Time Series Analysis : Univariate and Multivariate Methods
(William W. S. Wei)

Ch.2 Fundamental Concepts
Ch.3 Stationary Time Series Model

Analysis of Financial Time Series (Ruey S. Tsay)

Ch.2 Linear Time Series Analysis and Its Applications

Outline

Definition

1. Time Series : $Z_1, Z_2, \cdots, Z_n, \cdots$ (real value)
2. Stochastic process $Z_t(\omega)$: $t \in T =$ an index set,

   e.g. $T = \{0, \pm 1, \pm 2, \cdots\}$.

   $Z_t(\cdot)$ for fixed t, $Z_t(\cdot)$ is a r.v. defined on Ω.
   $Z_.(\omega)$ for fixed ω, $Z_.(\omega)$ is a realization of a time series.
3. n-dimensional distribution function of $\{Z_{t_1} Z_{t_2}, \cdots, Z_{t_n}\}$

$$F(z_{t_1}, \cdots, z_{t_n}) = P\{\omega : Z_{t_1} \leq z_{t_1}, \cdots, Z_{t_n} \leq z_{t_n}\}$$

4. If

$$P(Z_{t_1} < z_1, \cdots, Z_{t_n} < z_n) = P(Z_{t_1+k} < z_1, \cdots, Z_{t_n+k} < z_n),$$

where $z_i \in \mathbb{R}$ for all i, $\forall\,(t_1, \cdots, t_n)$ and $k \in \mathbb{N}$, then the process is called nth order stationary in distribution.

e.g. $n = 1$, $F(z_{t_1}) = F(z_{t_1+k}) \quad k = 2, \cdots$

If $\{Z_t\}_{t=1}^{n}$ is an nth order stationary process in distribution, $\forall\, n$, then $\{Z_t\}$ is called strictly stationary (strong stationary, completely stationary).

Remark: If $\{Z_t\}$ is nth order stationary in distribution, then $\{Z_t\}$ is also mth order stationary in distribution, $\forall\, m \leq n$.

e.g. $Z_1, Z_2, \cdots, Z_n$ *i.i.d.* : strictly stationary process.

5. Mean function : $\mu_t = E(Z_t)$.
Variance function : $\sigma_t^2 = E(Z_t - \mu)^2$.
Auto-Covariance function between Z_{t_1} and Z_{t_2},

$$\gamma(t_1, t_2) = E(Z_{t_1} - \mu_1)(Z_{t_2} - \mu_2).$$

Auto-Correlation function between Z_{t_1} and Z_{t_2},

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

6. For strictly stationary process, assuming $E(|Z_t|) < \infty$ and $E(Z_t^2) < \infty$, then

$$\gamma(t_1, t_2) = \gamma(t_1 + h, t_2 + h) = \gamma_{t_1 - t_2} = \gamma_k, (|t_1 - t_2| = k)$$

and

$$\rho(t_1, t_2) = \rho(t_1 + h, t_2 + h) = \rho_{t_1 - t_2} = \rho_k.$$

i.e. γ_k and ρ_k depends only on time lag k.

7. It is difficult to verify the joint distribution function of n random variables without stationarity assumption.
e.g. $Z_1, \cdots, Z_n \sim MN(\vec{\mu}, \Gamma)$, $\vec{\mu} = (\mu_1, \cdots, \mu_n)'$

$$\Gamma = \begin{pmatrix} \sigma_{11}^2 & \sigma_{12}^2 & \cdots & \sigma_{1n}^2 \\ \sigma_{21}^2 & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \sigma_{n1}^2 & \cdots & \cdots & \sigma_{nn}^2 \end{pmatrix}$$

參數個數 $= n + \frac{n(n+1)}{2} >$ sample size n 。然而就算是在 strictly stationary 假設下，若僅假設 multi-variate normal，參數個數 $= n + 1$，亦超過樣本個數，故一般而言，我們僅假設時序滿足下面的一種較弱的 stationary assumption 。

8. Covariance stationary (or weakly stationary),

$$\begin{cases} Cov(Z_{t_1}, Z_{t_2}) = Cov(Z_{t_1+k}, Z_{t_2+k}), \quad \forall\ (t_1, t_2) \text{ and } k \\ Var(Z_{t_1}) < \infty, \quad \forall\ t_1 \end{cases}$$

Remark A strictly stationary process with finite 2nd moment is also a weakly stationary process, 但若 2nd moment 不存在則未必 。

e.g. $Z_1, \cdots, Z_n$ *i.i.d.* Cauchy 是 strictly stationary 但不是 weakly stationary.

e.g. 丟 die 3 次 independently, $\Omega = \{1, 2, 3, 4, 5, 6\}$ sample space, 其中 $Z_i(\omega) = 2\omega$

$\because$ *i.i.d.* $\therefore$ strictly stationary.

$$e.g.\quad Z_t = A\sin(\omega t + \theta), \begin{cases} A: \text{ r.v.}, E(A) = 0, Var(A) = 1 \\ \theta: \text{ r.v.} \sim U(-\pi, \pi] \\ A \text{ and } \theta \text{ indep.} \end{cases}$$

$$E(Z_t) = 0$$

$$\begin{aligned} E(Z_t Z_{t+k}) &= E(A^2)E\{\frac{1}{2}[cos(\omega k) - cos(\omega(2t+k)+2\theta)]\} \\ &= \frac{1}{2}cos(\omega k) - \frac{1}{4\pi}\int_{-\pi}^{\pi} cos(\omega(2t+k)+2\theta)d\theta \\ &= \frac{1}{2}cos(\omega k) - \frac{1}{8\pi}[sin(\omega(2t+k)+2\theta)]|_{-\pi}^{\pi} \\ &= \frac{1}{2}cos(\omega k). \end{aligned}$$

Covariance stationary.

$$e.g.\quad Z_t = \begin{cases} X_t \sim N(0,1),\ t: \text{odd} \\ Y_t \sim P(Y_t = 1) = P(Y_t = -1) = \frac{1}{2},\ t: \text{even} \end{cases},$$

- $X_t \perp Y_t$
- $$E(Z_t) = 0, E(Z_t^2) = 1,\ \forall t$$
  $$E(Z_t Z_s) = \begin{cases} 0, & t \neq s \\ 1, & t = s \end{cases}$$
  $$\rho(t,s) = \frac{E(Z_t, Z_s)}{\sqrt{E(Z_t^2)}\sqrt{E(Z_s^2)}} = \begin{cases} 0, & t \neq s \\ 1, & t = s \end{cases}.$$
- Z_t is Covariance stationary, not strictly stationary.

Remark

- Gaussian (or normal) Process : any finite dimensional joint distribution is multivariate normal. That is

$$(Z_{t_1}, Z_{t_2}, \cdots, Z_{t_n}) \sim N(\vec{\mu}_n, \Sigma_n)$$

  for any $(t_1, t_2, \cdots, t_n)$.
- Since multivariate normal distribution is completely determined by its first two moments (mean vector and covariance matrix), covariance stationary = strictly stationary for Gaussian process.
- e.g. If (Z_1, Z_3, Z_5) and (Z_2, Z_4, Z_6) has the same mean vector and the same covariance structure, then they have the same distribution.

Financial TS: collection of a financial measurement over time.
Example: log return r_t.
Data: $\{r_1, r_2, \ldots, r_T\}$

- Stationarity:
  - strict: distribution are time-invariant.
  - Weak: first 2 moments are time-invariant. $E(r_t) = \mu$, Cov$(r_t, r_{t+k}) = \gamma_k$
- finite 2nd moment +strictly stationary $\Rightarrow$ weakly stationary
  weakly stationary $+$ normality $\Rightarrow$ strictly stationary

What does weak stationarity mean in practice?

Past: time plot of $\{r_t\}$ varies around a *fixed level* within a *finite range*. (fluctuate with constant variation around a constant level.)

Future: the first 2 moments of future r_t are the same as those of the data so that meaningful inferences can be made.

- Under weak stationarity assumption, we can define the ACVF and ACF of a time series model.

ACVF and ACF

1. Assume $\{Z_t\}$ is a stationary process, with $E(Z_t) = \mu$ and finite $Var(Z_t)$
   ACVF (autocovariance function):

$$\gamma_k = Cov(Z_t, Z_{t+k}) = E(Z_t - \mu)(Z_{t+k} - \mu).$$

   ACF (autocorrelation function) :

$$\rho_k = \frac{Cov(Z_t, Z_{t+k})}{\sqrt{Var(Z_t)}\sqrt{Var(Z_{t+k})}} = \frac{\gamma_k}{\gamma_0}.$$

(a) $\gamma_0 = Var(Z_t)$; $\rho_0 = 1$.

(b) $|\gamma_k| \leq \gamma_0$; $|\rho_k| \leq 1$, Cauchy-Schwarz inequality:

$$|E(Z_t - \mu)(Z_{t+k} - \mu)| \leq \sqrt{E(Z_t - \mu)^2}\sqrt{E(Z_{t+k} - \mu)^2} = \gamma_0.$$

(c) $\gamma_k = \gamma_{-k}$, $\rho_k = \rho_{-k}$, $\forall\ k \Rightarrow \gamma_k$ and ρ_k are even functions of k.

(d) γ_k : a positive semidefinite (non-negative definite) function.
$\sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j \gamma_{|t_i - t_j|} \geq 0$, for any set of $(t_1, t_2, \cdots, t_n)$, any $\alpha_i \in \Re$, $i = 1, 2, \cdots, n$.

pf: Define $X = \sum_{i=1}^{n} \alpha_i Z_{t_i}$, where Z_t is a stationary time series with zero mean. then

$$0 \leq Var(X) = \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j Cov(Z_{t_i}, Z_{t_j})$$
$$= \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j \gamma_{|t_i - t_j|}$$

(e) Similarly, we can prove that ρ_k is a positive semidefinite function.

PACF

1. $\{Z_t\}$: stationary process, $E(Z_t) = 0$.
2. Let $\hat{Z}_{t+k}$ be the best linear predictor of Z_{t+k} based on $\{Z_{t+1}, Z_{t+2}, \cdots, Z_{t+k-1}\}$, then

$$\hat{Z}_{t+k} = \alpha_1 Z_{t+k-1} + \alpha_2 Z_{t+k-2} + \cdots + \alpha_{k-1} Z_{t+1}$$

where $\alpha_i' s$ are obtained by minimizing

$$E(Z_{t+k} - \hat{Z}_{t+k})^2 = E(Z_{t+k} - \alpha_1 Z_{t+k-1} - \cdots - \alpha_{k-1} Z_{t+1})^2$$

by differentiation we have

$$\gamma_i = \alpha_1\gamma_{i-1} + \alpha_2\gamma_{i-2} + \cdots + \alpha_{k-1}\gamma_{i-k+1}, (1 \leq i \leq k-1)$$

$$\Rightarrow \rho_i = \alpha_1\rho_{i-1} + \alpha_2\rho_{i-2} + \cdots + \alpha_{k-1}\rho_{i-k+1},$$

$$\Rightarrow \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_{k-1} \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{k-2} & \rho_{k-3} & \rho_{k-4} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{k-1} \end{bmatrix}$$

3. 同理，令 $\hat{Z}_t$ be the best linear predictor of Z_t based on $\{Z_{t+1}, Z_{t+2}, \cdots, Z_{t+k-1}\}$, and $\hat{Z}_t = \beta_1 Z_{t+1} + \beta_2 Z_{t+2} + \cdots + \beta_{k-1} Z_{t+k-1}$ 其中 β_i's minimize

$$E(Z_t - \hat{Z}_t)^2 = E(Z_t - \beta_1 Z_{t+1} - \cdots - \beta_{k-1} Z_{t+k-1})^2$$

$$\Rightarrow \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_{k-1} \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{k-2} & \rho_{k-3} & \rho_{k-4} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{k-1} \end{bmatrix}$$

故 $\alpha_i = \beta_i$, $1 \leq i \leq k-1$.

4. Partial autocorrelation between Z_t and Z_{t+k}: the correlation between Z_t and Z_{t+k} after their mutual linear dependency on the interventing variables $Z_{t+1}, Z_{t+2}, \cdots$ and Z_{t+k-1} has been removed:
$P_1 = Corr(Z_1, Z_2)$ and
$P_k = Corr(Z_t - \hat{Z}_t, Z_{t+k} - \hat{Z}_{t+k}), k \geq 2$, i.e.

$$P_k = \frac{Cov[(Z_t - \hat{Z}_t), (Z_{t+k} - \hat{Z}_{t+k})]}{\sqrt{Var(Z_t - \hat{Z}_t)}\sqrt{Var(Z_{t+k} - \hat{Z}_{t+k})}}$$

$$\begin{aligned}
&Var(Z_{t+k} - \hat{Z}_{t+k}) \\
&= E[(Z_{t+k} - \alpha_1 Z_{t+k-1} - \cdots - \alpha_{k-1} Z_{t+1})^2] \\
&= E[Z_{t+k}(Z_{t+k} - \alpha_1 Z_{t+k-1} - \cdots - \alpha_{k-1} Z_{t+1})]
\end{aligned}$$

Hence

$$
\begin{aligned}
Var(Z_{t+k} - \hat{Z}_{t+k}) = Var(Z_t - \hat{Z}_t) \\
&= \gamma_0 - \alpha_1\gamma_1 - \cdots - \alpha_{k-1}\gamma_{k-1} \\
Cov[(Z_t - \hat{Z}_t), (Z_{t+k} - \hat{Z}_{t+k})] \\
&= E[(Z_t - \alpha_1 Z_{t+1} - \cdots - \alpha_{k-1} Z_{t+k-1}) \\
&\quad (Z_{t+k} - \alpha_1 Z_{t+k-1} - \cdots - \alpha_{k-1} Z_{t+1})] \\
&= E[(Z_t - \alpha_1 Z_{t+1} - \cdots - \alpha_{k-1} Z_{t+k-1}) Z_{t+k}] \\
&= \gamma_k - \alpha_1\gamma_{k-1} - \cdots - \alpha_{k-1}\gamma_1
\end{aligned}
$$

$$
\Rightarrow P_k = \frac{\gamma_k - \alpha_1\gamma_{k-1} - \cdots - \alpha_{k-1}\gamma_1}{\gamma_0 - \alpha_1\gamma_1 - \cdots - \alpha_{k-1}\gamma_{k-1}}
= \frac{\rho_k - \alpha_1\rho_{k-1} - \cdots - \alpha_{k-1}\rho_1}{1 - \alpha_1\rho_1 - \cdots - \alpha_{k-1}\rho_{k-1}}
$$

By 2. , we have

$$\alpha_i = \frac{\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_1 & \cdots & \rho_{k-2} \\ \rho_1 & 1 & & \vdots & & \vdots \\ \vdots & \vdots & & \vdots & & \vdots \\ \rho_{k-2} & \cdots & \cdots & \rho_{k-1} & \cdots & 1 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{i-1} & \cdots & \rho_{k-2} \\ \rho_1 & 1 & & \vdots & & \vdots \\ \vdots & \vdots & & \vdots & & \vdots \\ \rho_{k-2} & \cdots & \cdots & \rho_{k-i-1} & \cdots & 1 \end{vmatrix}}$$

