STAT 6104 - Financial Time Series

Chapter 4 - Estimation

Agenda

Introduction

- ARIMA(p, d, q) Model

$$\phi(B)(1-B)^d(Y_t-\mu)=\theta(B)Z_t \text{ where}, \quad Z_t \sim WN(0,\sigma^2).$$

- Unknown order: (p, d, q)
- Unknown parameters: $(\mu, \phi_1, \phi_2, \ldots, \phi_p, \theta_1, \theta_2, \ldots, \theta_q, \sigma^2)$

Question:
Given the order (p, d, q), how to estimate the unknown parameters ?

Agenda

Moment Estimators

- Basic idea:
  - Sample Moments ($\frac{1}{n}\sum_{i=1}^{n} Y_i^k$) can be computed from the data
  - Theoretical Moments ($E(Y^k)$) depends on unknown parameters

  $\Rightarrow$ Match the sample and theoretical moments to solve for the unknown parameters

- Toy example: i.i.d. Normal: $N(\mu, \sigma^2)$

	Sample	Theoretical
1^{st} moment	$\overline{Y}$	$E(Y) = \mu$
2^{nd} moment	$\frac{\sum_{i=1}^{n} Y_i^2}{n}$	$E(Y^2) = \mu^2 + \sigma^2$

Moment estimators:

- (match 1st moment): $\widehat{\mu} = \overline{Y}$
- (match 2nd moment): $\widehat{\sigma}^2 = \frac{\sum_{i=1}^{n} Y_i^2}{n} - \overline{Y}^2$

Moment Estimator for Time Series

Theoretical Moment	Sample Moment
μ	$\overline{Y} = \frac{1}{n}\sum_{t=1}^{n} Y_t$
$\gamma(0)$	$C_0 = \frac{1}{n}\sum(Y_t - \overline{Y})^2$
$\gamma(k)$	$C_k = \frac{1}{n}\sum_{t=1}^{n-k}(Y_t - \overline{Y})(Y_{t+k} - \overline{Y})$
$\rho(0)$	1
$\rho(k)$	$r_k = \frac{C_k}{C_0}$

Moment Estimator: AR(1)

Example 1

Find the moment estimator for the AR(1) model $(1-\phi B)(Y_t-\mu)=Z_t$, $Z_t \sim WN(0,1)$ in terms of observations $Y_1,\ldots,Y_n$. Give the value if the observations are $(1.2, 2.3, 2.1, 1.5, 0.8, 1.2)$.

Moment Estimator: MA models

MA(q): $Y_t = Z_t - \theta_1 Z_{t-1} - \cdots - \theta_q Z_{t-q} \quad Z_t \sim WN(0, \sigma^2)$.

Moment Estimator (method of moment)

- Express $\rho(k)$ in terms of unknown parameters θ_1, θ_2, ... , θ_q
- Solve $r_k = \rho(k)$, $k = 1, \ldots, q$, for unknown parameters
- Example - MA(1): $Y_t = Z_t - \theta Z_{t-1}$

  $\rho(1) = -\frac{\theta}{1+\theta^2} \qquad r_1 = \frac{\sum_{t=1}^{n-1}(Y_t - \overline{Y})(Y_{t+1} - \overline{Y})}{\sum_{t=1}^{n}(Y_t - \overline{Y})^2}$

  The estimate satisfies

  $r_1 = -\frac{\widehat{\theta}}{1+\widehat{\theta}^2} \Rightarrow r_1\widehat{\theta}^2 + \widehat{\theta} + r_1 = 0 \Rightarrow \widehat{\theta} = \frac{-1 \pm \sqrt{1-4r_1^2}}{2r_1}$

Agenda

Yule Walker Equation (Y.W.)

