi Cov(X_{t+h-1}, X_t)$

$= \varphi \gamma_X(t+h-1, t) = \varphi \gamma_X(h-1) = \varphi^h \gamma_X(0)$

Rmk. if $h < 0$, we still have $\gamma_X(t+h, t) = \varphi^{|h|} \gamma_X(0)$

$\gamma_X(0) = E[X_t^2] = E[(\varphi X_{t-1} + W_t)^2] = \varphi^2 E[X_{t-1}^2] + \sigma^2$

By stationary $E[X_t^2] = E[X_{t-1}^2] \Rightarrow \gamma_X(0) = \frac{\sigma^2}{1-\varphi^2}$

$\gamma_X(h) = \gamma_X(t+h, t) = \frac{\varphi^{|h|} \sigma^2}{1-\varphi^2}$

⑤ Linear Process

$X_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j W_{t-j} \quad W_j \overset{iid}{\sim} WN(0, \sigma^2)$

$E[X_t] = \mu$

$\gamma_X(t+h, t) = Cov(X_{t+h}, X_t)$

Lecture 3.

Estimators.

① $\sigma^2 = \gamma(0)$ known, $\frac{1}{n-|h|} \rightarrow$ unbiased

② σ^2 unknown, $\frac{1}{n}$ usually gives smaller bias

$\mu = E[X_t] \quad \hat{\mu} = \bar{x} = \frac{1}{n} \sum_t X_t$

Q: What is the unbiased estimator

$\gamma(h) = Cov(X_{t+h}, X_t)$, $\quad \hat{\gamma}(h) = \underline{\frac{1}{n} \sum_{t=1}^{n-|h|} (X_t - \bar{x})(X_{t+|h|} - \bar{x})}$

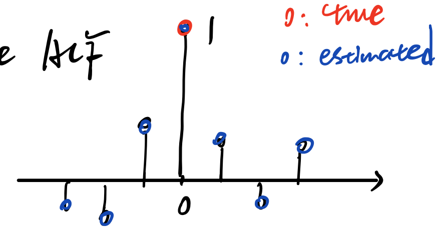
$\rho(h) = \frac{\gamma(h)}{\gamma(0)} \quad \hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$

Example:

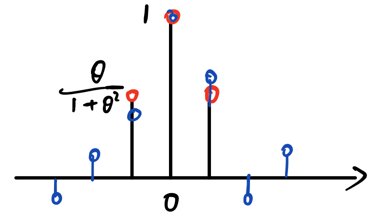
① White Gaussian Noise

$X_t = W_t \overset{iid}{\sim} WN(0, \sigma^2)$

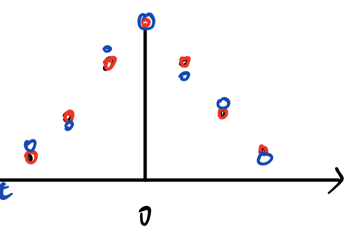
Sample ACF



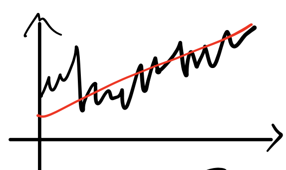
② Moving average $X_t = W_t + \theta W_{t-1}$



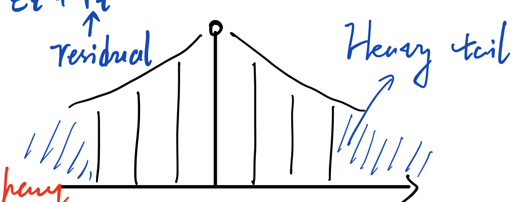
③ Auto regressive $X_t = \phi X_{t-1} + W_t$



④ Trends



$$\underbrace{X_t - X_{t-1}}_{\text{increment}} = Z_t + \underbrace{r_t}_{\text{residual}}$$



$X_{t+1} - X_t = \mu_t + f(\;)$

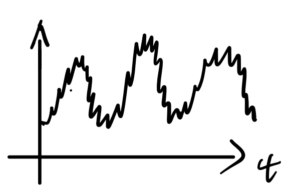
Gives the heavy tail

$X_{t+h} = X_t + \sum \mu_i + \cdots$ $\quad$ $Corr(X_t + \sum_i \mu_i + \cdots, X_t)$

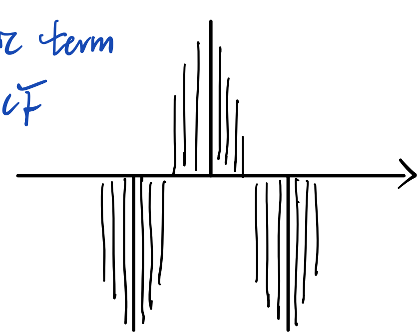
A more specific model: $X_t = \underbrace{\mu_t}_{\text{trend}} + \underbrace{e_t}_{\text{noise}}$

① $\mu_t = \beta_0 + \beta_1 t$ ② $\mu_t = \beta_0 + \beta_1 t + \beta_2 t^2$

⑤ Periodic.



Periodic term of ACF



Summary.

WN. Concentrate on 0.

MA. only several nontrivial values

AR. decays exponentially
Trend. Heavy tail.
Periodical Periodic

ACF and least square

$\min_c E(Y-c)^2$

$E(Y-c)^2 = Var(Y-c) + E^2(Y-c) = Var(Y) + E^2(Y-c) \geq Var(Y)$

The equality holds when $Y=c$

$\min E(X_{t+h} - f(X_t) \mid X_t)^2$: $f(X_t) = E[X_{t+h} \mid X_t]$. $\hat{P} = P(\cdot \mid X_t)$

$E(X_{t+h} - f(X_t) \mid X_t)^2 = Var(X_{t+h}) + E^2(X_{t+h} - f(X_t) \mid X_t)$

WTS: $E[X_{t+h} - E[X_{t+h} \mid X_t] \mid X_t] = 0$

$= E[X_{t+h} \mid X_t] - E[E[X_{t+h} \mid X_t] \mid X_t]$

$= E[X_{t+h} \mid X_t] - E[X_{t+h} \mid X_t] = 0$

For Gaussian process $\{X_t\}$: $\rho(h) \xlongequal{\text{Stationary}} \rho(h,t)$

$E[X_{t+h} \mid X_t] = \mu + \rho(h)(X_t - \mu)$

$MSE = E[(X_{t+h} - E[X_{t+h} \mid X_t])^2 \mid X_t] = Var(X_{t+h} \mid X_t) = \sigma^2(1 - \rho^2(h))$

$X_1, X_2 \sim N(\mu_i, \sigma_i^2)$ $X_1 \mid X_2 = x_2 \sim N(\mu + \rho\frac{\sigma_1}{\sigma_2}(x_2 - \mu_2), \sigma_1^2(1-\rho^2))$

Remove the Gaussian assumption:

$f(x) = a(x-\mu) + b$ for $\{X_t\}$ with $\mu, \sigma^2, \rho(h)$.

Prove $\min_{a,b} E[(X_{t+h} - f(X_t))^2 \mid X_t]$, $a = \rho(h)$, $b = \mu$

Lecture 4.

§. Trends $X_t = \mu_t + e_t$ μ_t: trend e_t: noise, Stationary

$E(e_t) = 0$

① $\mu_t \equiv \mu$ Constant.

$$\hat{\mu} = \frac{1}{n}\sum_{t=1}^{n} X_t := \bar{X}$$

$$Var(\hat{\mu}) = Var(\bar{X}) = Var(\frac{1}{n}\sum_i X_i) = \frac{1}{n^2}\sum_{i,j} Cov(X_i, X_j)$$

$$= \frac{1}{n^2}\left[\sum_i Var(X_i) + 2\sum_{i=1}^{n-1}\sum_{h=1}^{n-i} Cov(X_i, X_{i+h})\right]$$

$$= \frac{1}{n^2}\left[n\gamma(0) + 2\sum_{i=1}^{n-1}(n-i)\gamma(i)\right]$$

$$= \frac{\gamma(0)}{n}\left(1 + 2\sum_{i=1}^{n-1}(1-\frac{i}{n})\rho(i)\right)$$

Comparison: For $Y_1 \cdots Y_n$ iid (μ, σ^2)

$$Var(\bar{Y}) = \frac{\sigma^2}{n} = \frac{\gamma(0)}{n} \leq Var(\bar{X})$$

② $\mu_t = \beta_0 + \beta t$

In linear regression $Y = X\beta + \varepsilon$

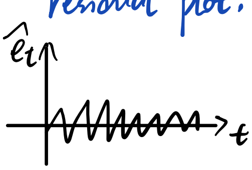
$$\min_{\beta} \|Y - X\beta\|_2^2 \Rightarrow \hat{\beta} = (X^T X)^{-1} X^T Y$$

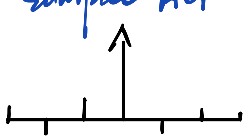
For $\mu_t = \beta_0 + \beta_t$ $\hat{\beta}_1 = \dfrac{\sum_t (y_t - \bar{y})(t - \bar{t})}{\sum_t (t - \bar{t})^2}$ $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{t}$, $\bar{t} = \frac{n+1}{2}$

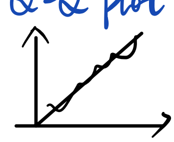
Interpretation: $R^2 = \dfrac{Var(\hat{Y})}{Var(Y)} = Corr^2(\hat{Y}, Y) = \cos^2(\theta)$?

③ Residuals $\hat{e}_t = X_t - \hat{\mu}_t$

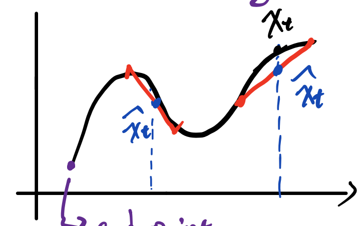
residual plot. Sampled ACF Histogram of $\hat{e}_t$ Q-Q plot







- Parametric fitting ⇒ $\mu_t = \beta_0 + \beta_1 t$ No assumption ⇒ Smoothing



- Smoothing: $\hat{\mu}_t = \frac{1}{2q+1} \sum_{h=-q}^{q} X_{t+h}$

  If μ_t is linear in $[t-q, t+q]$

  then $\hat{\mu}_t = \mu_t + \frac{1}{2q+1} \sum_{t=-q}^{q} e_t$

  Small $q \to$ high bias

  Large $q \to$ accurate estimation, but violates local linearity → large bias.

  Smoothing: Special case of linear filtering. $\hat{\mu}_t = \sum_{h=-q}^{q} a_h X_{t+h}$

- Issue: ① End point ② tunning parameter q

- Isotonic regression. $\min_{a_1 \cdots a_n} \sum_{t=1}^{n} (X_t - a_t)^2$ st $a_1, \cdots a_n$ monotonic

  merits: ① solved in $O(n\log n)$

  ② no tunning parameter

  ③ no end-point issue

  ④ non-parametric

  Drawback: Assuming monotonicity

Dif btw para & nonpara

para: # para << n

nonpara: # para ~ n

Lecture 5.

- Linear process & Convergence in L_2
- AR(1) and Causality
- MA(1) and invertibly

Def of WN:
mean square mean 0, finite var.
uncorrelated

1. Linear process and Convergence in L_2

Linear process $X_t = \sum_{j=-\infty}^{\infty} \psi_j W_{t-j}$ $W_t \sim WN(0,\sigma^2)$, $\psi_j \in \mathbb{R}$

Def. Convergence in L_2: $S_n \xrightarrow{L_2} Y$ iff $\lim_{n\to\infty} E(S_n - Y)^2 = 0$

Cauchy criterion: $\Leftrightarrow \lim_{n,m\to\infty} E(S_n - S_m)^2 = 0$

Proposition: Suppose $X_t = \sum_{j=-\infty}^{+\infty} \psi_j W_{t-j}$ with $\sum_{j=-\infty}^{+\infty} |\psi_j| < \infty$ $S_n = \sum_{j=-n}^{n} \psi_j W_{t-j}$

Then ① $|X_t| < \infty$ a.s. ② S_n Convergence to X_t in L_2

Def $|X_t| < \infty$ a.s. $\Leftrightarrow P(|X_t| > \lambda) \to 0$ as $\lambda \to \infty$

$$P(|X_t| > \lambda) \le \frac{1}{\lambda} E|X_t| = \frac{1}{\lambda} E\left|\sum_{j=-\infty}^{+\infty} \psi_j W_{t-j}\right| \le \frac{1}{\lambda} \sum_{j=-\infty}^{+\infty} |\psi_j| E|W_{t-j}|.$$

$$\le \frac{\sigma}{\lambda} \sum_{j=-\infty}^{+\infty} |\psi_j| \to 0 \quad \text{as } \lambda \to 0$$

$$E|W_{t-j}| = \sqrt{(E(|W_{t-j}| \cdot 1))^2} \le \sqrt{E|W_{t-j}|^2 \, E 1^2} = \sigma$$

$$E|W_{t-j}| = \sqrt{(E|W_{t-j}|)^2} \le \sqrt{E[W_{t-j}^2]} = \sqrt{\sigma^2}$$

Pf of ② $E(S_m - S_n) = E\left(\sum_{n\le|j|\le m} \psi_j W_{t-j}\right)^2$

$E(W_i W_j) = 0$

$$= \sum_{n\le|j|\le m} \psi_j^2 E W_{t-j}^2$$

$$= \sum_{n\le|j|\le m} \psi_j^2 \sigma^2$$

$$\le \sigma^2 \left(\sum_{n\le j\le m} |\psi_j|\right)^2 \to 0 \quad \text{as } m,n \to \infty$$

2. AR(1) and causality. AR 1: causal $\Leftrightarrow |\phi|<1 \Leftrightarrow \phi(z)\neq 0$ 若 $|z|>1$

AR(1): $X_t = \phi X_{t-1} + W_t$, $W_t \sim WN(0,\sigma^2)$

① 若 $|\phi|<1$, then $X_t = \sum_{j=0}^{\infty} \phi^j W_{t-j}$ is a solution

$\Downarrow$

$\sum |\phi^j| < \infty$

② $E X_t = 0$, $Cov(X_t, X_{t+h}) = \dfrac{\phi^{|h|}\sigma^2}{1-\phi^2}$

③ Uniqueness

assume there is another weakly stationary solution Y_t:

$$\lim_{n\to\infty} E\left(Y_t - \sum_{j=0}^{n} \phi^j W_{t-j}\right)^2 = \lim_{n\to\infty} E(\phi^n Y_{t-n})^2 = \lim_{n\to\infty} \phi^{n} E Y_{t-n}^2 \to 0 \text{ as } n\to\infty$$

$Y_t = X_t$ in L_2

Back-shift operator Def $BX_t = X_{t-1}$

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

$\Leftrightarrow X_t - \phi X_{t-1} = W_t$

$\Leftrightarrow \underbrace{(1-\phi B)}_{} X_t = W_t$

$\phi(B) X_t = W_t$

$X_t = \sum_{j=0}^{\infty} \phi^j W_{t-j}$ require: $|\phi|<1$

$= \underbrace{\sum_{j=0}^{\infty} \phi^j B^j}_{} W_t$

$= \pi(B) W_t$

$\phi(B)\pi(B) = 1$

Def $\{X_t\}$ is causal for W_t 若 X_t can be written as $X_t = \psi(B) W_t$

where $\psi(B) = \psi_0 + \psi_1 B + \psi_2 B^2 + \cdots$ $\sum_{j=0}^{\infty} |\psi_j| < \infty$

AR(1) is causal bcs $X_t = \pi(B) W_t$, $\pi(B) = \sum_{j=0}^{\infty} \phi^j B^j$

若 $|\phi| = 1 \Rightarrow X_t = \sum_{j=0}^{\infty} W_{t-j} \sim WN(0,\infty)$

若 $|\phi|>1 \Rightarrow X_{t-1} = \frac{1}{\phi} X_t - \frac{1}{\phi} W_t \Rightarrow X_t = -\sum_{j=0}^{\infty} \phi^{-j} W_{t+j}$

Depends on future w_j, not causal.

