1 Fundamental Concepts

1.1 Stationarity

A time series is a stochastic process $\{X_{(t)} : t \in T\}$. In this course, we consider the case with $T = \{0, \pm 1, \pm 2, \cdots\}$. Let $X_t = X(t)$, then $X_1, X_2, X_3, \cdots, X_n, \cdots$ is a sequence of time series.

Definition 1.1. Let $X_{t_1}, X_{t_2}, \cdots, X_{t_n}) \stackrel{d}{=} (X_{t_1+k}, X_{t_2+k}, \cdots, X_{t_n+k})$. If

$$F(x_{t_1}, x_{t_2}, \cdots, x_{t_n}) = F(x_{t_1+k}, x_{t_2+k}, \cdots, x_{t_n+k}),$$

for $\forall\, t_1, t_2, \cdots, t_n, k$, then we say that $\{X_t\}$ is strictly stationary.

Mean function of X_t is

$$\mu_t = EX_t$$

Variance function of X_t is

$$\sigma_t^2 = E(X_t - \mu_t)^2$$

Convariance function of X_t:

$$\gamma(t_1, t_2) = E[(X_{t_1} - \mu_{t_1})(X_{t_2} - \mu_{t_2})]$$

Correlation function of $X_{t_1} and\ X_{t_2}$ is

$$\rho(t_1, t_2) = \frac{\gamma(t_1, t_2)}{\sqrt{\sigma_{t_1}^2 \sigma_{t_2}^2}}$$

Definition 1.2. If the following three conditions are satisfied:

(i) $EX_t^2 < \infty$

(ii) $EX_t = m$ (m is a constant)

(iii) $E[(X_t - m)(X_s - m)] = R(|t - s|)$

then we say that X_t is weak (or convariance, or second-order) stationary.

Example 2.1

(1) if $X_1, X_2, \cdots, X_n$ be i.i.d r.v.s, then $\{X_t\}$ is strictly stationary and covariance stationary.

(2) if $\{X_t\}$ is strictly stationary and $\mathrm{E}X_t^2 < \infty$, then $\{X_t\}$ is covariance stationary.

1.2 Autocovariance and autocorrelation functions

Assume that $\{X_t\}$ is covariance stationary time series and denote

$$\gamma_k = cov(X_t, X_{t+k})$$

$$\rho_k = \frac{cov(X_t, X_{t+k})}{\sqrt{var(X_t)var(X_{t+k})}} = \frac{\gamma_k}{\gamma_0}$$

We call γ_k and ρ_k autocovariance function (ACV) and autocorrelation function (ACF), respectively.

Properties of γ_k and ρ_k:

(1)$\gamma_0 = \sigma^2,\ \rho_0 = 1,$

(2)$\gamma_k = \gamma_{-k},\ \rho_k = \rho_{-k},$

(3)$\gamma_k \leq \gamma_0,\ \rho_k \leq \rho_0.$

1.3 Partial Autocorrelation Function

Assume that $\{X_t\}$ is a stationary time series. Let

$$\phi_{kk} \equiv cov(X_t, X_{t+k} | X_{t+1}, \cdots, X_{t+k-1}) \qquad .$$

ϕ_{kk} is called partial autocorrelation function (PACF) of $\{X_t\}$. $\phi_{11} = \rho_1$ and

$$\phi_{kk} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_{k-2} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & 1 \end{vmatrix}}.$$

1.4 White Noise Processes

Definition 1.3. A time series $\{a_t\}$ is called white noise process if

$(1) Ea_t = 0 \text{ and } Ea_t^2 = \sigma_a^2 < \infty,$

$(2) \gamma_k = cov(a_t, a_{t+k}) = 0 \ \ if \ \ k \neq 0 \, \forall \, t.$

If a_t is a white noise, then

(1) (ACV): $\upsilon_k = \begin{cases} \sigma^2 & \text{k=0,} \\ 0 & \text{k}\neq\text{0;} \end{cases}$

(2) (ACF): $\rho_k = \begin{cases} 1 & \text{k=0,} \\ 0 & \text{k}\neq\text{0;} \end{cases}$.

(3) (PACF): $\phi_{kk} = \begin{cases} 1 & \text{k=0,} \\ 0 & \text{k}\neq\text{0.} \end{cases}$

If a time series $\{X_t\}$ does not satisfy (1)-(3) above, then $\{X_t\}$ is not white noise.

1.5 Estimation of ACV and ACF

Given $X_1, \cdots, X_n$, we want to find μ, σ^2, γ_{12} and ρ_{12}. Assume that X_t is strictly stationary and ergodic.

(1) if E $|X_t| < \infty$, then

$$\frac{1}{n}\sum_{t=1}^{n} X_t \xrightarrow{a.s.} EX_t = \mu, \qquad as\ \ n \to \infty.$$

so we can use 2 $\bar{X}_n = \frac{1}{n}\sum_{t=1}^{n} X_t$ as an estimator of μ.

(2) if $EX_t^2 < \infty$, then

$$\frac{1}{n}\sum_{t=1}^{n} X_t^2 \xrightarrow{a.s.} EX_t^2 \quad as\ \ n \to \infty;$$

$$\sigma_n^2 = \frac{1}{n}\sum_{t=1}^{n} X_t^2 - (\frac{1}{n}\sum_{t=1}^{n} X_t)^2 \xrightarrow{a.s.} EX_t^2 - (EX_t)^2 = \sigma^2.$$

Thus, $\hat{\sigma}_n^2$ can be used as an estimator of σ^2.

$$\hat{\gamma}_k = \frac{1}{n}\sum_{t=1}^{n-k}(X_t - \mu)(X_{t+k} - \mu) \xrightarrow{a.s.} \gamma_k \qquad ;$$

$$\hat{\gamma}_k \ = \ \frac{1}{n}\sum_{t=1}^{n-k}(X_t - \bar{X}_n)(X_{t+k} - \bar{X}_n) \xrightarrow{a.s.} \gamma_k \qquad as\ \ n \to \infty$$

$$\uparrow$$

$$\frac{1}{n-k}$$

$\hat{\gamma}_k$ is the estimator of γ_k. $\hat{\rho}_k = \hat{\gamma}_k/\hat{\gamma}_0$ is the estimator of ρ_k and it converges to ρ_k a.s. as n$\to \infty$.

We can estimate ϕ_{kk} by the following formula: $\hat{\phi}_{11} = \hat{\rho}_1$ and

$$\hat{\phi}_{kk} = \frac{\begin{vmatrix} 1 & \hat{\rho}_1 & \hat{\rho}_2 & \cdots & \hat{\rho}_{k-2} & \hat{\rho}_1 \\ \hat{\rho}_1 & 1 & \hat{\rho}_2 & \cdots & \hat{\rho}_{k-3} & \hat{\rho}_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \hat{\rho}_{k-1} & \hat{\rho}_{k-2} & \hat{\rho}_{k-3} & \cdots & \hat{\rho}_1 & \hat{\rho}_k \end{vmatrix}}{\begin{vmatrix} 1 & \hat{\rho}_1 & \hat{\rho}_2 & \cdots & \hat{\rho}_{k-2} & \hat{\rho}_{k-1} \\ \hat{\rho}_1 & 1 & \hat{\rho}_1 & \cdots & \hat{\rho}_{k-3} & \hat{\rho}_{k-2} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \hat{\rho}_{k-1} & \hat{\rho}_{k-2} & \hat{\rho}_{k-3} & \cdots & \hat{\rho}_1 & 1 \end{vmatrix}}$$

and $\hat{\phi}_{kk} \xrightarrow{a.s.} \phi_{kk}$. In particular,when X_t is white noise, we have $\hat{\gamma}_k \approx 0$, $\hat{\rho}_k \approx 0$ and $\hat{\phi}_{kk} \approx 0$ if $k \neq 0$.

1.6 Moving average and Autoregressive representation

Definition 1.4. We call a time series $\{X_t\}$ a MA(∞) process if

$$X_t = \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots = \mu + \sum_{j=0}^{\infty} \psi_j a_{t-j}$$

where $\psi_0 = 1$, a is white noise, and $\sum_{j=0}^{\infty} \psi_j^2 < \infty$. X_t is also called linear process or has Wold's representation.

Theorem 1.6.1. *Let* $X_{tn} = \mu + \sum_{j=0}^{n} \psi_j a_{t-j}$. *Then*

$$(1) \qquad P(\lim_{n\to\infty} X_{tn} = X_{t\infty} = \mu + \sum_{j=0}^{\infty} \psi_j a_{t-j}) = 1,$$

$$(2) \qquad \lim_{n\to\infty} E|X_{tn} - X_{t\infty}|^2 = 0.$$

Proof. Let $\mathcal{F}_n = \sigma(a_t, a_{t-1}, \cdots, a_{t-n})$. Then

$$E(X_{tn}|\mathcal{F}_{n-1}) = \mu + \textstyle\sum_{j=0}^{n-1} \psi_j a_{t-j} + E(\psi_n a_n|\mathcal{F}_{n-2}) + E(\psi_n a_{t-n}|\mathcal{F}_{n-1})$$

$$= \mu + \sum_{j=0}^{n-1} \psi_j a_{t-j} = X_{t,n-1} \qquad .$$

Thus, X_{tn} is a martingale with respect to $\mathcal{F}$

$$EX_{tn}^2 = \mu^2 + \sum_{j=0}^{n} \psi_j^2 E a_{t-j}^2 = \mu^2 + (\sum_{j=0}^{n} \psi_j^2)\sigma 2$$

$$\leq K = \mu^2 + (\sum_{j=0}^{\infty} \psi_j^2)\sigma^2 < \infty \text{ for all } n$$

By Martingale convergence theorem, we have

$$(1)\ P(\lim_{n\to\infty} X_{tn} = X_{t\infty}) = 1,$$

$$(2)\ \lim_{n\to\infty} E|X_{tn} - X_{t\infty}|^2 = 0.$$

We understand $X_t = X_{t\infty}$ a.s. This completes the proof.

Notation Backshift operator: $B_j x_t = x_{t-j}$.

$$X_t = \mu + \sum_{j=0}^{\infty} \psi_i B^j a_{t-j} = \mu + \sum_{j=0}^{\infty} \psi_j B^j) a_t \equiv \mu + \psi(B) a_t,$$

where $B^0 = 1$ and $\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j$.

Some properties:

$$EX_t = \mu + \sum_{j=1}^{\infty} \psi_j E a_{t-j} = \mu$$

$$Var(X_t) = E[\sum_{j=0}^{\infty} \psi_j a_{t-j}]^2 = \sum_{j=0}^{\infty} \psi_j^2 E a_{t-j}^2$$

$$= \sigma_a^2 \sum_{j=0}^{\infty} \psi_j^2 \qquad ;$$

$$E(a_t X_{t-j}) = \begin{array}{ll} \sigma_a^2 & for \;\; j = 0 \\ 0 & for \;\; j > 0; \end{array}$$

$$\gamma_k = E[(X_t - \mu)(X_{t+k} - \mu)] = \sigma_a^2 \sum_{j=0}^{\infty} \psi_i \psi_{i+k};$$

$$\rho_k = \frac{\sum_{i=0}^{\infty} \psi_i \psi_{i+k}}{\sum_{j=0}^{\infty} \psi_j^2}.$$

Definition: AR(∞) model.

$$X_t = \pi_1 X_{t-1} + \pi_2 X_{t-2} + \cdots + a_t = \sum_{i=1}^{\infty} \pi_i X_{t-i} + a_t$$

where $\sum_{j=1}^{\infty} |\pi_j| < \infty$ an $E|X_t|$ = a constant.

Theorem 1.6.2. *$\sum_{i=1}^{n} \pi_i X_{t-i}$ exists a.s. and $\sum_{i=1}^{n} \pi_i X_{t-i} + a_t$ converges to X_t in L^1 as $n \to \infty$.*

Proof. First, we note that $\sum_{i=1}^{n} |\pi_i||X_{t-i}(\omega)|$ exists and its limit is a finite number or infinity. Let $X(\omega) = \lim_{n\to\infty} \sum_{i=1}^{n} |\pi_i||X_{t-i}(\omega)|$. Then $\sum_{i=1}^{n} |\pi_i||X_{t-i}| \to X$ a.s.. By the Monotone convergence theorem,

$$E\sum_{i=1}^{n} |\pi_i||X_{t-i}| \to EX,$$

as $n \to \infty$. But $\max_n E(\sum_{i=1}^{n} |\pi_i||X_{t-i}|) \leq E|X| \sum_{i=1}^{\infty} |\pi_i| < \infty$. Thus.

$$EX < \infty \Longrightarrow P(X < \infty) = 1.$$

Thus, we can claim that $P(\lim_{n\to\infty} \sum_{i=1} \pi_i X_{t-1} \ exists\ ) = 1$, i.e., $X_t(\omega) = \sum_{i=1}^{\infty} \pi_i X_{t-i}(\omega) + a_t(\omega)$ is well defined.

Let $X_{tn} = \sum_{i=1}^{n} \pi_i X_{t-i}$, for $\forall\ n > m$.

$$E|X_{tn} - X_{tm}| = E|\sum_{i=m+1}^{n} \pi_i X_{t-i}| \leq c \sum_{i=m+1}^{n} |\pi_i| \longrightarrow 0$$

as n,m$\longrightarrow \infty$. By Cauchy criterion, there exists a r.v. $X_{t\infty}$ such that

$$X_{tn} \longrightarrow X_{t\infty}\ in\ L^1\ \ (X_{t\infty} \stackrel{?}{=} \sum_{i=1}^{\infty} \pi_i X_{t-i})$$

We want to prove that

$$P(X_{t\infty} = \sum_{i=1}^{\infty} \pi_i X_{t-i}) = 1.$$

It is sufficient to show that

$$E|X_{t\infty} - \lim_{n\to\infty} X_{tn}| = 0.$$

By the Factou's Lemma,

$$E|X_{t\infty} - \lim_{n\to\infty} X_{tn}| = E[\lim_{n\to\infty} |X_{t\infty} - X_{tn}|] \leq \lim_{n\to\infty} inf\ E|X_{t\infty} - X_{tn}| = 0.$$

So,we have $X_{t\infty} = \lim_{n\to\infty} X_{tn}$. Furthermore, we know that $X_{tn} + a_t$ converges to $X_{t\infty} + a_t$ in L^1. This completes the proof.

Relationship of MA(∞) and AR(∞):

(1) if the root of $\pi(B) = 0$ all lie outside the unit circle,then

$$\pi(B)x_t = a_t \Rightarrow x_t = \frac{1}{\pi(B)}a_t = \pi^{-1}(B)a_t$$

(2) if the root of $\psi(B)$=0 all lie outside the unit circle,then

$$a_t = \psi^{-1}(B)[X_t - \mu] = \psi^{-1}(B)X_t - \mu\phi^{-1}(1)$$

Time series models:

$$X_t = f(X_{t-1}, X_{t-2}, \cdots) + a_t$$

what is f ?

AR Model

MA Model

Threshold AR Model

GARCH model

vector ARMA-GAREM model

1.7 Appendix: Two Theorems

Strong Ergodic Theorem. *Let* $\{X_n : n = 0, 1, 2, \cdots\}$ *be a strictly stationary process with finite mean* $m = EX_n$*.Let* $\bar{X}_n = \sum_{i=0}^{n-1} X_i/n$*. Then, as* $n \to \infty$*,*

$$(1)\bar{X}_n \xrightarrow{a.s.} \bar{X},\ \text{a random variable} \qquad ,$$

$$(2) Furthermore,\, if\, \{X_n\}\, is\, ergodic,\, then \bar{X} = m.$$

Martingale Mean Square Convergence theorem. *Let* $\{X_n\}$ *be a martingale with respect to* $\mathcal{F}_n = \sigma\{Y_0, Y_1, \cdots, Y_n\}$*, and for some constant* K

$$EX_n^2 \leq K < \infty, for\, all\, n$$

Then $\{X_n\}$ *converges as* $n \to \infty$ *to a limit* X_∞ *with*

$$P(\lim_{n\to\infty} X_n = X_\infty) = 1 \text{ and } \lim_{n\to\infty} E|X_n - X_\infty|^2 = 0$$

Furthermore, $EX_0 = EX_n = EX_\infty$ *for* $\forall\ n$.

2 Stationary Time Series Models

2.1 Autoregressive Processes

2.1.1 AR(1) process

$$X_t = \phi X_{t-1} + a_t, \tag{2.1}$$

where a_t is white noise.