Recall

$$P_k = \frac{\rho_k - \alpha_1\rho_{k-1} - \cdots - \alpha_{k-1}\rho_1}{1 - \alpha_1\rho_1 - \cdots - \alpha_{k-1}\rho_{k-1}}$$

$$\Rightarrow \quad P_k = \frac{\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \cdots & \rho_{k-3} & \rho_2 \\ \vdots & \vdots & & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & \rho_1 & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \cdots & \rho_{k-3} & \rho_{k-2} \\ \vdots & \vdots & & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & \rho_1 & 1 \end{vmatrix}}$$

(對最後一行展行列式)

若將 Z_{t+k} 對 $Z_{t+k-1}, Z_{t+k-2}, \cdots, Z_{t+1}, Z_t$ 做 regression
i.e. $Z_{t+k} = \phi_{k1}Z_{t+k-1} + \phi_{k2}Z_{t+k-2} + \cdots + \phi_{kk}Z_t + \varepsilon_{t+k}$
求 $\phi_{k1}, \cdots, \phi_{kk}$，使得

$$\sum_{t=1}^{n-k}(Z_{t+k} - \phi_{k1}Z_{t+k-1} - \cdots - \phi_{kk}Z_t)^2$$

為最小 。 對這 k 個參數偏微，可得到下面 k 個線性聯立方程式

$$\rho_1 = \phi_{k1}\rho_0 + \phi_{k2}\rho_1 + \cdots + \phi_{kk}\rho_{k-1},$$

$$\rho_i = \phi_{k1}\rho_{i-1} + \phi_{k2}\rho_{i-2} + \cdots + \phi_{kk}\rho_{i-k},$$

$$\rho_k = \phi_{k1}\rho_{k-1} + \phi_{k2}\rho_{k-2} + \cdots + \phi_{kk}\rho_0,$$

The solution of ϕ_{kk} is exactly the same as the formula of P_k given on p.23. $\Longrightarrow$ $\phi_{kk} = P_k$

e.g $\{a_t\}$: $WN(0, \sigma_a^2)$

$$\Rightarrow \rho_k = \begin{cases} 1, & k = 0 \\ 0, & k \neq 0 \end{cases} \Rightarrow \phi_{kk} = \begin{cases} 1, & k = 0 \\ 0, & k \neq 0 \end{cases}$$

White noise: uncorrelated sequence (with finite variance), which is the building block of linear TS models.
White noise is not predictable, but has zero mean and finite variance.

Spectrum of a stationary time series

- Z_t : a real-valued stationary process with ACVF γ_k.
- If $\sum |\gamma_k| < \infty$, then the Fourier transform of γ_k exists and equals

$$
\begin{aligned}
f(\omega) &= \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_k e^{-i\omega k} = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_k \cos(\omega k) \\
&= \frac{1}{2\pi} \gamma_0 + \frac{1}{\pi} \sum_{k=1}^{\infty} \gamma_k \cos(\omega k), -\pi \leq \omega \leq \pi
\end{aligned}
$$

- $f(\omega)$ is called the spectrum of Z_t. (Hence the spectrum of a WN process is a constant.)
- By Inverse Fourier transform, we have

$$\gamma_k = \int_{-\pi}^{\pi} f(\omega) e^{i\omega k} \, d\omega$$

Estimation of mean, ACVF and ACF, Z_t stationary time series

1. Sample mean $\overline{Z}_n = \frac{1}{n}\sum_{t=1}^{n} Z_t$

$$E(\overline{Z}_n) = \frac{1}{n}\cdot n\mu = \mu \text{ unbiased}$$

$$\begin{aligned} Var(\overline{Z}_n) &= \frac{1}{n^2}\sum_{t=1}^{n}\sum_{s=1}^{n} Cov(Z_t, Z_s) \\ &= \frac{\gamma_0}{n^2}\sum_{t=1}^{n}\sum_{s=1}^{n}\rho_{t,s} \\ &= \frac{\gamma_0}{n^2}\sum_{k=-(n-1)}^{n-1}(n-|k|)\rho_k \\ &= \frac{\gamma_0}{n}\sum_{k=-(n-1)}^{n-1}(1-\frac{|k|}{n})\rho_k \end{aligned}$$

- If $\rho_k \to 0$, as $k \to \infty$, then

$$\frac{1}{n} \sum_{k=-(n-1)}^{n-1} \rho_k \underset{n\to\infty}{\longrightarrow} 0 \text{ (} \mathit{Cesàro} \text{ summable)}$$

and hence $Var(\bar{Z}_n) \to 0$ as $n \to \infty$.
$\Rightarrow \bar{Z}_n \to \mu$ in mean square & probability (by Chebyshev's inequality $P(|\bar{Z}_n - \mu| > \varepsilon) < \frac{\sigma_n^2}{\varepsilon^2} \to 0$), equivalently $\{Z_t\}$ is ergodic for the mean.

- Test $H_0 : \mu = 0$ vs $H_a : \mu \neq 0$. Compute

$$t = \frac{\bar{Z}_n}{std(\bar{Z}_n)} = \frac{\bar{Z}_n}{\sqrt{Var(Z_n)/n}}$$

Compare t ratio with $N(0,1)$ dist.
Decision rule: Reject H_0 of zero mean if $|t| > Z_{\alpha/2}$ or p-value is less than α.

Sample ACVF

$$\hat{\gamma}_k = \frac{1}{n}\sum_{t=1}^{n-k}(Z_t - \overline{Z}_n)(Z_{t+k} - \overline{Z}_n)$$

$$\hat{\hat{\gamma}}_k = \frac{1}{n-k}\sum_{t=1}^{n-k}(Z_t - \overline{Z}_n)(Z_{t+k} - \overline{Z}_n)$$

$$\begin{aligned}\sum_{t=1}^{n-k}(Z_t - \overline{Z}_n)(Z_{t+k} - \overline{Z}_n) &= \sum_{t=1}^{n-k}[(Z_t - \mu) - (\overline{Z}_n - \mu)][(Z_{t+k} - \mu) - (\overline{Z}_n - \mu)] \\ &= \sum_{t=1}^{n-k}(Z_t - \mu)(Z_{t+k} - \mu) - (\overline{Z}_n - \mu)\sum_{t=1}^{n-k}(Z_t - \mu) \\ &\quad - (\overline{Z}_n - \mu)\sum_{t=1}^{n-k}(Z_{t+k} - \mu) + (n-k)(\overline{Z}_n - \mu)^2 \\ &\approx \sum_{t=1}^{n-k}(Z_t - \mu)(Z_{t+k} - \mu) - (n-k)(\overline{Z}_n - \mu)^2\end{aligned}$$

$$\Rightarrow E(\hat{\gamma}_k) \approx \frac{n-k}{n}\gamma_k - \frac{n-k}{n}Var(\overline{Z}_n)$$
$$= \left(1 - \frac{k}{n}\right)(\gamma_k - Var(\overline{Z}_n))$$
$$E(\hat{\hat{\gamma}}_k) \approx \gamma_k - Var(\overline{Z}_n)$$

$\Rightarrow$ When we ignore the term $Var(\overline{Z}_n)$ representing the effect of estimating μ, $\hat{\hat{\gamma}}_k$ becomes unbiased but $\hat{\gamma}_k$ is still biased.

- In general $\hat{\gamma}_k$ has a larger bias than $\hat{\hat{\gamma}}$ especially when k is large w.r.t. n.
  (In practice we only compute $\hat{\gamma}_k$, for $k \leq \frac{n}{4}$.)
- $\hat{\gamma}_k$ and $\hat{\hat{\gamma}}_k$ both are asymptotically unbiased (but for finite sample both are biased)
  (We can compare $\hat{\gamma}_k$ and $\hat{\hat{\gamma}}_k$ by their mean squared errors, $MSE(\hat{\gamma}_k)$ and $MSE(\hat{\hat{\gamma}}_k)$.)

- $\hat{\gamma}_k$ is always positive semidefine but $\hat{\hat{\gamma}}_k$ is not necessary so

$$\because \quad \hat{\Gamma}_n = n^{-1} T T'$$

$$= \frac{1}{n} \begin{bmatrix} \cdots & 0 & 0 & Z_1 - \overline{Z}_n & Z_2 - \overline{Z}_n & \cdots & \cdots & Z_n - \overline{Z}_n \\ \cdots & 0 & Z_1 - \overline{Z}_n & Z_2 - \overline{Z}_n & Z_3 - \overline{Z}_n & \cdots & Z_n - \overline{Z}_n & 0 \\ \vdots & & & & & & \vdots & \vdots \\ Z_1 - \overline{Z}_n & \cdots & Z_{n-1} - \overline{Z}_n & Z_n - \overline{Z}_n & 0 & \cdots & \cdots & 0 \end{bmatrix}_{n \times (2n-1)}$$

$$\cdot \begin{bmatrix} 0 & 0 & Z_1 - \overline{Z}_n \\ \vdots & \vdots & \vdots \\ 0 & Z_1 - \overline{Z}_n & Z_{n-1} - \overline{Z}_n \\ Z_1 - \overline{Z}_n & Z_2 - \overline{Z}_n & Z_n - \overline{Z}_n \\ Z_2 - \overline{Z}_n & Z_3 - \overline{Z}_n & 0 \\ \vdots & \vdots & \vdots \\ \vdots & Z_n - \overline{Z}_n & \vdots \\ Z_n - \overline{Z}_n & 0 \quad \cdots & 0 \end{bmatrix}_{(2n-1) \times n}$$

$$\therefore \quad \vec{a}' \hat{\Gamma}_n \vec{a} = n^{-1} (\vec{a}' T)(T' \vec{a}) = n^{-1} \parallel T' \vec{a} \parallel^2 \geq 0, \quad \forall\ n.$$

- If $\{Z_t\}$ stationary Gaussian (Bartlett)

$$\begin{cases} Cov(\hat{\gamma}_k, \hat{\gamma}_{k+j}) \approx \frac{1}{n} \sum_{i=-\infty}^{\infty} (\gamma_i \gamma_{i+j} + \gamma_{i+k+j} \gamma_{i-k}) \\ Var(\hat{\gamma}_k) \approx \frac{1}{n} \sum_{i=-\infty}^{\infty} (\gamma_i^2 + \gamma_{i+k} \gamma_{i-k}) \end{cases}$$

$$\begin{cases} Cov(\hat{\hat{\gamma}}_k, \hat{\hat{\gamma}}_{k+j}) \approx \frac{1}{n-k} \sum_{i=-\infty}^{\infty} (\gamma_i \gamma_{i+j} + \gamma_{i+k+j} \gamma_{i-k}) \\ Var(\hat{\hat{\gamma}}_k) \approx \frac{1}{n-k} \sum_{i=-\infty}^{\infty} (\gamma_i^2 + \gamma_{i+k} \gamma_{i-k}) > Var(\hat{\gamma}_k) \end{cases}$$

especially when k is large.

Recall $\hat{\gamma}_k$ is an asymptotically unbiased estimator of γ_k and by Bartlett formula, we have

$$Var(\hat{\gamma}_k) \approx \frac{1}{n} \sum_{i=-\infty}^{\infty} (\gamma_i^2 + \gamma_{i+k}\gamma_{i-k}).$$

Hence

$$Var(\hat{\gamma}_k) < \frac{1}{n}(\sum_{i=-\infty}^{\infty} |\gamma_i|)^2.$$

- If $\sum_{i=-\infty}^{\infty} |\gamma_i| < \infty$, then $\hat{\gamma}_k \to \gamma_k$ in mean square.

$$(E(\hat{\gamma}_k - \gamma_k)^2 \underset{n\to\infty}{\longrightarrow} 0)$$

$\Rightarrow \hat{\gamma}_k \to \gamma_k$ in probability.
$\Rightarrow \{Z_t\}$ is ergodic for ACVF.

Sample ACF

- $\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} = \frac{\sum_{t=1}^{n-k}(Z_t - \overline{Z}_n)(Z_{t+k} - \overline{Z}_n)}{\sum_{t=1}^{n}(Z_t - \overline{Z}_n)^2}, \ k = 0, 1, 2, \cdots$
- Plot of $\hat{\rho}_k$ v.s. k is called sample correlogram.
- $\{Z_t\}$ stationary Gaussian (Bartlett).
- $$\begin{aligned} & Cov(\hat{\rho}_k, \hat{\rho}_{k+j}) \\ & \approx \frac{1}{n} \sum_{i=-\infty}^{\infty} (\rho_i \rho_{i+j} + \rho_{i+k+j} \rho_{i-k} - 2\rho_k \rho_i \rho_{i-k-j} \\ & - 2\rho_{k+j} \rho_i \rho_{i-k} + 2\rho_k \rho_{k+j} \rho_i^2), \quad k > 0, \ k + j > 0 \end{aligned}$$

- for large n, $\hat{\rho}_k \xrightarrow{D} N(\rho_k, Var(\hat{\rho}_k))$

$$Var(\hat{\rho}_k) \approx \frac{1}{n}\sum_{i=-\infty}^{\infty}(\rho_i^2 + \rho_{i+k}\rho_{i-k} - 4\rho_k\rho_i\rho_{i-k} + 2\rho_k^2\rho_i^2)$$

- If $\rho_k = 0$, for $k > m$, then

$$Var(\hat{\rho}_k) \approx \frac{1}{n}(1 + 2\rho_1^2 + \cdots + 2\rho_m^2) \text{ for } k > m$$

and

$$\hat{\rho}_k \to N(0, (1 + 2\sum_{i=1}^{q}\rho_i^2)/n) \text{ for } k > m.$$

- In practice，we can replace ρ_i by $\hat{\rho}_i$ to approximate the asymptotic stdev of $\hat{\rho}_k$, that is

$$S_{\hat{\rho}_k} = \sqrt{\frac{1}{n}(1 + 2\hat{\rho}_1^2 + \cdots + 2\hat{\rho}_m^2)}.$$

- For white noise process $\Longrightarrow \rho_k = 0$, for $k \neq 0$,

$$\hat{\rho}_k \to N(0, \frac{1}{n}).$$

  The large sample stdev of $\hat{\rho}_k$ is $S_{\hat{\rho}_k} = \sqrt{\frac{1}{n}}$
- Individual test: $H_0 : \rho_1 = 0$ vs $H_a : \rho_1 \neq 0$

$$t = \frac{\hat{\rho}_1}{\sqrt{1/n}} = \sqrt{n}\hat{\rho}_1 \xrightarrow{D} N(0,1)$$

  Decision rule: Reject H_0 if $|t| > Z_{\alpha/2}$ or p-value less than α.
- **Drawback of testing m individual tests ($H_0 : \rho_i = 0$ vs $H_a : \rho_i \neq 0$ for $i = 1, \ldots, m$):** The type I error probability $= 1 - (1-\alpha)^m > \alpha$.