AR is causal iff $|\phi|<1$ "$\Rightarrow$" $= -\sum_{j=0}^{\infty} \phi^{j} B^{j} w_t$

iff $\{\phi(z) \neq 0$ if $|z| \leq 1\}$

Lecture 6. «Plato and the nerd» Edward Lee

Paradigm ① Model ② estimation ③ evaluation ④ application.

I. AR(p)

Def. $\{X_t\}$. s.t. $X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} \cdots - \phi_p X_{t-p} = W_t$

$W_t \sim WN(0, \sigma^2)$ $\quad \phi(B) = 1 - [\phi_1 B + \phi_2 B^2 + \cdots + \phi_p B^p]$

$\Rightarrow \phi(B) X_t = W_t$ $\quad \phi(\ )$: polynomial

Thm Suppose $\{X_t\}$ w/ $X_t - \phi_1 X_{t-1} \cdots - \phi_p X_{t-p} = W_t$

$\{X_t\}$ is stationary $\Leftrightarrow |\phi(z)| \neq 0$ for $|z| = 1$

casual $\Leftrightarrow |\phi(z)| \neq 0$ for $|z| \leq 1$

II. MA(1) and Invertibility:

Def $\{X_t\}$ is invertible of $\{W_t\}$ if $W_t = \pi(B) X_t$

where $\pi(B) = \sum_{j=0}^{\infty} \pi_j B^j$ $\quad \sum_{j=0}^{\infty} |\pi_j| < \infty$

Def MA(1) $\{X_t\}$ stationary w/ $X_t = \theta W_{t-1} + W_t = \underbrace{(1+\theta B)}_{\theta(B)} W_t$

if $|\theta| < 1$. $\theta(B)^{-1} = 1 - \theta B + \theta^2 B^2 - \theta^3 B^3 \ldots$ invertible

Prop. MA(1) is invertible $\Leftrightarrow |\theta| < 1 \Leftrightarrow \theta(z) \neq 0$ for $|z| \leq 1$.

• 若 $|\theta|=1$, diverge 《女士品茶》. randomization experiment

• 若 $|\theta|>1$, $X_t = \theta W_{t-1} + W_t \iff \frac{1}{\theta} X_t = W_{t-1} + \frac{1}{\theta} W_t$

$\iff \frac{1}{\theta} X_t = W_{t+1} + \frac{1}{\theta} W_t \Rightarrow W_t = \sum_{j=0}^{\infty} (-\theta)^{-j} X_{t+j}$

$$\begin{cases} AR(p): \phi(B) X_t = W_t \\ MA(p): X_t = \theta(B) W_t \end{cases}$$

7u ARMA(p,q)

Def-ARMA $\{X_t\}$ stationary w/ $\phi(B) X_t = \theta(B) W_t$

i.e. $X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = W_t + \theta_1 W_{t-1} + \cdots + \theta_q W_{t-q}$

① $ARMA(1,0): \theta(B)=1 \Rightarrow AR(1)$

$ARMA(0,1): \phi(B)=1 \Rightarrow MA(1)$

② We need $\phi(B)$ $\theta(B)$ hue no common factors.

which implies it is not a lower order model

eg. $X_t = W_t \iff X_t - X_{t-3} = W_t - W_{t-3} \iff (1-B^3) X_t = (1-B^3) W_t$

Thm For a given stationary process w/ $\gamma(h)$

and any $k>0$, $\exists$ ARMA $\{X_t\}$. s.t. $\gamma_X(h) = \gamma(h)$. for $h = 0, 1, \ldots k$.

ARMA can approximate any stationary process.

Thm

① 若 $\phi(B)$, $\theta(B)$ hue no common factors, a unique, stationary solution $\{X_t\}$. $\phi(B) X_t = \theta(B) W_t$, $W_t \sim WN(0, \sigma^2)$

exists iff $|\phi(z)| \neq 0$ for $|z|=1$

① ARMA(p,q) $\begin{cases} \text{casual} \iff |\phi(z)| \neq 0 \text{ for } |z| \leq 1 \\ \text{invertible} \iff |\theta(z)| \neq 0 \text{ for } |z| \leq 1 \end{cases}$

② ARMA(p,q) $\phi(B)X_t = \theta(B)W_t$ if $|\theta(z)| \neq 0$ for $|z| = 1$

Then ∃ polynomials $\hat{\phi}$ $\hat{\theta}$ and white noise $\{\tilde{W}_t\}$ s.t.

$\tilde{\phi}(B)X_t = \tilde{\theta}(B)\tilde{W}_t$ is a casual and invertible ARMA

Rmk. we can only stick to casual and invertible MA.

Thm (Representations of ARMA)

For $\{X_t\}$: $\phi(B)X_t = \theta(B)W_t$

① if $\{X_t\}$ is causal, $X_t = \psi(B)W_t$ (linear combination of $\{W_t\}$)

② if $\{X_t\}$ is invertible, $\pi(B)X_t = W_t$

How to get $\psi(B)$: $\begin{cases} \phi(B)X_t = \theta(B)W_t \\ X_t = \psi(B)W_t \end{cases} \Rightarrow \phi(B)\psi(B)W_t = \theta(B)W_t$

$$(\phi_0 + \phi_1 B + \phi_2 B^2 + \cdots)(\psi_0 + \psi_1 B + \psi_2 B^2 + \cdots) = (\theta_0 + \theta_1 B + \theta_2 B^2 + \cdots)$$

$\begin{cases} \phi_0\psi_0 = \theta_0 \\ \phi_1\psi_0 + \phi_0\psi_1 = \theta_1 \\ \cdots \end{cases} \Rightarrow$ Solve the linear system

Lecture 7.

1. Computation for ARMA.

1.1. How to get ψ (causal)

Given ARMA(p,q) $\{X_t\}$ w/. $\phi(B)X_t = \theta(B)W_t$

If $\{X_t\}$ is causal then $X_t = \psi(B)W_t$

$\phi(B)\psi(B)W_t = \theta(B)W_t$ =>

$(1-\phi_1 B - \cdots - \phi_p B^p)(\psi_0 + \psi_1 B + \cdots) = (1+\theta_1 B + \cdots + \theta_q B^q)$

$$\begin{cases} \psi_0 = 1 \\ \psi_1 - \psi_0\phi_1 = \theta_1 \\ \cdots\cdots \end{cases}$$

Solve the linear system, then we obtain $\psi_0, \psi_1, \psi_2 \ldots$

Example $(1+\frac{1}{4}B^2)X_t = (1+\frac{1}{2}B)W_t$

$|\phi(z)| = 0$ => $z = \pm 2i$ $\quad |\theta(z)| = 0$ => $z = -2$ => invertible

$(1+\frac{1}{4}B^2)(\psi_0 + \psi_1 B + \psi_2 B^2 + \cdots) = 1 + \frac{1}{2}B$

$$\begin{cases} \psi_0 = 1 \\ \psi_1 = \frac{1}{2} \\ \psi_2 + \frac{1}{4}\psi_0 = 0 \Rightarrow \psi_2 = -\frac{1}{4} \\ \psi_3 + \frac{1}{4}\psi_1 = 0 \Rightarrow \psi_3 = -\frac{1}{8} \end{cases}$$

1.2 ACF of ARMA(p,q) $\{X_t\}$ (Don't memorize)

If $\{X_t\}$ causal then $X_t = \psi(B)W_t$

$\gamma(h) = Cov(X_{t+h}, X_t) = Cov(\psi(B)W_{t+h}, \psi(B)W_t)$

$= Cov(\psi_0 W_{t+h} + \psi_1 W_{t+h-1} + \cdots, \psi_0 W_t, \psi_1 W_{t-1}, \cdots)$

$= \left(\sum_{j\geq 0}^{\infty} \psi_j \psi_{j+h}\right)\sigma^2$

Another approach: Causal : $X_t = \psi(B) W_t \Rightarrow E[X_t] = 0$

$\phi(B) X_t = \theta(B) W_t$

$E[(\phi(B) X_t) X_{t-h}] = E[\theta(B) W_t X_{t-h}] = E[(\theta(B) W_t)(\psi(B) W_{t-h})]$

$E[(X_t - \phi X_{t-1} - \cdots - \phi_p X_{t-p}) X_{t-h}]$

$= E[(W_t + \theta_1 W_{t-1} + \cdots + \theta_q W_{t-q})(\psi_0 W_{t-h} + \psi_1 W_{t-h-1} + \cdots)]$

$\Rightarrow \gamma(h) - \phi_1 \gamma(h-1) - \cdots - \phi_p \gamma(h-p)$

$= (\theta_h \psi_0 + \theta_{h+1} \psi_1 + \cdots + \theta_q \psi_{q-h}) \sigma^2 \Rightarrow$ Linear equation system

Example. ARMA(2,1) $(1 + \frac{1}{4} B^2) X_t = (1 + \frac{1}{2} B) W_t$

$\psi = (1, \frac{1}{2}, -\frac{1}{4}, -\frac{1}{6}, \cdots)$ $\phi = (1, 0, -\frac{1}{4}, 0, \cdots)$ $\theta = (1, \frac{1}{2}, 0, \cdots)$

$$\gamma(h) + \frac{1}{4}\gamma(h-2) = \begin{cases} \sigma^2(\psi_0\theta_0 + \psi_1\theta_1) = \frac{5}{4}\sigma^2 \\ \sigma^2(\psi_0\theta_1) = \frac{1}{2}\sigma^2 \\ 0 \end{cases}$$

2. Non-Stationary time series

2.1 Trends model.

$X_t = \mu_t + W_t$, $W_t \sim WN(0, \sigma^2)$ $\begin{cases} \text{mean changes} \\ \text{Variance.} \\ \text{ACF usually not decay.} \end{cases}$

Deterministic: $\mu_t = C_0 + C_1 t + C_2 t^2$ mean chnges over time.

Can be converted to stationary by removing trends

Stochastic: $\mu_t = \mu_{t-1} + \underbrace{e_t}_{\text{Stochastic noise.}}$ $Var(X_t) \uparrow$. as $t \uparrow$

$Cov(X_t, X_{t-h})$ decays slowly.

Characteristics.

2.2 Differencing. $\nabla X_t = X_t - X_{t-1}$, $X_t = \mu_t + W_t$

(1) $\mu_t = C_0 + C_1 t$

$Y_t = \nabla X_t = C_1 + \underbrace{W_t - W_{t-1}}_{\text{Stationary}}$

estimate C_1 by $\hat{C}_1 = \frac{1}{n}\sum_{i=1}^{n} Y_i$

How to forecast X_{n+1}?

$\nabla X_{n+1} = X_{n+1} - X_n \Rightarrow X_{n+1} = X_n + \nabla X_{n+1} = X_n + \frac{1}{n}\sum_{i=1}^{n} Y_i$

(2) $\mu_t = C_0 + C_1 t + C_2 t^2$

$\nabla^2 X_t = (X_t - X_{t-1}) - (X_{t-1} - X_{t-2}) = X_t - 2X_{t-1} + X_{t-2}$

(3) Stochastic trend

$\mu_t = \mu_{t-1} + \delta + \underbrace{e_t}$ measurement error

$X_t = \mu_t + \underbrace{W_t}$ observation error

$\nabla X_t = \delta + e_t + W_t - W_{t-1}$ Stationary

(4) Other data transformation.