(1) when $|\phi_1| < 1$,

$$X_t = \underbrace{a_t + \phi a_{t-1} + \cdots + \phi^{t-1} a_1} + \phi^t x_0.$$

Let $X_{tn} = \sum_{i=0}^{\infty} \phi^i a_{t-i}$,then

$$X_{tn} = \sum_{i=0}^{\infty} \phi^i a_{t-i} \xrightarrow{a.s.} X_{t\infty} = \sum_{i=0}^{\infty} \phi^i a_{t-i}.$$

Put λ into $(2, 1)$, we have

$$S_t = \phi S_{t-1} + a_t.$$

So S_t is a solution of $(2, 1)$ and $S_t \equiv X_t$ is covariance stationary.

If there is another solution $\tilde{S}_t$ to model (2,1),then

$$\tilde{S}_t = \phi \tilde{S}_{t1} + a_t$$

$$\Rightarrow \qquad S_t - \tilde{S}_t = \phi(S_{t-1} - \tilde{S}_{t-1})$$

$$= \phi^2(S_{t-2} - \tilde{S}_{t-2})$$

$$= \phi^n(S_{t-n} - \tilde{S}_{t-n})$$

$$E|S_t - \tilde{S}_t|^2 = \phi^{2n} E|S_{t-n} - \hat{S}_{t-n}|^2 = c\,\phi^{2n} \longrightarrow 0$$

$$\Rightarrow \quad S_t = \tilde{S}_t \quad a.s.$$

So the solution S_t is unique.

(2) When $|\phi| = 1$, in particular,when $\phi = 1$,

$$X_t = \underline{a_t + a_{t-1} + \cdots + a_1} + X_0$$

$$\frac{1}{\sqrt{t}} \sum_{i=1}^{t} a_i \longrightarrow N(0, \sigma^2) = O_p(1)$$

$$\Rightarrow \sum_{i=1}^{t} a_i = O_p(\sqrt{t})$$

$$\Rightarrow X_t \longrightarrow \infty \quad in\ probability.$$

X_t is called the random walk or unstable process.

(3) When $|\phi| > 1$,

$$X_t = \sum_{i=0}^{m} \phi^i a_{t-i} + \phi^{m+1} X_{t-m+1} = \sum_{i=0}^{t} \phi^i a_{t-i} + \phi^t X_0 \quad (when\ \ m = t).$$

We call model (2.1) an explosive AR(1) model.

In this chapter, we only consider the case when the model is stationary.

For AR(1) model $(2,1)$, we have

$$\begin{aligned}
&\mu = EX_t = 0, \\
&\sigma^2 = \frac{\sigma_2^2 \phi^2}{1 - \phi^2}, \\
&\gamma_k = \sigma_a^2 \sum_{j=0}^{\infty} \phi_j^j \phi_{j+k}^{j+k} = \frac{\sigma_a^2 \phi^k}{1 - \phi^2}, \\
&\rho_k = \frac{\gamma_k}{\gamma_0} = \phi^k \qquad (ACF), \\
&\phi_{kk} = \left\{ \begin{array}{ccc} \rho_1 = \phi_1 & & k = 1 \\ 0 & for\ \ k \geq 2 & (PACF) \end{array} \right. .
\end{aligned}$$

2.1.2 AR (P) process

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + a_t \qquad (2.2)$$

$$\begin{bmatrix} X_t \\ X_{t-1} \\ X_{t-2} \\ \vdots \\ X_{t-p+1} \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_p \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_{t-1} \\ X_{t-2} \\ X_{t-3} \\ \vdots \\ X_{t-p} \end{bmatrix} + \begin{bmatrix} a_t \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Let A be the matrix in the previous equation, $Y_t = (X_t, \cdots, X_{t-p+1})'$ and $\varepsilon_t = (a_t, 0, \cdots, 0)'$. We can rewrite (2.2) in the vector form as follows.

$$Y_t = AY_{t-1} + \varepsilon_t. \qquad (2.3)$$

Thus,

$$\begin{aligned} Y_t &= AY_{t-1} + \varepsilon_t \\ &= \varepsilon_t + A\varepsilon_{t-1} + \cdots + A^m \varepsilon_{t-m} + A^{m+1} Y_{t-m-1} \\ &= S_{mt} + A^{m+1} Y_{t-m-1}, \end{aligned}$$

where $S_{nt} = \sum_{i=0}^{n} A^i \varepsilon_{t-i}$.

Theorem 2.1. *If all the roots of the polynomial $\lambda^p - \phi\lambda^{p-1} \cdots - \phi_p = 0$ lie inside the unit circle, then there exists a unique, and the second-order stationary, and $\mathcal{F}_t$-measurable solution to model* (2.2) *This solution has the following representation:*

$$X_t = \sum_{i-0}^{\infty} \psi_i a_{t-i} \text{ in m.s and a.s.}$$

where $\psi_i = o(\rho^i)$,with $\rho \in (0, 1)$.

Proof. First, we have

$$|I\lambda - A| = \begin{vmatrix} \lambda - \phi_1 & \cdots & \cdots & \cdots & \cdots & \phi_0 \\ -1 & \lambda & \cdots & \cdots & 0 & 0 \\ 0 & -1 & \lambda & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & \lambda \end{vmatrix} = \lambda^p - \phi_1 \phi^{p-1} - \cdots - \phi_p.$$

Thus, all the eigenvalues of A in absolute value are less than 1.

Denote $\rho = max\{|the\ \ eigenvalue|\ of\ \ A\}$. Then $0 < \rho < 1$. By corollary A.2 in Johansen $(1995, pp.220)$, there is a constant $c > 0$ such that

$$||A^n|| \leq c\rho^{\frac{n}{2}}.$$

Similar to the proof of Theorem 1.6.2, we can show that $\sum_{i=0}^{n} A^i\varepsilon_{t-i}$ exists a.s. as $n \to \infty$ and hence $\sum_{i=0}^{\infty} A^i\varepsilon_{t-i}$ is well defined. We can see that $\sum_{i=0}^{\infty} A^i\varepsilon_{t-i}$ is a solution of model (2.3). Denote $S_{nt} = \sum_{i=0}^{n} A^i\varepsilon_{t-i}$. Then

$$\begin{aligned}
&E||S_{n,t} - S_{n+m,t}||^2 \\
&\quad = E||A_t^{n+1}\varepsilon_{t-n-1} + \cdots + A^{n+m}\varepsilon_{t-m-n}||^2 \\
&\quad \leq (||A^{n+1}||^2 + \cdots + ||A^{n+m}||^2)E||\varepsilon_{t-m-n}||^2 \\
&\qquad\qquad \leq 2c(\rho^{n+1} + \cdots + \rho^{n+m})\sigma^2 \\
&\quad \longrightarrow 0 \quad as\ \ n, m \to \infty.
\end{aligned}$$

By Cauchy criterion, we have know that there is a r.v. S_t such that

$$S_{nt} \longrightarrow S_t \qquad in\ \ m.s.$$

By the Factou's lemma,

$$\begin{aligned}
E||S_t - \sum_{i=0}^{\infty} A^i\varepsilon_{t-i}|| &= E\lim_{h\to\infty} ||S_t - \sum_{i=0}^{n} A^i\varepsilon_{t-i}||^2 \\
&\leq \lim_{n\to\infty} inf E||S_t - S_t^{(n)}||^2 = 0 \\
&\Rightarrow P(S_t = \lim_{n\to\infty} S_{nt} = \sum_{i=0}^{\infty} A^i\varepsilon_{t-i}) = 1,
\end{aligned}$$

i.e. $S_t \overset{a.s}{=} \sum_{i=0}^{\infty} A^i\varepsilon_{t-i}$.

So, $S_t^{(n)} \longrightarrow S_t = \sum_{i=0}^{\infty} A^i\varepsilon_{t-i}$ in m.s..

It is easy to check that S_t in (1.8) is a solution of model (2.2)

Let $X_t = (1, 0, \cdots, 0)S_t$. Then

$$X_t \quad = \quad \sum_{i=0}^{\infty} \gamma' A^i \gamma \varepsilon_{t-i}$$

is a solution of model(2.2), where $\gamma = (1, 0, \cdots, 0)'$. y_t is strictly stationary and $\mathcal{F}_t$ is measurable. Since $Ey_t^2 =$ a constant, y_t is also weak stationary.

For uniqueness, assume that there are two solutions Y_t' and Y_t'' to model (1.2). Then

$$\begin{aligned}
&Y_t' = AY_{t-1}' + \tilde{\varepsilon}_t \qquad\qquad\qquad\qquad\qquad\qquad ,\\
&Y_t'' = AY_{t-1}'' + \tilde{\varepsilon}_t,\\
&Y_t' - Y_t'' = A(Y_{t-1}' - Y_{t-1}'') = \cdots = A^n(Y_{t-n}' - Y_{t-n}'').
\end{aligned}$$

Thus, we have

$$\begin{aligned}
E||Y_t' - Y_t''||^2 \quad &\leq \quad ||A^n||E||Y_{t-n}' - Y_{t-n}'||\\
&\leq \quad ||A^n||(E||Y_{t-n}''|| + E||Y_{t-n}''||)\\
&= \quad \lim_{n\to\infty} ||A^n||(E||Y_{t-n}''|| + E||Y_{t-n}''||) = 0.
\end{aligned}$$

Thus, $Y_t' \overset{a.s}{=} Y_t''$, i.e. the solution is unique. This completes the proof.

Theorem 2.2 *Suppose that model(1.1) has a covariance stationary solution y_t. Then it is necessary that all the roots of $z^p - \phi_1 z^{p-1} \cdots \phi_p = 0$ lie inside the unit circle.*

Proof. $Ey_t^2 =$ a constant $\Rightarrow E(Y_tY_t') = \Omega$ a constant matrix. By(2.2)

$$\begin{aligned}
E(Y_tY_t') \quad &= \quad AE(Y_{t-1}Y_{t-1}')A' + E\tilde{\varepsilon}_t\tilde{\varepsilon}_t'\\
\Omega \quad &= \quad A\Omega A' + \sigma^2\gamma\gamma',
\end{aligned}$$

where $\gamma = (1, 0, 0 \cdots 0)'$.

Assume that A has an eigenvalue λ with corresponding left eigenvector $Z = (z_1, z_2, \cdots, z_p)'$, i.e. $z'A = \lambda z$ and $\bar{z}'A = \bar{\lambda}\bar{z}$,where $\bar{z}$ is the conjugate vector of z. Thus,

$$\begin{aligned}
z'\Omega\bar{z} &= z'A\Omega A'\bar{z} + z'\gamma\gamma'\bar{z}\sigma^2 \\
&= \lambda z\Omega(\bar{z}'A)' + \sigma^2|z_1|^2 \\
&= \lambda z\Omega\bar{z}\bar{\lambda} + \sigma^2|z_1|^2 \\
&= |\lambda|^2 z\Omega\bar{z} + \sigma^2|z_1|^2.
\end{aligned}$$

If $z_1 \neq 0$, then $|\lambda| < 1$.(?). Note that

$$(z_1, \cdots, z_p)\begin{bmatrix} \phi_1 & \cdots & \cdots & \phi_p \\ 1 & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots \\ 0 & \cdots & 1 & 0 \end{bmatrix} = \lambda(z_1, \cdots, z_p)$$

If $z_1 = 0$, then using the previous equation, we have

$$\begin{aligned}
z_1\phi_1 + z_2 &= \lambda z_1, \\
z_1\phi_2 + z_3 &= \lambda z_2, \\
&\vdots \\
z_1\phi_{p-1} + z_p &= \lambda z_{p-1}, \\
z_1\phi_p &= \lambda z_p.
\end{aligned}$$

Thus, if $z_1 = 0$, then $z_1 = z_2 = \cdots = z_p = 0$. This is a contradiction since $z \neq 0$. This completes the proof.

In particular, when p=2, the roots of $z^2 - \phi_1 z - \phi_2 = 0$ lie inside the unit circle is equivalent to that

$$\begin{cases} \phi_2 + \phi_1 < 1 \\ \phi_2 - \phi_1 < 1 \\ -1 < \phi_2 < 1. \end{cases}$$

Basic Properties of AR(P) model: $\Omega = ES_tS_t' = \sigma^2\gamma\sum_{i=0}^{\infty} A^i\gamma\gamma' A^{i'}$

$$
\begin{aligned}
&(1)\quad Ey_t = 0\\
&(2)\quad Ey_t^2 = \sigma^2\sum_{i=0}^{\infty}(\gamma' A^i\gamma)^2\\
&(3)\quad ACF:\\
&\qquad \gamma_0 = \phi_1\gamma_1 + \cdots + \phi_p\gamma_p + \sigma_a^2,\\
&\qquad \gamma_k = \phi_1\gamma_{k-1} + \cdots + \phi_p\gamma_{k-p}, k > 0.\\
&\qquad \rho_k = \phi_1\rho_{k-1} + \cdots + \phi_p\rho_{k-p}, k > 0\\
&\qquad \left\{\begin{array}{l}\rho_1 = \phi\rho_0 + \cdots + \phi_p\rho_{p-1}\\ \rho_2 = \phi_1\rho_1 + \phi_2\rho_0 + \cdots + \phi_p\rho_{p-2}\\ \quad\vdots\\ \rho_p = \phi_1\rho_{p-1} + \cdots + \phi_p\rho_0\end{array}\right. \Rightarrow \begin{pmatrix}\rho_1\\ \rho_2\\ \vdots\\ \rho_p\end{pmatrix} =
\end{aligned}
$$

When $k \geq p+1$, $\rho_k - \phi_1\rho_{k-1} - \cdots - \phi_p\rho_{k-p}(\rho_k =?)$.

(4)PACF:

Using formula. In particular, $\psi_{kk} = 0$ when $k > p$.

For AR(2)model:

When $k = 1, 2$,

$$
\begin{aligned}
&\rho_1 = \frac{\phi_1}{1-\phi_2}\ and\ \rho_2 = \frac{\phi_1^2 + \phi_2 - \phi_2^2}{1-\phi_2},\\
&\phi_{11} = \rho_1 = \frac{\phi_1}{1-\phi_2},\\
&\phi_{22} = \phi_2,\\
&\phi_{kk} = 0\ as\ k \geq 2.
\end{aligned}
$$

2.2 Moving average processes

2.2.1 MA(1) Process

$X_t = a_t + \theta a_{t-1}$, where a_t is white noise.

Properties:

$$\mu = 0$$

$$Var(x_t) = \sigma_a^2(1+\theta^2)$$

$$\gamma_1 = -\theta_1\sigma_a^2$$

$$\gamma_k = 0, if\ k \geq 1.$$

$$ACF \begin{cases} \rho = -\frac{\theta}{1+\theta^2} \\ \rho_k = 0, if\ k \geq 1. \end{cases}$$

$$\phi_{11} \equiv \rho_1 = \frac{-\theta(1-\theta^2)}{1-\theta^4}$$

$$PACF \begin{cases} \phi_{22} = \frac{-\theta^2(1-\theta^2)}{1-\theta^6} \\ \phi_{kk} = \frac{-\theta^k(1-\theta^2)}{1-\theta^{2(k+1)}}, k \geq 1 \end{cases}$$

MA(1) process is always covariance stationary.

New problem:

Given $x_t, x_{t-1}, \cdots$,can we calculate a_t accurately from the model? Invertibility(invertible)

$$\begin{aligned} a_t &= x_t + \theta a_{t-1} \\ &= x_t + \theta x_{t-1} + \theta^2 a_{t-2} \\ &= \cdots \\ &= x_t + \theta x_{t-1} + \cdots + \theta^n x_{t-n} + \theta^{n+1} a_{t-n+1} \\ &\equiv S_{tn} + \theta^{n+1} a_{t-n+1}. \end{aligned}$$

We note that if $|\theta| < 1$, then

$$S_{tn} \overset{a.s}{\underset{m.s}{\longrightarrow}} \sum_{t=0}^{\infty} \theta^i x_{t-i},$$

$$E|\theta^{n+1} a_{t-n+1}| \leq |\theta|^{n+1} E|a_{t-n+1}| \to 0.$$

We can show that $\theta^{n+1}a_{t-n+1} \overset{a.s.}{\to} 0$. Thus, as $|\theta| < 1$, $a_t = \sum_{i=0}^{\infty} \theta^i x_{t-i}$ a.s.. In this case, we say that MA(1) model is invertible.

2.2.2 MA(q) process

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

where a_t is white noise.

(1) **Properties**:

$$\begin{array}{lrcl}
 & \mu & = & 0 \\
 & \gamma_0 & = & EX_t^2 = \sigma_a^2(1 + \theta_1^2 + \cdots + \theta_q^2) \\
 & \gamma_k & = & \begin{cases} \sigma_a^2(-\theta_k + \theta_1\theta_{k+1} + \cdots + \theta_{q-k}\theta_q) & ,k = 1, 2, \cdots, q \\ 0 & ,k > q. \end{cases} \\
ACF: & \rho_k & = & \begin{cases} \frac{-\theta_k + \theta_1\theta_{k+1} + \cdots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \cdots + \theta_q^2} & ,k = 1, 2, \cdots, q \\ 0 & ,k > q. \end{cases} \\
PACF: & & & \text{calculate from the formula.} \\
 & \phi_{kk} & = & 0(\rho^k)\ with\ |\rho| < 1
\end{array}$$

MA(q) model is always stationary (covariance).