- Joint test (Ljung-Box statistics):

$$H_0 : \rho_1 = \cdots = \rho_m = 0 \quad v.s. \quad H_a : \rho_i \neq 0$$

$$Q(m) = n(n+2) \sum_{k=1}^{m} \frac{\hat{\rho}_k^2}{n-k}$$

Asym. chi-squared dist with m degrees of freedom.

- **Decision rule:**
Reject H_0 if $Q(m) > \chi_m^2(\alpha)$ or p-value less than α.

- Application: Test market efficiency hypothesis by testing zero serial correlations hypothesis.
- 在資本資產定價模型 (CAPM) 的假設下，一個有效率的市場，資本資產 (e.g. 有價證券 or 投資組合) 的報酬率僅與系統風險 (i.e. 與市場報酬) 相關，為不可預測 。
- 當觀察到報酬率具相關性時 (i.e. 資本資產的報酬率具可預測性)，有部份的經濟學家即就此推論效率市場的假設是有爭議的 。
- However, significant sample ACF does not necessarily imply market inefficiency.
- Sources of serial correlations in financial TS: Nonsynchronous trading (非同步交易)(ch. 5); Bid-ask bounce (買賣彈跳)(ch. 5); Risk premium (風險溢酬) (ch. 3).

Sample PACF

- Recall

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

- Replace ρ_k by the sample ACF $\hat{\rho}_k \Longrightarrow \hat{\phi}_{kk}$.
- The trouble is : when k is large, it requires intensive computation effort.

Durbin recursive formula

- Alternatively, we consider the regression model

$$Z_{t+k} = \phi_{k1} Z_{t+k-1} + \cdots + \phi_{kk} Z_t + \varepsilon_t.$$

- Let $\hat{\phi}_{kk}$ be the best least squares estimator of the regression coefficient ϕ_{kk}.
- Durbin derive

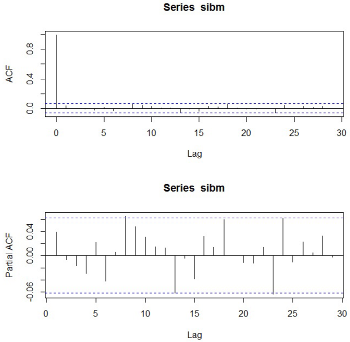
$$\begin{cases} \hat{\phi}_{k+1,k+1} = \frac{\hat{\rho}_{k+1} - \sum_{j=1}^{k} \hat{\phi}_{kj} \hat{\rho}_{k+1-j}}{1 - \sum_{j=1}^{k} \hat{\phi}_{kj} \hat{\rho}_j} \\ \hat{\phi}_{k+1,j} = \hat{\phi}_{kj} - \hat{\phi}_{k+1,k+1} \hat{\phi}_{k,k+1-j} \quad j = 1, 2, \cdots, k \end{cases}$$

(This recursive formulae also hold for theoretical PACF.)

- If $\{Z_t\}$ is a WN process, then $Var(\hat{\phi}_{kk}) \approx \frac{1}{n}$. Hence $\pm \frac{2}{\sqrt{n}}$ can be taken as the 95 % critical limits for $\hat{\phi}_k$.

Monthly returns of IBM stock from 1926 to 2008

- da=read.table("m-ibm3dx2608.txt",header=T) #Load data
  sibm=da[,2] #Get the IBM simple returns
  par(mfrow=c(2,1))
  acf(sibm)
  acf(sibm,type="partial")
  par(mfrow=c(1,1))
- Box.test(sibm,lag=5,type="Ljung-Box")
  - simple return R_t : $Q(5) = 3.37(0.64)$ and $Q(10) = 13.99(0.17)$
  - log return r_t : $Q(5) = 3.52(0.62)$ and $Q(10) = 13.39(0.20)$
  - qchisq(0.95,df=5) (ans=11.0705)
  - 1-pchisq(3.37,df=5) (ans=0.643)
- **Recall :** What is p-value? How to use it?



- The ACFs of both the simple returns and the log returns are within $\dfrac{1.96}{\sqrt{n}}$
- Implication of the Ljung-Box test and the above figures: Monthly IBM stock returns do not have significant serial correlations.

Monthly returns of CRSP value-weighted index from 1926 to 2008

- $R_t : Q(5) = 29.71$ and $Q(10) = 39.55$
- $r_t : Q(5) = 28.38$ and $Q(10) = 36.16$
  兩者的 p-value 均小於 0.0003 (All highly significant)，故顯示序列具相關性 。
- Implication: there exist significant serial correlations in the value-weighted index returns.
- Nonsynchronous trading might explain the existence of the serial correlations, among other reasons.
- Similar result is also found in equal-weighted index returns.

A proper perspective

- Data up to time t :$\{r_1, r_2, \ldots, r_{t-1}\}$
- Information available up to time $t-1 \equiv F_{t-1}$.
- The return is decomposed into two parts as

$$\begin{aligned} r_t &= \text{predictiable part} + \text{not predic. part} \\ &= \text{function of elements of } F_{t-1} + a_t \end{aligned}$$

- In other words, given information F_{t-1}

$$\begin{aligned} r_t &= \mu_t + a_t \\ &= E(r_t | F_{t-1}) + \sigma_t \epsilon_t \end{aligned}$$

(μ_t :conditional mean of r_t. a_t: shock or innovation at time t.
ϵ_t :an iid sequence with mean zero and variance 1.
σ_t :conditional standard deviation (commonly called volatility in finance)).

- Traditional TS modeling is concerned with μ_t:
- Model for μ: mean equation
- Model for σ_t^2 : **volatility equation**.
- Univariate TS analysis serves two purposes
  - a model for μ_t.
  - understanding models for σ_t^2 : properties, forecasting, etc.
- What are the important statistics in practice? Conditional quantities, not unconditional.

- **Linear time series:** Z_t is linear if
  - the predictable part is a linear function of F_{t-1}.
  - $\{a_t\}$ are indep. and have the same dist. (iid)
- Mathematically, it means Z_t can be written as

$$Z_t = \mu + \sum_{i=0}^{\infty} \psi_i a_{t-i}$$

  where μ is a constant, $\psi_0 = 1$ and $\{a_t\}$ is an iid sequence with mean zero and well-defined distribution.
- In the economic literature, a_t is the *shock* (or innovation) at time t and $\{\psi_i\}$ are the *impulse* responses of Z_t.

$MA(\infty)$ representation

- Linear process (infinite MA representation):

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

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

$$\sum \psi_j^2 < \infty \text{ (or } \sum |\psi| < \infty)$$

由 $\sum \psi_j^2 < \infty$ 的條件可推得

$$E[(Z_t - \mu - \sum_{j=0}^{n} \psi_j a_{t-j})^2] \to 0 \text{ as } n \to \infty.$$

i.e. $\mu + \sum_{j=0}^{\infty} \psi_j a_{t-j} \to Z_t$ in m.s.

- Wold: A nondeterministic stationary process can always be expressed in the form of $MA(\infty)$+deterministic term.
- A deterministic term is a process that can be predicted or forecast exactly from the past.

Back-shift (lag) operator

A useful notation in TS analysis.

- B (or L) means time shift.
- BZ_t is the value of the series at time $t-1$.
- Definition: $BZ_t = Z_{t-1}$ or $LZ_t = Z_{t-1}$
- $B^2Z_t = B(BZ_t) = BZ_{t-1} = Z_{t-2}$.

Suppose that the daily log returns are

Day	1	2	3	4
r_t	0.017	-0.005	-0.014	0.021

Answer the following questions:

- $r_2 = -0.005$
- $Br_3 = r_2 = -0.005$
- $B^2 r_5 = r_3 = -0.014$

Question: What is B^2?

- Recall the backshift operator $B^j Z_t = Z_{t-j}$, 故可將 MA(∞) 的 $\{Z_t\}$ 寫為

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

其中 $\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j$。

-

$$E(Z_t) = \mu,$$
$$Var(Z_t) = \sigma_a^2 \sum_{j=0}^{\infty} \psi_j^2,$$
$$E(a_t Z_{t-j}) = \begin{cases} \sigma_a^2, & \text{for } j = 0 \\ 0, & \text{for } j > 0 \end{cases},$$

$$\begin{aligned}\gamma_k &= E(Z_t-\mu)(Z_{t+k}-\mu)\\ &= E(\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\psi_i\psi_j a_{t-i}a_{t+k-j})\\ &= \sigma_a^2\sum_{i=0}^{\infty}\psi_i\psi_{i+k}\end{aligned}$$

$\Rightarrow$

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

- Note that $|\gamma_k| \le \gamma_0 = \sigma_a^2\sum\psi_j^2 < \infty$ 故 $\gamma_k < \infty\ \forall\ k$ 。

- Autocovariance generating function 定義為

$$\gamma(B) = \sum_{k=-\infty}^{\infty} \gamma_k B^k$$

- $\{Z_t\}$ linear process, 其 Autocovariance generating function 為

$$\begin{aligned} \gamma(B) &= \sum_{k=-\infty}^{\infty} \sum_{i=0}^{\infty} \psi_i \psi_{i+k} B^k \sigma_a^2 \\ &= \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} \psi_i \psi_j B^{j-i} \sigma_a^2 \\ &= \sigma_a^2 \psi(B) \psi(B^{-1}), \end{aligned}$$

($j = i + k$ and $\psi_j = 0$, for $j < 0$)

$\Rightarrow$

$$\rho(B) = \sum_{k=-\infty}^{\infty} \rho_k B^k = \frac{\gamma(B)}{\gamma_0}.$$

- $AR(\infty)$ representation : useful in understanding the mechanism of forecasting $\Rightarrow$ the process is invertible.

$$\dot{Z}_t = \pi_1 \dot{Z}_{t-1} + \pi_2 \dot{Z}_{t-2} + \cdots + a_t \quad \text{其中 } \dot{Z}_t = Z_t - \mu$$

$\Rightarrow$

$$\pi(B) Z_t = a_t, \quad \pi(B) = 1 - \sum_{j=1}^{\infty} \pi_j B^j$$

(註)

1. An invertible process is stationary

$\Longleftrightarrow$

$$Z_t = \frac{1}{\pi(B)} a_t = \psi(B) a_t \ s.t. \sum_{j=0}^{\infty} \psi_j^2 < \infty$$

$\Longleftrightarrow$

$$\pi(B) = 0 \text{ 的根落在單位圓外。}$$

- *2.* A linear process is invertible

  $\iff$

  $$a_t = \psi^{-1}(B)Z_t = \phi(B)Z_t \ s.t. \ \sum \phi_j^2 < \infty$$

  $\iff$ The roots of $\psi(B) = 0$ lie outside the unit circle.

- However, $AR(\infty)$ or $MA(\infty)$ contain infinite parameters which hinder model building.

Univariate linear time series models

1. *autoregressive (AR) models* : $AR(p)$
   $Z_t - \phi_1 Z_{t-1} - \cdots - \phi_p Z_{t-p} = a_t$
2. *moving-average (MA) models* : $MA(q)$
   $Z_t = a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}$
3. *mixed ARMA models* : $ARMA(p, q)$
   $Z_t - \phi_1 Z_{t-1} - \cdots - \phi_p Z_{t-p} = a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}$
4. *seasonal models*
5. *regression models with time series errors*
6. *fractionally differenced models (long-memory)*

Important properties of a model

- Stationarity condition
- Basic properties: mean, variance, serial dependence
- Empirical model building: specification, estimation, & checking
- Forecasting

AR(1) Model

1. Form: $Z_t = \phi_0 + \phi_1 Z_{t-1} + a_t$, where ϕ_0 and ϕ_1 are real numbers, which are referred to as "parameters" (to be estimated from the data in an application). For example,

$$Z_t = 0.005 + 0.2 Z_{t-1} + a_t$$

2. Stationarity: necessary and sufficient condition $|\phi_1| < 1$. Why? $(1 - \phi_1 B) Z_t = \phi_0 + a_t \Rightarrow$

$$Z_t = \phi_0 + (1 - \phi_1 B)^{-1} a_t = \phi_0 + \sum_{j=0}^{\infty} \phi_1^j a_{t-j}$$

is a stationary solution for $(1 - \phi_1 B) Z_t = \phi_0 + a_t$.

3. Mean: $E(Z_t) = \dfrac{\phi_0}{1 - \phi_1}$

4. Alternative representation: Let $E(Z_t) = \mu$ be the mean of Z_t so that $\mu = \phi_0/(1-\phi_1)$. Equivalently, $\phi_0 = \mu(1-\phi_1)$. Plugging in the model, we have

$$(Z_t - \mu) = \phi_1(Z_{t-1} - \mu) + a_t \tag{1}$$

This model also has two parameters (μ and ϕ_1). It explicitly uses the mean of the series. It is less commonly used in the literature, but is the model representation used in R.

5. Variance: $Var(Z_t) = \dfrac{\sigma_a^2}{1-\phi_1^2}$.

6. Forecast (minimum squared error): Suppose the forecast origin is n. For simplicity, we shall use the model representation in (1) and write $\dot{Z}_t = Z_t - \mu$. The model then becomes $\dot{Z}_t = \phi_1 \dot{Z}_{t-1} + a_t$. Note that forecast of Z_t is simply the forecast of $\dot{Z}_t$ plus μ.

   a. 1-step ahead forecast at time n:

$$\hat{\dot{Z}}_n(1) = \phi_1 \dot{Z}_n$$

   b. 1-step ahead forecast error:

$$e_n(1) = \dot{Z}_{n+1} - \hat{\dot{Z}}_n(1) = a_{n+1}$$

Thus, a_{n+1} is the *un-predictable* part of $\dot{Z}_{n+1}$. It is the shock at time $n+1$!

c. The variance of 1-step ahead forecast error:

$$Var[e_n(1)] = Var(a_{n+1}) = \sigma_a^2.$$

d. 2-step ahead forecast:

$$\hat{\dot{Z}}_n(2) = \phi_1 \hat{\dot{Z}}_n(1) = \phi_1^2 \dot{Z}_n.$$

e. 2-step ahead forecast error:

$$\begin{aligned} e_n(2) = \dot{Z}_{n+2} - \hat{\dot{Z}}_n(2) &= \phi_1 \dot{Z}_{n+1} + a_{n+2} - \phi_1 \hat{\dot{Z}}_n(1) \\ &= a_{n+2} + \phi_1 a_{n+1} \end{aligned}$$

f. The variance of 2-step ahead forecast error:

$$Var[e_n(2)] = (1 + \phi_1^2)\sigma_a^2$$

which is greater than or equal to $Var[e_n(1)]$, implying that uncertainty in forecasts increases as the number of steps increases.

g. Behavior of multi-step ahead forecasts. In general, for the l-step ahead forecast at n, we have

$$\hat{\dot{Z}}_n(l) = \phi_1^l \dot{Z}_n$$

the forecast error

- As $l \to \infty$,

$$\hat{\dot{Z}}_n(l) \to 0, \quad i.e., \quad \hat{Z}_n(l) \to \mu.$$

This is called the mean-reversion of the AR(1) process.