① percentage change $\underbrace{Y_t}_{\text{non-stationary}} = (1 + \underbrace{X_t}_{\text{stationary}}) Y_{t-1}$. we expect after trans, X_t is stationary

$\nabla \log Y_t = \log Y_t - \log Y_{t-1} = \log(1 + X_t) \approx X_t$

② $f(x) = \log(x)$ or $f(x) = \sqrt{x}$

$f(x) \approx f(t) + f'(t)(x - t)$ If $Var(X) \sim ct$ then $Var(\sqrt{X}) \sim \frac{c}{4}$

$Var f(x) = (f'(t))^2 Var(X)$ If $Var(X) \sim ct^2$ then $Var(\log(X)) \sim c$

③ Box-Cox λ: tunning parameter.

$$X_t \begin{cases} \frac{Y_t^{\lambda} - 1}{\lambda} & \lambda \neq 0 \\ \log(Y_t) & \lambda = 0 \end{cases}$$

3. ARIMA(p, d, q)

Differencing d times, has root 1

Def: $\phi(B)\nabla^d X_t = \theta(B) W_t$ ($\phi(B)(1-B)^d X_t = \theta(B) W_t$

Example: $X_t = X_{t-1} + W_t$ ARIMA(0.1.0)

Lecture 8

I. Examples of ARIMA(p.d.q)

① ARIMA(p, 1, q). X_t

$Y_t = X_t - X_{t-1}$

$Y_t - \phi_1 Y_{t-1} - \cdots - \phi_p Y_{t-p} = W_t + \theta_1 W_{t-1} + \cdots + \theta_q W_{t-q}$

Characteristic polynomial: $\tilde{\phi}(z) = (1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p)(1-z)$

② If there are no AR in ARIMA i.e. ARIMA(0, d, q)

we call it integral MA (IMA(d,q))

· IMA(1,1) $X_t - X_{t-1} = W_t - \theta W_{t-1}$

Assume $X_j = 0$ if $j \le -m-1$

Then $X_t = \sum_{j=-m}^{t} (X_j - X_{j-1})$

$= (W_t + \theta W_{t-1}) + (W_{t-1} + \theta W_{t-2}) + \cdots + (W_{-m} + \theta W_{-m-1})$

$= W_t + (1+\theta) W_{t-1} + (1+\theta) W_{t-2} + \cdots + (1+\theta) W_{-m} + \theta W_{-m-1}$

$Var(X_t) = [1 + \theta^2 + (t+m)(1+\theta)^2]\sigma^2$ ↑ Increasing variance.

$Corr(X_{t+k}, X_t) \approx \sqrt{\frac{t+m-k}{t+m}} \to 1$ For large t ↑ non-decay ACF.

(Check!)

③ ARIMA(p,d,0) (no MA in ARIMA)

we call it Autoregressive Integral. (ARI(p,d))

· ARI(1,1)

$$(X_t - X_{t-1}) - \phi(X_{t-1} - X_{t-2}) = W_t$$

$$\Rightarrow X_t = (1+\phi)X_{t-1} - \phi X_{t-2} + W_t$$

$$X_t - (1+\phi)X_{t-1} + \phi X_{t-2} = W_t$$

$$\tilde{\phi} = 1 - (1+\phi)z + \phi z^2 = (1-z)(1-\phi z)$$

root $= 1$: Implication.

II. Model Specification

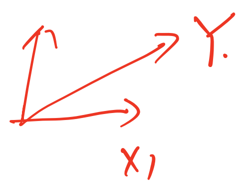
Q. Given the data, what model to fit it?

Excursion:

$Y \sim X_1 \Rightarrow R_1 = 0.1$

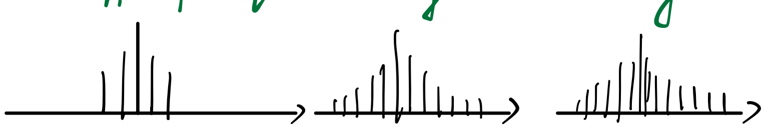
$Y \sim X_2 \Rightarrow R_2 = 0.2$

$Y \sim X_1 + X_2 \Rightarrow R_3$? A: $0.2 \le R_3 \le 1$.



1 ACF or Sample ACF.

MA(q)	AR(p)	ARMA(p, q)
cut off after q.	decay	decay.



Def of Sampled ACF

$$\hat{\gamma}(h) = \frac{1}{n}\sum_{t=1}^{n-|h|}(X_{t+|h|} - \bar{X})(X_t - \bar{X})$$

ACF: $\hat{\rho}(h) = \dfrac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$

Thm If $X_t = \sum_j \psi_j W_{t-j}$ stationary with $E|W_t|^4 < \infty$

Then $\begin{pmatrix} \hat\rho(1) \\ \vdots \\ \hat\rho(k) \end{pmatrix} \xrightarrow{\text{Asymptotic}} N\left(\begin{pmatrix} \rho(1) \\ \vdots \\ \rho(k) \end{pmatrix}, \frac{1}{n} V \right)$ as $n \to \infty$

where $V_{ij} = \sum_{h=1}^{\infty} (\rho(i+h) + \rho(i-h) - 2\rho(i)\rho(h))(\rho(j+h) + \rho(j-h) - 2\rho(j)\rho(h))$

(Bartlett's Formula)

Confidence Interval $\Leftrightarrow$ Hypothesis testing.

① White Noise

$$\begin{pmatrix} \hat\rho(1) \\ \vdots \\ \hat\rho(k) \end{pmatrix} \xrightarrow{\text{Asymptotic}} N\left(\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \frac{1}{n} I \right) \text{ as } n \to \infty$$

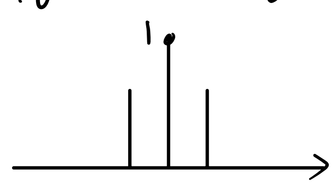
H_0: X_t is white noise $\hat\rho(h) \sim N(0, \frac{1}{n})$ (In most case)

Reject H_0 if $|\hat\rho(h)| > \frac{Z_{0.025}}{\sqrt{n}} \approx \frac{1.96}{\sqrt{n}}$ (pay attention to $\rho(1)$) ✓

② MA(1) $X_t = W_t + \theta W_{t-1}$ $\rho(1) = \frac{|\theta|}{1+\theta^2}$, $|\rho(1)| \le \frac{1}{2}$

$V_{11} = 1 - \rho^2(1) + 4\rho^4(1)$ (Check)

$V_{22} = 1 + 2\rho^2(1)$



Reject Region:

$h=1$: $|\hat\rho(h) - \rho(1)| \ge \frac{\sqrt{V_{11}}}{\sqrt{n}} Z_{0.025}$ (pay attention to $\rho(1)$ in most case)

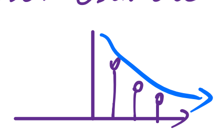
$h>1$: $|\hat\rho(h) - \rho(h)| = |\hat\rho(h)| > \frac{\sqrt{V_{hh}}}{\sqrt{n}} \times Z_{0.025}$

③ AR(1): $X_t - \phi X_{t-1} = W_t$. $\rho(h) = \phi^h$

$V_{ii} \overset{\approx}{=} \frac{(1+\phi^2)(1-\phi^{2i})}{1-\phi^2} \approx \frac{1+\phi^2}{1-\phi^2}$ for large i (Check)

$V_{11} = 1 - \phi^2$

Sequential test: estimate the paras first



④ MA(q) $V_{kk} = 1 + 2\sum_{i=1}^{q} \rho(i)^2$

Lecture 9. I. Partial ACF II. Parameter estimation III. MLE

I. Partial ACF. Generally we assume $\{X_t\}$ is stationary

1.1 Motivation: $X_t \sim AR(1)$ $X_t - \phi X_{t-1} = W_t$

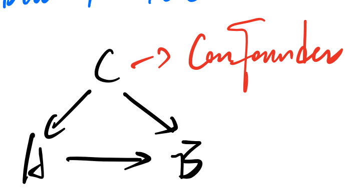
want to see $Cov(X_t, X_{t-2})$

$$
\begin{aligned}
Cov(X_t, X_{t-2}) &= Cov(\phi X_{t-1} + W_t, X_{t-2}) \\
&= Cov(\phi^2 X_{t-2} + \phi W_{t-1} + W_t, X_{t-2}) \\
&= \phi^2 \gamma(0)
\end{aligned}
$$

Therefore X_t depends on X_{t-2} because X_{t-1}.

But sometimes we want to see correlation btw X_t, X_{t-2} without X_{t-1}

· Remove Confounder: $\begin{cases} \text{Blocking} \\ \text{Fixing - residual} \end{cases}$



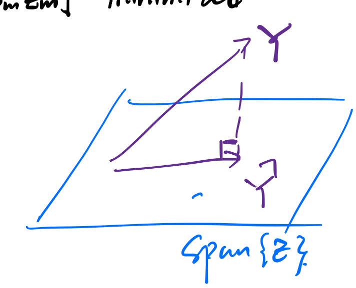
1.2 Best Linear Prediction

Let $Y, Z_1, \cdots Z_n$ r.v. we want to find:

$\beta_1, \cdots \beta_m$ s.t. $E[Y - \beta_1 Z_1 - \beta_2 Z_2 - \cdots - \beta_m Z_m]^2$ minimized

$$\frac{\partial E[Y - Z^T\beta]^2}{\partial \beta} = 2E ZZ^T\beta - 2E YZ = 0$$

$$\Rightarrow E ZZ^T \beta = E YZ$$



Example: $Y = X_{n+1}$, $Z_1 = X_n$, $Z_2 = 1$

$\beta_1 = \rho = Cov(X_{n+1}, X_n)$ $\beta_2 = (1-\rho)\mu$

① $Proj(Z_i | Z_1 \cdots Z_n) = Z_i$

② $Y - Proj(Y | Z_1 \cdots Z_n) \perp span(Z)$

Def PACF: $\phi_{hh} := Cov(X_{t+h} - X_{t+h}^{h-1}, X_t - \tilde{X}_t^{h-1})$

where $X_n^h = Proj(X_n | X_{n-1}, \cdots X_{n-h+1}) = \beta_{h1} X_{n-1} + \cdots + \beta_{hh} X_{n-1}$

$\tilde{X}_t^{h-1} = Proj(X_t | X_{t+1}, X_{t+2}, \cdots X_{t+h-1})$ (Interesting)

$$X_{t+h}^{h-1} = \mathrm{Proj}(X_{t+h} \mid X_{t+h-1}, X_{t+h-2}, \cdots X_{t+1})$$ (Decreasing)

Projection In expectation manner.

If $X_{t+h}^{h-1} = \beta_{h-1,1} X_{t+h-1} + \beta_{h-1,2} X_{t+h-2} + \cdots + \beta_{h-1,h-1} X_{t+1}$

Then $\tilde{X}_{t+h}^{h-1} = \beta_{h-1,1} X_{t+1} + \beta_{h-1,2} X_{t+2} + \cdots + \beta_{h-1,h-1} X_{t+h-1}$

The same coefficient if mean $=0$ and stationary.

Prof: Best linear prediction for Y given $Z = (Z_1 \cdots Z_m)^T$ is obtained by $\mathbb{E}(ZZ^T)\beta = \mathbb{E}YZ$

Crucial fact: ① No matter what n, $\beta_{h1}, \cdots \beta_{hh}$ are the same

② $\phi_{hh} = \beta_{hh}$ PACF = lm coefficient.

Last coefficient of X_{t+h+1}^{h}

Example:

① AR(p) $X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + W_t$

$$X_t^h = \mathrm{Proj}(X_t \mid X_{t-1}, \cdots X_{t-h})$$

$$= \mathrm{Proj}\Big(\sum_{j=1}^{p} \phi_j X_{t-j} \mid X_{t-1}, \cdots X_{t-h}\Big)$$

$$= \sum_{j=1}^{p} \phi_j \mathrm{Proj}(X_{t-j} \mid X_{t-1}, \cdots X_{t-h})$$

$$= \sum_{j=1}^{p} \phi_j X_{t-j} \quad \text{if } h \geq p$$

If $h<p$: $X_t^h = \sum_{j=1}^{p} \phi_j \mathrm{Proj}(X_{t-j} \mid X_{t-1}, \cdots X_{t-h})$

$$= \sum_{j=1}^{h} \phi_j X_{t-j} + \sum_{j=h+1}^{p} \mathrm{Proj}(X_{t-j} \mid X_{t-1}, \cdots X_{t-h})$$

$$\Rightarrow \phi_{hh} = \begin{cases} ? & \text{if } |h| < p \\ \beta_{hh} = \phi_h & \text{if } h = p \\ 0 & \text{if } h > p \end{cases}$$

② MA(q) $X_t = \sum_{j=1}^{q} \theta_j W_{t-j} + W_t$

· If Invertible:

$$X_t = \sum_{i=1}^{\infty} \pi_i X_{t-i} + W_t$$

$$X_t^h = \mathrm{Proj}(X_t \mid X_{t-1}, \cdots X_{t-h})$$

$$= \mathrm{Proj}\left(\sum_{j=1}^{\infty} \pi_i X_{t-i} + W_t \mid X_{t-1}, \cdots X_{t-h}\right)$$

$$= \sum_{j=1}^{h} \pi_i X_{t-i} + \sum_{i=h+1}^{\infty} \pi_i \mathrm{Proj}(X_{t-i} \mid X_{t-1}, \cdots X_{t-h})$$

usually $\phi_{hh} = \beta_{hh} \neq 0$

Summary:

	AR(p)	MA(q)	ARMA(p,q)
ACF	decay	cut off at q	decay
PACF	cut off at p	decay	decay

II. How to compute ACF.

Recall: Best Linear Prediction for Y given $Z = (Z_1 \cdots Z_m)^T$.