(2) Condition for invertibility:

All the roots of $\theta_q(z) = 1 - \theta_1 z - \cdots - \theta_q z^q$ lie ouside the unit circle.

Remark:AR(p) is also always invertible.

Summary for AR(p) and MA(q) models:

$$\begin{cases} stationarity & (?) \\ invertibility & (\surd) \end{cases} \qquad \begin{cases} stationarity & (\surd) \\ invertibility & (?) \end{cases}$$

2.3 ARMA(p,q)process

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} - \theta_1 a_{t-1} \cdots - \theta_q a_{t-q} + a_t,$$

where a_t is white noise.

The ARMA(p,q) model can be written as

$$\phi(B)X_t = \theta(B)a_t,$$

where

$$\phi(B) = 1 - \phi_1 B \cdots \phi_p B^p \; and \; \theta(B) = 1 - \theta_1 B \cdots \theta_q B^q.$$

A. Condition for the stationarity:

If all the roots of $\bar{\phi}(z) = \lambda^p - \phi_1\lambda^{p-1} \cdots \phi_p = 0$ and lie inside the unit circle, then there is a unique solution X_t

$$X_t = \sum_{i=0}^{\infty} \psi_i a_{t-i}, \qquad (\psi_0 = 1)$$

where $\psi_i = 0(\rho^i)$ with $|\rho| < 1$. [it is also necessary]

B. Condition for invertibility:

If all the roots of $\bar{\theta}(z) = \lambda^q - \theta_1\lambda^{q-1} \cdots \theta_q = 0$ lie inside the unit circle, then

$$a_t = \sum_{i=0}^{\infty} \pi_i X_{t-i} \qquad (\pi_0 = 1)$$

where $\pi_i = 0(\rho^i)$ with $|\rho| < 1$.

$$\frac{\phi(B)}{\theta(B)} = \sum_{i=0}^{\infty} \pi_i B^i,$$
$$\frac{\theta(B)}{\phi(B)} = \sum_{i=0}^{\infty} \phi_i B^i,$$

Another assumption:

$\phi(B)$ *and* $\theta(B)$ have no common root.

C. ACF:

$$\gamma_k = \phi_1 \gamma_{k-1} + \cdots + \phi_p \gamma_{k-p}, k \geq q+1,$$

$$\rho_k = \phi_1 \rho_{k-1} + \cdots + \phi_p \rho_{k-p}, k \geq q+1.$$

Important feature:

$$\rho_k = O(h^k) \; with \; |h| < 1.$$

D. PACF:

$$\phi_{kk} \text{ can be obtained from } \rho_1, \cdots, \rho_k.$$

$$\phi_{kk} = O(h^k) \; with \; |h| < 1.$$

E. ARMA(1,1) model.

$$X_t = \phi_1 X_{t-1} - \theta_1 a_{t-1} + a_t \; or \; (1 - \phi_1 B) X_t = (1 - \theta_1 B) a_t.$$

The condition for stationarity is $|\phi_1| < 1$.

$$\begin{aligned} \frac{1-\theta_1 B}{1-\phi_1 B} &= 1 + (\phi_1 - \theta_1) \sum_{i=1}^{\infty} \phi_1^{i-1} B^i \\ X_t &= a_t + (\phi_1 - \theta_1) a_{t-1} + (\phi_1 - \theta_1)\phi_1 a_{t-2} + \cdots . \end{aligned}$$

The condition for invertibility is $|\theta| < 1$

$$\begin{aligned} \frac{1-\phi_1 B}{1-\theta_1 B} &= 1 + (\theta_1 - \phi_1) \sum_{i=1}^{\infty} \theta_1^{i-1} B^i \\ a_t &= X_t + (\theta_1 - \phi_1) X_{t-1} + (\theta_1 - \phi_1)\theta_1 X_{t-2} + \cdots \end{aligned}$$

The ACF is

$$\begin{aligned}
\gamma_0 &= \phi_1\gamma_1 + \sigma_a^2 - \theta_1\sigma_a^2(\phi_1 - \theta_2) \\
\gamma_1 &= \phi_1\gamma_0 - \theta_1\sigma_a^2 \\
\gamma_k &= \phi_1\gamma_{k-1}, k \geq 2. \\
\rho_0 &= 1, k = 0 \\
\rho_1 &= \frac{(\phi_1 - \theta_1)(1 - \phi_1\theta_1)}{1 + \theta_1^2 - 2\phi_1\theta_1} \\
\rho_k &= \phi_1\gamma_{k-1}, k \geq 2, \\
\rho_0 &= 1, k = 0 \\
\rho_1 &= \frac{(\phi_1 - \theta_1)(1 - \phi_1\theta_1)}{1 + \theta_1^2 - 2\phi_1\theta_1} \\
\rho_k &= \phi_1\rho_{k-1}, k \geq 2.
\end{aligned}$$

The PACF is obtained from $\rho_0, \rho_1, \rho_2, \cdots$.

2.4. AR,MA, and ARMA models with a drift

$$AR(P): \quad X_t = \underset{\substack{\uparrow \\ drift}}{\theta_0} + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + a_t \quad (2.4.1)$$

If $\phi_1 + \cdots + \phi_p \neq 1$, let $\mu = \frac{\theta_0}{1-\phi_1 \cdots - \phi_p}$. Then

$$X_t - \mu = \phi_1(X_{t-1} - \mu) + \cdots + \phi_p(X_{t-p} - \mu) + a_t.$$

Let $y_t = X_t - \mu$. Then

$$y_t = \phi_1 y_{t-1} + \cdots + \phi_\rho y_{t-p} + a_t. \quad (2.4.2)$$

So (2.4.1)$\Rightarrow$(2.4.2).

$$X_t \quad \text{is stationary} \Longleftrightarrow y_t \quad \text{is stationary}$$

$$\Updownarrow \qquad\qquad \Updownarrow$$

$$\text{all the roots of}$$

$$\lambda^p - \phi_1\lambda^{p-1}\cdots - \phi_p = 0$$

$$\text{lie inside the unit circle}$$

$$\begin{cases} Ey_t = 0 & \Rightarrow & EX_t = \mu \\ Ey_t^2 & & E(X_t^2 - \mu)^2 \\ ACF & \Rightarrow & ACF \\ PACF & & PACF \end{cases}$$

MA(q): $X_t = \mu + a_t - \theta_1 a_{t-1}\cdots - \theta_q a_{t-q}$.

Let $y_t = X_{t-q} - \mu$, Then:

$$y_t = a_t - \theta_1 a_{t-1}\cdots - \theta_q a_{t-q}.$$

$$X_t \quad \text{is invertible} \Longleftrightarrow y_t \quad \text{is invertible}$$

$$\Updownarrow \qquad\qquad \Updownarrow$$

$$\text{all the roots of}$$

$$\lambda^q - \theta_1\lambda^{q-1}\cdots - \theta_q = 0$$

$$\text{lie inside the unit circle}$$

$$\begin{cases} Ey_t = 0 & \Rightarrow & EX_t = \mu \\ Ey_t^2 & & E(X_t - \mu)^2 \\ ACF & \Rightarrow & ACF \\ PACF & & PACF \end{cases}$$

ARMA(q):

$X_t = \theta_0 + \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-a} + a_t$.

If $\phi_1 + \cdots + \phi_p \neq 1$, let $\mu = \frac{\theta_0}{1-\phi_1-\cdots\phi_p}$. Then

$$X_t - \mu \;\; = \;\; \phi_1(X_{t-1} - \mu) + \cdots + \phi_p(X_{t-p} - \mu) - \theta_1 a_{t-1}\cdots - \theta_q a_{t-q} + a_t.$$

Let $y_t = X_t - \mu$. Then

$$y_t = \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q} + a_t.$$

Thus, the stationarity and invertibility conditions of $\{X_t\}$ are the same as those of $\{y_t\}$. They also have the same variance, ACF and PACF. But $Ey_t = 0$ and $EX_t = \mu$.

3 Non-stationary TS Models

3.1. **Nonstationarity in Mean**

3.1.1. Deterministic Trend Models

Let $\{x_t\}$ be a sequence of stationary time series. Z_t is called a Deterministic Trend Model, if

$$Z_t = \alpha_0 + \alpha_1 t + x_t, \quad \alpha_1 \neq 0.$$

Z_t is not stationary.

Feature: If Z_t is a Deterministic Trend Model. then after transformed:

$$\dot{Z}_t = Z_t - \alpha_0 - \alpha_1 t$$

Then $\dot{Z}_t = x_t$ and hence $\dot{Z}_t$ is stationary.

Other Deterministic Trend Models:

$$Z_t = \alpha_0 + \alpha_1 t + \alpha_1 t^2 + x_t$$
$$Z_t = \gamma_0 + \gamma_1 \cos(t\omega + \theta) + x_t$$
$$Z_t = \gamma_0 + \sum_{j=1}^{m} (\alpha_j \cos(t\omega_j) + \beta \sin(t\omega_j) + x_t.$$

3.1.2. Stochastic Trend Models

Let $\{x_t\}$ ba a sequence of stationary time series. Z_t is called a Stochastic Trend Model, if

$$Z_t = Z_{t-1} + x_t, \qquad \text{or} \qquad (1-B)Z_t = x_t$$

Z_t is not stationary.

In particular, when $x_t = a_t$, $Z_t = Z_{t-1} + a_t$ is a random walk.

Feature: If Z_t is a Stochastic Trend Model, then after differenced:

$$\dot{Z}_t = Z_t - Z_{t-1}$$

$\dot{Z}_t$ is stationary.

General Stochastic Trend Models:

$$(1-B)^d Z_t = x_t, \qquad d \geq 1.$$

Let $\dot{Z}_t = (1-B)^d Z_t$. Then $\dot{Z}_t$ is stationary.

3.2. **Autoregressive Integrated Moving-average Model**

3.2.1. The General ARIMA Model

Let Z_t is a General Stochastic Trend Model:

$$(1-B)^d Z_t = x_t, \qquad d \geq 1.$$

If x_t is a weakly stationary and invertible ARMA model,

$$\phi_p(B)x_t = \theta_q(B)a_t\,,$$

where $\phi_p(B)$ and $\theta_q(B)$ have no common roots. Then:

$$\phi_p(B)(1-B)^d Z_t = \theta_q(B)a_t\,,$$

Z_t is called ARIMA(p, d, q) model.(?)

If x_t is the following ARMA model,

$$\phi_p(B)x_t = \theta_0 + \theta_q(B)a_t\,,$$

then,

$$\phi_p(B)(1-B)^d Z_t = \theta_0 + \theta_q(B)a_t \, ,$$

Z_t is called ARIMA(p, d, q) model.

3.2.2. The Random Walk model

ARIMA(0, 1, 0) model:

$(1-B)Z_t = a_t$ or

$Z_t = Z_{t-1} + a_t$ (random walk)

$Z_t = \theta_0 + Z_{t-1} + a_t$

—- called the random walk with drift.

Example 3.1: Simulated 100 values from

$$(1-B)Z_t = a_t \, ,$$

and

$$(1-B)Z_t = 4 + a_t \, ,$$

Show the sample ACF and PACF.

3.2.3. The ARIMA(0, 1, 1) or IMA(1, 1) Model

ARIMA(0, 1, 1) model:

$$(1-B)Z_t = (1-\theta B)a_t \, .$$

or

$$Z_t = Z_{t-1} - \theta a_{t-1} + a_t \, ,$$

where $|\theta| < 1$.

Expansion:

$$a_t = Z_t - \alpha \sum_{j=1}^{\infty} (1-\alpha)^{j-1} Z_{t-j}.$$

or

$$Z_t = \alpha \sum_{j=1}^{\infty} (1-\alpha)^{j-1} Z_{t-j} + a_t \,,$$

where $\alpha = 1 - \theta$.

Example 3.2: Simulated 100 values from three models:

ARIMA(1, 1, 0) model: $(1 - 0.8B)(1 - B)Z_t = a_t$,

ARIMA(0, 1, 1) model: $(1 - B)Z_t = (1 - 0.75B)a_t$,

ARIMA(1, 1, 1) model: $(1 - 0.9B)(1 - B)Z_t = (1 - 0.5B)a_t$,

a. Show the sample ACF and PACF.

b. Let $W_t = (1 - B)Z_t$. Show the sample ACF and PACF of W_t.

4 Model Identification

Suppose we have real data: $y_1, y_2, \cdots, y_n$. How to identify the model?

4.1 Steps for model identification

Step 1. Plot data y_t and choose proper transformations.

some common transformation:

$$Z_t = \ln y_t \,, \quad \text{or} \quad Z_t = \sqrt{y_t} \,, \quad \text{or} \quad Z_t = \frac{y_t^{\lambda} - 1}{\lambda}.$$

Denote $Z_t = f(y_t)$, where f is the transformation function.

Step 2. Compute and examine the sample ACF of Z_t.

Observe whether or not Z_t is stationary or nonstationary.

If it is stationary, then go to **Step 4**. Otherwise go to **Step 3**.

Step 3. Differenced data.

Let $x_t = Z_t - Z_{t-1}$ [i.e. $d = 1, x_t = (1-B)Z_t$]. Now, we have data x_t, and go back **Step 2**. If x_t is stationary, then go to **Step 4**, otherwise:

Let $x_t' = x_t - x_{t-1}$ [i.e. $d = 2, x_t' = (1-B)^2 Z_t$]. Now, we have data x_t', and go back **Step 2**. If x_t' is stationary, then go to **Step 4**, otherwise:

Let $x_t'' = x_t' - x_{t-1}'$ [i.e. $d = 3, x_t'' = (1-B)^3 Z_t$]. Now, we have data x_t'', and go back **Step 2**. If x_t is stationary, then go to **Step 4**, otherwise:

$\cdots$.

Usually, when $d = 1$ or 2, we can obtain a stationary data.

Step 4. Identify p and q in **ARMA**(p, q) model.

Now, we have stationary data x_t. Model is:

$$\phi_p(B)x_t = \theta_q(B)a_t\,, \quad \text{or} \quad \phi_p(B)(x_t - \mu) = \theta_q(B)a_t.$$

The criteria for identifying p and q:

	ACF	PACF
AR(p)	Tails off as exponential decay or damped sine wave	Cuts off after lag p
MA(q)	Cuts off after lag q	Tails off as exponential decay or damped sine wave
ARMA(p, q)	Cuts off after lag $q - p$	Cuts off after lag $q - p$

For the stationary **ARMA**(p, q) model, there are three important things:

1. Parameter estimation.
2. Diagnostic checking.
3. Model selection.

We will study these in other chapters .

Step 5. Final Model

1. When $x_t = Z_t$ (i.e. $d = 0$), model is

$$\phi_p(B)[f(y_t)] = \theta_q(B)a_t$$

or

$$\phi_p(B)[f(y_t) - \mu] = \theta_q(B)a_t$$

2. When $x_t = (1 - B)^d Z_t$, model is:

$$\phi_p(B)(1 - B)^d[f(y_t)] = \theta_q(B)a_t$$

or

$$\phi_p(B)[(1 - B)^d f(y_t) - \mu] = \theta_q(B)a_t$$

4.2 Empirical Examples

Example 4.1. Series W1.

Bun (1976, p.134).

daily average number of defects per truck found in inspection at the end of the assembly line of a truck manufacturing plant between 04/11-10/01 (45 observations)

Model: $(1 - \phi B)(Z_t - \mu) = a_t$.

Example 4.2. Series W2.

Yule (1927), Bartlett (1950), Whittle (1954), Brillinger and Rosenblatt (1967), $\cdots$.

This data set is the classic series of the Wolf yearly sunspot numbers from 1700-1955. Scientists believe that the sunspot numbers affect the weather of Earth and hence human activities such as agriculture, telecommunications and others.

Model:

$$(1 - \phi_1 B - \phi_2 B^2)(\sqrt{Z_t} - \mu) = a_t.$$

Box and Jenkins (1976):

$$(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3)(\sqrt{Z_t} - \mu) = a_t.$$
$$(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_9 B^9)(\sqrt{Z_t} - \mu) = a_t.$$

Example 4.4. Series W4.

Data: the U.S. monthly series of unemployed females between ages 16 and 19 from Jan. 1961 Dec. 1985.

Model: $(1 - B)(Z_t - \mu) = (1 - \theta B)a_t$.