- The l-step forecast error

$$
\begin{aligned}
e_n(l) = \dot{Z}_{n+l} - \hat{\dot{Z}}_n(l) \\
&= \phi_0 + \phi_1 \dot{Z}_{n+l-1} + a_{n+l} - \phi_0 - \phi_1 \hat{\dot{Z}}_n(l-1) \\
&= a_{n+l} + \phi_1 e_n(l-1) \\
&= a_{n+l} + \phi_1 a_{n+l-1} + \cdots + \phi_1^{l-1} a_{n+1}
\end{aligned}
$$

- The variance of forecast error

$$
Var(e_n(l)) = \sum_{l=0}^{l-1} \phi_1^{2l} \sigma_a^2 \to \frac{\sigma_a^2}{1-\phi_1^2} = Var(Z_t)
$$

- In practice, it means that for the long-term forecasts serial dependence is not important.
- The forecast is just the sample mean and the uncertainty is simply the uncertainty about the series.

7. A compact form: $(1 - \phi_1 B)Z_t = \phi_0 + a_t$.
8. **Half-life:** A common way to quantify the speed of mean reversion is the half-life, which is defined as the number of periods needed so that the magnitude of the forecast becomes half of that of the forecast origin. For an AR(1) model, this mean

$$|\dot{Z}_n(k)| = |\frac{1}{2}\dot{Z}_n|.$$

Thus, $\phi_1^k \dot{Z}_n = \frac{1}{2}\dot{Z}_n$. Consequently, the half-life of the AR(1) model is $k = \frac{\ln(0.5)}{\ln(|\phi_1|)}$. For example, if $\phi_1 = 0.5$, the $k = 1$. If $\phi = 0.9$, then $k \approx 6.58$.

9. **ACF:** $E(\dot{Z}_{t-k}\dot{Z}_t) = E(\phi_1 \dot{Z}_{t-k}\dot{Z}_{t-1}) + E(\dot{Z}_{t-k}a_t)$

$$\begin{aligned} \gamma_k &= \phi_1 \gamma_{k-1}, k \geq 1 \\ \rho_k &= \phi_1 \rho_{k-1} = \phi_1^k, k \geq 1 \end{aligned}$$

ACF decay exponentially
$(i)\ 0 < \phi < 1 \quad positive\ correleted$
$(ii)\ -1 < \phi < 0 \quad negative\ correleted$

10. **PACF:**

$$\phi_{kk} = \begin{cases} \rho_1 = \phi, & k = 1 \\ 0, & k \geq 2 \end{cases}$$

The PACF of AR(1) models cuts off abruptly after lag 1.

AR(2) Model

1. From: $Z_t = \phi_0 + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + a_t$, or

$$
\begin{aligned}
(1 - \phi_1 B - \phi_2 B^2) Z_t &= \phi_0 + a_t, \\
\dot{Z}_t &= \phi_1 \dot{Z}_{t-1} + \phi_2 \dot{Z}_{t-2} + a_t, \dot{Z}_t = Z_t - \mu
\end{aligned}
$$

3. Mean: $\mu = E(Z_t) = \dfrac{\phi_0}{1 - \phi_1 - \phi_2}$.

   Variance: Can be solved from the following two equations

$$
\begin{aligned}
\gamma_0 &= \phi_1^2 \gamma_0 + \phi_2^2 \gamma_0 + \sigma_a^2 + 2\phi_1\phi_2\gamma_1 \\
\gamma_1 &= \phi_1 \gamma_0 + \phi_2 \gamma_1
\end{aligned}
$$

4. ACF: $\rho_0 = 1,\ \rho_1 = \frac{\phi_1}{1-\phi_2}$,

$$
\rho_l = \phi_1 \rho_{l-1} + \phi_2 \rho_{l-2}, \quad l \geq 2.
$$

Homogenous Linear Difference Equations

Assume X_t satisfies the following homogenous linear difference equations (齊次差分方程):

$$C(B)X_t = 0, \text{ where } C(B) = 1 + C_1 B + C_2 B^2 + \cdots + C_n B^n,$$

If $C(B) = \prod_{i=1}^{N}(1 - R_i B)^{m_i}$, then

$$X_t = \sum_{i=1}^{N} R_i^t \left(\sum_{j=0}^{m_i - 1} b_{ij} t^j \right)$$

where R_i^{-1}, $i = 1, 2, \ldots, N$ are the N distinct roots of $C(B) = 0$, with multiplicity m_i.

a. 當根均為相異 (no multiple roots)，則 $X_t = \sum_{i=1}^{n} b_i R_i^t$

b. 當有重根時 (there are multiple roots) e.g. $C(B) = (1-B)^2$, $X_t = b_1 + b_2 t$

c. 當根的倒數為複數根 (complex root):
If $R = c + di = \alpha(\cos\theta + i\sin\theta)$，where $\alpha = \sqrt{c^2 + d^2}$, $\theta = \tan^{-1}\left(\frac{d}{c}\right)$. Since

$$(c + di)^t = \alpha^t(\cos\theta t + i\sin\theta t),$$

the solutions of the difference equation include the following items

$$\alpha^t\cos\theta t, \alpha^t\sin\theta t, t\alpha^t\cos\theta t, t\alpha^t\sin\theta t, \cdots, t^{m-1}\alpha^t\cos\theta t, t^{m-1}\alpha^t\sin\theta t.$$

E.g.

$$X_t - 2X_{t-1} + X_{t-2} = 0$$

$$\because \quad (1 - 2B + B^2) = (1 - B)^2 = 0$$

$$\therefore \quad X_t = 1^t(b_1 + b_2 t) = b_1 + b_2 t$$

E.g.

$$(1 - \phi B)(1 - B)^2 X_t = 0$$

$$\because R_1 = \phi,\ R_2 = 1(\text{重根})$$

$$\therefore\ X_t = b_1\phi^t + b_2 + b_3 t$$

E.g.

$$X_t - 2X_{t-1} + 1.5X_{t-2} - 0.5X_{t-3} = 0$$

$$\because (1 - 2B + 1.5B^2 - 0.5B^3) = (1 - B + 0.5B^2)(1 - B)$$

$$R_1^{-1} = 1 \;\Rightarrow R_1 = 1$$

$$R_2^{-1} = \frac{1 + \sqrt{1 - 4(0.5)}}{2(0.5)} = 1 + i \quad (R_3^{-1} = 1 - i)$$

$$= \sqrt{2}\left(\cos\frac{\pi}{4} + i\sin\frac{\pi}{4}\right)$$

$$R_2 = \frac{1}{\sqrt{2}}\left(\cos\left(\frac{\pi}{4}\right) - i\sin\left(\frac{\pi}{4}\right)\right)$$

$$\therefore \; X_t = b_1 + b_2\left(\sqrt{\frac{1}{2}}\right)^t \cos\left(\frac{\pi}{4}t\right) + b_3\left(\sqrt{\frac{1}{2}}\right)^t \sin\left(\frac{\pi}{4}t\right)$$

Autocorrelation Function of AR(2) models

- The ACF of an AR(2) model satisfies the following second-order difference equation

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}, \quad k \geq 2 \tag{2}$$

with starting values $\rho_0 = 1$ and $\rho_1 = \phi_1/(1-\phi_2)$.

- If the roots of the characteristic equation are distinct real numbers, then the general solution of the difference equation is

$$\begin{aligned}\rho_k &= aG_1^k + bG_2^k \\ &= \frac{G_1(1-G_2^2)G_1^k - G_2(1-G_1^2)G_2^k}{(G_1-G_2)(1+G_1G_2)}\end{aligned}$$

where G_1^{-1} and G_2^{-1} are the roots of the characteristic equation $1 - \phi_1 B - \phi_2 B^2 = 0$.

The roots of the characteristic function $1 - \phi_1 x - \phi_2 x^2 = 0$

$$x = \frac{-\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{2\phi_2}.$$

If $\phi_1^2 + 4\phi_2 > 0$, then

$$\begin{aligned}\rho_k &= a\left(\frac{-\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{2\phi_2}\right)^{-k} + b\left(\frac{-\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}}{2\phi_2}\right)^{-k} \\ &= a\left(\frac{\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{2}\right)^{k} + b\left(\frac{\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}}{2}\right)^{k}\end{aligned}$$

其中 a, b 選擇滿足

$$\rho_1 = \phi_1/(1-\phi_2),\ \rho_2 = (\phi_1^2 + \phi_2 - \phi_2^2)/(1-\phi_2)$$

的條件 。

5. Stochastic business cycle:
When $\phi_1^2 + 4\phi_2 < 0$, the AR(2) model shows characteristics of business cycles.
ACF 所滿足的差分方程:

$$(1 - \phi_1 B - \phi_2 B^2)\rho_l = 0, \quad \text{for } l \geq 2$$

若 $\phi_1^2 + 4\phi_2 < 0$，則有兩個共軛根

$$x = \frac{-\phi_1 \pm \sqrt{-(\phi_1^2 + 4\phi_2)}i}{2\phi_2},$$

$$x^{-1} = \frac{2\phi_2}{-\phi_1 \pm \sqrt{-(\phi_1^2 + 4\phi_2)}i} = \frac{\phi_1 \pm \sqrt{-(\phi_1^2 + 4\phi_2)}i}{2} = \alpha e^{i\theta}, \alpha e^{-i\theta}$$

故 $\rho_k = c\alpha^k \cos k\theta + d\alpha^k \sin k\theta$ 。

$$\alpha = \sqrt{-\phi_2}, \quad \cos\theta = \frac{\phi_1}{2\sqrt{-\phi_2}}$$

where the cosine inverse is stated in radian.

若要求

$$\cos(k+p)\theta = \cos k\theta$$

$$\sin(k+p)\theta = \sin k\theta$$

$$\Rightarrow p\theta = 2\pi$$

$$\Rightarrow p = \frac{2\pi}{\theta} = \frac{2\pi}{\cos^{-1}[\phi_1/2\sqrt{-\phi_2}]}.$$

則

$$\rho_k \sim \rho_{k+p}.$$

此 p 即稱為 AR(2) 模型的 average length of the stochastic business cycle。

- If we denote the solutions of the polynomial

$$x^{-1} = \frac{\phi_1 \pm \sqrt{-(\phi_1^2 + 4\phi_2)}i}{2} = a \pm bi,$$

then we have

$$\phi_1 = 2a, \phi_2 = -(a^2 + b^2)$$

so that the average length of the stochastic business cycle

$$p = \frac{2\pi}{\cos^{-1}(a/\sqrt{a^2 + b^2})}.$$

In R or S-Plus, one can obtain $\sqrt{a^2 + b^2}$ using the command **Mod**.

6. Forecasts: Similar to AR(1) models.

Simple AR models:

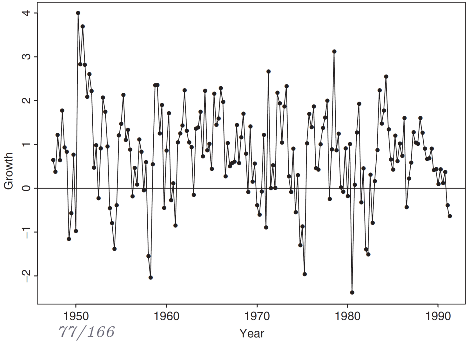
(Regression with lagged variables.)
Motivating example: The growth rate of U.S. quarterly real GNP from the second quarter of 1947 to the first quarter of 1991.

$$r_t = 0.0047 + 0.348 r_{t-1} + 0.179 r_{t-2} - 0.142 r_{t-3} + a_t, \quad \hat{\sigma}_a = 0.0097.$$

This is called an AR(3) model because the growth rate r_t depend on the growth rates of the past three quarters. Why is this model adequate? How do we specify this model from the data? These are the questions we shall address in this lecture.

Example

- Consider the quarterly growth rate of U.S. real gross national product (GNP Gross National Product 國民生產毛額), seasonally adjusted, from the second quarter of 1947 to the first quarter of 1991.
- Time plot of this series is shown below.



- Here we simply employ an AR(3) model for the data. Denoting the growth rate by r_t.
- The fitted model is

$$r_t = 0.0047 + 0.348 r_{t-1} + 0.179 r_{t-2} - 0.142 r_{t-3} + a_t, \quad \hat{\sigma}_a = 0.0097.$$

- Rewriting the model as

$$r_t - 0.348 r_{t-1} - 0.179 r_{t-2} + 0.142 r_{t-3} = 0.0047 + a_t,$$

we obtain a corresponding third-order difference equation

$$1 - 0.348B - 0.179B^2 + 0.141B^3 = 0,$$

which can be factored approximately as

$$(1 + 0.521B)(1 - 0.869B + 0.274B^2) = 0.$$

- The first factor $(1 + 0.521B)$ shows an exponentially decaying feature of the GNP growth rate.
- The second-order factor $1 - 0.869B + 0.274B^2 = 0$, we have

$$\phi_1^2 + 4\phi_2 = 0.869^2 + 4(-0.274) = -0.341 < 0.$$

  The second factor of the AR(3) model confirms the existence of stochastic business cycles in the quarterly growth rate of U.S. real GNP.
- The average length of the stochastic cycles is approximately

$$p = \frac{2(3.14159)}{\cos^{-1}[\phi_1/(2\sqrt{-\phi_2})]} = 10.62 \text{ quarters},$$

  which is about 3 years.