Yuler-Walker's equation:

$$(E ZZ^T)\beta = E YZ \Leftrightarrow \Gamma_h \beta_h = \gamma_h$$

In our case $Z = \begin{pmatrix} X_{t-h} \\ \vdots \\ X_t \end{pmatrix} \Rightarrow \Gamma_h = E(ZZ^T) = \begin{pmatrix} \gamma(0), & \gamma(1) & \cdots & \gamma(h-1) \\ \gamma(1) & \gamma(0) & \ddots & \\ \vdots & & & \\ \gamma(h-1) & & & \gamma(0) \end{pmatrix}$

If solve linear system

$\Rightarrow$ Complexity $O(h^3)$

$$\gamma_h = E YZ = (\gamma(1) \cdots \gamma(h))^T$$

1.1. Durbin - Levinson.

$\beta_h = (\beta_{h1} \cdots \beta_{hh})^T \qquad \tilde{\beta}_h = (\beta_{hh}, \cdots \beta_{h1})^T$

$\gamma_h = (\gamma(1), \cdots \gamma(h)) \qquad \tilde{\gamma}_h = (\gamma(h), \cdots \gamma(1))^T$

Given β_{h-1}, $\beta_h = \begin{pmatrix} \beta_{h-1} - \beta_{hh} \tilde{\beta}_{h-1} \\ \beta_{hh} \end{pmatrix} \qquad \beta_{hh} = \dfrac{\gamma(h) - \beta_{h-1}^T \cdot \tilde{\gamma}_{h-1}}{\gamma(0) - \beta_{h-1}^T \cdot \gamma_{h-1}}$

Complexity: $O(n^2)$

Example: AR(2)

$$\beta_1 = \beta_{11} = \phi_{11} = \rho(1) = \frac{\gamma(1)}{\gamma(0)}$$

$$\beta_{22} = \frac{\gamma(2) - \rho(1)\gamma(1)}{\gamma(0) - \rho(1)\gamma(1)} = \frac{\rho(2) - \rho(1)^2}{1 - \rho^2(1)}$$

$$\beta_2 = \begin{pmatrix} \rho(1) - \beta_{22}\rho(1) \\ \beta_{22} \end{pmatrix}$$

Lecture 10:

I: Parameter Estimation

II: Methods for estimation: MLE | LSE | Moment.

Roadmap:

① Plot: trend, Seasonality. outlier:

② Data transformation: → residual:

a) Differencing.

b) Nonlinear: log. sqrt.

③ Forecasting, inverting transformation.

Model Fitting:

① Plot ACF, PACF → serval possible (p, q)

② For each (p, q) fit ARMA(p, q)

③ Compare fitted models by metrics like MSE. AIC. BIC

Assumptions: ① known (p, q) ② mean 0. (if not, demean)

II. Method for parameter estimation.

1. MLE. $\{X_t\}$. Stationary, Gaussian

there is $\phi(B)X_t = \theta(B)W_t$ w/ $W_t \overset{iid}{\sim} N(0,\sigma^2)$

$$f(\theta,\phi,\sigma^2) = \frac{1}{(\sqrt{2\pi})^n |\Gamma_n|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}X^T\Gamma_n^{-1}X\right)$$

$$X = (X_1 \cdots X_n)^T, \quad \Gamma_n = \begin{pmatrix} \gamma(0) & \cdots & \gamma(n-1) \\ \vdots & \ddots & \\ \gamma(n-1) & & \gamma(0) \end{pmatrix}$$

Advantages:

① Consistency. MLE $\to$ true value

② Efficiency. reach CR lower bound

③ Even not Gaussian. asymptotically Gaussian.

Disadvantage

① Hard to optimize

② need a good initial point.

Examples of MLE: AR(1) (is a Markov Chain)

$E X_t = \mu$, $X_t - \mu = \phi(X_{t-1} - \mu) + W_t$. $W_t \overset{iid}{\sim} N(0,\sigma^2)$

Observe $X_1, \cdots X_n \xrightarrow{\text{estimate}} \phi, \mu, \sigma^2$

① Given X_1:

$$f(X_2 \cdots X_n | X_1) = f(X_n | X_{n-1}, \cdots X_1) f(X_{n-1} | X_{n-2}, \cdots X_1) \cdots f(X_2 | X_1)$$
$$= f(X_n | X_{n-1}) f(X_{n-1} | X_{n-2}) \cdots f(X_2 | X_1)$$

$X_n | X_{n-1} \sim N(\mu + \phi(X_{t-1} - \mu), \sigma^2)$

$$f(X_2,\cdots X_n|X_1) = (\sqrt{2\pi\sigma^2})^{1-n} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=2}^{n}(X_i-\mu-\phi(X_{i-1}-\mu))^2\right\}$$

② $Var(X_t) = \frac{\sigma^2}{1-\phi^2} \quad X_1 \sim N(\mu, \frac{\sigma^2}{1-\phi^2})$

③ ①+②: $f(X_1,\cdots X_n) = f(X_1)\cdot f(X_2,\cdots X_n|X_1)$

$$= \frac{\sqrt{1-\phi^2}}{(\sqrt{2\pi\sigma^2})^n} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=2}^{n}(X_i-\mu-\phi(X_{i-1}-\mu))^2 + (1-\phi^2)(X_i-\mu)^2\right)$$

$\ell(\mu,\phi,\sigma^2) = \log f(X_1,\cdots X_n)$

$\frac{\partial \ell}{\partial \sigma^2} = 0 \Rightarrow \hat{\sigma}^2 = \frac{S(\hat{\phi},\hat{\mu})}{n}$ $\quad \hat{\phi},\hat{\mu}$ satisfies $\begin{cases} \frac{\partial \ell}{\partial \mu} = 0 \\ \frac{\partial \ell}{\partial \phi} = 0 \end{cases}$

$S(\phi,\mu) = \sum_{i=1}^{n}(X_i-\mu-\phi(X_i-\mu))^2 + (1-\phi^2)(X_i-\mu)^2$

2. Lease Square · minimize $\sum_{i=2}^{n} W_i^2$

AR(1): $\frac{\partial S_c}{\partial \mu} = 0 \Rightarrow \hat{\mu} = \bar{X}$

$\frac{\partial S_c}{\partial \phi} = 0 \Rightarrow \hat{\phi} = \rho(1)$

3. Moment Method: (matching)

the moment $\overset{set}{=}$ sampled moment.

$\mu = \bar{X}. \quad \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$

Example AR(1)

$\rho(k) = |\phi|^k$

$\rho(1) = \hat{\rho}(1) \Rightarrow \hat{\phi} = \hat{\rho}(1)$ Sample ACF.

Example AR(p)

Yule Walker $\Gamma_n \beta_n = \gamma_n \Rightarrow h=p, \ \beta_p = (\phi_1, \cdots \phi_p)^T$

$\hat{\Gamma}_p \beta_p = \hat{\gamma}_p$ $\quad\downarrow\quad$ Stands only when $h=p$

in AR. equivalent to MLE

Lecture 11. 4/7/2023

EDA (Exploratory Data Analysis)
John Tukey

I. Roadmap

Given a dataset of time series.

① Plot the data $\begin{cases} \text{trend} \\ \text{seasonality} \\ \text{outlier} \\ \text{stepchange} \end{cases}$ → need transformation for stationary.

② Data transformation. $\begin{cases} \text{Differencing} \\ \text{nonlinear} \begin{cases} \log \\ \sqrt{\ } \end{cases} \\ \text{Smoothing} \end{cases}$

③ Identify hyper-parameters p. q by acf or pacf.

⇒ several candidate pairs

ARMA(p, q). $\phi(B)$. $\theta(B)$. ACF s.t.

$$\frac{\rho(h)}{\rho(h-1)} = 0.3,\ h \geq 3,\quad \frac{\rho(2)}{\rho(1)} \neq 0.3$$

$$\rho(h) = 0.3\,\rho(h-1)$$

$$\gamma(h) = 0.3\,\gamma(h-1)$$

$$\gamma(h) - \phi_1\gamma(h-1) - \cdots - \phi_p\gamma(h-p) = \begin{cases} \sigma^2 \sum_{j=0}^{q-h} \theta_{h+j}\psi_j & h \leq q \\ 0, & h > q \end{cases}$$

$\phi_1 = 0.3$. $\phi_j = 0$ $j \geq 2$

相当数目>p的线性方程.
有解则解唯一

④ estimate params for each (p, q)

Yuler-Walker equation $\hat{\Gamma}_n \hat{\beta}_n = \hat{\gamma}_n$

has value $\phi_i = \beta_i$ when $h = p$

⑤ diagnostic

⑥ Model selection: choose p, q.

Distribution of MLE

ARMA(p, q) $\phi(B) X_t = \theta(B) W_t$. $\underline{W_t \sim WN(0, \sigma^2)}$ $\phi(\,)\ \theta(\,)$ cannot uniquely specify model.

$\downarrow$ MLE

$\hat{\phi}_{MLE}\ \hat{\theta}_{MLE}$

$$\phi(B) Z_t = W_t \rightarrow \Gamma_{\phi\phi} = \begin{pmatrix} \gamma_\phi(0) & & \gamma_\phi(p-1) \\ & \ddots & \\ \gamma_\phi(p-1) & & \gamma_\phi(0) \end{pmatrix}$$

$$\theta(B) Y_t = W_t \rightarrow \Gamma_{\theta\theta}$$

$$\text{Then } \begin{pmatrix} \hat{\phi}_{MLE} \\ \hat{\theta}_{MLE} \end{pmatrix} \sim AN\left(\begin{pmatrix} \phi \\ \theta \end{pmatrix}, \frac{\sigma^2}{n} \begin{pmatrix} \Gamma_{\phi\phi} & \Gamma_{\phi\theta} \\ \Gamma_{\theta\phi} & \Gamma_{\theta\theta} \end{pmatrix}^{-1} \right)$$

II. Forecast. Given $\hat{\phi}, \hat{\theta}$ for an ARMA(p, q)

$$\text{Compute } \hat{\Gamma}_n, \hat{\gamma}_n, \Gamma_n = \begin{pmatrix} \gamma(0) & \cdots & \gamma(h-1) \\ \vdots & \ddots & \\ \gamma(h-1) & & \gamma(0) \end{pmatrix}$$

Yuler-Walker $\Rightarrow \hat{\beta}_n = \hat{\Gamma}_n^{-1} \hat{\gamma}_n$

Forecast. $X_t^h = \hat{\beta}_{h1} X_{t-1} + \cdots + \hat{\beta}_{hh} X_{t-h}$

Exercise. Check $P_t^h = E[X_t - X_t^h]^2 = \gamma(0) - \gamma(h)^T \Gamma_n^{-1} \gamma(h)$

7.7 Diagnostic: goodness of fit.

Time Series = trend + Seasonality + Stationary + WN.

1 Good Fit: WN has iid, mean 0, WN.

① plot residuals: any pattern?

② Check acf, pacf: if decay slowly ⇒ may not stationary

③ Stationary check: Dicker-Fuller: unit-root test:

ARMA stationary iff $\phi(z)$ has no root $=1$

④ Mean 0 check: ANOVA

⑤ equal variance check: F-test

⑥ If residuals are uncorrelated:

- Ljung-Box Pierce test test uncorrelated

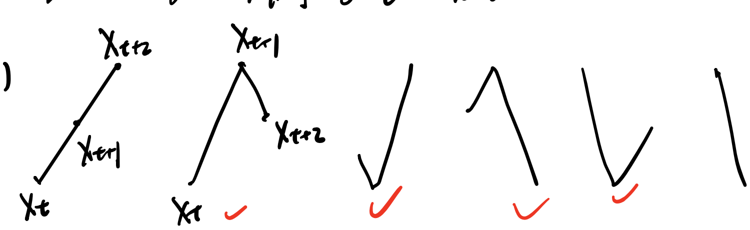
$r_e^2(h) = \sum_{t=h+1}^{n} (X_t - X_t^h)^2 \quad h = 1, \dots, H$ usually $H = 20$

$Q = n(n+2) \sum_{h=1}^{H} \frac{r_e^2(h)}{n-h}$ $H_0: \rho(1) = \rho(2) = \cdots = \rho(H) = 0$

$Q \sim \chi^2(H-p-q)$ Before: check mean 0 and equal variance.

- Turning-point test: test $\{X_t\}$ uncorrelation

(X_t, X_{t+1}, X_{t+2})



$T = \#(\text{turning points})$

$E(T) = \frac{2}{3}(n-2)$ $T \overset{asy}{\sim} N(\frac{2}{3}n, \frac{8}{45}n)$

- Difference-signed test: uncorrelation

  $N = \#\{t: X_{t+1} > X_t\}$ $\quad E[N] = \frac{n-1}{2}$ $\quad N \sim AN(\frac{n}{2}, \frac{n}{12})$

- Rank test: $Z = \#\{(i,j): X_i > X_j\}$

  Ex: $EZ = ?$ $\quad Z \sim AN(\frac{n^2}{4}, \frac{n^3}{36})$

⑦ Normal : Q-Q plot

Lecture 12 & 13 ⅠⅠ Model selection ⅠⅡ Forecasting.

Ⅰ. Model selection:

1. AIC, BIC (Information Criteria)

Def $AIC = -2\log ML + 2k$

$BIC = -2\log ML + k\log n$. Heavier penalty.

ML: maximum likelihood.

n: sample size, $k = p+q$ (number of parameters).

2. Overfitting as a selection tool.

Given a data, suppose AR(1) can fit it well.

If we use AR(2) to fit it, we will observe:

① $\hat{\phi}_2 \approx 0$

② $\hat{\phi}_1$ does not change much.

Stepwise selection: Sequentially fit AR(1), AR(2) ⋯ AR(k) until we observe the above property.

Remark: When using overfitting as a selection tool, we need to be as simple as possible.

For example: $AR(1) \not\to ARMA(2,1)$

Reason: 若 $\{X_t\} \sim AR(1)$ $X_t - \phi X_{t-1} = W_t$

Then for any ϕ_2: $(1-\phi_2 B)(X_t - \phi_1 X_{t-1}) = (1-\phi_2)W_t$

Exercise: $AR(1) \to AR(2)$. Why not $ARMA(1,1)$?

$X_t - \phi_1 X_{t-1} = W_t + \theta_1 W_{t-1}$

Remark: when we use overfitting as a tool, we always have find a satisfying (p, q)

3. Cross-validation: Data → training | validation | testing

- k-fold CV: for limited data

① CV for hyperparameters → fit on all data → overfitting risk

② residuals are correlated between folds

- In time series:

only care about future observations. For test

0	1	2	3	4

① 0 train, 1 val

② 0, 1 training, 2 val

③ 0, 1, 2 training, 3 val.