- Yule-Walker (Y.W.) equations is a system of equations connecting AR parameters ϕ_is and ACVF/ACFs:
  - Find covariance of Y_{t-k} with $Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + Z_t$

$$\Rightarrow \gamma(k) = \phi_1\gamma(k-1) + \cdots + \phi_p\gamma(k-p)$$
$$\Rightarrow \rho(k) = \phi_1\rho(k-1) + \cdots + \phi_p\rho(k-p)$$

where $k = 1, 2, ..., p$

- Given parameters ϕ_is, we have used Yule-Walker equation to find ACVF/ACFs (Chapter 3)
- Given **sample ACFs**, we can use Yule-Walker equation to estimate ϕ_is

Parameter Estimators in AR Model: Yule Walker

$$
\begin{aligned}
\rho(1) &= \phi_1\rho(0) + \cdots + \phi_p\rho(p-1) \\
\rho(2) &= \phi_1\rho(1) + \cdots + \phi_p\rho(p-2) \\
\cdots &= \cdots + \cdots + \cdots \\
\rho(p) &= \phi_1\rho(p-1) + \cdots + \phi_p\rho(0)
\end{aligned}
$$

Matrix Form

$$
\begin{pmatrix} \rho(1) \\ \vdots \\ \rho(p) \end{pmatrix} = \begin{pmatrix} \rho(0) & \rho(1) & \cdots & \rho(p-1) \\ \rho(1) & \rho(0) & \cdots & \rho(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ \rho(p-1) & \rho(p-2) & \cdots & \rho(0) \end{pmatrix} \begin{pmatrix} \phi_1 \\ \vdots \\ \phi_p \end{pmatrix}
$$

Yule-Walker Estimator (estimate $\rho(k)$ by r_k)

$$
\widehat{\phi} = \begin{pmatrix} \widehat{\phi}_1 \\ \vdots \\ \widehat{\phi}_p \end{pmatrix} = \begin{pmatrix} 1 & r_1 & \cdots & r_{p-1} \\ r_1 & 1 & \cdots & r_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p-1} & r_{p-2} & \cdots & 1 \end{pmatrix}^{-1} \begin{pmatrix} r_1 \\ \vdots \\ r_p \end{pmatrix}
$$

Parameter Estimators in AR Model: Yule Walker

Yule-Walker Estimator

$$\widehat{\phi} = \begin{pmatrix} \widehat{\phi}_1 \\ \vdots \\ \widehat{\phi}_p \end{pmatrix} = \begin{pmatrix} 1 & r_1 & \cdots & r_{p-1} \\ r_1 & 1 & \cdots & r_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p-1} & r_{p-2} & \cdots & 1 \end{pmatrix}^{-1} \begin{pmatrix} r_1 \\ \vdots \\ r_p \end{pmatrix}$$

Example: If the time series is $(-1.4, 0.39, 0.97, 1.5, 0.59, -2.4, -2.2, -1.5, -0.42, 0.10)$, find the Yule Walker estimate of an AR(2) model.

Agenda

Parameter Estimation in AR Model: LSE

$$\text{AR}(p)\text{: } Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + Z_t$$

- Express as a regression problem

$$Y_t = \left(Y_{t-1} \cdots Y_{t-p} \right) \begin{pmatrix} \phi_1 \\ \vdots \\ \phi_p \end{pmatrix} + Z_t = \mathbf{Y}'_{t-1}\boldsymbol{\phi} + Z_t \, ,$$

  where $\boldsymbol{\phi} = (\phi_1 \cdots \phi_p)'$ and $\mathbf{Y}_{t-1} = (Y_{t-1} \cdots Y_{t-p})'$.
- Recall in linear regression model with p predictors, we have

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

  Least squares estimate of β is $\widehat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$.
  Least squares estimate of σ^2 is $\hat{\sigma}^2 = (\mathbf{Y} - \mathbf{X}\widehat{\beta})'(\mathbf{Y} - \mathbf{X}\widehat{\beta})/(n-p)$.

Least Squares Estimate (LSE) for AR Model

- AR(p): $Y_t = \mathbf{Y}'_{t-1}\boldsymbol{\phi} + Z_t$
- Regression: $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$
  Results: $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$, $\hat{\sigma}^2 = \frac{1}{n-p}\sum_{i=p+1}^{n}(Y_i - \widehat{Y}_i)^2$

Least squares estimate of $\boldsymbol{\phi}$:

$\mathbf{Y} = (Y_{p+1}, \ldots, Y_n)'$, $\mathbf{X} = (\mathbf{Y}_p \; \mathbf{Y}_{p+1} \cdots \mathbf{Y}_{n-1})'$, $\boldsymbol{\varepsilon} = (Z_{p+1}, \ldots, Z_n)$, $\boldsymbol{\beta} = \boldsymbol{\phi}$

$$\begin{aligned} \widehat{\boldsymbol{\phi}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \\ &= \left(\sum_{t=p+1}^{n} \mathbf{Y}_{t-1}\mathbf{Y}'_{t-1}\right)^{-1} \left(\sum_{t=p+1}^{n} \mathbf{Y}_{t-1} Y_t\right) \end{aligned}$$