R Demonstration

```
> gnp=scan(file='dgnp82.txt') % Load data
 % To create a time-series object
> gnp1=ts(gnp,frequency=4,start=c(1947,2))
> plot(gnp1)
> points(gnp1,pch='*')

> m1=ar(gnp,method=''mle'') % Find the AR order
> m1$order    % An AR(3) is selected based on AIC
[1] 3
```

```
> m2=arima(gnp,order=c(3,0,0)) % Estimation
> m2
Call:
arima(x = gnp, order = c(3, 0, 0))

Coefficients:
         ar1     ar2      ar3   intercept
      0.3480  0.1793  -0.1423      0.0077
s.e.  0.0745  0.0778   0.0745      0.0012

sigma^2 estimated as 9.427e-05: log likelihood=565.84,
   aic=-1121.68
```

```
% In R, ''intercept'' denotes the mean of the series.
% Therefore, the constant term is obtained below:
> (1-.348-.1793+.1423)*0.0077
[1] 0.0047355
> sqrt(m2$sigma2)  % Residual standard error
[1] 0.009709322

> p1=c(1,-m2$coef[1:3]) % Characteristic equation
> roots=polyroot(p1) % Find solutions
> roots
[1] 1.590253+1.063882i -1.920152+0.000000i 1.590253-1.063882i
> Mod(roots) % Compute the absolute values of the solutions
[1] 1.913308 1.920152 1.913308
 % To compute average length of business cycles:
> k=2*pi/acos(1.590253/1.913308)
> k
[1] 10.65638
```

6. AR(2) 平穩的參數區間: characteristic equation
$(1-\phi_1 B-\phi_2 B^2)=0$ 的根落在單位圓外

$$B_1=\frac{-\phi_1+\sqrt{\phi_1^2+4\phi_2}}{2\phi_2},\quad B_2=\frac{-\phi_1-\sqrt{\phi_1^2+4\phi_2}}{2\phi_2}$$

$$B_1^{-1}=\frac{\phi_1+\sqrt{\phi_1^2+4\phi_2}}{2},\quad B_2^{-1}=\frac{\phi_1-\sqrt{\phi_1^2+4\phi_2}}{2}$$

Z_t 有平穩解 $\leftrightarrow |B_i^{-1}|<1, \quad i=1,2$。

(a) $\begin{cases} 1 > |B_1^{-1} \cdot B_2^{-1}| & = |\phi_2|, \text{(考慮}x^2 - \phi_1 x - \phi_2 = 0\text{的根)} \\ 2 > |B_1^{-1} + B_2^{-1}| & = |\phi_1| \end{cases}$,

故 $\begin{cases} -1 & < \phi_2 < 1 \\ -2 & < \phi_1 < 2 \end{cases}$ 為平穩的必要條件 。

(b) 有實根時，$\phi_1^2 + 4\phi_2 \geq 0$，則

$$-1 < B_2^{-1} \leq B_1^{-1} < 1 \Rightarrow \begin{cases} \phi_2 + \phi_1 & < 1 \\ \phi_2 - \phi_1 & < 1 \end{cases} 。$$

(c) 有虛根時，$\phi_1^2 + 4\phi_2 < 0.$

Homework: Derive the stationary region of ϕ_1 and ϕ_2 for AR(2) Model: $(1 - \phi_1 B - \phi_2 B^2)\dot{Z}_t = a_t$

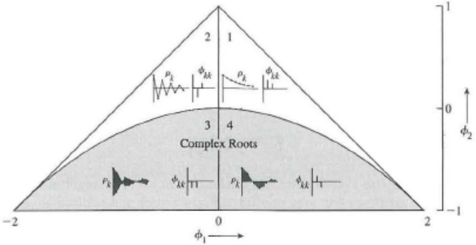


Figure: Typical autocorrelation and partial autocorrelation functions ρ_k and ϕ_{kk} for various stationary AR(2) models.

Stationary region of AR(2) models:

$$\begin{aligned} \phi_2 + \phi_1 &< 1 \\ \phi_2 - \phi_1 &< 1 \\ -1 < \phi_2 &< 1 \end{aligned} \tag{3}$$

shown in Figure 1.

8. MA(∞) representation of AR(2):

$$(1-\phi_1 B-\phi_2 B^2)Z_t=\phi_0+a_t$$

- $$Z_t=(1-\phi_1 B-\phi_2 B^2)^{-1}(\phi_0+a_t)=\left(\sum_{j=0}^{\infty}\psi_j B^j\right)(\phi_0+a_t)$$

  其中 $1=(1-\phi_1 B-\phi_2 B^2)(\psi_0+\psi_1 B+\psi_2 B^2+\psi_3 B^3+\cdots)$

- $$\psi_0=1$$
  $$\psi_1-\phi_1=0$$
  $$\psi_k-\phi_1\psi_{k-1}-\phi_2\psi_{k-2}=0,\quad k\geq 2$$

AR(p) model

- $$(1-\phi_1 B-\phi_2 B^2\cdots-\phi_p B^p)\dot{Z}_t=a_t$$
- **ACF:**

$$E(\dot{Z}_{t-k}\dot{Z}_t)=E(\phi_1\dot{Z}_{t-k}\dot{Z}_{t-1}+\cdots+\phi_p\dot{Z}_{t-k}\dot{Z}_{t-p}+\dot{Z}_{t-k}a_t)$$

$$\gamma_k=\phi_1\gamma_{k-1}+\cdots+\phi_p\gamma_{k-p},\quad k=1,2,\ldots,p$$

$$\Rightarrow \rho_k=\phi_1\rho_{k-1}+\cdots+\phi_p\rho_{k-p},\quad k=1,2,\ldots,p$$

- 此 p 個聯立方程式，稱為 AR(p) 係數所滿足的 Yule Walker Equation 。

- Suppose

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p = \prod_{i=1}^{m} (1 - G_i B)^{d_i}, \quad \sum_{i=1}^{m} d_i = p$$

$$\Rightarrow \rho_k = \sum_{i=1}^{m} G_i^k \sum_{j=0}^{d_i - 1} A_{ij} k^j$$

(i) If all G_i are distinct, then $\rho_k = \sum_{i=1}^{p} A_i G_i^k, k > 0$
(ii) Stationary, if $|G_i^{-1}| > 1$

- ACF of an AR(p) model exponentially decays to 0 as the lag increases.

- **PACF:** Recall

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

- 由於

$$\rho_k = \phi_1 \rho_{k-1} + \cdots + \phi_p \rho_{k-p}$$

- 因此當 $k > p$ 時，行列式中的最後一行為前 p 行的線性組合，故 $\phi_{kk} = 0$ 。

- 從另一個觀點：
  將 Z_{t+k} 對 $Z_{t+k-1}, Z_{t+k-2}, \cdots, Z_{t+1}, Z_t$ 建立迴歸模型

$$Z_{t+k} = \phi_{k1} Z_{t+k-1} + \phi_{k2} Z_{t+k-2} + \cdots + \phi_{kk} Z_t + \varepsilon_{t+k}$$

  則第 k 個迴歸係數 ϕ_{kk} 的值，即為 lag k 的 PACF 值 。
- AR(p) 模型：

$$\dot{Z}_{t+k} = \phi_1 \dot{Z}_{t+k-1} + \phi_2 \dot{Z}_{t+k-2} \cdots + \phi_p \dot{Z}_{t+k-p} + a_t$$

  即 Z_{t+k} 可表為前 p 個觀測值 $Z_{t+k-1}, Z_{t+k-2}, \cdots, Z_{t+1}, Z_t$ 的線性組合 。故若 $k > p, \phi_{kk} = 0$ 。
- PACF of AR(p) models cut off after lag p.

$AR(p)$ 模型的 l 步預測

- Find $\hat{Z}_k(l)$ (a function of $Z_1, Z_2, \ldots, Z_k$) to minimize the mean squared prediction error

$$\min_{\hat{Z}_k(l)} E[Z_{k+l} - \hat{Z}_k(l)]^2$$

$\Longrightarrow$

$$\hat{Z}_k(l) = E(Z_{k+l}|Z_k, Z_{k-1}, \cdots) = \phi_0 + \sum_{i=1}^{p} \phi_i \hat{Z}_k(l-i)$$

where $\hat{Z}_k(i) = Z_{k+i}$ if $i \leq 0$.

- l 步預測誤差，$e_k(l) = Z_{k+l} - \hat{Z}_k(l)$ 。

對一個平穩的 AR(p) 模型，可證明當 $\to \infty$

1. $\hat{Z}_k(l) \to E(Z_t)$, 表示長期的預測趨近於無條件平均，此性質在財務上稱為 mean reversion 的現象。
2. $Var(e_k(l)) \to Var(Z_t)$。

Building an AR(p) model

Order specification:

Two general approaches are available for determining the value of p.

- The first approach is to use the partial autocorrelation function.
- The second approach uses some information criteria, e.g. AIC.

1. Partial ACF: (naive, but effective)

- **Key feature**: PACF cuts off at lag p for an AR(p) model.
- Use consecutive fittings.

2. Information criteria

- There are several information criterion available to determine the order p of an AR process.
- All of them are likelihood based.
- The well known Akaike (赤池) information criterion (AIC) (Akaike, 1973)

$$\text{AIC}(p) = \frac{-2}{T}\ln(\text{likelihood}) + \frac{2}{T} \times (\text{number of parameters})$$

where

- the likelihood function is evaluated at the mle;
- T is the sample size.

Information criteria

- For a Gaussian AR(p) model, AIC reduces to

$$\text{AIC}(p) = \ln(\hat{\sigma}_p^2) + \frac{2p}{T},$$

where $\hat{\sigma}_p^2$ is the MLE of the residual variance.

- **Goodness of Fit + Penalty for Model Complexity**
- Goal: Find the AR order p with minimum AIC.

Building an AR(p) model

- 實務上可先藉由 PACF 決定一個上界 p，求所有的 AIC(k), $k \leq p$ 的值，挑選使 AIC 值最小的 k。 或是直接使用以下的 R codes to find the order of the AR model which has the minimum AIC.
- >m1=ar(gnp, method="mle")(% Find the AR order)
- >m1$order(% show the order with the minimum AIC)
- 此外亦需考慮研究的目的來選取模型 。

Example 1

- The PACF of the U.S. quarterly growth rate of GNP (gross national product 國民生產毛額或國民生產總值).
- The AIC obtained from R also identifies an AR(3) model.

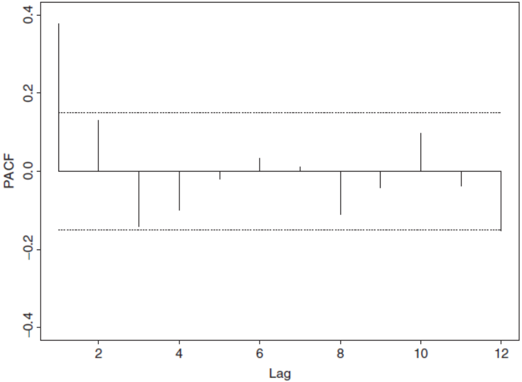


Figure 2.6 Sample partial autocorrelation function of U.S. quarterly real GNP growth rate from 1947.II to 1991.I. Dotted lines give approximate pointwise 95% confidence interval.

- Note that the AIC value of the ar command in R has been adjusted so that the minimum AIC is zero.

```
> gnp=scan(file='q-gnp4791.txt')
> ord=ar(gnp,method=''mle'')
> ord$aic
 [1] 27.847 2.742 1.603 0.000 0.323 2.243
 [7]  4.052 6.025 5.905 7.572 7.895 9.679
> ord$order
[1] 3
```

Asymptotic efficiency

- The mean-square prediction error is defined as

$$H(p) = E_Z(Y_{n+1} - \hat{\phi}_{p1}Y_n - \cdots - \hat{\phi}_{pp}Y_{n+1-p})^2$$

其中 $\hat{\phi}_{p1}, \ldots, \hat{\phi}_{pp}$ 為配適 AR(p) 模型的 Yule-Walk estimate，$\{Y_1, \ldots, Y_n\}$ 為一組與 $\{Z_1, \ldots, Z_n\}$ 獨立但具相同分佈的 r.v.'s.

- An efficient order selection procedure chooses an AR model which achieves the optimal rate of convergence of the mean-square prediction error.
- That is 所選擇的 $\hat{p}_n$ 會滿足 $\frac{H(p_n^*)}{H(\hat{p}_n)} \to 1$, as $n \to \infty$, 其中 p_n^* 會使 $H(p)$ 達到最小值.

Consistency

- AIC 是一個 asymptotically efficient 的選模準則, 通常會高估階次，所以後來有學者提出修正的選模準則，稱為 AICC 。

$$\text{AICC} = \text{AIC} + \frac{2p(p+1)}{T-p-1}$$

- AICC: an asymptotically unbiased estimator.
- Schwarz-Bayesian information criterion (BIC) for Gaussian AR(p):

$$\text{BIC}(p) = \ln(\hat{\sigma}_p^2) + \frac{p\ln(T)}{T}.$$

- BIC is a consistent criterion (所選擇的 $\hat{p}_n$ 會收斂到真值, i.e. $\hat{p}_n \to p_0, n \to \infty$).

Example 2

- Monthly simple returns of CRSP Value-Weighted Index, Jan. $1926 \sim$ Dec. 2008.
- The AIC obtained from R identifies an AR(9) model.

p	1	2	3	4	5	6
PACF	0.115	-0.030	-0.102	0.033	0.062	-0.050
AIC	-5.838	-5.837	-5.846	-5.845	-5.847	-5.847
BIC	-5.833	-5.827	-5.831	-5.825	-5.822	-5.818
p	7	8	9	10	11	12
PACF	0.031	0.052	0.063	0.005	-0.005	0.011
AIC	-5.846	-5.847	-5.849	-5.847	-5.845	-5.843
BIC	-5.812	-5.807	-5.805	-5.798	-5.791	-5.784

Parameter Estimation

- Needs a constant term? Check the sample mean.
- Conditioning on the first p observations, we have

$$Z_t = \phi_0 + \phi_1 Z_{t-1} + \cdots + \phi_p Z_{t-p} + a_t, \quad t = p+1, \ldots, T,$$

which is in the form of a multiple linear regression and can be estimated by the least-squares method.

- Denote the estimate of ϕ_i by $\hat{\phi}_i$. The fitted model is

$$\hat{Z}_t = \hat{\phi}_0 + \hat{\phi}_1 Z_{t-1} + \cdots + \hat{\phi}_p Z_{t-p},$$

and the associated residual is

$$\hat{a}_t = Z_t - \hat{Z}_t.$$

- The series $\{\hat{a}_t\}$ is called the residual series, from which we obtain

$$\hat{\sigma}_a^2 = \frac{\sum_{t=p+1}^{T} \hat{a}_t^2}{T - 2p - 1}.$$

- If the conditional-likelihood method is used, the estimates of ϕ_i remain unchanged, but the estimate of σ_a^2 becomes

$$\tilde{\sigma}_a^2 = \hat{\sigma}_a^2 \times \frac{T-2p-1}{T-p}.$$

- In some packages, $\tilde{\sigma}_a^2$ is defined as

$$\hat{\sigma}_a^2 \times \frac{T-2p-1}{T}.$$

Model Checking

1. Residual: obs minus the fit, i.e. 1-step ahead forecast errors at each time point.
2. Residual should be close to white noise if the model is adequate. Use Ljung-Box statistics of residuals, but degrees of freedom is $m - p$, where p is the number of AR coefficients used in the model.

Goodness of Fit

- A commonly used statistic to measure goodness of fit of a stationary model is the R square (R^2) defined as

$$R^2 = 1 - \frac{\text{residual sum of squares}}{\text{total sum of squares}}.$$

- For a stationary AR(p) time series model with T observations $\{r_t | t = 1, \ldots, T\}$, the measure becomes

$$R^2 = 1 - \frac{\sum_{t=p+1}^{T} \hat{a}_t^2}{\sum_{t=p+1}^{T} (r_t - \bar{r})^2},$$

  where $\bar{r} = \sum_{t=p+1}^{T} r_t / (T - p)$.
- It is easy to show that $0 \leq R^2 \leq 1$.

- Typically, a larger R^2 indicates that the model provides a closer fit to the data.
- However, this is only true for a stationary time series.
- For the unit-root nonstationary series discussed later in this chapter, R^2 of an AR(1) fit converges to one when the sample size increases to infinity, regardless of the true underlying model of r_t.

- For a given data set, it is well known that R^2 is a nondecreasing function of the number of parameters used.
- To overcome this weakness, an adjusted R^2 is proposed, which is defined as

$$\begin{aligned}\text{Adj-}R^2 &= 1 - \frac{\text{variance of residuals}}{\text{variance of } r_t} \\ &= 1 - \frac{\hat{\sigma}_a^2}{\hat{\sigma}_r^2},\end{aligned}$$

  where $\hat{\sigma}_r^2$ is the sample variance of r_t.
- This new measure takes into account the number of parameters used in the fitted model. However, it is no longer between 0 and 1.