VI. Forecasting $\begin{cases} \text{forecast future } X. \\ \text{evaluation.} \end{cases}$

① Best linear prediction: $h \in \mathcal{L}$

$E(X_t - h(X_{t-1}, \cdots X_{t-n}))^2$ minimized

$X_t^h = \beta_{n1} X_{t-1} + \beta_{n2} X_{t-2} + \cdots + \beta_{nh} X_{t-h}$

Yule-Walker $\Gamma_h \beta_h = \gamma_h$

$P_t^h = E(X_t - X_t^h)^2 = \gamma(0) - \gamma_h^T \Gamma_h^{-1} \gamma_h$

$X_t^h \pm Z_{1-\alpha/2} \sqrt{P_h^t}$

②. Best prediction: f is any function.

$E(X_t - f(X_{t-1}, \cdots, X_{t-n}))^2$ minimized:

$\Rightarrow f(X_{t-1}, \cdots X_{t-n}) = E(X_t \mid X_{t-1}, \cdots X_{t-n})$

In general, we consider:

Filtration

$E(X_t \mid X_{t-1}, \cdots X_1) = E(X_t \mid \mathcal{F}_{t-1})$

Best predictor for X_{t+h} is still:

$E(X_{t+h} \mid X_1, \cdots X_t)$

Example

① Deterministic model $X_t = \mu_t + W_t \sim WN(0, \sigma^2)$

$\hat{X}_t(1) = E(X_{t+1} \mid \mathcal{F}_t) = \mu_{t+1}$

$err_t(1) = W_{t+1}$ $\quad E(err_t(1)) = 0.$ $\quad Var(err_t(1)) = \sigma^2.$

unbiased.

$\mathcal{F}_t = \sigma(X_t, W_t)$

② MA(1). $X_t = W_t + \theta W_{t-1}$, $W_t \sim WN(0, \sigma^2)$ we may assume $W_0 = 0$.

$\hat{X}_t(1) = E(X_{t+1} \mid \mathcal{F}_t) = E(W_{t+1} + \theta_1 W_t \mid \mathcal{F}_t) = \theta_1 W_t$

$err_t(1) = X_{t+1} - \hat{X}_t(1) = W_{t+1}$. $\quad E(err_t(h)) = 0$ $\quad$ unbiased

$X_t = W_t + \theta W_{t-1}$

$X_{t-1} = W_{t-1} + \theta W_{t-2}$ $\cdots$ $X_1 = W_1 + \theta W_0 = W_1$

$Var(err_t(1)) = \sigma^2$ $Var(err_t(2)) = (1+\theta_1^2)\sigma^2$

$\hat{X}_t(\ell) = \cdots$

③ AR(1) $X_t = \mu + \phi_1 X_{t-1} + W_t$ $W_t \sim WN(0, \sigma^2)$

$\hat{X}_t(1) = \hat{\mathbb{E}}[X_{t+1} | \mathcal{F}_t] = \mu + \phi_1 X_{t-1}$

$\hat{X}_t(2) = \hat{\mathbb{E}}[X_{t+2} | \mathcal{F}_t] = \mu + \phi_1 \hat{\mathbb{E}}[X_{t+1} | \mathcal{F}_t] = \mu + \phi_1\mu + \phi_1^2 X_{t-1}$

$\hat{X}_t(\ell) = \hat{\mathbb{E}}(X_{t+\ell} | \mathcal{F}_t] = \mu + \phi_1 \hat{X}_t(\ell-1) = \frac{\mu(1-\phi_1^\ell)}{1-\phi_1} + \phi_1^\ell X_t \to \frac{\mu}{1-\phi_1}$

Rmk: $\hat{\mathbb{E}}(X_t) = \frac{\mu}{1-\phi_1} = \lim_{\ell \to \infty} \hat{X}_t(\ell)$ Best prediction leads to mean.

④ ARMA (p, q)

(1). ARMA(1,1) $X_t = \mu + \phi_1 X_{t-1} + W_t + \theta_1 W_{t-1}$. $|\phi_1| < 1$, $\mathbb{E}(X_t) = \frac{\mu}{1-\phi_1}$

$\hat{X}_t(1) = \hat{\mathbb{E}}[\mu + \phi_1 X_t + W_{t+1} + \theta_1 W_t | \mathcal{F}_t]$

$= \mu + \phi_1 X_t + \theta_1 W_t$ (W_t can be written as $g(\mathcal{F}_t)$?).

$\hat{X}_t(2) = \hat{\mathbb{E}}[X_{t+2} | \mathcal{F}_t] = \mu + \phi_1 \hat{X}_t(1)$

$\hat{X}_t(\ell) = \mu + \frac{\mu(1-\phi_1^\ell)}{1-\phi_1} + \phi_1^\ell X_t + \theta_1 \phi^{\ell-1} W_t \longrightarrow \frac{\mu}{1-\phi_1} = \hat{\mathbb{E}}(X_t)$.

$e_t(1) = X_{t+1} - \hat{X}_t(1) = W_{t+1}$ $Var(e_t(1)) = \sigma^2$.

$\Rightarrow$ Prediction Interval $\hat{X}_t(1) \pm Z_{\alpha/2}\sigma$.

(2) ARMA(p,q) $X_t = \mu + \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + W_t + \theta_1 W_{t-1} + \cdots + \theta_q W_{t-q}$

$\hat{X}_t(1) = \mu + \phi_1 X_t + \cdots + \phi_p X_{t+1-p} + 0 + \theta_1 W_t + \cdots + \theta_q W_{t+1-q}$.

$\hat{X}_t(\ell) = \mu + \sum_{i=1}^{p} \phi_i \hat{X}_t(\ell-i) + \sum_{j=1}^{q} \theta_j W_t(\ell-j)$

we define $\hat{X}_t(\ell-i) = X_{t+\ell-i}$ if $\ell-i \le 0$

$$W_t(\ell-j) = \begin{cases} 0 & \text{if } \ell-j > 0 \\ W_{t+\ell-j} & \text{if } \ell-j \le 0 \end{cases}$$

$e_t(\ell) = X_{t+\ell} - \hat{X}_t(\ell) = \sum_{i=1}^{\min(\ell,p)-1} \phi_i e_t(i) + \sum_{j=0}^{\min(\ell,q)-1} W_{t+\ell-j} \uparrow$ as $\ell \uparrow$?

$Var(e_t(\ell)) < \infty$ for $\ell \to \infty$. Suppose $\{X_t\}$ is causal

$X_t = \mu + \sum_{j=0}^{\infty} \psi_j W_{t-j}$ $(\psi_0 = 1)$

$e_t(\ell) = W_{t+\ell} + \psi_1 W_{t+\ell-1} + \cdots + \psi_{\ell-1} W_{t+1}$

$Var(e_t(\ell)) = \sum_{j=0}^{\infty} \psi_j^2 < \infty$

Prediction interval: $\hat{X}_t(\ell) \pm Z_{\alpha/2} \sqrt{Var(e_t(\ell))}$ (Assume Gaussian)

3) Transformed time series.

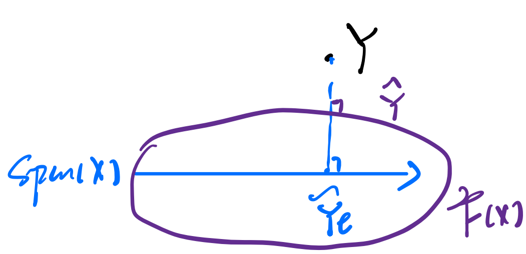
Exercise: ① ARIMA(0,1,1) $\Delta X_t = W_t + \theta_1 W_{t-1}$

$X_t = W_t + \theta_1 W_{t-1} + X_{t-1}$

$\hat{X}_t(1) = \hat{E}[X_{t+1} | \mathcal{F}_t] = X_t + \theta_1 W_t$

$\hat{X}_t(2) = \hat{E}[X_{t+2} | \mathcal{F}_t] = \hat{E}[X_{t+1} | \mathcal{F}_t] = X_t + \theta_1 W_t$

$\hat{X}_t(\ell) = \hat{E}[X_{t+\ell} | \mathcal{F}_t] = X_t + \theta_1 W_t$ for all ℓ



Exercise 2. Log transformation $\quad \mathbb{E}(\exp(W_t)) = \exp(\frac{\sigma^2}{2})$

$Z_t = \phi_1 Z_{t-1} + \cdots + \phi_p Z_{t-p} + W_t, \quad W_t \sim N(0, \sigma^2)$

$Z_t = \log(X_t) \quad X_t = \exp(Z_t) \quad C_t(\ell) = \phi_1 Z_{t+\ell-1} + \cdots + \phi_p Z_{t+\ell-p}$

$\hat{X}_t(1) = \mathbb{E}[X_{t+1} | \mathcal{F}_t] = \exp\{\phi_1 Z_{t-1} + \cdots + \phi_p Z_{t-p}\}$

$$\begin{aligned}\hat{X}_t(\ell) &= \mathbb{E}[X_{t+\ell} | \mathcal{F}_t] = \mathbb{E}[\exp\{Z_{t+\ell}\} | \mathcal{F}_t] \\ &= \mathbb{E}[\exp(C_t(\ell)) | \mathcal{F}_t] \cdot \mathbb{E}[\exp(W_{t+\ell})] \\ &= \mathbb{E}[\exp(C_t(\ell)) | \mathcal{F}_t] \exp(\tfrac{\sigma^2}{2}) \quad \begin{cases} \ell > p \\ \ell < p \end{cases}\end{aligned}$$

Exercise 3. $Z_t = \phi_0 + Z_{t-1} + \phi_2 Z_{t-2} + \phi_3 Z_{t-3} + W_t \quad W_t \sim N(0, \sigma^2)$

$Z_t = \log(X_t)$

$$\begin{aligned}\hat{X}_t(2) &= \mathbb{E}[X_{t+2} | \mathcal{F}_t] = \mathbb{E}[\exp(Z_{t+2}) | \mathcal{F}_t] \\ &= \mathbb{E}[\exp(\phi_0 + Z_{t+1} + \phi_2 Z_t + \phi_3 Z_{t-1} + W_{t+2}) | \mathcal{F}_t] \\ &= \exp(\phi_0 + \tfrac{\sigma^2}{2}) \exp(\phi_2 Z_t + \phi_3 Z_{t-1}) \underbrace{\mathbb{E}[\exp(Z_{t+1}) | \mathcal{F}_t]}_{\hat{X}_t(1)}\end{aligned}$$

$\hat{X}_t(1) = \mathbb{E}[\exp(Z_{t+1}) | \mathcal{F}_t] = \exp(\phi_0 + \frac{\sigma^2}{2}) \exp\{Z_t + \phi_2 Z_{t-1} + \phi_3 Z_{t-2}\}$.

Lecture 14. (II) Multiplicative Seasonal ARMA model.

(1) Pure Seasonal ARMA model.

Def. For $P, Q \in \mathbb{N}$, $S \in \mathbb{N}^+$, we say $\{X_t\}$ is an $ARMA(P,Q)_S$ process iff $\Phi(B^S) X_t = \Theta(B^S) W_t$.
$W_t \sim WN(0, \sigma^2)$. $\Phi(B^S) = 1 - \sum_{i=1}^{P} \Phi_i B^{Si}$, $\Theta(B^S) = 1 + \sum_{j=1}^{Q} \Theta_j B^{Sj}$

① $ARMA(P,Q)_S$ stationary iff roots of $\Phi(z^S)$ are not on the unit circle

② Causal iff roots of $\Phi(z^S)$ are outside the unit circle.

③ Invertible iff ... $\Theta(B^S)$ .. outside . — —

Example. $MA(1)_{12}$

$P=0$, $Q=1$, $S=12$.

$X_t = W_t + \Theta_1 W_{t-12}$

$\gamma(0) = (1+\Theta_1^2)\sigma^2$

$\gamma(12) = \Theta_1 \sigma^2$

$\gamma(h) = 0$. $h \neq 0, 12$

Example ($AR(1)_{12}$)

$P=1$, $Q=0$, $S=12$

$X_t = \Phi_1 X_{t-12} + W_t$

$\gamma(0) = \dfrac{\sigma^2}{1-\Phi^2}$

$\gamma(12i) = \dfrac{\sigma^2 \Phi^i}{1-\Phi^2}$

$\gamma(h) = 0$ otherwise

Summary.

Model	ACF(h)	PACF(h)	h = is
$AR(P)_S$	decay	0 for $i > P$	
$MA(Q)_S$	0 for $i > Q$	decay	
$ARMA(P,Q)_S$	decay.	decay	

Seasonal + nonseasonal

(1.) $X_t = W_t + \theta_1 W_{t-1} + \Theta_1 W_{t-4}$

(2.) $X_t = \phi_1 X_{t-1} + \Phi_1 X_{t-4} + W_t$

Exercise:

(3). $X_t = W_t - \theta_1 W_{t-1} - \Theta W_{t-12} + \theta\Theta W_{t-13}$

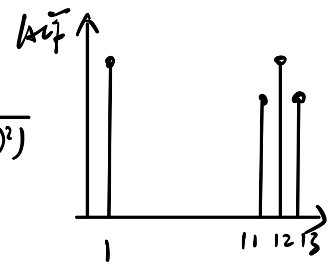
$\gamma(0) = (1+\theta_1^2+\Theta^2+\theta^2\Theta^2)\sigma^2 = (1+\theta^2)(1+\Theta^2)\sigma^2$

$\gamma(1) = -\theta_1\sigma^2 - \theta\Theta^2 W_{t-13}$ $\quad \rho(1) = \frac{-\theta}{1+\theta^2}$

$\gamma(11) = \theta_1\Theta\sigma^2$ $\quad \rho(11) = \rho(13) = \frac{\theta\Theta}{(1+\theta^2)(1+\Theta^2)}$

$\gamma(12) = -\Theta\sigma^2$ $\quad \rho(12) = \frac{-\Theta}{1+\Theta^2}$

$\gamma(13) = \theta\Theta\sigma^2$



② Multiplicative Seasonal Models

Def. $P, Q, p, q \geq 0$, $s>0$. we say $\{X_t\}$ is a multiplicative seasonal ARMA process $ARMA(p,q)\times(P,Q)_s$ if $\Phi(B^s)\phi(B)X_t = \Theta(B^s)\theta(B)W_t$

Additionally $D, d>0$. we say $\{X_t\}$ is a multiplicative seasonal ARIMA process $ARIMA(p,d,q)\times(P,D,Q)_s$ if $\Phi(B^s)\phi(B)\nabla_s^D\nabla^d X_t = \Theta(B^s)\theta(B)W_t$

where $\nabla_s^D = (1-B^s)^D$

How to choose paras:

① Differencing → stationary

② ACF, PACF to give potential candidates

Example $P=q=1,\ p=\Theta=0,\ s=12.$

$$X_t = \Phi X_{t-12} + W_t - \theta W_{t-1}$$

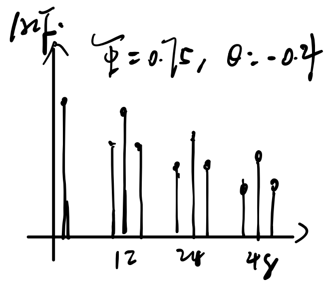
$$\gamma(0) = \frac{(1+\theta^2)\sigma^2}{1-\Phi^2} \qquad \gamma(1) = \Phi^2\gamma(1) - \theta\sigma^2 \Rightarrow \gamma(1) = \frac{-\theta\sigma^2}{1-\Phi^2}$$

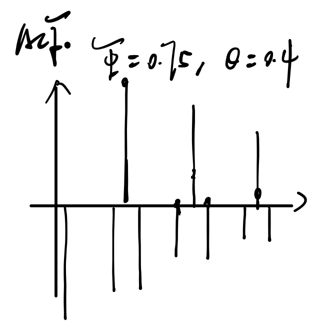
$$\rho(1) = -\frac{\theta}{1+\theta^2}$$

$$\gamma(12k) = \Phi^k\gamma(0)$$

$$\gamma(12k-1) = \Phi^k\gamma(1)$$

$$\gamma(12k+1) = \Phi^k\gamma(1)$$





Exercise: AR characteristic function. $(1-1.6x+0.7x^2)(1-0.8x^{12})$

① Is it stationary?