Example 4.5. Series W5.

Data: yearly accidental death rate (per 100,000 population) of Pennsylvania (1950-1984).

Model: $(1 - B)Z_t = \theta_0 + a_t$ or $(1 - \phi B)(Z_t - \mu) = a_t$.

Example 4.6. Series W6.

Yearly U.S. tobacco production (1981-1984) published in the 1985 Agricultural Statistics by the United States Department of Revenue.

Model: $(1 - B)\ln Z_t = (1 - \theta_1 B)a_t$.

Example 4.7. Series W7.

Yearly number of lynx pelts sold by the Hudson's Bay Company in Canada between 1857 and 1911.

Model:

1. More (1953): $(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3)(\ln Z_t - \mu) = a_t$.

2. Nicholls and Quin (1982): Random coefficient AR(2) model. (rank 1).

3. Subba-Rao and Gabr (1984): bilinear model. (rank 1)

4. Davis and Blockwell (1986): AR(7) model. (rank 3)

5. Tong (1990): STAR(2; 7, 2) model. (rank 2)

5 Parameter Estimation, Diagnostic checking and Model selection

5.1 AR Model

Assume that $X_1, \cdots, X_n$ are generated by the AR(p) model:

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

where $a_t \sim$ i.i.d. $N(0, \sigma^2)$ and the true parameters are ϕ_{0i}.

Let $\tilde{X}_t = (X_t, \cdots, X_{t-p+1})$. Then the density of X_t given $X_{t-1}, \cdots, X_1$ is

$$\begin{aligned}
f(X_{p+1}|\tilde{X}_p) &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X_{p+1}-\phi'\tilde{X}_p)'}{2\sigma^2}}, \\
f(X_{p+2}|\tilde{X}_{p+1}) &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X_{p+2}-\phi'\tilde{X}_{p+1})'}{2\sigma^2}}, \\
&\vdots \\
f(X_n|\tilde{X}_{n-1}) &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X_n-\phi'\tilde{X}_{n-1})^2}{2\sigma^2}},
\end{aligned}$$

where $\phi = (\phi_1, \cdots, \phi_p)'$. Thus, the joint density of $\{X_n, X_{n-1}, \cdots, X_p + 1\}$ given $\tilde{X}_p$ is

$$f(X_n, \cdots, X_{p+1}|\tilde{X}_p) = [\frac{1}{\sqrt{2\pi\sigma^2}}]^{n-p} e^{-\sum_{t=p+1}^{n} \frac{(X_n-\phi'\tilde{X}_{n-1})^2}{2\sigma^2}}.$$

We call $f(X_n, \cdots, X_{p+1}|\tilde{X}_p)$ the conditional likelihood function.

The conditional log-likelihood function is

$$\mathcal{L}(\lambda) = -(n-p)\log(2\pi\sigma^2)^{\frac{1}{2}} - \sum_{=p+1}^{n} \frac{(X_t - \phi'\tilde{X}_{t-1})^2}{2\sigma^2},$$

where $\lambda = (\phi', \sigma^2)'$. The maximizer of $\mathcal{L}(\lambda)$ is called the (conditional) maximum likelihood estimator (MLE), denoted by $\hat{\lambda}_n$, of λ_0, where λ_0 is the true parameter of λ. $\hat{\lambda}_n$ should satisfy the equation:

$$\frac{\partial \mathcal{L}(\hat{\lambda}_n)}{\partial \lambda} = 0.$$

Thus, to find $\tilde{\lambda}_n$, we first need to solve the equation:

$$\frac{2\mathcal{L}(\lambda)}{2\lambda} = 0 \qquad .$$

Let

$$\frac{\partial \mathcal{L}(\hat{\lambda}_n)}{\partial \phi} = -\frac{1}{\sigma^2} \sum_{t=p+1}^{n} (X_t - \phi' \tilde{X}_{t-1}) \tilde{X}_{t-1} = 0.$$

We get the solution

$$\hat{\phi}_n = (\sum_{t=p+1}^{n} \tilde{X}_{t-1} \tilde{X}'_{t-1})^{-1} (\sum_{t=p+1}^{n} X_t \tilde{X}_{t-1}). \qquad (5.2)$$

Here, $\hat{\phi}_n$ is the MLE of ϕ_0. and it does not depend on σ^2.

Remark: The least squares estimator (LSE) of ϕ_0 is the minimizer of the following function:

$$L(\phi) = \sum_{t=p+1}^{n} (X_t - \phi' \tilde{X}_{t-1})^2$$

Let $\tilde{\phi}_n$ be the LSE. Then

$$\frac{\partial L(\tilde{\phi}_n)}{\partial \phi} = -2 \sum_{t=p+1}^{n} (X_t - \tilde{\phi}'_n \tilde{X}_{t-1}) \tilde{X}_{t-1} = 0.$$

$$\tilde{\phi}_n = (\sum_{t=p+1}^{n} \tilde{X}_{t-1} \tilde{X}_{t-1})^{-1} (\sum_{t=p+1}^{n} \tilde{X}_{t-1} X_t).$$

So LSE = MLE in this case.

Let $\varepsilon_t(\phi) = X_t - \phi' \tilde{X}_{t-1}$. Then

$$L(\phi) = \sum_{t=p+1}^{n} \varepsilon_t^2(\phi).$$

If a_t is not normal, MLE $\neq$ LSE.

Question: $\hat{\phi}_n$ is MLE or LSE ?

$$\begin{aligned}
\frac{\partial^2 \mathcal{L}(\lambda)}{\partial \phi \partial \phi'} &= -\frac{1}{\sigma^2}\sum_{t=1}^{n}(\tilde{X}_{t-1}\tilde{X}'_{t-1}). \\
\frac{\partial^2 L(\phi)}{\partial \phi \partial \phi'} &= \sum_{t=1}^{n}(\tilde{X}_{t-1}\tilde{X}'_{t-1}).
\end{aligned}$$

Lemma 5.1. *Assume that $X_1, \cdots, X_n$ are generated by AR(P) model and all the roots of $z^p - \phi_1 z^{p-1} \cdots \phi_p = 0$ lie inside the unit circle. Then*

$$\frac{1}{n}\sum_{t=p+1}^{n}(\tilde{X}_{t-1}\tilde{X}'_{t-1}) \xrightarrow{a.s.} E(\tilde{X}_t\tilde{X}'_t),$$

and $E(\tilde{X}_t\tilde{X}'_t) > 0$ (i.e positive definite).

Proof. By Theorem 2.1. $\{X_t\}$ is strictly stationary and ergodic,and $EX_t^2 < \infty$. Thus, $\tilde{X}_t\tilde{X}'_t$ is also strictly stationary and ergodic, and

$$E|X_{t-i}X_{t-j}| \leq \sqrt{EX_{t-i}^2 EX_{t-j}^2} < \infty$$

By ergodic theorem, for $\forall i, j$,

$$\frac{1}{n}\sum_{t=p+1}^{n} X_{t-i}X_{t-j} \xrightarrow{a.s.} \quad E(X_{t-i}X_{t-j}).$$

So

$$\frac{1}{n}\sum_{t=p+1}^{n} \tilde{X}_{t-1}\tilde{X}'_{t-1} \xrightarrow{a.s.} \quad E(\tilde{X}_{t-1}\tilde{X}'_{t-1}) = E(\tilde{X}_t\tilde{X}'_t).$$

For $\forall c = (c_1, c_2, \cdots c_p)'$,

$$c'E(\tilde{X}_t\tilde{X}'_t)c = E(c'\tilde{X}_t\tilde{X}'_t c) = E(c'\tilde{X}_t)^2 \geq 0.$$

So, $E(\tilde{X}_t\tilde{X}'_t) \geq$, i.e. semi-positive definite.

If $E(\tilde{X}_t\tilde{X}'_t)$ is not positive definite, $\exists c \neq 0$ such that

$$c'E(\tilde{X}_t\tilde{X}'_t)c = E(c'\tilde{X}_t)^2 = 0$$

$\Rightarrow c'\tilde{X}_t = 0$ a.s. $\Rightarrow c_1 X_t + c_2 X_{t-1} + \cdots + c_p X_{t-p+1} = 0.$

For simplicity, we assume that $c_1 \neq 0$. Then

$$\begin{aligned}
X_t &= -\frac{c_2}{c_1} X_{t-1} + \cdots + \frac{c_p}{c_1} X_{t-p+1} \\
&\Rightarrow \; a_t = (-\frac{c_2}{c_1} + \phi_1) X_{t-1} - \cdots - (\frac{c_p}{c_1} + \phi_{p-1}) X_{t-p+1} - \phi_p X_{t-p}.
\end{aligned}$$

This is a contradiction with the model (2.1). So $E(\tilde{X}_t \tilde{X}'_t) > 0$. $\square$

By Lemma 5.1

$$\begin{aligned}
\frac{1}{n} \frac{\partial^2 \mathcal{L}(\lambda)}{\partial \phi \partial \phi'} &\xrightarrow{a.s.} -\frac{1}{\sigma^2} E(\tilde{X}_t \tilde{X}'_t) < 0. \\
\frac{1}{n} \frac{\partial^2 L(\phi)}{\partial \phi \partial \phi'} &\xrightarrow{a.s.} E(\tilde{X}_t \tilde{X}'_t) > 0.
\end{aligned}$$

Theorem 5.1. *Under the assumptions of Lemma 2.1,*

$$\hat{\phi}_n \xrightarrow{a.s.} \phi_0, \; as \; n \to \infty.$$

Proof. First

$$\begin{aligned}
\tilde{\phi}_n - \phi_0 &= (\sum_{t=p+1}^{n} \tilde{X}_{t-1} \tilde{X}'_{t-1})^{-1} [\sum_{t=p+1}^{n} [X_{t-1} \tilde{X}'_{t-1} - \tilde{X}_{t-1} \tilde{X}'_{t-1} \phi_0]] \\
&= (\sum_{t=p+1}^{n} \tilde{X}_{t-1} \tilde{X}'_{t-1})^{-1} [\sum_{t=p+1}^{n} (\tilde{X}_{t-1} | X_t - \tilde{X}'_{t-1} \phi_0)] \\
&= (\sum_{t=p+1}^{n} \tilde{X}_{t-1} \tilde{X}'_{t-1})^{-1} \sum_{t=p+1}^{n} \tilde{X}_{t-1} a_t \\
&= (\frac{1}{n} \sum_{t=p+1}^{n} \tilde{X}_{t-1} \tilde{X}'_{t-1})^{-1} (\frac{1}{n} \sum_{t=p+1}^{n} \tilde{X}_{t-1} a_t) \quad (5.3)
\end{aligned}$$

By Lemma 5.1,

$$\frac{1}{n} \sum_{t=p+1}^{n} \tilde{X}_{t-1} \tilde{X}'_{t-1} \xrightarrow{a.s.} E(X_t X'_t) > 0.$$

By the ergodic theorem,

$$\frac{1}{n} \sum_{t=1}^{n} \tilde{X}_{t-1} a_t \xrightarrow{a.s.} E(X_t a_t) = 0.$$

So, $\tilde{\phi}_n \xrightarrow{a.s.} \phi_0$. $\square$

Definition 5.1. (consistency)

(i) If $\hat{\lambda}_n \xrightarrow{p} \lambda_0$*, we call that* $\hat{\lambda}_n$ *is a consistent estimator of* λ_0 *in probability.*

(ii) If $\hat{\lambda}_n \xrightarrow{a.s.} \lambda_0$*, we call that* $\hat{\lambda}_n$ *is a consistent estimator of* λ_0*,almost surely,(or strong consistency).*

Definition 5.2.(Asymptotic Normality) *If* $\sqrt{n}(\hat{\lambda}_n - \lambda_0 \xrightarrow{\mathcal{L}} N(0, \Omega)$*, We say that the estimator* $\hat{\lambda}_n$ *of* λ_0 *is asymptotically normal with the asymptotic covariance matrix* Ω.

Theorem 5.2. *Under the assumptions of Lemma 5.5,*

$$\sqrt{n}(\hat{\lambda}_n - \lambda_0 \xrightarrow{\mathcal{L}} N(0, \sigma^2 E^{-1}(\tilde{X}_{t-1}\tilde{X}'_{t-1})).$$

Proof. By (5.3),

$$\sqrt{n}(\hat{\lambda}_n - \lambda_0) \;=\; \underset{\hookrightarrow \Omega^{-1}}{\left(\frac{1}{n}\sum_{t=p+1}^{n} \tilde{X}_{t-1}\tilde{X}'_{t-1}\right)^{-1}} \underset{\hookrightarrow ?}{\left(\frac{1}{\sqrt{n}}\sum_{t=p+1}^{n} \tilde{X}_{t-1}a_t\right)}$$

Theorem 5.3. (Central Limit Theorem). *Suppose that* $\zeta_1, \zeta_2, \cdots$ *is an ergodic and strictly stationary time series with*

$$\sigma^2 = E\zeta_n^2 < \infty \text{ and } E(\zeta_n|\mathcal{F}_{n-1}) = 0,$$

where $\mathcal{F}_n = \sigma\{\zeta_t : t \le n\}$*. Let* $S_n = \sum_{i=1}^{n} \zeta_i$*. Then*

$$\frac{1}{\sqrt{n}} S_n \longrightarrow_{\mathcal{L}} N(0, \sigma^2) \text{ as } n \to \infty.$$

Theorem 5.4 (Grame'r-Wold device). *Let* $\{w_n\}$ *be a sequence of random vector. Then:*

$$w_n \xrightarrow{L} w \text{ if and only if } \lambda' w_n \xrightarrow{L} \lambda' w$$

for any constant vector $\lambda = (\lambda_1, \cdots, \lambda_m)'$ *with* $\lambda\lambda' \neq 0$ *where* m *is the dimension of* w_n.

Back to (5.4): Let $w_n = \frac{1}{\sqrt{n}} \sum_{t=p+1}^{n} \tilde{X}_{t-1} a_t$, For $\forall$ constant c with $c'c \neq 0$, consider

$$c' w_n = \frac{1}{\sqrt{n}} \sum_{t=p+1}^{n} \quad c' \tilde{X}'_{t-1} a_t.$$

Let $\zeta_t = c'\tilde{X}'_{t-1} a_t$. Then ζ_t is strictly stationary and ergodic, and

$$\begin{aligned} E(\zeta_t | \mathcal{F}_{t-1}) &= 0, \\ E(\zeta_t^2 |) &= c' E(\tilde{X}_{t-1} \tilde{X}'_{t-1} a_t^2) c = \sigma^2 c' E(\tilde{X}_{t-1} \tilde{X}'_{t-1}) c. \end{aligned}$$

By Theorem 5.3.

$$c' w_n \xrightarrow{L} N(0, \sigma^2 c' E(\tilde{X}_t \tilde{X}'_t) c).$$

Furthermore,by Theorem 5.4.

$$w_n \xrightarrow{L} N(0, \sigma^2 c' E(\tilde{X}_t \tilde{X}'_t) c) \qquad (5.5)$$

By (5.4),(5.5) and Lemma 5.1, we have

$$\sqrt{n}(\hat{\phi}_n - \phi_0) \xrightarrow{L} N(0, \sigma^2 c' E(\tilde{X}_t \tilde{X}'_t) c).$$

□

5.2. MA Models.

Assume that $X_1, \cdots, X_n$ are generated by the MA(1) model:

$$X_t = a_t - \theta a_{t-1},$$

where $a_t \sim N(0, \sigma^2)$ and $|\theta| < 1$ and the true value of θ is θ_0.