Example - AR(1): $Y_t = \phi_1 Y_{t-1} + Z_t$, $(\mathbf{Y}_{t-1} = Y_{t-1})$

$$\widehat{\phi}_1 = \frac{\sum_{t=2}^{n} Y_{t-1} Y_t}{\sum_{t=2}^{n} Y_{t-1}^2} \qquad \text{and} \qquad \widehat{\sigma}^2 = \frac{1}{n-1}\sum_{t=2}^{n}(Y_t - \widehat{\phi}_1 Y_{t-1})^2$$

Statistical Inference in AR Model: LSE

Asymptotic distribution of the estimator :

$$\sqrt{n}(\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}) \to N(0, \sigma^2 \Gamma_p^{-1})$$

where

$$\Gamma_p = \begin{pmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(p-1) \\ \gamma(1) & \gamma(0) & \cdots & \gamma(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(p-1) & \gamma(p-2) & \cdots & \gamma(0) \end{pmatrix}$$

Confidence Intervals or hypothesis tests can be applied to make inference on the parameter $\boldsymbol{\phi}$

- $\mathrm{Var}(\hat{\phi}_k)$ is $\frac{\sigma^2}{n}$ times k-th diagonal entry of Γ_p^{-1}.
- $\widehat{\mathrm{Var}}(\hat{\phi}_k)$ is $\frac{\hat{\sigma}^2}{n}$ times k-th diagonal entry of $\widehat{\Gamma}_p^{-1}$ (Replace $\gamma(k)$ by $\hat{\gamma}_k = C_k$).
- 95% C.I.: $\hat{\phi}_k \pm 2\sqrt{\widehat{\mathrm{Var}}(\hat{\phi}_k)}$
- Test for $\phi_k = 0$ at 5% sig. level.: Compare $\hat{\phi}_k / \sqrt{\widehat{\mathrm{Var}}(\hat{\phi}_k)}$ with 2.

Statistical Inference in AR Model: LSE

Asymptotic distribution of the estimator:

$$\sqrt{n}(\widehat{\phi} - \phi) \to N(0, \sigma^2 \Gamma_p^{-1}) \qquad \Gamma_p = \begin{pmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(p-1) \\ \gamma(1) & \gamma(0) & \cdots & \gamma(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(p-1) & \gamma(p-2) & \cdots & \gamma(0) \end{pmatrix}$$

Confidence Intervals or hypothesis tests to make inference on ϕ:

Example: Estimate AR(2) model: $n = 200$,

$Y_t = 0.3\,Y_{t-1} + 0.04\,Y_{t-2} + Z_t$; $\hat{\gamma}(0) = 1.11$, $\hat{\gamma}(1) = 0.347$, $\hat{\sigma}^2 = 1$.

$$\hat{\sigma}^2 \widehat{\Gamma}_p^{-1} = \begin{pmatrix} 1.11 & 0.347 \\ 0.347 & 1.11 \end{pmatrix}^{-1} = \begin{pmatrix} 0.998 & -0.312 \\ -0.312 & 0.998 \end{pmatrix}$$

- C.I. for ϕ_2: $[0.04 \pm 2\sqrt{0.998/n}] = [-0.101, 0.181]$. (Not significantly different from 0)
- Testing for $\phi_1 = 0$: $z = \dfrac{\hat{\phi}_1}{\sqrt{\widehat{\text{Var}}(\hat{\phi}_1)}} = 0.3/\sqrt{0.998/200} = 4.24 > 2$. (significantly different from 0)

Agenda

MA(q): $Y_t = Z_t - \theta_1 Z_{t-1} - \cdots - \theta_q Z_{t-q} \quad Z_t \sim WN(0, \sigma^2)$.