② identify the model as certain seasonal

Solution ①. Stationary. Because the roots are $z_1 = \frac{\sqrt{6}}{7}i$ $z_2 = -\frac{\sqrt{6}}{7}i$ $z_3 = 1.02$ $z_4 \approx -1.02$

② $s=12$ $\Phi(B^s)\phi(B)X_t = W_t$

$\Phi(B^s) = 1-0.8B^{12}$ $\phi(B) = 1-1.6B+0.7B^2$

Exercise: $\{X_t\}$ satisfies $X_t = a+bt+S_t+Z_t$

S_t deterministic and periodic with period s

and $\{Z_t\}$ is a seasonal ARIMA$(p,0,q)\times(P,1,Q)_s$

What is $Y_t = X_t - X_{t-s}$?

$$Y_t = a+bt+S_t+Z_t - a - b(t-s) - S_{t-s} - Z_{t-s}$$

$$= bs + Z_t - Z_{t-s}$$

$$\Phi(B^s)\phi(B)(1-B^s)Z_t = \Theta(B^s)\theta(B)W_t$$

$\Phi(B^s)\phi(B)(Z_t - Z_{t-s}) = \Theta(B^s)\theta(B)W_t$

$Z_t - Z_{t-s}$ is ARMA $(p,q)\times(P,Q)_s$ stationary.

Exercise: Identify the following multiplicative seasonal ARIMA

(a) $X_t = X_{t-1} + X_{t-12} - X_{t-13} + W_t - 0.5W_{t-1} - 0.5W_{t-1} - 0.5W_{t-2} + 0.25W_{t-3}$

(b) $X_t = 0.5X_{t-1} + X_{t-4} - 0.5X_{t-5} + W_t - 0.3W_{t-1}$

$\Rightarrow (X_t - X_{t-4}) - 0.5(X_{t-1} - X_{t-5}) = W_t - 0.3W_{t-1}$

(a) $X_t - X_{t-1} - X_{t-12} + X_{t-13} = W_t - 0.5W_{t-1} - 0.5W_{t-2} + 0.25W_{t-3}$

ARMA $(1,3)\times(1,0)_{12}$

(b) $(1-0.5B)(1-B^4)(X_t) = (1-0.3B)W_t$

ARIMA $(1,0,1)\times(0,1,0)_4$

Lecture 15:

① Forecast Seasonal models. Best prediction.

① $ARIMA(0,1,1)\times(1,0,1)_{12}$

$$X_t - X_{t-1} = \Phi(X_{t-12} - X_{t-13}) + W_t - \theta W_{t-1} - \Theta W_{t-12} + \theta\Theta W_{t-13}$$

$$\Rightarrow X_t = X_{t-1} + \Phi X_{t-12} - \Phi X_{t-13} + W_t - \theta W_{t-1} - \Theta W_{t-12} + \theta\Theta W_{t-13}$$

$$\hat{X}_t(1) = E(X_{t+1} \mid F_t) = X_t + \Phi X_{t-11} - \Phi X_{t-12} - \theta W_t - \Theta W_{t-11} + \theta\Theta W_{t-12}$$

$$\hat{X}_t(\ell) = \hat{X}_t(\ell-1) + \Phi \hat{X}_t(\ell-12) - \Phi \hat{X}_t(\ell-13) \quad \text{if } \ell > 13.$$

② $AR(1)_{12}$.

$$X_t = \Phi X_{t-12} + W_t, \quad W_t \sim WN(0, \sigma^2)$$

$$\hat{X}_t(\ell) = \begin{cases} \Phi \hat{X}_t(\ell-12), & \ell > 12 \\ \Phi X_{t+\ell-12}, & 1 \le \ell \le 12 \end{cases}$$

$$\frac{1-\Phi^{2k+2}}{1-\Phi^2}\sigma^2 \ ?$$

$$\hat{X}_t(\ell) = \Phi^{k+1} X_{t+r-11} \quad \text{where } \ell = 12k + r + 1, \quad 0 \le r \le 12$$

$$e_t(\ell) = X_{t+\ell} - \hat{X}_t(\ell)$$

Ex. Compute $Var(e_t(\ell))$. where $\ell = 12k + r + 1$

$$\gamma(0) = \Phi^2\gamma(0) + \sigma^2 \quad (1-\Phi^2)\gamma(0) = \sigma^2 \quad \gamma(0) = \frac{\sigma^2}{1-\Phi^2}$$

$$\gamma(12k) = Cov(\Phi^k X_t, X_t) = \Phi^k \gamma(0)$$

$$Var(e_t(\ell)) = \gamma(0) + \Phi^{2k+2}\gamma(0) - 2Cov(X_{t+\ell}, \Phi^{k+1} X_{t+r-11})$$

$$= (1+\Phi^{2k+2})\gamma(0) - 2\Phi^{k+1}\gamma(\ell - r + 11)$$

$$= (1+\Phi^{2k+2})\gamma(0) - \Phi^{k+1}\gamma(12(k+1))$$

$$= (1+\Phi^{2k+2})\gamma(0) - 2\Phi^{k+1}\Phi^{k+1}\gamma(0)$$

$$= \frac{1-\Phi^{2k+2}}{1-\Phi^2}\sigma^2$$

② $MA(1)_{12}$ $X_t = W_t + \Theta W_{t-12}$

$\tilde{X}_t(1) = \Theta W_{t-11}$

$$\tilde{X}_t(\ell) = \begin{cases} 0 & \ell > 12 \\ \Theta W_{t+\ell-12} & 1 \le \ell \le 12 \end{cases}$$

$e_t(\ell) = X_{t+\ell} - \tilde{X}_t(\ell)$ $\quad Var(e_t(\ell)) = \begin{cases} \sigma^2, & 1 \le \ell \le 12 \\ (1+\Theta^2)\sigma^2, & \ell > 12 \end{cases}$

Exercise: $X_t = \frac{1}{2} X_{t-4} + W_t - \theta_1 W_{t-1} - \theta_2 W_{t-2}$

① Find the first four ψ-weights for the model.

② Suppose $\theta_1 = 0.5$, $\theta_2 = -0.25$, $\sigma^2 = 1$

Find the forecast for next four quarters:

Sales	25	20	25	40
residual	2	1	2	3

③ Give 95% prediction interval

Answer:

① $(1 - \frac{1}{2}B^4) X_t = (1 - \theta_1 B - \theta_2 B^2) W_t$

$(1 - \frac{1}{2}B^4)(\psi_0 + \psi_1 B + \psi_2 B^2 + \cdots) W_t = (1 - \theta_1 B - \theta_2 B^2) W_t$

$\psi_0 = 1$ $\quad \psi_2 = -\theta_2$ $\quad \psi_1 = -\theta_1$ $\quad \psi_3 = 0$ $\quad \psi_4 - \frac{1}{2} = 0$ $\quad \psi_4 = \frac{1}{2}$

Another approach:

$X_t = \frac{1}{2} X_{t-4} + W_t - \theta_1 W_{t-1} - \theta_2 W_{t-2}$

$= \frac{1}{2}(\frac{1}{2} X_{t-8} + W_{t-4} - \theta_1 W_{t-5} - \theta_2 W_{t-6}) + W_t - \theta_1 W_{t-1} - \theta_2 W_{t-2}$.

$= W_t - \theta_1 W_{t-1} - \theta_2 W_{t-2} + \frac{1}{2} W_{t-4} - \frac{1}{2}\theta_1 W_{t-5} - \frac{1}{2}\theta_2 W_{t-6} + \frac{1}{4} X_{t-8}$

From the coefficients we can get:

$\psi_0 = 1$ $\quad \psi_1 = -\theta_1$ $\quad \psi_2 = -\theta_2$ $\quad \psi_3 = 0$. $\quad \psi_4 = \frac{1}{2}$

② $\tilde{E}(X_5 | \mathcal{F}_4) = \frac{1}{2}X_1 - \theta_1 W_4 - \theta_2 W_3 = 11.5$

$\tilde{E}(X_6 | \mathcal{F}_4) = \frac{1}{2}X_2 - \theta_2 W_4 = 10.75$

$\tilde{E}(X_7 | \mathcal{F}_4) = \frac{1}{2}X_3 = 12.5$

$\tilde{E}(X_8 | \mathcal{F}_4) = \frac{1}{2}X_4 = 20$

③ $e_t(1) = W_{t+1}$ $\quad Var(e_t(1)) = \sigma^2 = 1$

$e_t(2) = W_{t+2} - \theta_1 W_{t+1} \Rightarrow Var(e_t(2)) = (1+\theta_1^2)\sigma^2 = \frac{5}{4}$

$e_t(3) = X_{t+3} - \hat{X}_t(3) = W_{t+3} - \theta_1 W_{t+2} - \theta_2 W_{t+1}$

$Var(e_t(3)) = (1+\theta_1^2+\theta_2^2)\sigma^2 = \frac{21}{16}$

$e_t(4) = W_{t+4} - \theta_1 W_{t+3} - \theta_2 W_{t+2}$

$\Rightarrow Var(e_t(4)) = (1+\theta_1^2+\theta_2^2) = \frac{21}{16}$

The prediction interval for $\hat{X}_t(i)$ is $\hat{X}_t(i) \pm Z_{\alpha/2}\sqrt{Var(e_t(i))}$

(4) Models for Seasonality (Seasonality)

$X_t = \underset{\sim}{S_t} + \underset{\sim}{Z_t}$ stationary

strong seasonality: ($S_{t+d} = S_t$) d is the period

① Parameter Model:

$$S_t = a_0 + \sum_{f=1}^{k} [a_f \sin(\frac{2\pi f}{d}t) + b_f \cos(\frac{2\pi f}{d}t)]$$

(Fourier Expansion)

a_f, b_f: amplitude. d/f: period for sin, cos

f/d: frequency for sin, cos

k: Hyperparameter

① Non parametric Seasonality function.

$\hat{s}_i = \text{mean}(s_i, s_{i+d}, \cdots, s_{i+kd}, \cdots)$

Small sample size => overfitting.

② Differencing (seasonal differencing)

$X_t - X_{t-d} = Z_t - Z_{t-d}$ stationary.

Models for trend + Seasonality:

$X_t = \underbrace{m_t}_{\text{trend}} + \underbrace{s_t}_{\text{seasonality}} + \underbrace{Z_t}_{\text{stationary}}$

① Parametric:

Linear regression => m_t

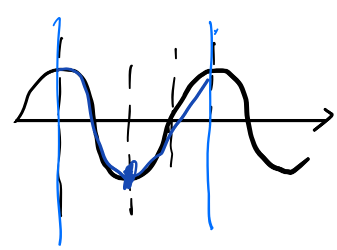
Fourier series => $\hat{s}_t$

② Smoothing.

2021.1. - 2021.6. -> ~~spring~~. summer, winter

2021.7 ~ 2021.12 -> summer. fall, winter. Heterogeneous

$$\hat{m}_t = \frac{X_{t-d/2} + X_{t-d/2+1} + \cdots + X_{t+d/2}}{d}$$



The time interval for smoothing should match the seasonal period.

③. Differencing $X_t - X_{t-d} = s_t - s_{t-d} + m_t - m_{t-d} + Z_t - Z_{t-d}$

$= m_t - m_{t-d} + Z_t - Z_{t-d}$

Can be done by ARIMA (seasonal.)

Coefficients will be simpler.

Lecture 16. Financial time series.

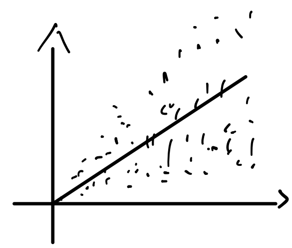
price: p_t not comparable across different stocks

Return on the t-th day: $r_t = \log(p_t) - \log(p_{t-1}) \approx \frac{p_t - p_{t-1}}{p_{t-1}}$

Advantage: scaless.

① Features of Returns.