By the invertibility of MA(1) model,

$$a_t = \sum_{i=0}^{\infty} \theta^i X_{t-i}, \qquad (5.6)$$

$$\textit{or} \qquad X_t = -\sum_{i=1}^{\infty} \theta^i X_{t-i} + a_t. \qquad (5.7)$$

Denote $\tilde{X}_t = \{X_t, X_{t-1}, \cdots\}$. Then

$$\begin{aligned}
X_t|\tilde{X}_t &\sim N(-\sum_{i=1}^{\infty}\theta^i X_{t-i}, \sigma^2). \\
f(X_n, X_{n-1}, \cdots, X_1|\tilde{X}_0) &= \Pi_{i=1}^n f(X_i|X_{i-1}, \cdots, X_0) \\
&= [\frac{1}{\sqrt{2\pi\sigma^2}}]^n e^{-\frac{(X_t+\sum_{i=1}^{\infty}\theta^i X_{t-i})^2}{2\sigma^2}}.
\end{aligned}$$

The conditional log-likelihood function is

$$L(\theta) = -\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{n}(X_t + \sum_{i=1}^{\infty}\theta^i X_{t-i})^2.$$

Let $a_t(\theta) = X_t + \sum_{i=1}^{\infty}\theta^i X_{t-i}$. $a_t(\theta)$ is called residual.

$$L(\theta) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{n} a_t^2(\theta). \quad (5.8)$$

To maximize $L(\theta)$, we first need to maximize

$$S_n(\theta) = -\sum_{t=1}^{n} a_t^2(\theta).$$

The minimizer is called CMLE or MLE. Or to minimize

$$\tilde{S}_n(\theta) = \sum_{t=1}^{n} a_t^2(\theta) \qquad (5.9).$$

Denote

$$\hat{\theta}_n = argmax_{|\theta|\le c<1} S_n(\theta), \qquad (5.10)$$

i.e. $S_n(\hat{\theta}_n) = max_{|\theta|\le c<1} S_n(\theta)$. Suppose that $X_i = 0$ as $i \le 0$. Then $a_0(a) = 0$.

$$\begin{aligned}
a_1(\theta) &= X_1 \equiv a_1(\theta) \\
a_2(\theta) &= X_2 + \theta X_1 = X_2 + Q a_1^*(\theta) \equiv a_2^*(\theta) \\
a_3(\theta) &= X_3 + \theta^2 X_1 = X_3 + \theta a_2^*(\theta) \equiv a_3^*(\theta) \\
&\vdots \\
a_n(\theta) &= X_n + \theta X_{n-1} + \cdots + \theta^n X_1 = X_n + \theta a_{n-1}^*(\theta) \equiv a_n^*(\theta).
\end{aligned}$$

$$S_A^*(\theta) = -\sum_{t=1}^{n}[a_t^*(\theta)]^2. \qquad (5.11)$$

Denote $\hat{\theta}_n^* = argmax_{|\theta|\leq c<1} S_n^*(\theta)$. $a_t(\theta)$ and $a_t^*(\theta)$ are nonlinear functions in terms of θ.

Theorem 5.5. *If* $\theta_0, \theta \in [-c, c]$ *with* $0 < c < 1$, *then, as* $n \to \infty$,

$$(1) \quad \hat{\theta}_n \xrightarrow{p} \theta_0,$$
$$(2) \quad \hat{\theta}_n^* \xrightarrow{p} \theta_0.$$

Proof. For(1), we only need to verify the Assumption A.1.

(i) $\Theta = [-c, c]$ is compact and $f(\theta, Y_t) = -a_t^2(\theta)$ is continuous and θ_0 is an interior point in Θ. Thus, (i) holds.

(ii)

$$\begin{aligned} -Ea_t^2(\theta) &= -E[X_t + \theta a_{t-1}(\theta)]^2 \\ &= -E[-\theta_0 a_{t-1} + \theta a_{t-1}(\theta) + a_t]^2 \\ &= -Ea_t^2 - 2E[a_t(-\theta_a a_{t-1} + \theta a_{t-1}(\theta))] - E[-\theta_0 a_{t-1} + \theta a_{t-1}(\theta)]^2 \\ &= -\sigma_a^2 - E[-\theta_0 a_{t-1} + \theta a_{t-1}(\theta)]^2. \end{aligned}$$

Thus, $Ef(\theta, X_t) = -Ea_t^2(\theta)$ achieves its maximizer if and only if $E[\theta_0 a_{t-1} - \theta a_{t-1}(\theta)]^2 = 0$,

$$i.e. \qquad \theta_0 a_{t-1} - \theta a_{t-1}(\theta) = 0 \ as \quad (5.12)$$

By Taylor's expansion:

$$a_{t-1}(\theta) = a_{t-1}(\theta_0) - (\theta - \theta_0)\frac{\partial a_{t-1}(\theta^*)}{\partial \theta}.$$

By (5.12), we have

$$(\theta - \theta_0)[a_{t-1} - \theta\frac{\partial a_{t-1}(\theta^*)}{\partial \theta}] = 0. \quad (5.13)$$

If $\theta - \theta \neq 0$, then

$$\begin{aligned}
a_{t-1} &= \theta \frac{\partial a_{t-1}(\theta^*)}{\partial \theta}, \qquad\qquad (5.14) \\
\frac{\partial a_t(\theta)}{\partial \theta} &= a_{t-1}(\theta) + \theta \frac{\partial a_{t-1}(\theta)}{\partial \theta}, \\
\frac{\partial a_t(\theta)}{\partial \theta} &= \frac{1}{1-\theta B} a_{t-1}(\theta) = \sum_{i=0}^{\infty} \theta^i a_{t-1-i}(\theta) \\
&= a_{t-1}(\theta) + \theta a_{t-1}(\theta) + \cdots \\
&= g(\theta, X_{t-1}, X_{t-2}, \cdots) \qquad (5.15)
\end{aligned}$$

By (5.14)-(5.15), we know that

$$a_{t-1} = g(\theta, X_{t-2}, X_{t-3}, \cdots)$$

This is impossible since a_{t-1} is independent of $X_{t-2}, X_{t-3}, \cdots$. So, $\theta = \theta_0$ and hence $-Ea_t^2(\theta)$ achieves its maximum at and only at $\theta = \theta_0$. Thus (ii) d holds.

Now, we prove that (iii) holds.

$$\begin{aligned}
\max_{\theta \in [-c,c]} a_t^2(\theta) &\leq [\max_{\theta \in [-c,c]} |a_t(\theta)|]^2 \qquad (5.16) \\
\max_{\theta \in [-c,c]} |a_t(\theta)| &= \max_{\theta \in [-c,c]} |X_t - \sum_{i=1}^{\infty} \theta^i X_{t-i}| \\
&\leq |X_t| + \sum_{i=1}^{\infty} c^i |X_{t-i}|. \qquad (5.17)
\end{aligned}$$

By (5.16)-(5.17), we have

$$E[\max_{\theta \in [-c,c]} a_t(\theta)]^2 \leq E(|X_t| + \sum_{i=1}^{\infty} c^i |X_{t-i}|)^2 < \infty.$$

By Theorem A.1.

$$\tilde{\theta}_n \stackrel{a.s.}{\rightarrow} \theta_0, \ as \ n \rightarrow \infty,$$

i.e. (1) holds. □

Appendix 5.1. A General Asymptotic Theory

Let $\{y_t, t = 0, \pm 1, \cdots\}$ be a strictly stationary and ergodic time series.

$$L_n(\lambda) = \sum_{t=1}^{n} f(\lambda, Y_t),$$

where $Y_t = \{y_t, y_{t-1}, \cdots\}$, λ is the unknown parameter and its true value is λ_0. Assume that parameter space is Θ and $\lambda_0 \in \Theta$ and $\lambda \in \Theta$.

Let $\hat{\lambda}_n$ be the maximizer of $L_n(\lambda)$ in Θ. Denote

$$\hat{\theta}_n \quad = \quad argmax_{\lambda \in \Theta} L_n(\lambda)$$

Assumption A.1.

(i) Θ is compact and f is continuous, and λ_0 is an interior point in Θ,

(ii) $Ef(\lambda, Y_t)$ has a unique maximizer at λ_0.

(iii) $Emax_{\Theta}|f(\lambda, Y_t)| < \infty$.

Theorem A.1. *If Assumption A.1 holds, then*

$$\hat{\lambda}_n \xrightarrow{p} \lambda_0,$$

as $n \to \infty$.

Assumption A.2.

(i) $f(\lambda, Y_t)$ is twice differentiable and $\frac{\partial^2 f(\lambda, Y_t)}{\partial\lambda\partial\lambda'}$ is continuous in terms of λ;

(ii) $E[\frac{\partial f(\lambda_0, Y_t)}{\partial\lambda}|\mathcal{F}_{t-1}] = 0$ and $0 < E[\frac{\partial f(\lambda_0, Y_t)}{\partial\lambda}\frac{\partial f(\lambda_0, Y_t)}{\partial\lambda'}] < \infty$;

(iii) $\exists$ a neighbor of λ_0, $V_{\lambda 0}(\delta) = \{\lambda : ||\lambda - \lambda_0|| < \delta\}$, such that

$$E \sup_{\lambda \in V_{\lambda_0}(\delta)} ||\frac{\partial^2 f(\lambda, Y_t)}{\partial\lambda\partial\lambda}|| < \infty \textit{ and } E[\frac{\partial^2 f(\lambda_0, Y_t)}{\partial\lambda\partial\lambda'}] < 0.$$

Theorem A.2. *If Assumptions A.1, A.2 hold, then*

$$\sqrt{n}(\hat{\lambda}_n - \lambda_0) \xrightarrow{L} N(0, B^{-1}AB^{-1}),$$

where

$$A = E[\frac{\partial f(\lambda, Y_t)}{\partial \lambda}\frac{\partial f(\lambda, Y_t)}{\partial \lambda'}] \text{ and } B = E[\frac{\partial^2 f(\lambda, Y_t)}{\partial \lambda \partial \lambda'}].$$

Given $y_n, y_{n-1}, \cdots, y_1$, let $Y_t^* = \{y_t, \cdots, y_1, c_0, c_{-1}, \cdots\}$ and

$$L_n^*(\lambda) = \sum_{t=1}^{n} f(\lambda, Y_t^*),$$

$$\hat{\lambda}_n^* = argmax_{\lambda\in\Theta} L_n^*(\lambda),$$

where Θ is the parameter space.

Theorem A.3. *If Assumption A.1 holds and*

$$\max_{\lambda\in\Theta} |\frac{1}{n}\sum_{t=1}^{n}[f(\lambda, Y_t) - f(\lambda, Y_t^*)]| = o_p(1),$$

then $\hat{\lambda}_n^* \xrightarrow{p} \lambda_0$, *as* $n \to \infty$.

Theorem A.4. *If Assumptions A.1-A.2 hold, and*

$$\max_{\lambda\in\Theta} |\frac{1}{n}\sum_{t=1}^{n}[f(\lambda, Y_t) - f(\lambda, Y_t^*)]| = o_p(1),$$
$$\max_{\lambda\in\Theta} |\frac{1}{n}\sum_{t=1}^{n}[\frac{\partial^2 f(\lambda, Y_t)}{\partial \lambda \partial \lambda'} - \frac{\partial^2 f(\lambda, Y_t^*)}{\partial \lambda \partial \lambda'}]| = o_p(1),$$
$$\max_{\lambda\in\Theta} |\frac{1}{n}\sum_{t=1}^{n}[\frac{\partial f(\lambda_0, Y_t)}{\partial \lambda} - \frac{\partial f(\lambda, Y_t^*)}{\partial \lambda}]| = o_p(1).$$

Then

$$\sqrt{n}(\hat{\lambda}_n^* - \lambda_0) \xrightarrow{L} N(0, B^{-1}AB^{-1}).$$

To prove (2), we only need to prove that

$$\max_{\theta\in[-c,c]} \left| \frac{1}{n}\sum_{t=1}^{n} \left[a_t^2(\theta) - a_t^{*2}(\theta)\right] \right| = o_p(1) \tag{5.18}$$

$$\begin{aligned} a_t(\theta) - a_t^*(\theta) &= \sum_{i=0}^{\infty} \theta^i X_{t-i} - \sum_{i=0}^{t-1} \theta^i X_{t-i} \\ &= \sum_{i=t}^{\infty} \theta^i X_{t-i}. \end{aligned}$$

Thus,

$$\max_{\theta\in[-c,c]}\left|\frac{1}{n}\sum_{t=1}^{n}\left[a_t^2(\theta)-a_t^{*2}(\theta)\right]\right| \le \left[\frac{1}{n}\sum_{t=1}^{n}\left(|a_t(\theta)|+|a_t^*(\theta)|\right)\left(|a_t(\theta)-a_t^*(\theta)|\right)\right]$$
$$\le 2\left(\sum_{i=1}^{\infty}c^i|X_{t-i}|\right)\left(\sum_{i=t}^{\infty}c^i|X_{t-i}|\right) \tag{5.19}$$

For $\forall\varepsilon>0$,

$$\begin{aligned}
&\mathbb{P}\left(\max_{\theta\in[-c,c]}\left|\frac{1}{n}\sum_{t=1}^{n}\left[a_t^2(\theta)-a_t^{*2}(\theta)\right]\right|>\varepsilon\right)\\
\le\ & \frac{1}{\varepsilon}\mathbb{E}\left\{\max_{\theta\in[-c,c]}\left|\frac{1}{n}\sum_{t=1}^{n}\left[a_t^2(\theta)-a_t^{*2}(\theta)\right]\right|\right\}\\
\le\ & \frac{1}{n\varepsilon}\sum_{t=1}^{n}\mathbb{E}\left\{\max_{\theta\in[-c,c]}\left|a_t^2(\theta)-a_t^{*2}(\theta)\right|\right\}\\
\le\ & \frac{2}{n\varepsilon}\sum_{t=1}^{n}\mathbb{E}\left\{\left(\sum_{i=1}^{\infty}c^i|X_{t-i}|\right)\left(\sum_{i=t}^{\infty}c^i|X_{t-i}|\right)\right\}\\
\le\ & \frac{2\,\triangle}{n\varepsilon}\sum_{t=1}^{n}c^t\to 0 \quad \text{as } n\to\infty.
\end{aligned}$$

where $\triangle$ is some constant. Thus, (2) holds.

Theorem 5.6. *If $\theta_0,\theta\in[-c,c]$ with $0<c<1$ and $\theta_0\neq -c,c$, then, as $n\to\infty$,*

(1). $$\sqrt{n}\left(\widehat{\theta}_n-\theta_0\right)\xrightarrow{\mathscr{L}}\mathcal{N}\left(0,\ \sigma^2\mathbb{E}^{-1}\left[\frac{\partial a_t(\theta_0)}{\partial\theta}\right]^2\right)$$

(2). $$\sqrt{n}\left(\widehat{\theta}_n^*-\theta_0\right)\xrightarrow{\mathscr{L}}\mathcal{N}\left(0,\ \sigma^2\mathbb{E}^{-1}\left[\frac{\partial a_t(\theta_0)}{\partial\theta}\right]^2\right).$$

Remark. In Theorem A.2,

$$\begin{aligned}
f(\lambda,Y_t) &= -a_t^2(\theta)\\
\frac{\partial f(\lambda,Y_t)}{\partial\lambda} &= -2a_t(\theta)\frac{\partial a_t(\theta)}{\partial\theta}
\end{aligned}$$

Thus,

$$A=4\mathbb{E}\left[a_t^2(\theta)\frac{\partial a_{t-1}(\theta)}{\partial\theta}\right]^2\Bigg|_{\theta=\theta_0}=4\sigma^2\mathbb{E}\left[\frac{\partial a_t(\theta_0)}{\partial\theta}\right]^2 \tag{5.20}$$

$$\frac{\partial^2 f(\lambda, Y_t)}{\partial\lambda\partial\lambda'} = -2\frac{\partial a_{t-1}(\theta)}{\partial\theta}\frac{\partial a_t(\theta)}{\partial\theta} - 2a_t(\theta)\frac{\partial^2 a_t(\theta)}{\partial\theta\partial\theta}.$$

$$\begin{aligned} B &= -2\mathbb{E}\left[\frac{\partial a_{t-1}(\theta)}{\partial\theta}\right]^2\Bigg|_{\theta=\theta_0} - 2\mathbb{E}\left[a_t(\theta)\frac{\partial^2 a_t(\theta)}{\partial\theta\partial\theta}\right]^2\Bigg|_{\theta=\theta_0} \\ &= -2\mathbb{E}\left[\frac{\partial a_t(\theta_0)}{\partial\theta}\right]^2. \end{aligned} \tag{5.21}$$

By (5.20)—-(5.21),

$$B^{-1}AB^{-1} = \sigma^2\mathbb{E}^{-1}\left[\frac{\partial a_t(\theta_0)}{\partial\theta}\right]^2. \tag{5.22}$$

Proof of Theorem 5.6. We only need to verify the condition in Theorems A.2 and A.4.

(i) of Assumption A.2 holds. Now, we consider (ii) of Assumption A.2,

$$\mathbb{E}^{-1}\left[a_t(\theta_0)\frac{\partial a_t(\theta_0)}{\partial\theta}\middle|\mathscr{F}_{t-1}\right] = 0. \tag{5.23}$$

By (5.20), we only need to prove that

$$0 < \mathbb{E}\left(\frac{\partial a_t(\theta_0)}{\partial\theta}\right)^2 < \infty. \tag{5.24}$$

First, if $\mathbb{E}\left(\frac{\partial a_t(\theta_0)}{\partial\theta}\right)^2 = 0$, then,

$$\frac{\partial a_t(\theta_0)}{\partial\theta} = 0, \quad \text{a.s..}$$

But

$$\frac{\partial a_t(\theta)}{\partial\theta} = a_{t-1}(\theta) + \theta\,\frac{\partial a_{t-1}(\theta)}{\partial\theta},$$

we have $a_{t-1}(\theta_0) = 0$ a.s.. This is impossible.
Thus, $\mathbb{E}\left(\frac{\partial a_t(\theta_0)}{\partial\theta}\right)^2 > 0$.

$$\mathbb{E}\left[\frac{\partial a_t(\theta_0)}{\partial\theta}\right] = \mathbb{E}\left[\sum_{i=0}^{\infty}\theta_0^i a_{t-i}(\theta_0)\right]^2 < \infty.$$

Thus, (ii) of Assumption A.2 holds.