Conditional Least Squares (CLS) Method

(also called Conditional Sum of Squares (CSS))

Idea -

- Find the noise sequence $\{Z_t\}$ from the observation $\{Y_t\}$ (so we require the MA model to be *invertible*!)
- Minimize the sum of squares

$$S_*(\theta) = \sum_{t=1}^{n} Z_t^2$$

Example - MA(1): $Y_t = Z_t + \theta Z_{t-1}$

- Conditional on $Z_0 = 0$, then
  $Z_1 = Y_1$, $Z_2 = Y_2 - \theta Z_1$,..., $Z_k = Y_k - \theta Z_{k-1}$
- Minimize $S_*(\theta) = \sum_{t=1}^{n} Z_t^2$ by numerical optimization

Conditional Least Squares: Moving Average Models

Example - MA(1): $Y_t = Z_t + \theta Z_{t-1}$

- Conditional on $Z_0 = 0$, then
  $Z_1 = Y_1,\ Z_2 = Y_2 - \theta Z_1, \ldots,\ Z_k = Y_k - \theta Z_{k-1}$
- Minimize $S_*(\theta) = \sum_{t=1}^{n} Z_t^2$ by numerical optimization

Implementation:

```
set.seed(6104)
y=arima.sim(2000,model=list(ar=c(0.3)))
get.z=function(theta){
   z=rep(0,length(y))
   z[1]=y[1]
   for (k in 2:length(y)){ z[k]=y[k]-theta*z[k-1]}
   return(sum(z^2))
}
optim(0.1,get.z)
```

Conditional Least Squares: ARMA Model

- Method of moment become very tedious on ARMA models due to the complicated expression of $\gamma(k)$
- Conditional Least Squares can be applied similarly as in MA models

Example (CLS) - ARMA(1,1): $Y_t - \phi Y_{t-1} = Z_t - \theta Z_{t-1}$

- Conditional on $Z_0 = 0,\ Y_0 = 0$
  $\Rightarrow Z_1(\phi, \theta) = Y_1,\ Z_2(\phi, \theta) = Y_2 - \phi Y_1 + \theta Z_1, \ldots,$
  $Z_k(\phi, \theta) = Y_k - \phi Y_{k-1} + \theta Z_{k-1}$
- Minimizes $S_*(\theta) = \sum_{t=1}^{n} Z_t^2(\phi, \theta)$ by numerical optimization

Agenda

Maximum Likelihood Estimator (MLE)

- MLE $=$ Most probable parameter value of the statistical model inferred from the observed data

- Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with probability density function $f(x, \theta)$
  - $f(X_i, \theta)$ is the probability of observing X_i
  - The likelihood function $L(\mathbf{X}, \theta) = \prod_{i=1}^{n} f(X_i, \theta)$ is the probability of observing the data set, where $\mathbf{X} = (X_1, \ldots, X_n)$.
- The MLE $\widehat{\theta}$ maximizes $L(\mathbf{X}, \theta)$
  - The most probable parameter value such that the current data set is obtained

MLE for Time Series

- **Main Idea**: Write down the **joint p.d.f** of $Y_1, Y_2, ..., Y_n$
- Example: AR(1): $Y_t = \phi Y_{t-1} + Z_t$
  - Given $Y_1, (Y_2, ..., Y_n)$ and $(Z_2, ..., Z_n)$ are of 1-1 correspondence

$$\begin{aligned} f(Y_2, \ldots, Y_n | Y_1) &= f(Z_2, \ldots, Z_n | Y_1) = f(Z_2, \ldots, Z_n) \\ &= \left(\tfrac{1}{2\pi\sigma^2}\right)^{\frac{n-1}{2}} \exp\left(-\tfrac{1}{2\sigma^2} \textstyle\sum_{t=2}^n Z_t^2\right) \\ &= \left(\tfrac{1}{2\pi\sigma^2}\right)^{\frac{n-1}{2}} \exp\left(-\tfrac{1}{2\sigma^2} \textstyle\sum_{t=2}^n (Y_t - \phi Y_{t-1})^2\right) \end{aligned}$$

  - Likelihood:

$$L(Y_1, Y_2, \ldots, Y_n) = f(Y_2, \ldots, Y_n | Y_1) f(Y_1)$$

$$= f(Y_2, \ldots, Y_n | Y_1) \frac{\sqrt{1-\phi^2}}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(1-\phi^2) Y_1^2\right)$$

$$= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \sqrt{1-\phi^2} \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{t=2}^n (Y_t - \phi Y_{t-1})^2 + (1-\phi^2) Y_1^2\right]\right\}$$

$\Rightarrow$ $(\hat{\phi}, \hat{\sigma})$ are obtained by maximizing $L(Y_1, \ldots, Y_n)$, or equivalently $\log L(Y_1, \ldots, Y_n)$, with respect to (ϕ, σ) by numerical methods