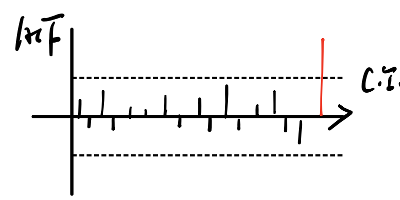
① Heteroscedasticity 异方差性



Volatility clustering

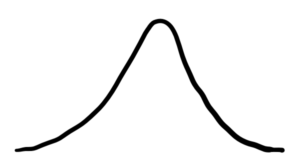
② Serially uncorrelated.

③ Sometimes significantly uncorrelated



④ Peak and heavy tail.

Leptokurtosis. (尖峰厚尾)



Kurtosis (峰度) $= E(X - EX)^4$

Meaning: compare tail to standard normal (=3)

>3: Heavy tail

Def. For a r.v. X, if there exists c_1, c_2 s.t.

① $P(|X| \geq t) \leq c_1 \exp(-\frac{t^2}{c_2}) \Rightarrow X$ is called sub-Gaussian.

② $P(|X| \geq t) \leq c_1 \exp(-\frac{t}{c_2}) \Rightarrow$ called sub-exponential

Fact: A sub-Gaussian X, $\mu = \mathbb{E}X$. $\sigma^2 = Var(X)$.
$m \in$ medians of X then $|\mu - m| < C \cdot \sigma$ for some $C > 0$

H-D: $Y = X^{\top}\beta + \varepsilon$ $X \in \mathbb{R}^{n \times p}$ $p = h(n)$ $\hat{\beta} \not\to \beta$

$K(t) = \log(\mathbb{E}(e^{tX})) \overset{\text{Taylor}}{=} \sum_{n=1}^{\infty} K_n \frac{t^n}{n!}$

$K(1) =$ mean. $K(2) =$ variance $K(3) =$ 3rd central moment
$K(4) = \mathbb{E}[X - \mathbb{E}(X)]^4 - 3[\mathbb{E}(X - \mathbb{E}X)^2]^2$

For normal distribution: $K(n) = 0$ for $n \geq 3$. and normal is the only distribution with the property.

Summary.
① $\mathbb{E}X$. mean ② $\mathbb{E}(X - \mathbb{E}X)^2$ variance
③ $\mathbb{E}(X - \mathbb{E}X)^3$: skewness 偏度
④ $\mathbb{E}(X - \mathbb{E}X)^4$: kurtosis: 峰度.

⑪ Volatility Models. $Y = f(X_1, \cdots X_k) + \varepsilon$
To give ε heteroscedasticity and heavy tail.
We model $\varepsilon = \sigma(X_1, \cdots X_k)V$. $V \overset{iid}{\sim} WN(0,1)$ (produce model).

ARCH: $\varepsilon_t = \sqrt{h_t}\, V_t$ $h_t = a_0 + a_1 \varepsilon_{t-1}^2 + \cdots + a_q \varepsilon_{t-q}^2$

$V_t \overset{iid}{\sim} WN(0,1)$ $V_t \perp\!\!\!\perp \varepsilon_{t-j}$. $j \geq 1$ $\mathbb{E}(\varepsilon_t^2) = \mathbb{E}[h_t^2 V_t^2]$

Properties:
$Var(\varepsilon_t | \mathcal{F}_{t-1}) = \mathbb{E}[\varepsilon_t^2 | \mathcal{F}_{t-1}] - \underbrace{\mathbb{E}^2(\varepsilon_t | \mathcal{F}_{t-1})}_{=0}$
$= \mathbb{E}V_t^2\, \mathbb{E}[h_t | \mathcal{F}_{t-1}] = a_0 + a_1 \varepsilon_{t-1}^2 + \cdots + a_q \varepsilon_{t-q}^2 = h_t$

$\eta_t = \varepsilon_t^2 - h_t$ η_t: serially uncorrelated with mean 0

Example ARCH(1) $\varepsilon_t = \sqrt{h_t}\, v_t$

① $Var(\varepsilon_t | \mathcal{F}_{t-1}) = h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2$ (depicts volatility)

$\sqrt{h_t}$ is scale of ε_t => scale of ε_t depends on ε_{t-1}^2

② $Var(\varepsilon_t) = E(\varepsilon_t^2) - E^2(\varepsilon_t) = E\varepsilon_t^2 = \alpha_0 + \alpha_1 E\varepsilon_{t-1}^2 + \cdots + \alpha_q E\varepsilon_{t-q}^2$

stationary $\Rightarrow$ $Var(\varepsilon_t) = E(\varepsilon_t^2) = \dfrac{\alpha_0}{1-\alpha_1-\cdots-\alpha_q}$ Why?

weak stationary does not preclude the nonconstant conditional variance

③ ARCH(q) $E(\varepsilon_t) = 0$ $E\varepsilon_t\varepsilon_{t'} = 0$ => ε_t is white noise

④ $\varepsilon_t^2 \sim AR(1)$ for ARCH(1) $\varepsilon_t^2 = \alpha_1\varepsilon_{t-1}^2 + w_t + \alpha_0$

$\eta_t \triangleq \varepsilon_t^2 - E(\varepsilon_t^2) = \varepsilon_t^2 - h_t$ => $\varepsilon_t^2 = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \eta_t$

$E\eta_t = 0$ $E\eta_t\eta_{t'} = E[(\varepsilon_t^2 - E\varepsilon_t^2)(\varepsilon_{t'}^2 - E\varepsilon_{t'}^2)] = 0$

⑤ Kurtosis ARCH(1) v_t iid $N(0,1)$

$E(\varepsilon_t^2) = \dfrac{\alpha_0}{1-\alpha_1}$ $E(\varepsilon_t - E\varepsilon_t)^4 = E\varepsilon_t^4 = E(h_t^2 v_t^4) = 3E(h_t^2)$

$= 3E(\alpha_0^2 + \alpha_1^2\varepsilon_{t-1}^4 + 2\alpha_0\alpha_1\varepsilon_{t-1}^2)$

stationary $\Rightarrow$ $(1-3\alpha_1^2)E\varepsilon_t^4 = 3\alpha_0^2 + \dfrac{6\alpha_0^2\alpha_1}{1-\alpha_1} = \dfrac{3\alpha_0^2(1+\alpha_1)}{1-\alpha_1}$

$\Rightarrow$ $E\varepsilon_t^4 = \dfrac{3\alpha_0^2(1+\alpha_1)}{(1-\alpha_1)(1-3\alpha_1^2)}$ $\dfrac{E\varepsilon_t^4}{(E\varepsilon_t^2)^2} - 3 > 0$

(iii) ARCH (Lagrange Multiplier) testing

$H_0: \alpha_0 = \alpha_1 = \cdots = \alpha_q$

$H_1: \alpha_0, \alpha_1, \cdots \alpha_q \neq 0$

Forecasting:

$h_t = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \cdots + \alpha_q\varepsilon_{t-q}^2$

$h_t(1) = E(h_{t+1} | \mathcal{F}_t) = \alpha_0 + \alpha_1\varepsilon_t^2 + \cdots + \alpha_q\varepsilon_{t+1-q}^2$

$h_t(\ell) = E[h_{t+\ell} | \mathcal{F}_t) = \alpha_0 + \alpha_1 h_t(\ell-1) + \cdots + \alpha_q h_t(\ell-q)$ $\ell > q$.

Lecture 18.

⑧ Generalized ARCH: GARCH(p, q)

Motivation: If ε_t depends on large lag of noise. ε_{t-j}

Then ARCH(q) has issues: $\begin{cases} \text{parameter estimation not acc} \\ \text{calculation of } h_t \text{ unacc} \end{cases}$

Solution: introduce lags of conditional variance (h_{t-j})

Def. $\varepsilon_t = \sqrt{h_t}\, V_t$. V_t uncorrelated, indpt of ε_{t-j}

$EV_t = 0$. $Var\, V_t = 1$

$$h_t = \alpha_0 + \beta_1 h_{t-1} + \cdots + \beta_p h_{t-p} + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_q \varepsilon_{t-q}^2$$

Rmk. ① Looks like ARMA(p, q)

② More suitable for higher-order ARCH process, with less computation (less paras)

③ GARCH(1,1) <=> ARCH(∞) $\beta_1 + \alpha_1 < 1 \Rightarrow$ Stationary

Properties of GARCH(1,1) V_t uncorrelated, indpt of ε_{t-1}

$$\begin{cases} \varepsilon_t = \sqrt{h_t}\, V_t \\ h_t = \alpha_0 + \beta_1 h_{t-1} + \alpha_1 \varepsilon_{t-1}^2 \\ EV_t = 0. \quad Var(V_t) = 1. \end{cases}$$

$E\varepsilon_t = 0$

$Var\, \varepsilon_t = \dfrac{\alpha_0}{1-\alpha_1-\beta_1}$

① $E[h_t] = \alpha_0 + \beta_1 E h_{t-1} + \alpha_1 \underbrace{E\varepsilon_{t-1}^2}_{\alpha_1 E h_{t-1}}$

$= \alpha_0 + (\beta_1 + \alpha_1) E h_t$ (Stationary)

$\Rightarrow E h_t = \dfrac{\alpha_0}{1-\alpha_1-\beta_1}$ $Var(\varepsilon_t) = \dfrac{\alpha_0}{1-\alpha_1-\beta_1}$

② GARCH: equivalent to ARCH($+\infty$)

$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}$.

Ex. $\varepsilon_t \sim ARCH(1)$. we know $\varepsilon_t^2 \sim AR(1)$

③ $\varepsilon_t \sim GARCH(p,q)$: $\varepsilon_t^2 \sim ARMA(r,p)$. $r = \max(p,q)$.

Let $\eta_t = \varepsilon_t^2 - h_t \Rightarrow h_t = \varepsilon_t^2 - \eta_t$

$$\Rightarrow \varepsilon_t^2 = \eta_t + h_t = \eta_t + \alpha_0 + \beta_1 h_{t-1} + \cdots + \beta_p h_{t-p} + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_q \varepsilon_{t-q}^2$$

$$= \eta_t + \alpha_0 + \beta_1(\varepsilon_{t-1}^2 - \eta_{t-1}) + \cdots + \beta_p(\varepsilon_{t-p}^2 - \eta_{t-p})$$
$$+ \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_q \varepsilon_{t-q}^2$$

$$= \alpha_0 + \eta_t - \beta_1 \eta_{t-1} - \cdots - \beta_p \eta_{t-p} + \sum_{j=1}^{r} (\alpha_j + \beta_j)\varepsilon_{t-j}^2$$

where $\alpha_j = 0$ if $j > q$. $\beta_j = 0$ if $j > p$

$$E\eta_t = E\varepsilon_t^2 - Eh_t = 0$$

$$Cov(\eta_t, \eta_s) = E\eta_t\eta_s - E\eta_t E\eta_s$$
$$= E\eta_t\eta_s = E(\varepsilon_t^2 - h_t)E(\varepsilon_s^2 - h_s)$$
$$= E(1-\nu_t^2)(1-\nu_s^2)h_t h_s = 0$$
$$= E(1-\nu_t^2)E(1-\nu_s^2)h_t h_s = 0$$

$Var(\eta_t) = E(\eta_t^2) = E(\varepsilon_t^2 - h_t)^2 = Const$ (4-th order stationary)

④ $E\varepsilon_t^2 = \alpha_0 + \sum_{j=1}^{r}(\alpha_j + \beta_j) E\varepsilon_{t-j}^2$

$$E\varepsilon_t^2 = \frac{\alpha_0}{1 - \sum_{j=1}^{r}(\alpha_j + \beta_j)} \quad r = \max(p,q)$$

Prop. For GARCH(p,q): $\sum_{j=1}^{r}(\alpha_j + \beta_j) < 1 \Rightarrow$ Stationary

Ex ARCH(q) Stationary if $1-\alpha_1-\cdots-\alpha_q>0$

GARCH(q) Stationary if $1-\sum_{j=1}^{r}(\alpha_j+\beta_j)>0$ $r=\max(p,q)$

Ex. ARCH(1): $r_t=\sigma_t v_t$, $\sigma_t^2=\alpha_0+\alpha_1 r_{t-1}^2$

α_0, α_1 unknown parameters. V_t iid WN(0,1)

① verify $\{\eta_t\}$ serially uncorrelated. w/ mean 0

② verify $Cov(\eta_t, r_{t-k}^2)=0$, $k>0$

③ verify $r_t^2=\alpha_0+\alpha_1 r_{t-1}^2+\eta_t \sim AR(1)$

Implies that if $r_t^2\sim AR(1)$ we can use ARCH(1) to model it.

Lecture 17. ① Properties of GARCH.
② Estimation & Forecasting.

① Properties of GARCH:

⑤ Kurtosis of GARCH. $\frac{E[\varepsilon_t^4]}{E[\varepsilon_t^2]^2} - 3$

Assume: $V_t \sim N(0,1) \Rightarrow E[V_t^{2k}] = (2k-1)!!$

$E\varepsilon_t^4 = E h_t^2 V_t^4 = 3 E h_t^2$

$E h_t^2 = E(\alpha_0^2 + \beta_1^2 h_{t-1}^2 + \alpha_1^2 \varepsilon_{t-1}^4 + 2\alpha_0\beta_1 h_{t-1} + 2\alpha_0\alpha_1\varepsilon_{t-1}^2 + 2\alpha_1\beta_1 h_{t-1}\varepsilon_{t-1}^2)$

$= \cdots$

Assume 4-th order stationary: $E h_t^2 = \frac{(1+\alpha_1+\beta_1)\alpha_0^2}{(1-3\alpha_1^2-\beta_1^2-2\alpha_1\beta_1)(1-\alpha_1-\beta_1)}$

$E\varepsilon_t^4 = \frac{3\alpha_0^2(1+\alpha_1+\beta_1)}{(1-3\alpha_1^2-\beta_1^2-2\alpha_1\beta_1)(1-\alpha_1-\beta_1)}$

excess kurtosis $= \frac{E\varepsilon_t^2}{(E\varepsilon_t^2)^2} - 3 = \frac{6\alpha_1^2}{(1-3\alpha_1^2-\beta_1^2-2\alpha_1\beta_1)(1-\alpha_1-\beta_1)} > 0$

Summary: ε_t has properties of financial time series:

① Heteroscedasticity, volatility, clustering.
② Serially uncorrelated
③ Squares of ε_t show significant correlation. ✓
④ Heavy-tail.