Example: monthly value-weighted simple returns, 1926.01-2008.12

- Although for the monthly value-weighted simple returns, the minimum AIC order is 9, here we fit the data by the AR(3) model suggested by the PACF.
- Consider the residual series of the fitted AR(3) model for the monthly value-weighted simple returns from January 1926 to December 2008.
- We have $Q(12) = 16.35$ with a p value 0.060 (based on χ^2_9, degrees of freedom is $m - p = 12 - 3 = 9$, where p is the number of AR coefficients used in the model).
- The null hypothesis of no residual serial correlation in the first 12 lags is barely not rejected at the 5% level.

Example - continue

- However, since the lag-2 AR coefficient is not significant at the 5% level, one can refine the model as

$$r_t = 0.0088 + 0.114 r_{t-1} - 0.106 r_{t-3} + a_t, \quad \hat{\sigma}_a = 0.0536,$$

  where all the estimates are now significant at the 1% level.
- The residual series gives $Q(12) = 16.83$ with a p value 0.078 (based on χ^2_{10}). The model is adequate in modeling the dynamic linear dependence of the data.

R Demonstration

```
> vw=read.table('m-ibm3dx2608.txt',header=T)[,3]
> m3=arima(vw,order=c(3,0,0))
> m3
Call:
arima(x = vw, order = c(3, 0, 0))

Coefficients:
         ar1      ar2      ar3   intercept
      0.1158  -0.0187  -0.1042      0.0089
s.e.  0.0315   0.0317   0.0317      0.0017

sigma^2 estimated as 0.002875: log likelihood=1500.86,
   aic=-2991.73

> (1-.1158+.0187+.1042)*mean(vw) % Compute
   the intercept phi(0).
[1] 0.00896761
> sqrt(m3$sigma2)  % Compute standard error of residuals
[1] 0.0536189
```

```
% Approach 1:
> Box.test(m3$residuals, lag=12, type="Ljung", fitdf=3)
        Box-Ljung test
data: m3$residuals      % R use 9 degrees of freedom
X-squared = 16.3525,    df = 9,    p-value = 0.05988

% Approach 2:
> Box.test(m3$residuals, lag=12, type="Ljung")
        Box-Ljung test
data: m3$residuals      % R use 12 degrees of freedom
X-squared = 16.3525,    df = 12,    p-value = 0.1756
> pv=1-pchisq(16.3525,9)      % Compute p-value using 9 degrees of freedom
> pv
[1] 0.05987551
```

```
% To fix the AR(2) coef to zero:
> m3=arima(vw,order=c(3,0,0),fixed=c(NA,0,NA,NA))
% The subcommand 'fixed' is used to fix parameter values,
% where NA denotes estimation and 0 means fixing the parameter to 0.
% The ordering of the parameters can be found using m3$coef.
> m3
```

```
Call:
arima(x = vw, order = c(3, 0, 0), fixed = c(NA, 0, NA, NA))

Coefficients:
         ar1  ar2      ar3  intercept
      0.1136    0  -0.1063     0.0089
s.e.  0.0313    0   0.0315     0.0017

sigma^2 estimated as 0.002876: log likelihood=1500.69,
   aic=-2993.38
> (1-.1136+.1063)*.0089 % Compute phi(0)
[1] 0.00883503
> sqrt(m3$sigma2)  % Compute residual standard error
[1] 0.05362832

> Box.test(m3$residuals,lag=12,type='Ljung')

      Box-Ljung test

data: m3$residuals
X-squared = 16.8276, df = 12, p-value = 0.1562

> pv=1-pchisq(16.83,10)
> pv
[1] 0.0782113
```

- Table 2.2 contains the $1-$step to $12-$step ahead forecasts and the standard errors of the associated forecast errors at the forecast origin 984 for the monthly simple return of the value-weighted index using an AR(3) model that was reestimated using the first 984 observations.
- The fitted model is

$$r_t = 0.0098 + 0.1024r_{t-1} - 0.0201r_{t-2} - 0.1090r_{t-3} + a_t,$$

where $\hat{\sigma}_a^2 = 0.054$.

- The actual returns of 2008 are also given in Table 2.2.
- Because of the weak serial dependence in the series, the forecasts and standard deviations of forecast errors converge to the sample mean and standard deviation of the data quickly.

> reEstimateModel=arima(vw[1:984], order=c(3,0,0))

> reEstimateModel

```
Call:
arima(x = vw[1:984], order = c(3, 0, 0))

Coefficients:
         ar1      ar2      ar3  intercept
      0.1034  -0.0201  -0.1089     0.0095
s.e.  0.0317   0.0318   0.0317     0.0017

sigma^2 estimated as 0.00284:  log likelihood = 1488.85,  aic = -2967.71
```

> forecast=predict(reEstimateModel, n.ahead=12)

> forecast

```
$pred
Time Series:
Start = 985
End = 996
Frequency = 1
 [1] 0.007455686 0.015952488 0.011701663 0.009806653 0.008770886 0.009164644
 [7] 0.009432520 0.009565095 0.009530555 0.009495151 0.009477748 0.009480420

$se
Time Series:
Start = 985
End = 996
Frequency = 1
 [1] 0.05328902 0.05357340 0.05357574 0.05390675 0.05392026 0.05392029
 [7] 0.05392456 0.05392493 0.05392493 0.05392499 0.05392500 0.05392500
```

TABLE 2.2 Multistep Ahead Forecasts of an AR(3) Model for Monthly Simple Returns of CRSP Value-Weighted Index

Step	1	2	3	4	5	6
Forecast	0.0076	0.0161	0.0118	0.0099	0.0089	0.0093
Std. Error	0.0534	0.0537	0.0537	0.0540	0.0540	0.0540
Actual	−0.0623	−0.0220	−0.0105	0.0511	0.0238	−0.0786
Step	7	8	9	10	11	12
Forecast	0.0095	0.0097	0.0096	0.0096	0.0096	0.0096
Std. Error	0.0540	0.0540	0.0540	0.0540	0.0540	0.0540
Actual	−0.0132	0.0110	−0.0981	−0.1847	−0.0852	0.0215

[a] The forecast origin is $h = 984$.

- Figure 2.7 shows the corresponding out-of-sample prediction plot for the monthly simple return series of the value-weighted index.
- The forecast origin $t = 984$ corresponds to December 2007.
- The prediction plot includes the two standard error limits of the forecasts and the actual observed returns for 2008.
- The forecasts and actual returns are marked by $\circ$ and $\bullet$, respectively.
- From the plot, except for the return of October 2008, all actual returns are within the 95% prediction intervals.

% Figure 2.7

```
> time=seq(2007+1/12, 2009, 1/12)
> upperDashed=c(vw[984], forecast$pred+2*forecast$se)
> lowerDashed=c(vw[984], forecast$pred-2*forecast$se)

> plot(time,vw[973:996], pch=16, xlab="Time", ylab="Simple return", xlim=c(2007.09,2009), ylim=c(-0.2,0.2))
> par(new=T)
> plot(time, vw[973:996], type="l", xlab="", ylab="", xlim=c(2007.09,2009), ylim=c(-0.2,0.2))
> par(new=T)
> plot(time[13:24], forecast$pred, xlab="", ylab="", xlim=c(2007.09,2009), ylim=c(-0.2,0.2))
> par(new=T)
> plot(time[12:24], c(vw[984], forecast$pred), type="l", xlab="", ylab="", xlim=c(2007.09,2009), ylim=c(-0.2,0.2))
> par(new=T)
> plot(time[12:24], upperDashed, type="l", lty=2, xlab="", ylab="", xlim=c(2007.09,2009), ylim=c(-0.2,0.2))
> par(new=T)
> plot(time[12:24], lowerDashed, type="l", lty=2, xlab="", ylab="", xlim=c(2007.09,2009), ylim=c(-0.2,0.2))
```

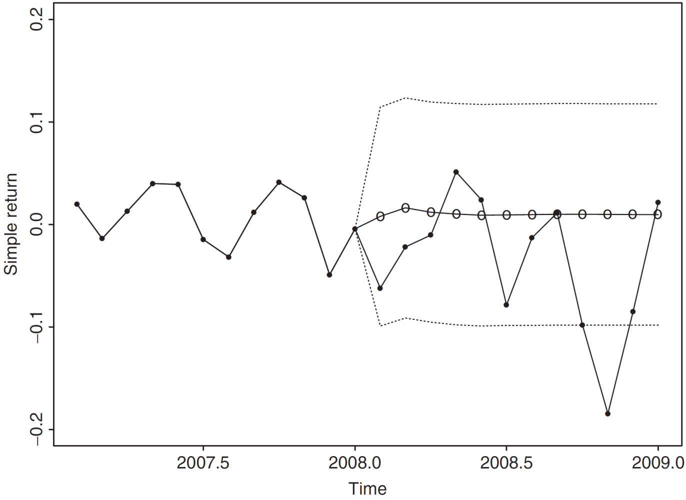


Figure 2.7 Plot of 1- to 12-step-ahead out-of-sample forecasts for monthly simple returns of CRSP value-weighted index. Forecast origin is $t = 984$, which is December 2007. Forecasts are denoted by "o" and actual observations by "•". Two dashed lines denote two standard error limits of the forecasts.

MA processes

1. $\dot{Z}_t = a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q} = \theta(B)a_t$, 不論係數 $(\theta_i's)$ 為何，$\{\dot{Z}_t = Z_t - \mu\}$ is always covariance stationary 。
2. Is an MA processes invertible $\theta(B)^{-1}\dot{Z}_t = a_t$?
   - Concept: $\dot{Z}_t$ is a proper linear combination of a_t and the past observations $\{\dot{Z}_{t-1}, \dot{Z}_{t-2}, \cdots\}$.
   - Why is it important? It provides a simple way to obtain the shock a_t.
   - For an invertible model, the dependence of $\dot{Z}_t$ on $\dot{Z}_{t-q}$ converges to zero as q increase.
3. MA processes is invertible $\Leftrightarrow \theta(B) = 0$ 的根落在單位圓外 。
4. Invertibility of MA models is the dual property of stationarity for AR models.

- MA processes are useful in describing phenomena in which events produce an immediate effect that only lasts for short periods of time.
- Model with finite memory!
- Some daily stock returns have minor serial correlations and can be modeled as MA or AR models.

MA(1)

$$\dot{Z}_t = (1-\theta_1 B)a_t, \quad a_t \sim WN(0, \sigma_a^2)$$

- Invertible $\Leftrightarrow \frac{1}{|\theta_1|} > 1 \Leftrightarrow |\theta_1| < 1$
- Autocovariance generating function

$$\gamma(B) = \sigma_a^2(1-\theta_1 B)(1-\theta_1 B^{-1}) = \sigma_a^2\{-\theta_1 B^{-1} + (1+\theta_1^2) - \theta_1 B\}$$

$$\Rightarrow \gamma_k = \begin{cases} (1+\theta_1^2)\sigma_a^2, & k=0 \\ -\theta_1\sigma_a^2, & k=1 \\ 0, & k>1 \end{cases}$$

$$\Rightarrow \rho_k = \begin{cases} \frac{-\theta_1}{1+\theta_1^2}, & k=1 \\ 0, & k>1 \end{cases}$$

Remark:

(i) $\dot{Z}_t = (1-\theta_1 B)a_t$ 與 $\dot{Z}_t = (1-\frac{1}{\theta_1}B)a_t$ 有相同的 ACF，故為了使 ACF 與模型有唯一的對應關係，一般採取 invertible 的模型，即 $|\theta_1| < 1$ 的解 。

(ii) Since $\rho_1 = -\frac{\theta_1}{1+\theta_1^2}$ 故 $|\rho_1| < 0.5$ 。

- **PACF:**

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

$$\phi_{11} = \rho_1 = \frac{-\theta_1}{1+\theta_1^2} = \frac{-\theta_1(1-\theta_1^2)}{1-\theta_1^4}$$

$$\phi_{22} = \frac{-\rho_1^2}{1-\rho_1^2} = \frac{-\theta_1^2(1-\theta_1^2)}{1-\theta_1^6}$$

$$\phi_{33} = \frac{\rho_1^3}{1-2\rho_1^2} = \frac{-\theta_1^3(1-\theta_1^2)}{(1-\theta_1^8)}$$

$$\vdots$$

$$\phi_{kk} = \frac{-\theta_1^k(1-\theta_1^2)}{1-\theta_1^{2(k+1)}}, k \geq 1$$

$\Rightarrow$ tails off exponentially.

e.g. 模擬 MA(1) 模型 $\dot{Z}_t = (1 - 0.5B)a_t$ 的樣本 ACF 及 PACF

Table 3.5 Sample ACF and sample PACF for a simulated series from $Z_t = (1 - 0.5B)a_t$

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	-0.44	0.00	0.02	-0.03	-0.01	-0.05	0.04	-0.03	-0.03	0.02
St.E.	0.06	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.08	0.08
$\hat{\phi}_{kk}$	-0.44	-0.24	-0.11	-0.08	-0.07	-0.12	-0.06	-0.07	-0.10	-0.08
St.E.	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06

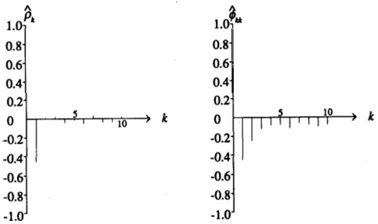


Fig. 3.11 Sample ACF and sample PACF of a simulated MA(1) series: $Z_t = (1 - .5B)a_t$.