Next, we prove that (iii) of A.2 holds.

$$\mathbb{E}\sup_{|\theta|\le c}\left|\frac{\partial^2 f(\lambda, Y_t)}{\partial\lambda\partial\lambda'}\right| \le 2\mathbb{E}\sup_{|\theta|\le c}\left(\frac{\partial a_{t-1}(\theta)}{\partial\theta}\right)^2 + 2\mathbb{E}\sup_{|\theta|\le c}\left|a_t(\theta)\frac{\partial a_{t-1}(\theta)}{\partial\theta}\right|. \quad (5.25)$$

$$\begin{aligned}
\left|\frac{\partial a_t(\theta)}{\partial\theta}\right| &\le \sum_{i=0}^{\infty}|\theta|^i|a_{t-i-1}(\theta)|,\\
\left|\frac{\partial a_t(\theta)}{\partial\theta}\right|^2 &\le \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}|\theta|^i|\theta|^j|a_{t-i-1}(\theta)||a_{t-j-1}(\theta)|,\\
a_t^2(\theta) &\le \left(\sum_{i=0}^{\infty}|\theta|^i|X_{t-i}|\right)^2,\\
\mathbb{E}\sup_{|\theta|\le c} a_t^2(\theta) &\le \mathbb{E}\left(\sum_{i=0}^{\infty}c^i|X_{t-i}|\right)^2 < \infty.
\end{aligned}$$

So

$$\begin{aligned}
\mathbb{E}\sup_{|\theta|\le c}\left|\frac{\partial a_{t-1}(\theta)}{\partial\theta}\right|^2 &\le \mathbb{E}\left(\sum_{i=0}^{\infty}c^i\sup_{|\theta|\le c}|a_t^2(\theta)|\right)^2 < \infty.\\
\mathbb{E}\sup_{|\theta|\le c}\left|a_t(\theta)\frac{\partial a_{t-1}(\theta)}{\partial\theta}\right| &\le \left(\mathbb{E}\sup_{|\theta|\le c}|a_t(\theta)|^2\,\mathbb{E}\sup_{|\theta|\le c}\left|\frac{\partial a_{t-1}(\theta)}{\partial\theta}\right|^2\right)^{1/2} < \infty.
\end{aligned}$$

By (5.21),

$$\mathbb{E}\sup_{|\theta|\le c}\left|\frac{\partial^2 f(\lambda_0, Y_t)}{\partial\lambda\partial\lambda'}\right| < \infty.$$

Thus, (iii) of Assumption A.2 holds. Hence, (1) holds.

To prove (2), we need to verify conditions in Theorem A.4.. The first condition was verified on page 46. The second condition can be verified similarly as that for the third one. We only prove that the third condition is satisfied, i.e.,

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(a_t(\theta_0)\frac{\partial a_t(\theta_0)}{\partial\theta} - a_t^*(\theta_0)\frac{\partial a_t^*(\theta_0)}{\partial\theta}\right) = o_p(1) \quad (5.26)$$

$$a_t(\theta) - a_t^*(\theta) = \sum_{i=t}^{\infty}\theta^i X_{t-i}, \quad (5.27)$$

$$\begin{aligned}
\frac{\partial a_t(\theta)}{\partial\theta} - \frac{\partial a_t^*(\theta)}{\partial\theta} &= \sum_{i=0}^{\infty}\theta^i a_{t-1-i}(\theta) - \sum_{i=0}^{\infty}\theta^i a^*_{t-1-i}(\theta)\\
&= \sum_{i=0}^{t}\theta^i\left[a_{t-1-i}(\theta) - a^*_{t-1-i}(\theta)\right] + \sum_{i=t+1}^{\infty}\theta^i a_{t-1-i}(\theta). \quad (5.28)
\end{aligned}$$

For $\forall \varepsilon > 0$,

$$\begin{aligned}
&\mathbb{P}\left(\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(a_t(\theta_0)\frac{\partial a_t(\theta_0)}{\partial\theta} - a_t^*(\theta_0)\frac{\partial a_t^*(\theta_0)}{\partial\theta}\right)\right| > \varepsilon\right) \\
&\quad \le \frac{1}{\sqrt{n}\varepsilon}\mathbb{E}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(a_t(\theta_0)\frac{\partial a_t(\theta_0)}{\partial\theta} - a_t^*(\theta_0)\frac{\partial a_t^*(\theta_0)}{\partial\theta}\right)\right| \\
&\quad \le \frac{1}{\sqrt{n}\varepsilon}\sum_{t=1}^{n}\left\{\mathbb{E}\left[|a_t(\theta_0) - a_t^*(\theta_0)|\left|\frac{\partial a_t^*(\theta_0)}{\partial\theta}\right|\right] + \right. \\
&\qquad\qquad\qquad\qquad \left. + \ \mathbb{E}\left[\left|\frac{\partial a_t(\theta_0)}{\partial\theta} - \frac{\partial a_t^*(\theta_0)}{\partial\theta}\right| |a_t(\theta_0)|\right]\right\}. \qquad (5.29)
\end{aligned}$$

By (5.27),

$$\begin{aligned}
\mathbb{E}\left(a_t(\theta_0) - a_t^*(\theta_0)\right)^2 &= \mathbb{E}\left(\sum_{i=t}^{\infty}\theta_0^i X_{t-i}\right)^2 \\
&= \sum_{i=t}^{\infty}\sum_{j=t}^{\infty}\theta_0^i\theta_0^j\mathbb{E}(X_{t-i}X_{t-j}) \\
&\le \left(\mathbb{E}(X_t)^2\right)\sum_{i=t}^{\infty}\sum_{j=t}^{\infty}\theta_0^i\theta_0^j \\
&\le c_0\theta_0^{2t},
\end{aligned}$$

where c_0 is a constant. Note that $\mathbb{E}\left|\frac{\partial a_t^*(\theta_0)}{\partial\theta}\right|^2 \le c$, where c is a constant independent of t. Thus,

$$\begin{aligned}
&\mathbb{E}\left[|a_t(\theta_0) - a_t^*(\theta_0)|\left|\frac{\partial a_t^*(\theta_0)}{\partial\theta}\right|\right] \\
&\qquad \le \left(\mathbb{E}\left|a_t(\theta_0) - a_t^*(\theta_0)\right|^2\mathbb{E}\left|\frac{\partial a_t^*(\theta_0)}{\partial\theta}\right|^2\right)^{1/2} \le c_1\theta_0^t. \qquad (5.30)
\end{aligned}$$

By (5.28),

$$\begin{aligned}
&\mathbb{E}\left|\frac{\partial a_t(\theta_0)}{\partial\theta} - \frac{\partial a_t^*(\theta_0)}{\partial\theta}\right| \\
&\le \sum_{i=0}^{t}\theta_0\mathbb{E}\left|a_{t-1-i}(\theta_0) - a_{t-1-i}^*(\theta_0)\right| + \sum_{i=t+1}^{\infty}\theta_0^i\mathbb{E}|a_{t-1-i}(\theta_0)| \\
&\le c_2\theta_0^t,
\end{aligned}$$

where c_2 is a constant.

$$\mathbb{E}\left[\left|\frac{\partial a_t(\theta_0)}{\partial\theta} - \frac{\partial a_t^*(\theta_0)}{\partial\theta}\right| |a_t(\theta_0)|\right] \le c_3\theta_0^t. \qquad (5.31)$$

By (5.29), (5.30) and (5.31), we can show that

$$(5.29) \leq \frac{1}{\sqrt{n}\varepsilon} \sum_{t=1}^{n} (c_2 + c_3)\, \theta_0^t = o_p(1).$$

Hence, (5.26) holds. $\square$

Assume that $X_1, \cdots, X_n$ are generated by the MA(q) model:

$$X_t = a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}$$

where $a_t \sim i.i.d.\mathcal{N}(0, \sigma^2)$ and all the roots of $\lambda^q - \theta_1 \lambda^{q-1} - \cdots - \theta_q = 0$ lie inside the unit circle.

Let $\theta = (\theta_1, \cdots, \theta_q)'$ and $\lambda = (\theta', \sigma^2)$. By the invertibility of MA(q) model,

$$a_t = \sum_{i=0}^{\infty} \psi_i X_{t-i}, \qquad \psi_i = O(\rho^i) \text{ with } |\rho| < 1,$$

or,

$$X_t = -\sum_{i=1}^{\infty} \psi_i X_{t-i} + a_t.$$

Denote $\tilde{X}_t = \{X_t, X_{t-1}, \cdots, \}$, then,

$$X_t | \tilde{X}_{t-1} \sim \mathcal{N}(-\sum_{i=1}^{\infty} \psi_i X_{t-i},\ \sigma^2),$$

$$f(X_n, X_{n-1}, \cdots, X_1 | \tilde{X}_0) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left\{ -\frac{\sum_{t=1}^{n} \left(X_t + \sum_{i=1}^{\infty} \psi_i X_{t-i}\right)^2}{2\sigma^2} \right\}$$

The conditional log-Likelihood function is

$$\mathcal{L}(\lambda) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{t=1}^{n} \left(X_t + \sum_{i=1}^{\infty} \psi_i X_{t-i}\right)^2.$$

Let $a_t(\theta) = X_t + \sum_{i=1}^{\infty} \psi_i X_{t-i} = \theta^{-1}(B)X_t$.

$$
\begin{aligned}
\mathcal{L}(\lambda) &= -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{n} a_t^2(\theta) \\
S_n(\theta) &= -\sum_{t=1}^{n} a_t^2(\theta) \\
\widehat{\theta}_n &= \arg\max_{\theta\in\Theta}\left(-\sum_{t=1}^{n} a_t^2(\theta)\right)
\end{aligned}
$$

Let $X_i = 0$, as $i \le 0$.

$$\theta(B)a_t(\theta) = X_t.$$

$$
\begin{aligned}
&a_t(\theta) - \theta_1 a_{t-1}(\theta) - \cdots - \theta_q a_{t-q}(\theta) = X_t, \\
&a_t(\theta) = X_t + \theta_1 a_{t-1}(\theta) + \cdots + \theta_q a_{t-q}(\theta).
\end{aligned}
$$

$$
\begin{aligned}
&a_1^*(\theta) = a_1(\theta) = X_1; \\
&a_2^*(\theta) = X_1 + \theta_1 a_1^*(\theta); \\
&a_3^*(\theta) = X_2 + \theta_1 a_2^*(\theta) + \theta_2 a_1^*(\theta); \\
&\vdots \\
&a_n^*(\theta) = X_n + \theta_1 a_{n-1}^*(\theta) + \cdots + \theta_q a_{t-q}^*(\theta).
\end{aligned}
$$

$$S_n^*(\theta) = -\sum_{t=1}^{n} a_t^*(\theta).$$

Denote

$$\theta_n^* = \arg\max_{\theta\in\Theta} S_n^*(\theta).$$

$\widehat{\theta}_n$ and θ_n^* are called the (conditional) MLE of θ_0.

Theorem 5.7. *Suppose that the parameter space Θ is compact subset of $\mathbb{R}^q$, and for each $\theta \in \Theta$, all the roots of $\lambda^q - \theta_1\lambda^{q-1} - \cdots - \theta_q = 0$ lie* *inside the unit circle. Then,*

$$(i) \qquad \widehat{\theta}_n,\ \theta_n^* \xrightarrow{\mathbb{P}} \theta_0.$$

$$(ii) \qquad \sqrt{n}\left(\widehat{\theta}_n - \theta_0\right) \xrightarrow{\mathscr{L}} \mathcal{N}(0,\Omega),$$

$$\sqrt{n}\,(\theta_n^* - \theta_0) \xrightarrow{\mathscr{L}} \mathcal{N}(0,\Omega),$$

where

$$\Omega = \sigma^2 \mathbb{E}^{-1}\left[\frac{\partial a_t(\theta_0)}{\partial \theta}\frac{\partial a_t(\theta_0)}{\partial \theta'}\right].$$

5.3. ARMA Models

$$\phi(B)X_t = \theta(B)a_t,$$

where $\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$ and $\theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$. Denote $\lambda = (\phi_1, \phi_2, \cdots, \phi_p, \theta_1, \cdots, \theta_q)'$.

$$\begin{aligned} X_t &= \phi^{-1}(B)\theta(B)a_t, \\ a_t &= \theta^{-1}\phi(B)X_t \\ &= X_t - \sum_{i=1}^{\infty} \pi_i X_{t-i}, \\ X_t &= a_t + \sum_{i=1}^{\infty} \pi_i X_{t-i}. \end{aligned}$$

Given $\tilde{X}_t = \{X_t, X_{t-1}, \cdots\}$.

$$X_t|\tilde{X}_{t-1} \sim \mathcal{N}\left(\sum_{i=1}^{\infty} \pi_i X_{t-i},\ \sigma^2\right),$$

$$f(X_n, X_{n-1}, \cdots, X_1|\tilde{X}_0) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{n}\left(X_t - \sum_{i=1}^{\infty}\pi_i X_{t-i}\right)^2\right\}$$

The conditional log-Likelihood function is

$$\mathcal{L}(\lambda) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{n}\left(X_t - \sum_{i=1}^{\infty}\pi_i X_{t-i}\right)^2. \qquad (5.32)$$

Let

$$
\begin{aligned}
a_t(\lambda) &= X_t - \sum_{i=1}^{\infty} \pi_i X_{t-i}, \quad \text{——— residuals,} \\
\mathcal{L}(\lambda) &= -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{n} a_t^2(\lambda), \\
S_n(\lambda) &= -\sum_{t=1}^{n} a_t^2(\lambda).
\end{aligned}
$$

Denote

$$
\widehat{\lambda}_n = \arg\max_{\lambda\in\Theta} S(\lambda) \quad \text{——— } \textbf{MLE} \text{ of } \lambda_0.
$$

When $X_i = 0$, as $i \le 0$.

$$
a_t^*(\lambda) \equiv a_t(\lambda)\Big|_{\substack{X_i=0 \\ \text{as } i\le 0}} = X_t - \sum_{i=1}^{t-1} \pi_i X_{t-i}, \tag{5.33}
$$

$$
a_t^* = X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} - \theta_1 a_{t-1}^*(\lambda) - \cdots - \theta_q a_{t-q}^*(\lambda), \tag{5.34}
$$

$$
\mathcal{L}(\lambda) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{n} a_t^*(\lambda), \tag{5.35}
$$

$$
S_n^*(\lambda) = -\sum_{t=1}^{n} a_t^*(\lambda),
$$

$$
\widehat{\lambda}_n^* = \arg\max_{\lambda\in\Theta} S_n^*(\lambda) \quad \text{——— } \textbf{MLE} \text{ of } \lambda_0.
$$

Theorem 5.8. *Suppose that the parameter space Θ is compact subset of $\mathbb{R}^{p+q}$, and for each $\theta \in \Theta$, all the roots of $\lambda^p - \phi_1\lambda^{p-1} - \cdots - \phi_p = 0$ and $\lambda^q - \theta_1\lambda^{q-1} - \cdots - \theta_q = 0$ lie inside the unit circle, and they have not common roots. Then,*

(i) $$\widehat{\lambda}_n,\ \widehat{\lambda}_n^* \xrightarrow{\mathbb{P}} \theta_0.$$

(ii) $$\sqrt{n}\left(\widehat{\lambda}_n - \lambda_0\right) \xrightarrow{\mathscr{L}} \mathcal{N}(0,\Omega),$$

$$\sqrt{n}\left(\widehat{\lambda}_n^* - \lambda_0\right) \xrightarrow{\mathscr{L}} \mathcal{N}(0,\Omega),$$

where

$$
\Omega = \sigma^2 \mathbb{E}^{-1}\left[\frac{\partial a_t(\lambda_0)}{\partial \lambda}\frac{\partial a_t(\lambda_0)}{\partial \lambda'}\right].
$$