Maximum Likelihood Estimator: Two standard approaches

- Approach 1: Iterative Conditioning
  - $f(Y_1,\ldots,Y_n) = \frac{f(Y_1,\ldots,Y_n)}{f(Y_1,\ldots,Y_{n-1})}\frac{f(Y_1,\ldots,Y_{n-1})}{f(Y_1,\ldots,Y_{n-2})}\cdots\frac{f(Y_1,Y_2)}{f(Y_1)}f(Y_1)$
    $= \{\prod_{t=2}^{n} f(Y_t|Y_{t-1},\ldots,Y_1)\} f(Y_1)$
  - $f(Y_t|Y_{t-1},\ldots,Y_1)$ may be easy to write down:
    e.g. AR(p) model: $Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + Z_t$, $Z_t \sim N(0,\sigma^2)$
    $Y_t|Y_{t-1},\ldots,Y_1 \sim N(\phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p}, \sigma^2)$
- Approach 2: Multivariate Normal Distribution
  - In **causal** ARMA models, if $\{Z_t\}$ are normally distributed, then $\{Y_t\}$ are also normal (as $Y_t = \sum_{j=1}^{\infty}\psi_j Z_{t-j}$ is a sum of normal noises)
  - $\{Y_1,\ldots,Y_n\}$ follows multivariate normal distribution, which is characterized by the ACVFs:

$$f_{Y_1,\ldots,Y_n}(y_1,\ldots,y_n) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}\mathbf{y}'\Sigma^{-1}\mathbf{y}},$$

where $\mathbf{y} = (y_1,\ldots,y_n)$ and

$$\Sigma = (\gamma(|i-j|)) = \begin{pmatrix} \gamma(0) & \gamma(1) & \ldots & \gamma(n-1) \\ \gamma(1) & \gamma(0) & \ldots & \gamma(n-2) \\ & & \ddots & \\ \gamma(n-1) & \gamma(n-2) & \ldots & \gamma(0) \end{pmatrix}$$

Maximum Likelihood Estimator using R

- ARMA(p, q) model:
  $Y_t = \mu + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}$,
  $Z_t \sim WN(0, \sigma^2)$
- Except for some simple models such as AR(1), we have to employ numerical optimization for the estimation.
-
```
arima(x, order = c(0L, 0L, 0L),
seasonal = list(order = c(0L, 0L, 0L), period = NA),
xreg = NULL, include.mean = TRUE,
fixed = NULL, init = NULL,
method = c("CSS-ML", "ML", "CSS"))
```
- Examples:
  - `set.seed(6104)`
  - `x=arima.sim(n=1000, model=list(ar=c(0.7,-0.12),ma=c(0.7)))`
  - `arima(x,order=c(2,0,1))`
  - `arima(x,order=c(2,0,1),include.mean=F)`

Agenda

Partial ACF (PACF)

- PACF measures the correlation between Y_{k+1} and Y_1 that is not explained by $Y_k, \ldots, Y_2$

- Formally, ϕ_{kk} is the coefficient in the representation

$$Y_{k+1} = \phi_{k1} Y_k + \phi_{k2} Y_{k-1} + \cdots + \phi_{kk} Y_1 + Z_{k+1} \,. \tag{1}$$

  (Note that this is not the "true model")
- Representation (1) can be found by searching for the minimum of

$$E[(Y_{k+1} - \beta_1 Y_k - \cdots - \beta_k Y_1)^2]$$

- Differentiating $\beta_1, \beta_2, \ldots, \beta_k$ in terms and set the equations to zero, $\phi_{k1}, \ldots, \phi_{kk}$ can be obtained by solving

$$\begin{pmatrix} \rho(0) & \cdots & \rho(k-1) \\ \vdots & \ddots & \vdots \\ \rho(k-1) & \cdots & \rho(0) \end{pmatrix} \begin{pmatrix} \phi_{k1} \\ \vdots \\ \phi_{kk} \end{pmatrix} = \begin{pmatrix} \rho(1) \\ \vdots \\ \rho(k) \end{pmatrix}$$