- Forecast (at origin $t = n$):
  MA(1) process: $Z_{n+1} = \mu + a_{n+1} - \theta_1 a_n$.
  1. 1-step ahead: $\hat{Z}_n(1) = \mu - \theta_1 a_n$. Why? Because at time n, a_n is known, but a_{n+1} is not.
  2. 1-step ahead forecast error: $e_n(1) = a_{n+1}$ with variance σ_a^2.

$$Z_{n+1} = \mu + a_{n+1} - \theta_1 a_n$$

3. Multi-step ahead, $l > 1$: $\hat{Z}_n(l) = \mu$ for $l > 1$. Thus, for an MA(1) model, the multi-step ahead forecasts are just the mean of the series. Why? Because the model has memory of 1 time period.
4. Multi-step ahead forecast error:

$$e_n(l) = a_{n+l} - \theta_1 a_{n+l-1}$$

5. Variance of multi-step ahead forecast error:

$$(1 + \theta_1^2)\sigma_a^2 = \text{ variance of } Z_t$$

MA(2)

1. 模型 $\dot{Z}_t = (1 - \theta_1 B - \theta_2 B^2) a_t$
2. invertible $\Leftrightarrow 1 - \theta_1 B - \theta_2 B^2 = 0$ 的根在單位圓外

$$\Leftrightarrow \begin{cases} \theta_2 + \theta_1 < 1 \\ \theta_2 - \theta_1 < 1 \\ -1 < \theta_2 < 1 \end{cases}$$

- **ACF:** Autocovariance generating function

$$\begin{aligned}
\gamma(B) &= \psi(B)\psi(B^{-1})\sigma_a^2 \\
&= \sigma_a^2(1 - \theta_1 B - \theta_2 B^2)(1 - \theta_1 B^{-1} - \theta_2 B^{-2}) \\
&= \sigma_a^2\{-\theta_2 B^{-2} - \theta_1(1 - \theta_2)B^{-1} \\
&\qquad + (1 + \theta_1^2 + \theta_2^2) - \theta_1(1 - \theta_2)B - \theta_2 B^2\}
\end{aligned}$$

$$\Rightarrow \gamma_0 = (1 + \theta_1^2 + \theta_2^2)\sigma_a^2$$

$$\gamma_1 = -\theta_1(1 - \theta_2)\sigma_a^2$$

$$\gamma_2 = -\theta_2\sigma_a^2$$

$$\gamma_k = 0, \quad k > 2$$

$$\Rightarrow \rho_k = \begin{cases} \frac{-\theta_1(1-\theta_2)}{1+\theta_1^2+\theta_2^2}, & k = 1 \\ \frac{-\theta_2}{1+\theta_1^2+\theta_2^2}, & k = 2 \\ 0, & k > 2 \end{cases}$$

- **PACF:**

$$\phi_{11} = \rho_1$$

$$\phi_{22} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$$

$$\phi_{33} = \frac{\rho_1^3 - \rho_1\rho_2(2 - \rho_2)}{1 - \rho_2^2 - 2\rho_1^2(1 - \rho_2)}$$

$$\vdots$$

$\Rightarrow$ tails off exponentially.

e.g. MA(2) 模型 $\dot{Z}_t = (1 - 0.65B - 0.24B^2)a_t$, $a_t \sim N(0,1)$ 的樣本 ACF 及 PACF

Table 3.6 Sample ACF and sample PACF for a simulated MA(2) series from $Z_t = (1 - 0.65B - 0.24B^2)a_t$

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	-0.35	-0.17	0.09	-0.06	0.01	-0.01	-0.04	0.07	-0.07	0.09
St.E.	0.06	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07
$\hat{\phi}_{kk}$	-0.35	-0.34	-0.15	-0.18	-0.11	-0.12	-0.14	-0.05	-0.14	0.00
St.E.	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06

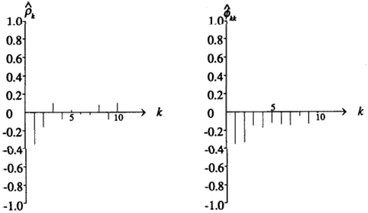


Fig. 3.13 Sample ACF and sample PACF of a simulated MA(2) series: $Z_t = (1 - .65B - .24B^2)a_t$.

- Forecasts of an MA(2) model go to the mean of the series after two steps.
-

$$
\begin{aligned}
Z_{n+l} &= \mu + a_{n+l} - \theta_1 a_{n+l-1} - \theta_2 a_{n+l-2} \\
\hat{Z}_n(1) &= \mu - \theta_1 a_n - \theta_2 a_{n-1} \\
\hat{Z}_n(2) &= \mu - \theta_2 a_n \\
\hat{Z}_n(l) &= \mu, \quad l \geq 2
\end{aligned}
$$

- a_t's $(t \geq 2)$ can be obtained recursively from

$$a_t = (1 - \theta_1 B - \theta_2 B^2)^{-1}(Z_t - \mu)$$

setting $\ldots, a_{-1} = 0,\ a_0 = 0,\ a_1 = Z_1 - \mu$.

$$Z_{n+l} = \mu + a_{n+l} - \theta_1 a_{n+l-1} - \theta_2 a_{n+l-2}$$
$$\hat{Z}_n(1) = \mu - \theta_1 a_n - \theta_2 a_{n-1}$$
$$\hat{Z}_n(2) = \mu - \theta_2 a_n$$
$$\hat{Z}_n(l) = \mu, \quad l \geq 2$$

- $e_n(1) = Z_{n+1} - \hat{Z}_n(1) = a_{n+1}$
  $\Longrightarrow Var(e_n(1)) = \sigma_a^2$
- $e_n(2) = a_{n+2} - \theta_1 a_{n+1}$
  $\Longrightarrow Var(e_n(2)) = (1 + \theta_1^2)\sigma_a^2$
- $e_n(l) = a_{n+l} - \theta_1 a_{n+l-1} - \theta_2 a_{n+l-2}$
  $\Longrightarrow Var(e_n(l)) = (1 + \theta_1^2 + \theta_2^2)\sigma_a^2, \quad l > 2$

MA(q)

$$\dot{Z}_t = (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q)a_t \Rightarrow \gamma_0 = \sigma_a^2 \sum_{j=0}^{q} \theta_j^2, \quad \theta_0 = 1$$

$$\gamma_k = \begin{cases} \sigma_a^2(-\theta_k + \theta_1\theta_{k+1} + \cdots + \theta_{q-k}\theta_q), & k = 1, 2, \ldots, q, \\ 0, & k > q. \end{cases}$$

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

(i) cut off after lag q.

(ii) PACF tails off exponentially.

Building an MA model

- Specification: Use sample ACF
  Sample ACFs are all small after lag q for an MA(q) series.
  (See test of ACF.)
- Constant term? Check the sample mean.
- Estimation: use maximum likelihood method
  - Conditional: Assume $a_t = 0$ for $t \leq 0$
  - Exact: Treat a_t with $t \leq 0$ as parameters, estimate them to obtain the likelihood function.

  Exact method is preferred, but it is more computing intensive.
- Model checking: examine residuals (to be white noise)
- Forecast: use the residuals as $\{a_t\}$ (which can be obtained from the data and fitted parameters) to perform forecasts.

Model form in R

R parameterizes the MA(q) model as

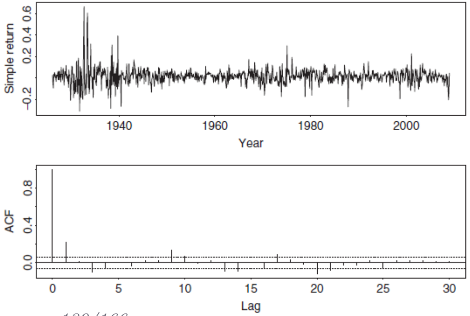
$$Z_t = \mu + a_t + \theta_1 a_{t-1} + \cdots + \theta_q a_{t-q},$$

instead of the usual *minus* sign in θ. Consequently, care needs to be exercised in writing down a fitted MA parameter in R. For instance, an estimate $\hat{\theta}_1 = -0.5$ of an MA(1) in R indicates the model is $Z_t = a_t - 0.5a_{t-1}$.

Example

- CRSP 的 equal weighted index, $1926/1 \sim 2008/12$.
- Fig 2.8 中的 ACF 在 lag $1, 3, 9$ 均顯著。雖然大於 9 的 lag 也有部分 ACF 是顯著的，但屬於 marginal case，故我們在此不予考慮 lag 大於 9 的部分。建議考慮下列模型：

$$r_t = \mu + a_t - \theta_1 a_{t-1} - \theta_3 a_{t-3} - \theta_9 a_{t-9}$$



```
> exactMLE=arima(ew,order=c(0,0,9),fixed=c(NA,0,NA,0,0,0,0,0,NA,NA),method="ML")
> exactMLE$coef
        ma1         ma2         ma3         ma4         ma5         ma6 
 0.19092655  0.00000000 -0.11986223  0.00000000  0.00000000  0.00000000 
        ma7         ma8         ma9   intercept 
 0.00000000  0.00000000  0.12265345  0.01220892 
> sqrt(exactMLE$sigma2)
[1] 0.07139368
> Box.test(exactMLE$residuals,lag=12,type="Ljung")

        Box-Ljung test

data:  exactMLE$residuals
X-squared = 17.6029, df = 12, p-value = 0.1283
> Box.test(exactMLE$residuals,lag=12,type="Ljung",fitdf=3)

        Box-Ljung test

data:  exactMLE$residuals
X-squared = 17.6029, df = 9, p-value = 0.04007
```

- conditional MLE

$$r_t = 0.012 + a_t + 0.189a_{t-1} - 0.121a_{t-3} + 0.122a_{t-9},\ \hat{\sigma}_a = 0.0714$$

$$Q(12) = 17.5,\ p\text{-value} = 0.041 \text{ (based on an asymptotic } \chi^2(9))$$
$$p\text{-value} = 0.132 \text{ (based on an asymptotic } \chi^2(12))$$

Implications of the model?

- exact MLE

$$r_t = 0.012 + a_t + 0.191a_{t-1} - 0.120a_{t-3} + 0.123a_{t-9},\ \hat{\sigma}_a = 0.0714 \tag{2.24}$$

$$Q(12) = 17.6,\ p\text{-value} = 0.04 \text{ (based on an asymptotic } \chi^2(9))$$
$$p\text{-value} = 0.128 \text{ (based on an asymptotic } \chi^2(12))$$

- For this particular instance, the difference between the conditional- and exact-likelihood methods is negligible.

- Consider the out-of-sample multi-step ahead forecasts for the monthly return of the CRSP equal weighted index.
- The forecast origin the forecast origin t $=$ 986 (February 2008).
- We re-estimate the MA(9) model using exact MLE and the data from $t = 1 \sim 986$.

```
> reEstimateModel=arima(ew[1:986],order=c(0,0,9),fixed=c(NA,0,NA,0,0,0,0,0,NA,NA),method="ML")
> reEstimateModel
Call:
arima(x = ew[1:986], order = c(0, 0, 9), fixed = c(NA, 0, NA, 0, 0, 0, 0, 0,
    NA, NA), method = "ML")

Coefficients:
         ma1  ma2      ma3  ma4  ma5  ma6  ma7  ma8     ma9  intercept
      0.1844    0  -0.1206    0    0    0    0    0  0.1218     0.0128
s.e.  0.0295    0   0.0338    0    0    0    0    0  0.0312     0.0027

sigma^2 estimated as 0.005066:  log likelihood = 1206.44,  aic = -2402.88
> forecast=predict(reEstimateModel,n.ahead=10)
> forecast

$pred
Time Series:
Start = 987
End = 996
Frequency = 1
 [1] 0.004284404 0.013560426 0.015025565 0.014453940 0.012047087 0.001806804
 [7] 0.012211891 0.005515984 0.008514008 0.012792573

$se
Time Series:
Start = 987
End = 996
Frequency = 1
 [1] 0.07117457 0.07237450 0.07237450 0.07288140 0.07288140 0.07288140
 [7] 0.07288140 0.07288140 0.07288140 0.07339524
```

- Table 2.3 gives some out-of-sample forecasts of an MA(9) model in the form of Eq. (2.24) for the monthly simple returns of the equal-weighted index at the forecast origin $h = 986$ (February 2008).
- The model parameters are reestimated using the first 986 observations.

TABLE 2.3 Out-of-Sample Forecast Performance of an MA(9) Model for Monthly Simple Returns of CRSP Equal-Weighted Index[a]

Step	1	2	3	4	5
Forecast	0.0043	0.0136	0.0150	0.0144	0.0120
Std. Error	0.0712	0.0724	0.0729	0.0729	0.0729
Actual	−0.0260	0.0312	0.0322	−0.0871	−0.0010
Step	6	7	8	9	10
Forecast	0.0019	0.0122	0.0056	0.0085	0.0128
Std. Error	0.0729	0.0729	0.0729	0.0729	0.0734
Actual	0.0141	−0.1209	−0.2060	−0.1366	0.0431

[a] The forecast origin is February 2008 With $h = 986$. The model is estimated by the exact maximum-likelihood method.

- The sample mean and standard error of the estimation subsample are 0.0128 and 0.0736, respectively.
- As expected, the table shows that
  - (a) The 10-step-ahead forecast is the sample mean.
  - (b) The standard deviations of the forecast errors converge to the standard deviation of the series as the forecast horizon increases.
- In this particular case, the point forecasts deviate substantially from the observed returns because of the worldwide financial crisis caused by the subprime mortgage problem (次貸風暴) and the collapse of Lehman Brothers.
- The crisis had severe, long-lasting consequences for the U.S. and European economies. The U.S. entered a deep recession, with nearly 9 million jobs lost during 2008 and 2009, roughly 6% of the workforce.

The dual relationship between AR(p) and MA(q)

1. A finite order stationary AR(p) process $\Rightarrow$ infinite order MA process.
2. A finite order invertible MA(q) process $\Rightarrow$ infinite order AR process.
3. AR(p) has ACF tail off and PACF cut off $\Leftrightarrow$ MA(q) has ACF cut off and PACF tail off.

- AR(2)

$$\dot{Z}_t = \frac{1}{1 - \phi_1 B - \phi_2 B^2} a_t = \psi(B) a_t,$$

$$\Rightarrow (1 - \phi_1 B - \phi_2 B^2)(1 + \psi_1 B + \psi_2 B^2 + \psi_3 B^3 + \cdots) = 1,$$

$$\Rightarrow \psi_1 = \phi_1, \qquad \psi_2 = \psi_1 \phi_1 + \phi_2 = \phi_1^2 + \phi_2$$

$$\psi_3 = \psi_2 \phi_1 + \psi_1 \phi_2$$

$$\Rightarrow \psi_j = \psi_{j-1} \phi_1 + \psi_{j-2} \phi_2 \quad \text{for } j \geq 2$$

- MA(2)

$$a_t = \frac{1}{1 - \theta_1 B - \theta_2 B^2} \dot{Z}_t = (1 - \phi_1 B - \phi_2 B^2 - \cdots) \dot{Z}_t$$

$$\Rightarrow \phi_1 = -\theta_1, \quad \phi_2 = \theta_1^2 - \theta_2$$

$$\phi_j = \phi_{j-1} \theta_1 + \phi_{j-2} \theta_2, \quad \text{for } j \geq 3$$

ARMA(p, q)

(can reduce the number of parameter efficiently)

$$\phi_p(B)\dot{Z}_t = \theta_q(B)a_t$$
$$\phi_p(B) = 1 - \phi_1 B - \cdots - \phi_p B^p, \&$$
$$\theta_q(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$$

- invertible: the roots of $\theta_q(B) = 0$ lie outside the unit circle
- stationary: the roots of $\phi_p(B) = 0$ lie outside the unit circle

Assume $\phi_p(B) = 0$ and $\theta_q(B) = 0$ have no common roots

1. AR(∞) representation $\pi(B)\dot{Z}_t = a_t$, where $\pi(B) = \frac{\phi_p(B)}{\theta_q(B)}$
2. MA(∞) representation $\dot{Z}_t = \psi(B)a_t$, where $\psi(B) = \frac{\theta_q(B)}{\phi_q(B)}$

- Find the ACF $\times \dot{Z}_{t-k}$

$$\dot{Z}_t = \phi_1 \dot{Z}_{t-1} + \cdots + \phi_p \dot{Z}_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}$$

$$\gamma_k = \phi_1 \gamma_{k-1} + \cdots + \phi_p \gamma_{k-p} + E(\dot{Z}_{t-k} a_t) - \cdots - \theta_q E(\dot{Z}_{t-k} a_{t-q})$$

- Since $E(\dot{Z}_{t-k} a_{t-q}) = 0,$ for $k \geq q+1$

$$\Rightarrow \gamma_k = \phi_1 \gamma_{k-1} + \cdots + \phi_p \gamma_{k-p} \qquad \text{for } k \geq q+1$$

$$\Rightarrow \rho_k = \phi_1 \rho_{k-1} + \cdots + \phi_p \rho_{k-p} \qquad \text{for } k \geq q+1,$$

difference equations

$\Rightarrow$ ACF of ARMA(p, q) tails off after lag q

- PACF: mixture of exponential decays and/or damped sine waves

ARMA$(1,1)$ Model

- A compact form for flexible models.
  1. simplicity
  2. useful for understanding GARCH models in Ch. 3 for volatility modeling.
- $(1-\phi_1 B)Z_t = \phi_0 + (1-\theta_1 B)a_t$
  $\Longrightarrow Z_t = \phi_1 Z_{t-1} + \phi_0 + a_t - \theta_1 a_{t-1}$. Need to require $\phi_1 \neq \theta_1$, otherwise
  $Z_t = (1-\phi_1 B)^{-1}(\phi_0 + (1-\theta_1 B)a_t) = \frac{\phi_0}{1-\phi_1} + a_t$ 。
- Stationarity: same as AR(1), $|\phi_1| < 1$.
- Invertibility: same as MA(1), $|\theta_1| < 1$.