$\square$

How to find the solution for the following equation

$$\frac{\partial S_n^*(\lambda)}{\partial \lambda} = 0 \quad ?$$

$$\begin{aligned}
&\frac{\partial S_n^*(\widehat{\lambda}_n^*)}{\partial \lambda} = 0, \\
&\frac{\partial S_n^*(\widehat{\lambda}_n^*)}{\partial \lambda} - \frac{\partial S_n^*(\lambda_0^{(0)})}{\partial \lambda} = \left(\widehat{\lambda}_n^* - \lambda_0^{(0)}\right)' \frac{\partial^2 S_n^*(\xi^{(0)})}{\partial \lambda \partial \lambda'}, \\
&\widehat{\lambda}_n^* - \lambda_0^{(0)} = -\left[\frac{\partial^2 S_n^*(\xi^{(0)})}{\partial \lambda \partial \lambda'}\right]^{-1} \left[\frac{\partial S_n^*(\lambda_0^{(0)})}{\partial \lambda}\right].
\end{aligned}$$

$$\begin{aligned}
&\lambda^{(1)} - \lambda^{(0)} = -\left[\frac{\partial^2 S_n^*(\lambda^{(0)})}{\partial \lambda \partial \lambda'}\right]^{-1} \left[\frac{\partial S_n^*(\lambda^{(0)})}{\partial \lambda}\right], \\
&\widehat{\lambda}_n^* - \lambda^{(1)} = ?
\end{aligned}$$

$$\begin{aligned}
&\frac{\partial S_n^*(\widehat{\lambda}_n^*)}{\partial \lambda} - \frac{\partial S_n^*(\lambda^{(1)})}{\partial \lambda} = \left(\widehat{\lambda}_n^* - \lambda^{(1)}\right)' \frac{\partial^2 S_n^*(\xi^{(1)})}{\partial \lambda \partial \lambda'}, \\
&\widehat{\lambda}_n^* - \lambda^{(1)} = -\left[\frac{\partial^2 S_n^*(\xi^{(1)})}{\partial \lambda \partial \lambda'}\right]^{-1} \left[\frac{\partial S_n^*(\lambda^{(1)})}{\partial \lambda}\right], \\
&\lambda^{(2)} - \lambda^{(1)} = -\left[\frac{\partial^2 S_n^*(\lambda^{(1)})}{\partial \lambda \partial \lambda'}\right]^{-1} \left[\frac{\partial S_n^*(\lambda^{(1)})}{\partial \lambda}\right]. \\
&\widehat{\lambda}_n^* - \lambda^{(2)} = ?
\end{aligned}$$

$$\lambda^{(m)} - \lambda^{(m-1)} = -\left[\frac{\partial^2 S_n^*(\lambda^{(m-1)})}{\partial \lambda \partial \lambda'}\right]^{-1} \left[\frac{\partial S_n^*(\lambda^{(m-1)})}{\partial \lambda}\right].$$

$\lambda^{(0)}, \lambda^{(1)}, \cdots, \lambda^{(m)} \longrightarrow \widehat{\lambda}_n^*$, as $m \to \infty$. $\square$

5.4. Diagnostic Checking

Let $a_1, a_2, \cdots, a_n$ be i.i.d. white noises.

$$\begin{aligned}
\widehat{\gamma}_k &= \frac{1}{n}\sum_{t=1}^{n-k} a_t a_{t+k} \xrightarrow{\text{a.s.}} \mathbb{E}(a_t a_{t+k}) = 0; \\
\widehat{\rho}_k &= \frac{\widehat{\gamma}_k}{\widehat{\gamma}_0} = \frac{\sum_{t=1}^{n-k} a_t a_{t+k}}{\sum_{t=1}^{n} a_t^2} \xrightarrow{\text{a.s.}} 0 \quad \text{if } k \geq 1.
\end{aligned}$$

Denote $\widehat{\rho} = (\widehat{\rho}_1, \widehat{\rho}_2, \cdots, \widehat{\rho}_{_M})'$. Then $\widehat{\rho} \xrightarrow{\text{a.s.}} 0$.

Theorem 5.9. *If a_t is i.i.d. white noises with $\mathbb{E}a_t^2 = \sigma^2 < \infty$, then,*

$$\sqrt{n}\widehat{\rho} \xrightarrow{\mathscr{L}} \mathcal{N}(0, I_{_M}),$$

as $n \to \infty$.

Proof.

$$\sqrt{n}\widehat{\rho} = \frac{1}{\frac{1}{n}\sum\limits_{t=1}^{n} a_t^2} \cdot \frac{1}{\sqrt{n}} \left(\sum_{t=1}^{n-1} a_t a_{t+1}, \cdots, \sum_{t=1}^{n-M} a_t a_{t+M} \right)'$$

$$\begin{aligned} \frac{1}{\sqrt{n}} \left(\sum_{t=1}^{n-1} a_t a_{t+1}, \cdots, \sum_{t=1}^{n-M} a_t a_{t+M} \right)' &= \frac{1}{\sqrt{n}} \left(\sum_{t=1}^{n-M} a_t a_{t+1}, \cdots, \sum_{t=1}^{n-M} a_t a_{t+M} \right)' \\ &\quad + \frac{1}{\sqrt{n}} \left(\sum_{t=n-M+1}^{n-1} a_t a_{t+1}, \cdots, \sum_{t=n-M+1}^{n-M+1} a_t a_{t+M}, 0 \right)' \\ &\equiv A + B. \end{aligned}$$

$$\mathbb{E}\left(\frac{1}{\sqrt{n}} \sum_{t=n-M+1}^{n-1} a_t a_{t+1} \right)^2 = \frac{1}{n} \sum_{t=n-M+1}^{n-1} \mathbb{E}(a_t^2 a_{t+1}^2) = \frac{\sigma^4}{n} \cdot M \to 0.$$

So,

$$\frac{1}{\sqrt{n}} \sum_{t=n-M+1}^{n-1} a_t a_{t+1} = o_p(1).$$

Similarly, we can show that other elements in B are $o_p(1)$, and hence $B = o_p(1)$.

We now consider A:

$$A = \frac{1}{\sqrt{n}} \left(\sum_{t=1}^{n-M} a_t a_{t+1}, \cdots, \sum_{t=1}^{n-M} a_t a_{t+M} \right)'.$$

Let $\xi_t = c'(a_t a_{t+1}, \cdots, a_t a_{t+M})'$. Then $\{\xi_t\}$ is a sequence of strictly stationary and ergodic time series, $\mathbb{E}(\xi_t | \mathscr{F}_{t-1}) = 0$.

$$\begin{aligned} \mathbb{E}\xi_t^2 &= \sum_i \sum_j \mathbb{E}(c_i c_j a_{t+i} a_{t+i+1} a_{t+j} a_{t+j+1}) \\ &= \sum_{i=1}^{M} c_i^2 \mathbb{E}(a_t^2 a_{t+i}^2) = (c' I_{_M} c)\sigma^4. \end{aligned}$$

By the Central Limit Theorem, we have

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\xi_t \xrightarrow{\mathscr{L}} \mathcal{N}(0,(c'I_{M}c)\sigma^4).$$

Further, by Cramer-Wold device, we know that

$$A \xrightarrow{\mathscr{L}} \mathcal{N}(0,\sigma^4 I_{M}).$$

Thus,

$$\frac{1}{\sqrt{n}}\left(\sum_{t=1}^{n-1}a_t a_{t+1},\cdots,\sum_{t=1}^{n-M}a_t a_{t+M}\right)' \xrightarrow{\mathscr{L}} \mathcal{N}(0,\sigma^4 I_{M}). \tag{5.36}$$

$$\frac{1}{n}\sum_{t=1}^{n}a_t^2 \xrightarrow{\text{a.s.}} \sigma^2. \tag{5.37}$$

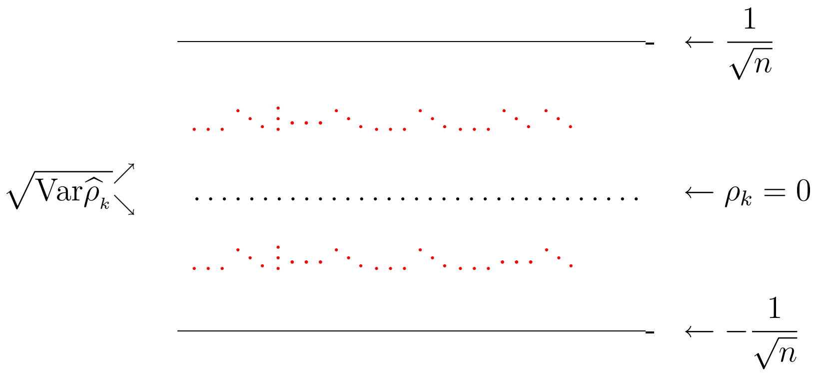
By (5.36)-(5.37), we know that

$$\sqrt{n}\widehat{\rho} \xrightarrow{\mathscr{L}} \mathcal{N}(0,I_{M}). \qquad (5.38)\square$$

By Theorem 5.9,

$$\text{Var}\left(\sqrt{n}\widehat{\rho}_{k}\right) \approx 1, \qquad \text{as } n \text{ is large enough.}$$

Thus, $\sqrt{\text{Var}\left(\widehat{\rho}_{k}\right)} \approx \frac{1}{\sqrt{n}}$.



Hypothesis:

$$H_0: \quad \rho_1 = \rho_2 = \cdots = \rho_{M} = 0;$$

$$H_1: \quad H_0 \text{ does not hold.}$$

Test statistics:

$$Q = n \sum_{k=1}^{M} \widehat{\rho}_k^2 \sim \chi^2(M)$$

or

$$Q = \frac{n(n+2)}{n-M} \sum_{k=1}^{M} \widehat{\rho}_k^2 \sim \chi^2(M).$$

Q is called portmanteau test (or Ljung-Box test).

If $Q < \chi_{0.05}$, we accept H_0, (5% possible error).

If $Q \geq \chi_{0.05}$, we reject H_0.

Do you believe this test?

We can verify this test by simulation method. From computer, we can generate a data set.

$$\begin{aligned} \bar{X}_1 &\equiv \{a_1, a_2, \cdots, a_n\} \quad \text{——— calculate } Q \text{ values.} \\ \bar{X}_2 &\equiv \{a_1, a_2, \cdots, a_n\} \quad \text{——— calculate } Q \text{ values.} \\ &\vdots \\ \bar{X}_m &\equiv \{a_1, a_2, \cdots, a_n\} \quad \text{——— calculate } Q \text{ values.} \end{aligned}$$

n ——— call the sample size.

$\bar{X}_1, \bar{X}_2, \cdots, \bar{X}_m$ ——— replication.

m ——— the number of replications.

k denotes the number of Q values greater than $\chi_{0.05}$.

The reject frequency:

$\frac{k}{m}$ ——— size if $a_1, a_2, \cdots, a_n$ are white noises.

$\frac{k}{m}$ ——— power if $a_1, a_2, \cdots, a_n$ are not white noises.

$\frac{k}{m} \approx 0.05$ as n is large and $a_1, a_2, \cdots, a_n$ are white noises.

The power depends on data. If the data are far away from white noise, then power becomes large, and $\to 1$ as $n \to \infty$.

Given a sequence of data: $X_1, X_2, \cdots, X_n$, we use an AR model to fit the data:

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + a_t, \tag{5.39}$$

where $\{a_t\}$ are i.i.d. white noises. Then, we obtain the residuals:

$$\widehat{a}_t \equiv a_t(\widehat{\phi}_n) = X_t - \widehat{\phi}_{1n} X_{t-1} - \cdots - \widehat{\phi}_{pn} X_{t-p}$$

where $\widehat{\phi}_n = (\widehat{\phi}_{1n}, \cdots, \widehat{\phi}_{pn})'$. Define:

$$\begin{aligned} \widehat{\gamma}_k &= \frac{1}{n} \sum_{t=1}^{n-k} \widehat{a}_t \widehat{a}_{t+k}, \\ \widehat{\rho}_k &= \frac{\widehat{\gamma}_k}{\widehat{\gamma}_0} = \frac{\sum_{t=1}^{n-k} \widehat{a}_t \widehat{a}_{t+k}}{\sum_{t=1}^{n} \widehat{a}_t^2}. \end{aligned}$$

Theorem 5.10. *If model (5.39) is correct for the data $\{X_1, X_2, \cdots, X_n\}$ with a_t being i.i.d. white noise and $\mathbb{E} a_t^2 = \sigma^2 < \infty$, then*

$$\widehat{\rho}_k \xrightarrow{a.s.} 0 \qquad \text{as } n \to \infty,$$

where $\widehat{\phi}_n$ is the LSE or MLE of ϕ_0 in Theorem 5.1..

Proof.

$$\widehat{a}_t = X_t - \widehat{\phi}_n' \tilde{X}_{t-1},$$

where

$$\tilde{X}_{t-1} = (X_{t-1}, \cdots, X_{t-p})'$$

Note that

$$X_t = \phi_0' \tilde{X}_{t-1} + a_t \tag{5.40}$$

$$\widehat{a}_t = a_t - \left(\widehat{\phi}_n - \phi_0\right)' \tilde{X}_{t-1}.$$

Thus,

$$\widehat{a}_t^2 = a_t^2 - 2\left(\widehat{\phi}_n - \phi_0\right)' \tilde{X}_{t-1} a_t + \left(\widehat{\phi}_n - \phi_0\right)' \tilde{X}_{t-1}\tilde{X}_{t-1}' \left(\widehat{\phi}_n - \phi_0\right). \quad (5.41)$$

By (5.3),

$$\widehat{\phi}_n - \phi_0 = \left(\frac{1}{n}\sum_{t=p+1}^{n} \tilde{X}_{t-1}\tilde{X}_{t-1}'\right)^{-1} \left(\frac{1}{n}\sum_{t=p+1}^{n} \tilde{X}_{t-1} a_t\right). \quad (5.42)$$

By (5.41)—-(5.42),

$$\begin{aligned}\frac{1}{n}\sum_{t=p+1}^{n} \widehat{a}_t^2 = \frac{1}{n}\sum_{t=p+1}^{n} a_t^2 &- 2\left(\frac{1}{n}\sum_{t=p+1}^{n} \tilde{X}_{t-1} a_t\right)' \left(\frac{1}{n}\sum_{t=p+1}^{n} \tilde{X}_{t-1}\tilde{X}_{t-1}'\right)^{-1} \left(\frac{1}{n}\sum_{t=p+1}^{n} \tilde{X}_{t-1} a_t\right) \\ &+ \left(\frac{1}{n}\sum_{t=p+1}^{n} \tilde{X}_{t-1} a_t\right)' \left(\frac{1}{n}\sum_{t=p+1}^{n} \tilde{X}_{t-1}\tilde{X}_{t-1}'\right)^{-1} \left(\frac{1}{n}\sum_{t=p+1}^{n} \tilde{X}_{t-1} a_t\right). \\ &\xrightarrow{\text{a.s.}} \sigma^2. \end{aligned} \quad (5.43)$$

(So σ^2 can be estimated by $\frac{1}{n}\sum_{t=p+1}^{n} \widehat{a}_t^2$.)

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n-k} \widehat{a}_t \widehat{a}_{t+k} = \frac{1}{n}\sum_{t=p+1}^{n} a_t a_{t+k} &- \left(\widehat{\phi}_n - \phi_0\right)' \left(\frac{1}{n}\sum_{t=1}^{n-k} \tilde{X}_{t-1} a_{t+k}\right) \\ &- \left(\widehat{\phi}_n - \phi_0\right)' \left(\frac{1}{n}\sum_{t=1}^{n-k} \tilde{X}_{t+k-1} a_t\right) \\ &+ \left(\widehat{\phi}_n - \phi_0\right)' \left(\frac{1}{n}\sum_{t=1}^{n-k} \tilde{X}_{t-1}\tilde{X}_{t+k-1}'\right) \left(\widehat{\phi}_n - \phi_0\right) \\ &\xrightarrow{\text{a.s.}} 0. \end{aligned} \quad (5.44)$$

by ergodic theorem. Thus, we know that

$$\widehat{\rho}_k \xrightarrow{\text{a.s.}} 0 \qquad \text{as } n \to \infty. \qquad \square$$

Theorem 5.11. Under the assumption of Theorem 5.1,

$$\sqrt{n}\left(\widehat{\rho}_1, \widehat{\rho}_2, \cdots, \widehat{\rho}_M\right) \xrightarrow{\mathscr{L}} \mathcal{N}(0, \Sigma),$$

as $n \to \infty$, where $\Sigma = I - D'(\Omega^{-1} \otimes I_{_M})D$, $\Omega = \mathbb{E}\left(\tilde{X}_t \tilde{X}_t'\right)$, $D = \sigma^{-2}(D_1, D_2, \cdots, D_{_M})'$ and $D_k = \mathbb{E}\left(\tilde{X}_{t+k-1} a_t\right)$.