Exercise: For a stock return data, we model an AR-GARCH to forecast this stock's return and risk. The model estimation is as follows:

$r_t = 0.1 + 0.2 r_{t-1} + 0.02 r_{t-2} + \varepsilon_t, \quad \varepsilon_t = \sqrt{h_t} V_t$

p-val 0.001 0.0002 0.23

$h_t = 0.12 + 0.8 h_{t-1} + 0.21 \varepsilon_{t-1}^2 + 0.08 \varepsilon_{t-2}^2$

$\frac{\hat{\varepsilon}_t}{\sqrt{\hat{h}_t}}$: expected to be uncorrelated

Ljung-Box Q test on $\underline{\text{standardized } \hat{\varepsilon}_t}$ p-val = 0.2

squares of standardized $\hat{\varepsilon}_t$, p-val = 0.003

Is this a good model?

① r_{t-2} is not significant

② $\hat{\alpha}_1 + \hat{\beta}_1 = 0.8 + 0.21 = \underline{1.01 > 1 \Rightarrow \text{unstationary}}$. (By previous exercise)

③ Ljung-Box. squares p-val = 0.003 < 0.05

$\Rightarrow$ Noise term correlated

(VI). Estimation & Forecasting:

Estimation.

MLE of GARCH:

$$\text{GARCH}(1,1) \begin{cases} r_t = \sqrt{h_t}\, V_t \\ h_t = \alpha_0 + \alpha r_{t-1}^2 + \beta \cdot h_{t-1} \end{cases}$$

with $h_1 = \sigma^2 = \frac{\alpha_0}{1-\alpha-\beta}$

conditional pdf: $f(r_t, \cdots, r_n \mid r_{t-1}, \cdots, r_1) = \frac{1}{\sqrt{2\pi h_t}} \exp\left(\frac{r^2}{-2h_t}\right)$

joint pdf: $f(r_n, \cdots r_1) = f(r_{n-1}, \cdots r_1)\, f(r_n \mid r_{n-1}, \cdots r_1)$

Likelihood: $L(\alpha_0, \alpha_1, \beta) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{n}\left\{\log(h_t) + \frac{r_t^2}{h_t}\right\}$

Maximize $L(\alpha_0, \alpha_1, \beta) \Rightarrow \hat{\theta}_{mle}$

$\hat{\theta}_{mle} \overset{asy}{\sim} N(0, \hat{J}_n(\hat{\theta}_{mle}))$ $\quad \hat{J}_n(\hat{\theta}_{mle}) = \left[-\frac{\partial^2 L(\theta)}{\partial^2 \theta^2}\Big|_{\theta = \hat{\theta}_{mle}}\right]^{-1}$

· Consistency. $\hat{\theta}_{mle} \xrightarrow{P} \theta_0$

· Asymptotically normal $\hat{\theta}_{mle} \xrightarrow{d} N(0, J_n(\hat{\theta}_{mle}))$

· Efficiency: C-R lower bound.

Forecasting: GARCH(1,1)

$\varepsilon_t = \sqrt{h_t}\, V_t \qquad h_t = \alpha_0 + \beta_1 h_{t-1} + \alpha_1 \varepsilon_{t-1}^2$

$h_{T+1} = \alpha_0 + \alpha_1 \varepsilon_T^2 + \beta_1 h_T$

$h_T(1) = E(h_{T+1} \mid \mathcal{F}_T) = \alpha_0 + \alpha_1 \varepsilon_T^2 + \beta h_T$

$h_{t+1} = \alpha_0 + \alpha_1 h_t V_t^2 + \beta_1 h_t = \alpha_0 + (\alpha_1 + \beta_1) h_t + \alpha_1 h_t (V_t^2 - 1)$

$h_T(2) = E(h_{T+2} \mid \mathcal{F}_T) = \alpha_0 + (\alpha_1 + \beta_1) h_T(1)$

$h_T(\ell) = \alpha_0 + (\alpha_1 + \beta_1) h_T(\ell - 1) \quad \ell > 1. \quad (\alpha_1 + \beta_1 < 1)$

$$= \frac{\alpha_0 (1 + (\alpha_1 + \beta_1)^{\ell - 1})}{1 - \alpha_1 - \beta_1} + (\alpha_1 + \beta_1)^{\ell - 1} h_T(1)$$

$$\to \frac{\alpha_0}{1 - \alpha_1 - \beta_1} \quad \ell \to \infty.$$

Remark ① tends to unconditional var of GARCH(1,1).

② This condition holds for general GARCH(p,q)

Review: Properties of GARCH(p,q)

① $E \varepsilon_t = 0$

$Var(\varepsilon_t \mid \mathcal{F}_{t-1}) = h_t$

$$Var(\varepsilon_t) = \frac{\alpha_0}{1 - \sum_j^r (\alpha_j + \beta_j)}$$

where $r = \max(p, q)$ $\alpha_j = 0$ if $j > q$. $\beta_j = 0$ if $j > p$.

② GARCH(1,1) equivalent to ARCH(∞).

③ $\varepsilon_t \sim GARCH(p,q) \Rightarrow \varepsilon_t^2 \sim ARMA(r,p)$ $r = \max(p,q)$

④ Excess Kurtosis of GARCH > 0 ⇒ heavy tail

5.23 Review: Exercises

Ex. 1. Consider an ARCH(2) model with normal shocks.

$$r_t = \sqrt{h_t}\, V_t \qquad V_t \overset{iid}{\sim} N(0,1)$$

$$h_t = 1 + 0.2 r_{t-1}^2 + 0.2 r_{t-2}^2$$

① Compute the unconditional variance

② $r_5 = 1$, $r_6 = -1$. Compute 95% C.I. for r_7 and h_8

① $Var(r_t) = \dfrac{1}{1-0.2-0.2} = \dfrac{5}{3}$

$N(0,1)$

② $\mathbb{E}(r_7 \mid r_5, r_6) = 0 \qquad r_7 = \sqrt{1.4}\, V_7$

$\mathbb{E}(r_7^2 \mid r_5, r_6) = 1 + 0.2 r_6^2 + 0.2 r_5^2 = 1.4$

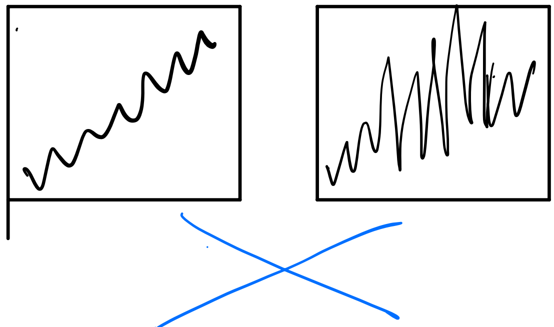
C.I.: $0 \pm Z_{\alpha/2}\sqrt{1.4}$

$h_8 = 1.2 + 0.2 h_7 V_7^2 = 1.2 + 0.28 V_7^2 \sim \chi^2(1)$

③ 95% CI of r_8?

④ C.I. for general GARCH?

Ex2. Data 1. Data 2.



① $X_t = W_1 + \cdots + W_t$ Model 1. $X_t = at + b + W_t$ Model 2. which model?

② Model 1 stationary? No. Var diff $X_t(1-B) = W_t$ $\Rightarrow$ root $= 1$

③ Best linear prediction of $X_t = W_1 + \cdots + W_t$

$$\mathbb{E}(X_{n+1} \mid X_1, \dots X_n) = \mathbb{E}(X_n + W_{n+1} \mid \mathcal{F}_n) = X_n$$

Ex3. $\{X_t\} \sim AR(2)$. process with $X_t(1 + 0.75B - 0.125B^2) = W_t$

Find PACF

$$X_t = -0.75 X_{t-1} + 0.125 X_{t-2} + W_t$$

① $PACF(h) = 0$ for $h > p$

② $PACF(p) = \phi_p = 0.125$

③ $PACF(1) = ACF(1) = \rho(1)$

$$X_t = -0.75 X_{t-1} + 0.125 X_{t-2} + W_t$$

Trick: Compute Cov w/ X_{t-1} in both sides.

$Cov(X_t, X_{t-1}) = -0.75\rho(0) + 0.125\rho(1) \Rightarrow \rho(1) = 6/7$

Ex 4 $Y_t = S_t + W_t$, $S_t = S_{t+d}$

① Show $X_t = Y_t - Y_{t-d}$ is a stationary process

$X_t = S_t + W_t - S_{t-d} - W_{t-d} = W_t - W_{t-d}$ $\quad W_t = X_t + W_{t-d}$

$\mathbb{E}(X_t) = 0$ $\quad Var(X_t) = 2$ $\quad \gamma(h) = 0$ for $h \neq |d|$

$\gamma(d) = Cov(W_{t+d} - W_t, W_t - W_{t-d}) = -1 \Rightarrow$ weak stationary

② Best linear predictor for X_{nd+1} in terms of $X_1, \cdots X_{nd}$

equals $\frac{1}{n+1}(-X_1 - 2X_{1+d} - \cdots - nX_{1+(n-1)d})$

verify $\hat{X}_{nd+1} - X_{nd+1} = \hat{\gamma} \perp X_i$ $\quad Cov(\hat{\gamma}, X_i) = 0$

$\mathbb{E}(X_{nd+1} | \mathcal{F}_{nd}) = -W_{t-d} = -X_{t-d} - W_{t-d}$

$= -X_{t-d} - X_{t-2d} - W_{t-2d}$

③ Compute PACF of $\{X_t\}$.

Best linear prediction of X_t given $X_{t-1}, \cdots$

$$\phi_{hh} = \begin{cases} -\frac{1}{n+1}. & h = nd. \\ 0, & h \neq nd \end{cases}$$

Final exam:

① One double-side cheating sheet

② Coverage: everything in this semester

Lecture 21: 5/30/2023.

Review session:

* Concepts.

① mean variance. kurtosis. (excess).

② acf. pacf (sampled)

③ stationary

④. Causal./invertible

* EDA. { scatter plot, histogram, sampled acf. } $\Rightarrow$ understand.

what can you read from plots?

* Non-stationary pattern. $\Rightarrow$ stationary.

① Trending { OLS, smoothing, differencing }

① Seasonality
- modelling
- smoothing
- seasonal differencing
- ARIMA: differencing

1 Modeling
- MA(q)
- AR(p) } properties: ACF, PACF
- ARMA(p, q)
  - ① roots of characteristic functions.
  - ② any stationary model can be approximated by ARMA

1 Fitting:
- MLE
  - consistency
  - asymptotic normality
- Least square
- Moment method

1 Diagnostic
- residual analysis
  - scatter plot
  - acf/pacf
  - Ljung-Box
- QQ plot
- overfitting as a tool

2. Forecasting
- Best linear predictor
- Best prediction

2. Inference
- Estimators asymptotic distn. (C.I)
- Prediction limits

1 Heterogeneous noise
- ARCH
- GARCH
(Kurtosis)

Format of Final Exam:

① T / F

② Results → Model. ACF → ARMA (relationship)

③ Graph Analysis

④ Computation: acf / pacf

Exercise:

①. True / False.

(a) Estimating the parameters of a seasonal $ARMA(P,Q)_s$ model is equivalent to estimating is equivalent to estimate $ARMA(p,q)$, where $p = P_s$, $q = Q_s$

False

(b)

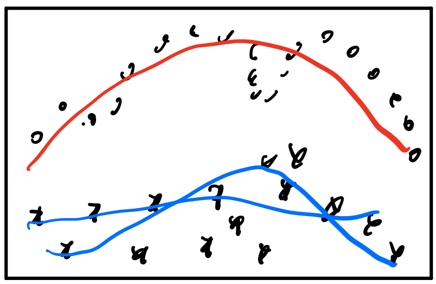
② Give two polynomials $\phi(z)$ and $\theta(z)$ s.t. the ARMA $\phi(B)X_t = \theta(B)W_t$ is stationary, causal, invertible and whose acf $\rho_X(h)$ satisfies $\rho_X(h) = 0.4$ for $|h| = 1$ and $\rho_X(h) = 0$ for $|h| > 1$

(b) How about $\rho_X(h) \neq 0$ for $|h| \in \{0, 1, 11, 12\}$ and $\rho_X(h) = 0$. o.w. $X_t = W_t + W_{t-1} + W_{t-12}$

$z(1+z^{11})+1=0$ $\quad X_{t-11}=W_{t-11}+W_{t-12}+\cdots$

$z(1+z^{11})+1=0$ $\quad W_{t-12}=W_{t-13}-$

③ 0 US 8 OCEP.



(a) Find two properties suggesting non-stationarity

(b) For each properties, suggest a way to modify $\{X_t\}$ to increase the property

(c) Suppose that the estimate of the mean function in the figure was computed using a kernel smoothing method. what the function would look like if bandwidth ↓

④ Computation.

(a) Let $\{X_t\}$ be an AR(1) process w/ $|\phi|<1$. what is the best linear prediction of X_{51} in terms of $X_1 \cdots X_{50}$?

(b) Write an expression for the mean squared error of the prediction of X_{51} computed in (a)

稳定随机顺序

*. Take-away of Time-series analysis.

- Stationary <-> generalize
- Direction: time follows forward.
- AI / ML: overfitting. Do too much in one domain.

① Computer vision: Portrait pictures $\xrightarrow{\text{transfer model}}$ skeleton pictures

Fit ↓ (above "Portrait pictures")

Domain is stable

NLP: harder

② Physics: Newtown: overfitting

③ Time Series: domain always changes

Distribution shift very quickly

believe stationary parts remain unchanged

Finance:

$$\text{return} \approx \frac{P_t}{P_{t-1}} - 1$$

$$r_t = \text{Country return}(t) + \text{Industry return}_t + \text{Style return} + \underline{\text{Stationary}}$$

Like some Axiom.

Style Exposure: 风格暴露

偏好某种 style，受 non-stationary part 影响.