$ARMA(1,1)$ model

- MA(∞) 的表式：

$$\begin{aligned} Z_t &= (1+\phi_1 B+\phi_1^2 B^2+\cdots)(\phi_0+(1-\theta_1 B)a_t) \\ &= \frac{\phi_0}{1-\phi_1}+(1+(\phi_1-\theta_1)B+\phi_1(\phi_1-\theta_1)B^2+\cdots)a_t \\ &= \frac{\phi_0}{1-\phi_1}+\sum_{j=0}^{\infty}\psi_j B^j a_t, \end{aligned}$$

其中 $\psi_0=1$ 且 $\psi_j=\phi_1^{j-1}(\phi_1-\theta_1)$ for $j\geq 1$ 。

- Mean：same as AR(1), i.e. $E(Z_t)=\dfrac{\phi_0}{1-\phi_1}$.

- **Variance:**Assume $\phi_0 = 0$,

$$
\begin{aligned}
Z_{t-k}Z_t &= \phi_1 Z_{t-k}Z_{t-1} + Z_{t-k}a_t - \theta_1 Z_{t-k}a_{t-1} \\
\gamma_k &= \phi_1\gamma_{k-1} + E(Z_{t-k}a_t) - \theta_1 E(Z_{t-k}a_{t-1}) \\
\gamma_0 &= \phi_1\gamma_1 + E(Z_t a_t) - \theta_1 E(Z_t a_{t-1}) \\
Z_t &= a_t + \psi_1 a_{t-1} + \cdots \Rightarrow E(Z_t a_t) = \sigma_a^2
\end{aligned}
$$

$$
\begin{aligned}
k = 0 \Rightarrow \gamma_0 &= \phi_1\gamma_1 + \sigma_a^2 - \theta_1 E(Z_t a_{t-1}) \\
&= \phi_1\gamma_1 + \sigma_a^2 - \theta_1(\phi_1 - \theta_1)\sigma_a^2 \\
k = 1 \Rightarrow \gamma_1 &= \phi_1\gamma_0 - \theta_1\sigma_a^2 \qquad \text{代入 } k = 0 \\
\gamma_0 &= \phi_1(\phi_1\gamma_0 - \theta_1\sigma_a^2) + \sigma_a^2 - \theta_1(\phi_1 - \theta_1)\sigma_a^2 \\
\gamma_0 &= \frac{1 + \theta_1^2 - 2\phi_1\theta_1}{1 - \phi_1^2}\sigma_a^2
\end{aligned}
$$

- **ACF:**

$$\gamma_1 = \phi_1\gamma_0 - \theta_1\sigma_a^2 = \frac{(\phi_1-\theta_1)(1-\phi_1\theta_1)}{1-\phi_1^2}\sigma_a^2$$

$$\gamma_k = \phi_1\gamma_{k-1} \quad \text{for} \quad k \geq 2$$

$$\rho_k = \begin{cases} 1, & k=0 \\ \frac{(\phi_1-\theta_1)(1-\phi_1\theta_1)}{1+\theta_1^2-2\phi_1\theta_1}, & k=1 \\ \phi_1\rho_{k-1}, & k\geq 2 \end{cases}$$

- **PACF:** does not cut off at finite lags.

e.g: ARMA$(1,1)$ $(1-0.9B)Z_t = (1-0.5B)a_t$, $a_t \sim N(0,1)$
Simulate 250 observations. Sample ACF and PACF

Table 3.7 Sample ACF and sample PACF for a simulated ARMA(1, 1) series from $(1-.9B)Z_t = (1-.5B)a_t$.

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	.57	.50	.47	.35	.31	.25	.21	.18	.10	.12
St.E.	.06	.08	.09	.10	.11	.11	.11	.11	.11	.11
$\hat{\phi}_{kk}$	.57	.26	.18	−.03	.01	−.01	.01	.01	−.08	.05
St.E.	.06	.06	.06	.06	.06	.06	.06	.06	.06	.06

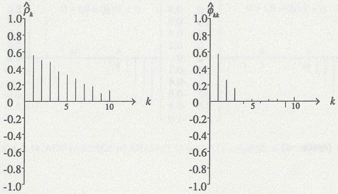


Fig. 3.15 Sample ACF and sample PACF of a simulated ARMA(1,1) series: $(1-.9B)Z_t = (1-.5B)a_t$.

- Both $\hat{\rho}_k$ and $\hat{\phi}_{kk}$ tails off $\Rightarrow$ a mixed ARMA(p,q) model ; We can choose p, q by model selection criterion.
- $\hat{\phi}_{kk}$ cut off after lag 4 $\Rightarrow$ an AR(4) model 。
- Since all the significant PACF are positive, it is not a pure MA model

e.g: ARMA(1,1) $(1-0.6B)Z_t=(1-0.5B)a_t$, $a_t \sim N(0,1)$
模擬 250 筆資料的 ACF 及 PACF

Table 3.8 Sample ACF and sample PACF for a simulated series of the ARMA(1,1) process: $(1-.6B)Z_t=(1-.5B)a_t$.

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	.10	.05	.09	.00	−.02	.02	−.02	.04	−.04	.01
St.E.	.06	.06	.06	.06	.06	.06	.06	.06	.06	.06
$\hat{\phi}_{kk}$	.10	.04	.08	−.02	−.02	.01	−.02	.05	−.05	.02
St.E.	.06	.06	.06	.06	.06	.06	.06	.06	.06	.06

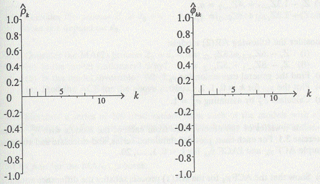


Fig. 3.16 Sample ACF and sample PACF of a simulated ARMA(1,1) series: $(1-.6B)Z_t=(1-.5B)a_t$.

Recall the ACF of an ARMA$(1,1)$ model:

$$\rho_k=\phi_1\rho_{k-1}=\frac{\phi_1^{k-1}(\phi_1-\theta_1)(1-\phi_1\theta_1)}{1+\theta_1^2-2\phi_1\theta_1},\ k\geq 2$$

When $\phi_1\approx\theta_1$, $\rho_k\approx 0$.

- When the sample ACF behave like white noise proces (i.e. $\hat{\rho}_k \approx 0$), then there are two possibilities:
  - it's a random noise process or
  - it's an ARMA process with its AR and MA polynomial being nearly equal.
- Hence we usually assume $\phi_p(B) = 0$ and $\theta_q(B) = 0$ have no common rootes.

- **ARMA(p,q)($\mu \neq 0$) another representation**

$$(1-\phi_1 B-\cdots-\phi_p B^p)(Z_t-\mu)=(1-\theta_1 B-\cdots-\theta_q B^q)a_t$$

$$\Rightarrow (1-\phi_1 B-\cdots-\phi_p B^p)Z_t=\theta_0+(1-\theta_1 B-\cdots-\theta_q B^q)a_t$$

- $$\theta_0=(1-\phi_1 B-\cdots-\phi_p B^p)\mu$$

$$\Rightarrow \mu=\frac{\theta_0}{(1-\phi_1-\cdots-\phi_p)}$$

- For pure MA model, $\mu=\theta_0$.

Building an $ARMA(1,1)$ model

- Specification: use EACF(Extended ACF, Tsay & Tiao, 1984) or AIC
- What is EACF? How to use it?
- The output of EACF is a two-way table, where the rows correspond to AR order p and the columns to MA order q.
- The theoretical version of EACF for an ARMA$(1,1)$ model is given in Table 2.4.

- Table 2.4: Theoretical EACF table for an ARMA$(1,1)$ model, where X denotes nonzero, O denotes zero, and $*$ denotes either zero or nonzero.

	MA							
AR	0	1	2	3	4	5	6	7
0	X	X	X	X	X	X	X	X
1	X	O	O	O	O	O	O	O
2	$*$	X	O	O	O	O	O	O
3	$*$	$*$	X	O	O	O	O	O
4	$*$	$*$	$*$	X	O	O	O	O
5	$*$	$*$	$*$	$*$	X	O	O	O

- The key feature of the table is that it contains a triangle of O with the upper left vertex located at the order $(1,1)$. This is the characteristic we use to identify the order of an ARMA process.
- In general, for an ARMA(p,q) model, the triangle of O will have its upper left vertex at the (p,q) position.

Example

- Consider the monthly log stock returns of the 3M Company from February 1946 to December 2008.
- There are 755 observations. The return series and its sample ACF are shown in Figure 2.9.

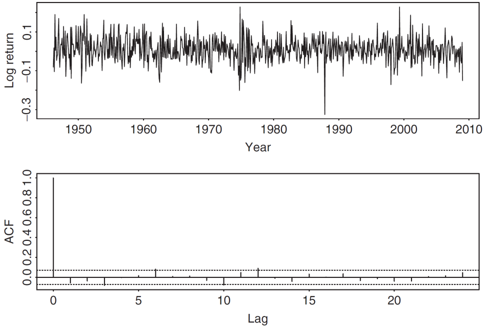


Figure 2.9 Time plot and sample autocorrelation function of monthly log stock returns of 3M Company from February 1946 to December 2008.

- The ACF indicates that there are no significant serial correlations in the data at the 1% level.
- Table 2.5 shows the sample EACF and a corresponding simplified table for the series. The simplified table is constructed by using the following notation:
  1. X denotes that the absolute value of the corresponding EACF is greater than or equal to $2/\sqrt{T}$, which is twice of the asymptotic standard error of the EACF.
  2. O denotes that the corresponding EACF is less than $2/\sqrt{T}$ in modulus.

TABLE 2.5 Sample Extended Autocorrelation Function and a Simplified Table for the Monthly Log Returns of 3M Stock from February 1946 to December 2008

Sample Extended Autocorrelation Function

	MA Order: q												
p	0	1	2	3	4	5	6	7	8	9	10	11	12
0	−0.06	−0.04	−0.08	−0.00	0.02	0.08	0.01	0.01	−0.03	−0.08	0.05	0.09	−0.01
1	−0.47	0.01	−0.07	−0.02	0.00	0.08	−0.03	0.00	−0.01	−0.07	0.04	0.09	−0.02
2	−0.38	−0.35	−0.07	0.02	−0.01	0.08	0.03	0.01	0.00	−0.03	0.02	0.04	0.04
3	−0.18	0.14	0.38	−0.02	0.00	0.04	−0.02	0.02	−0.00	−0.03	0.02	0.01	0.04
4	0.42	0.03	0.45	−0.01	0.00	0.00	−0.01	0.03	0.01	0.00	0.02	−0.00	0.01
5	−0.11	0.21	0.45	0.01	0.20	−0.01	−0.00	0.04	−0.01	−0.01	0.03	0.01	0.03
6	−0.21	−0.25	0.24	0.31	0.17	−0.04	−0.00	0.04	−0.01	−0.03	0.01	0.01	0.04

Simplified EACF Table

	MA Order: q												
p	0	1	2	3	4	5	6	7	8	9	10	11	12
0	O	O	X	O	O	X	O	O	O	X	O	X	O
1	X	O	O	O	O	X	O	O	O	O	O	X	O
2	X	X	O	O	O	X	O	O	O	O	O	O	O
3	X	X	X	O	O	O	O	O	O	O	O	O	O
4	X	O	X	O	O	O	O	O	O	O	O	O	O
5	X	X	X	O	X	O	O	O	O	O	O	O	O
6	X	X	X	X	X	O	O	O	O	O	O	O	O

- R-code:

```
> library(TSA)
> simpleReturn=read.table("m-3m4608.txt",header=T)[,2]
> logReturn=log(simpleReturn+1)
> eacf(logReturn)
```

- The simplified table exhibits a triangular pattern of O with its upper left vertex at the order $(p, q) = (0, 0)$.
- A few exceptions of X appear when $q = 2, 5, 9$, and 11. However, the EACF table shows that the values of sample ACF corresponding to those X are around 0.08 or 0.09.
- These ACFs are only slightly greater than $2/\sqrt{755} = 0.073$. Indeed, if 1% critical value is used, those X would become O in the simplified EACF table.
- The EACF suggests that the monthly log returns of 3M stock follow an ARMA$(0, 0)$ model (i.e., a white noise series). This is in agreement with the result suggested by the sample ACF in Figure 2.9.

Three model representation

- ARMA form: compact, useful in estimation and forecasting.
- AR representation: (by long division)

$$Z_t = \frac{\phi_0}{1 - \theta_1 - \cdots - \theta_q} + a_t + \pi_1 Z_{t-1} + \pi_2 Z_{t-2} + \cdots .$$

  It tells how Z_t depends on its past values.
- MA representation: (by long division)

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

  It tells how Z_t depends on the past shocks.

- For a stationary series, ψ_i converges to zero as $i \to \infty$. thus, the effect of any shock is transitory.
- The MA representation is particularly useful in computing variances of forecast errors.
- For a l-step ahead forecast, the l-step-ahead point forecast is

$$\hat{Z}_n(l) = \mu + \psi_l a_n + \psi_{l+1} a_{n-1} + \cdots$$

and the forecast error is

$$e_n(l) = a_{n+l} + \psi_1 a_{n+l-1} + \cdots + \psi_{l-1} a_{n+1}.$$

The variance of forecast error is

$$Var[e_n(l)] = (1 + \psi_1^2 + \cdots + \psi_{l-1}^2)\sigma_a^2.$$