Proof. By (5.44),

$$
\begin{aligned}
\frac{1}{\sqrt{n}} \sum_{t=1}^{n-k} \widehat{a}_t \widehat{a}_{t+k} &= \frac{1}{\sqrt{n}} \sum_{t=p+1}^{n} a_t a_{t+k} - \sqrt{n}\left(\widehat{\phi}_n - \phi_0\right)' \left(\frac{1}{n}\sum_{t=1}^{n-k} \tilde{X}_{t-1} a_{t+k}\right) \\
&\quad -\sqrt{n}\left(\widehat{\phi}_n - \phi_0\right)' \left(\frac{1}{n}\sum_{t=1}^{n-k} \tilde{X}_{t+k-1} a_t\right) \\
&\quad +\sqrt{n}\left(\widehat{\phi}_n - \phi_0\right)' \left(\frac{1}{n}\sum_{t=1}^{n-k} \tilde{X}_{t-1} \tilde{X}_{t+k-1}'\right)\left(\widehat{\phi}_n - \phi_0\right) \\
&= \frac{1}{\sqrt{n}} \sum_{t=1}^{n-k} a_t a_{t+k} - \sqrt{n}\left(\widehat{\phi}_n - \phi_0\right)' D_k + o_p(1) \\
&= \frac{1}{\sqrt{n}} \sum_{t=1}^{n-k} a_t a_{t+k} - D_k' \left(\frac{1}{n}\sum_{t=p+1}^{n} \tilde{X}_{t-1} \tilde{X}_{t-1}'\right)^{-1} \left(\frac{1}{\sqrt{n}}\sum_{t=p+1}^{n} \tilde{X}_{t-1} a_t\right) + o_p(1) \\
&= \frac{1}{\sqrt{n}} \sum_{t=1}^{n-k} a_t a_{t+k} - D_k'\Omega^{-1} \left(\frac{1}{\sqrt{n}}\sum_{t=p+1}^{n} \tilde{X}_{t-1} a_t\right) + o_p(1) \\
&= \frac{1}{\sqrt{n}} \sum_{t=1}^{n-M} a_t a_{t+k} + \frac{1}{\sqrt{n}} \sum_{t=n-M+1}^{n-k} a_t a_{t+k} + \frac{1}{\sqrt{n}} \sum_{t=k+1}^{n-M+k} \left(D_k'\Omega^{-1}\tilde{X}_{t-1} a_t\right) \\
&\quad + \frac{1}{\sqrt{n}} \sum_{t=n-M+k+1}^{n} \left(D_k'\Omega^{-1}\tilde{X}_{t-1} a_t\right) + \frac{1}{\sqrt{n}} \sum_{t=p+1}^{k} \left(D_k'\Omega^{-1}\tilde{X}_{t-1} a_t\right) + o_p(1) \\
&= \frac{1}{\sqrt{n}} \sum_{t=1}^{n-M} a_t a_{t+k} + \frac{1}{\sqrt{n}} \sum_{t=1}^{n-M} D_k'\Omega^{-1}\tilde{X}_{t+k-1} a_{t+k} + o_p(1) \\
&= \frac{1}{\sqrt{n}} \sum_{t=1}^{n-M} \left(a_t - D_k'\Omega^{-1}\tilde{X}_{t+k-1}\right) a_{t+k} + o_p(1).
\end{aligned}
$$

So,

$$
\begin{aligned}
&\frac{1}{\sqrt{n}} \left(\sum_{t=1}^{n-k} \widehat{a}_t \widehat{a}_{t+k}, \cdots, \sum_{t=1}^{n-M} \widehat{a}_t \widehat{a}_{t+M}\right)' \\
&\quad = \frac{1}{\sqrt{n}} \sum_{t=1}^{n-M} \left(\left(a_t - D_1'\Omega^{-1}\tilde{X}_t\right) a_{t+1}, \left(a_t - D_2'\Omega^{-1}\tilde{X}_{t+1}\right) a_{t+2}, \cdots,\right. \\
&\qquad\qquad \left. \cdots, \left(a_t - D_{_M}'\Omega^{-1}\tilde{X}_{t+M-1}\right) a_{t+M}\right) + o_p(1) \\
&\quad \equiv \frac{1}{\sqrt{n}} \sum_{t=1}^{n-M} Y_t + o_p(1).
\end{aligned}
$$

Let $\xi_t \equiv c\,'Y_t = \sum_{k=1}^{M} c_k \left(a_t - D_k'\Omega^{-1}Y_{t+k-1}\right) a_{t+k}$. Then,

$$\begin{aligned}
\mathbb{E}\left(\xi_t|\mathscr{F}\right) &= 0. \\
\mathbb{E}\left(\xi_t^2|\mathscr{F}\right) &= \sigma^2 \sum_{k=1}^{M} c_k^2 \mathbb{E}\left(a_t - D_k'\Omega^{-1}Y_{t+k-1}\right)^2 \\
&= \sigma^2 \sum_{k=1}^{M} c_k^2 \left(\sigma^2 - D_k'\Omega^{-1}D_k\right)^2 \\
&= \sigma^4 \left(\sum_{k=1}^{M} c_k^2 - \frac{1}{\sigma^2}\sum_{k=1}^{M}(c_k^2 D_k'\Omega^{-1}D_k)\right) \\
&= \sigma^4 \left(c\,'Ic - \frac{1}{\sigma^2}c\,'\widehat{\Omega}c\right) \\
&= \sigma^4 c\,'\left(I - \widehat{\Omega}\right)c,
\end{aligned}$$

where

$$\begin{aligned}
\widehat{\Omega} &= \mathbf{diag}\left\{\frac{D_1'\Omega^{-1}D_1}{\sigma^2}, \frac{D_2'\Omega^{-1}D_2}{\sigma^2}, \cdots, \frac{D_{_M}'\Omega^{-1}D_{_M}}{\sigma^2}\right\} \\
&= D'\left(\Omega^{-1}\otimes I_{_M}\right)D.
\end{aligned}$$

Hence,

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\xi_t \xrightarrow{\mathscr{L}} \mathcal{N}\left(0, \quad \sigma^4 c\,'\left(I - \widehat{\Omega}\right)c\right).$$

By Cramer-Wold device,

$$\sqrt{n}\widehat{\rho} \xrightarrow{\mathscr{L}} \mathcal{N}\left(0, \quad I - \widehat{\Omega}\right), \qquad \text{as } n \to \infty.$$

□

Hypothesis:

H_0 : Model (5.3) is corrected.

H_1 : H_0 does not hold.

Test Statistics:

$$\begin{aligned}
Q(M) &= n\widehat{\rho}\,'\left(I - \widehat{\Omega}\right)\widehat{\rho} \\
&= \frac{n}{\sigma^2}\sum_{k=1}^{M}\widehat{\rho}_k^{\,2}\left(1 - D_k'\Omega^{-1}D_k\right) \sim \chi^2(M).
\end{aligned}$$

$Q(M)$ is called portmanteau test for diagnostic checking the adequacy of Model (5.39).

Box and Pierce (1970):

$$Q_1(M) = n\widehat{\rho}'\widehat{\rho} \approx n\sum_{k=1}^{M}\widehat{\rho}_k^2.$$

Ljung and Box (1978) defined:

$$Q_2(M) = n(n+2)\sum_{k=1}^{M}(n-k)^{-1}\widehat{\rho}_k^2.$$

$$Q(M) > Q_2(M) > Q_1(M).$$

6 Forecasting

Objective of this Chapter:

Given you a sequence of data: $Z_1, Z_2, \cdots, Z_n$, ARMA or ARIMA model, you can forecast $Z_{t+1}, \cdots, Z_{n+l}$ and give its forecasting interval.

6.1. **Minimum Mean Square Error Forecasts for ARMA models**

Let Z_t be a stationary and invertible ARMA model,

$$\phi(B)Z_t = \theta(B)a_t\,.$$

Given the observations: $Z_n, Z_{n-1}, \cdots,$

how to forecast $Z_{n+1}, \cdots, Z_{n+l}, \cdots$?

Notation:

$\widehat{Z}_n(l)$ denotes the forecast value of Z_{n+l} and is called the l-step ahead of the forecast of Z_{n+l} at the forecast origin n.

Simply say, l-step forecasting.

Forecasting function:

$$\widehat{Z}_n(l) = g(Z_n, Z_{n-1}, \cdots).$$
$$\widehat{Z}_n(l) = g(a_n, a_{n-1}, \cdots).$$

Linear Predictors (LP):

$$\widehat{Z}_n(l) = \psi_l^* a_n + \psi_{l+1}^* a_{n-1} + \psi_{l+2}^* a_{n-2} + \cdots ,$$

where ψ_j^* are to be determined.

Criterion of the best LP (BLP):

$\widehat{Z}_n(l)$ is said to be a BLP

if $E[Z_{n+l} - \widehat{Z}_n(l)]^2$ is the smallest among all the LP.

What is the BLP of Z_{n+l}?

Note that

$$Z_{n+l} = \frac{\theta(B)}{\phi(B)} a_{n+l} = \sum_{j=0}^{\infty} \psi_j a_{n+l-j}.$$

According to the above criterion, the BLP is that

$$\widehat{Z}_n(l) = \psi_l a_n + \psi_{l+1} a_{n-1} + \psi_{l+2} a_{n-2} + \cdots .(?)$$

Forecasting error:

$$e_n(l) = Z_{n+l} - \widehat{Z}_n(l) = \sum_{j=0}^{l-1} \psi_j a_{n+l-j}.$$

Forecasting variance:

$$\mathrm{Var}[e_n(l)] = \sigma_a^2 \sum_{j=0}^{l-1} \psi_j^2.$$

Forecast interval (limit) (FI):

$$\left[\widehat{Z}_n(l) - N_{\frac{\alpha}{2}} \sigma_a \sqrt{\sum_{j=0}^{l-1} \psi_j^2},\ \widehat{Z}_n(l) + N_{\frac{\alpha}{2}} \sigma_a \sqrt{\sum_{j=0}^{l-1} \psi_j^2} \right]$$

where $N_{\frac{\alpha}{2}}$ is the $\alpha/2$-quantile of the standard normal distribution,

i.e., $P(N > N_{\frac{\alpha}{2}}) = \alpha/2$.

When $\alpha = 0.05$, $N_{\frac{\alpha}{2}} = 1.96$.

Important Fact:

$$\widehat{Z}_n(l) = E(Z_{n+l}|Z_n, Z_{n-1}, \cdots).$$

The Formulas of Computation of Forecasts For ARMA Model:

$$\begin{aligned}\widehat{Z}_n(l) &= \phi_1\widehat{Z}_n(l-1) + \phi_2\widehat{Z}_n(l-2) + \cdots + \phi_p\widehat{Z}_n(l-p) \\ &\quad +\widehat{a}_n(l) - \theta_1\widehat{a}_n(l-1) - \cdots - \theta_q\widehat{a}_n(l-q).\end{aligned}$$

where

$$\begin{aligned}\widehat{Z}_n(j) &= \begin{cases} E(Z_{n+j}|Z_n, Z_{n-1}, \cdots) & \text{if} \quad j = 1, 2, \cdots, l. \\ Z_{n+j} & \text{if} \quad j = 0, -1, -2, \cdots . \end{cases} \\ \widehat{a}_n(j) &= \begin{cases} 0 & \text{if} \quad j = 1, 2, \cdots, l. \\ a_{n+j} & \text{if} \quad j = 0, -1, -2, \cdots . \end{cases}\end{aligned}$$

6.2 **Minimum Mean Square Error Forecasts for ARIMA models**

A. Model:

Let Z_t be ARIMA(p, d, q) model with $d \neq 0$,

$$\phi(B)(1-B)^d Z_t = \theta(B)a_t.$$

where all the roots of $\phi(z) = 0$ and $\theta(z) = 0$ lie outside the unit circle.

Given the observations: $Z_n, Z_{n-1}, \cdots$,

how to forecast $Z_{n+1}, \cdots, Z_{n+l}, \cdots$?

B. **Minimum Mean Square Error Forecasts**:

$$\widehat{Z}_n(j) = E(Z_{n+l}|Z_n, Z_{n-1}, \cdots).$$

C. Computation of forecast:

Denote

$$\begin{aligned}
\Psi(B) &= \phi(B)(1-B)^d \\
&= 1 - \Psi_1 B - \Psi_2 B^2 - \cdots - \Psi_{p+d} B^{p+d}. \\
Z_t &= \Psi_1 Z_{t-1} + \Psi_2 Z_{t-2} + \cdots + \Psi_{p+d} Z_{t-p-d} \\
&\quad + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q}.
\end{aligned}$$

Formulas:

$$\begin{aligned}
\widehat{Z}_n(l) &= \Psi_1 \widehat{Z}_n(l-1) + \cdots + \Psi_{p+d} \widehat{Z}_n(l-p-d) \\
&\quad + \widehat{a}_n(l) - \theta_1 \widehat{a}_n(l-1) - \cdots - \theta_q \widehat{a}_n(l-q).
\end{aligned}$$

where

$$\begin{aligned}
\widehat{Z}_n(j) &= \begin{cases} E(Z_{n+j}|Z_n, Z_{n-1}, \cdots) & \text{if} \quad j = 1, 2, \cdots, l. \\ Z_{n+j} & \text{if} \quad j = 0, -1, -2, \cdots. \end{cases} \\
\widehat{a}_n(j) &= \begin{cases} 0 & \text{if } j = 1, 2, \cdots, l. \\ a_{n+j} & \text{if } j = 0, -1, -2, \cdots. \end{cases}
\end{aligned}$$

D. Forecast error:

$$e_n(l) = Z_{n+l} - \widehat{Z}_n(l) = \sum_{j=0}^{l-1} \psi_j a_{n+l-j} \, ,$$

where ψ_i can be calculated, recursively:

$$\psi_j = \sum_{j=0}^{l-1} \pi_{j-i} \psi_i \, , \quad j = 1, 2, \cdots, l-1.$$

π_j is the coefficients of the expansion:

$$\pi(B) = \frac{\phi(B)(1-B)^d}{\theta(B)} = 1 - \sum_{j=1}^{\infty} \pi_j B^j.$$

$$Z_{t+l} = \sum_{j=1}^{\infty} \pi_j Z_{t+l-j} + a_{t+l}.$$

E. Forecast variance:

$$\mathrm{Var}[e_n(l)] = \sigma_a^2 \sum_{j=0}^{l-1} \psi_j^2.$$

F. Forecast interval (limit) (FI):

$$\left[\widehat{Z}_n(l) - N_{\frac{\alpha}{2}}\sigma_a\sqrt{\sum_{j=0}^{l-1}\psi_j^2},\ \widehat{Z}_n(l) + N_{\frac{\alpha}{2}}\sigma_a\sqrt{\sum_{j=0}^{l-1}\psi_j^2}\right]$$

where $N_{\frac{\alpha}{2}}$ is the $\alpha/2$-quantile of the standard normal distribution,

6.3 Some practical forecasts

Assume we have real data: $y_n, y_{n-1}, \cdots$.

Let $Z_t = \ln y_t$, we have data: $Z_n, Z_{n-1}, \cdots$.

If Z_t ia an **ARMA** or **ARIMA** model, then we can forecast Z_{n+l}.

Denote the forecast value of Z_{n+l} by $\widehat{Z}_n(l)$ and the forecast FI by

$$\left[\widehat{Z}_n(l) - N_{\frac{\alpha}{2}}\,\sigma_a\sqrt{\sum_{j=0}^{l-1}\psi_j^2},\ \widehat{Z}_n(l) + N_{\frac{\alpha}{2}}\sigma_a\sqrt{\sum_{j=0}^{l-1}\psi_j^2}\right]$$

Then the forecast value of y_{n+l} is $e^{\widehat{Z}_n(l)}$ and FI is

$$\left[e^{\left\{\widehat{Z}_n(l) - N_{\frac{\alpha}{2}}\sigma_a\sqrt{\sum_{j=0}^{l-1}\psi_j^2}\right\}},\ e^{\left\{\widehat{Z}_n(l) + N_{\frac{\alpha}{2}}\sigma_a\sqrt{\sum_{j=0}^{l-1}\psi_j^2}\right\}}\right]$$

Similarly, if $Z_t = \sqrt{y_t}$, then the forecast value and FI of y_{n+l} are, respectively, $\widehat{Z}_n^2(l)$ and

$$\left[\left\{\widehat{Z}_n(l) - N_{\frac{\alpha}{2}}\,\sigma_a\sqrt{\sum_{j=0}^{l-1}\psi_j^2}\right\}^2,\quad \left\{\widehat{Z}_n(l) + N_{\frac{\alpha}{2}}\,\sigma_a\sqrt{\sum_{j=0}^{l-1}\psi_j^2}\right\}^2\right].$$