PACF for AR Models

If the data follows AR(p) model:

$$Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + Z_t$$

- $\phi_{11}, \phi_{22}, \ldots, \phi_{pp}$ are non-zero, i.e. Y_t is linearly related to Y_{t-k} for $k \leq p$
- $Y_k - \phi_1 Y_{k-1} - \cdots - \phi_p Y_{k-p} = Z_k$ is uncorrelated with all $\{Y_j\}_{j<k-p}$, thus, **additional** $Y_{k-p-1}, Y_{k-p-2}, \cdots, Y_1$ **cannot explain more information about** Y_k, so

$$\phi_{k,k} = 0 \qquad \text{for all } k > p$$

Computational formula for PACF

$$\begin{pmatrix} \phi_{k1} \\ \vdots \\ \phi_{kk} \end{pmatrix} = \begin{pmatrix} \rho(0) & \cdots & \rho(k-1) \\ \vdots & \ddots & \vdots \\ \rho(k-1) & \cdots & \rho(0) \end{pmatrix}^{-1} \begin{pmatrix} \rho(1) \\ \vdots \\ \rho(k) \end{pmatrix}$$

Example: AR(1): $Y_t = \phi Y_{t-1} + Z_t$. Recall that $\rho(k) = \phi^{|k|}$.
1st lag PACF is:

$$\phi_{11} = \rho(0)^{-1}\rho(1) = \rho(1) = \phi$$

For 2nd lag PACF,

$$\begin{aligned} \begin{pmatrix} \phi_{21} \\ \phi_{22} \end{pmatrix} &= \begin{pmatrix} \rho(0) & \rho(1) \\ \rho(1) & \rho(0) \end{pmatrix}^{-1} \begin{pmatrix} \rho(1) \\ \rho(2) \end{pmatrix} \\ &= \frac{1}{1-\phi^2} \begin{pmatrix} 1 & -\phi \\ -\phi & 1 \end{pmatrix} \begin{pmatrix} \phi \\ \phi^2 \end{pmatrix} = \begin{pmatrix} \phi \\ 0 \end{pmatrix} \end{aligned}$$

The 2nd lag PACF is $\phi_{22} = 0$.

Computational formula for PACF

$$\begin{pmatrix}\phi_{k1}\\ \vdots \\ \phi_{kk}\end{pmatrix} = \begin{pmatrix}\rho(0) & \cdots & \rho(k-1)\\ \vdots & \ddots & \vdots \\ \rho(k-1) & \cdots & \rho(0)\end{pmatrix}^{-1}\begin{pmatrix}\rho(1)\\ \vdots \\ \rho(k)\end{pmatrix}$$

Example: MA(1): $Y_t = Z_t + \theta Z_{t-1}$. Recall $\rho(1) = \frac{\theta}{1+\theta^2}$, $\rho(k) = 0$ $(k \geq 2)$.
1st lag PACF is:

$$\phi_{11} = \rho(0)^{-1}\rho(1) = \rho(1) = \frac{\theta}{1+\theta^2}$$

For 2nd lag PACF,

$$\begin{aligned}\begin{pmatrix}\phi_{21}\\ \phi_{22}\end{pmatrix} &= \begin{pmatrix}1 & \rho(1)\\ \rho(1) & 1\end{pmatrix}^{-1}\begin{pmatrix}\rho(1)\\ 0\end{pmatrix}\\ &= \frac{1}{1-\rho(1)^2}\begin{pmatrix}1 & -\rho(1)\\ -\rho(1) & 1\end{pmatrix}\begin{pmatrix}\rho(1)\\ 0\end{pmatrix} = \frac{1}{1-\rho(1)^2}\begin{pmatrix}\rho(1)\\ -\rho^2(1)\end{pmatrix}\end{aligned}$$

The 2nd lag PACF is $\phi_{22} = \frac{-\rho^2(1)}{1-\rho^2(1)} = -\frac{\theta^2}{1+\theta^2+\theta^4}$.

$$\begin{pmatrix} \phi_{k1} \\ \vdots \\ \phi_{kk} \end{pmatrix} = \begin{pmatrix} \rho(0) & \cdots & \rho(k-1) \\ \vdots & \ddots & \vdots \\ \rho(k-1) & \cdots & \rho(0) \end{pmatrix}^{-1} \begin{pmatrix} \rho(1) \\ \vdots \\ \rho(k) \end{pmatrix}$$

Example: MA(1): $Y_t = Z_t + \theta Z_{t-1}$. Recall $\rho(1) = \frac{\theta}{1+\theta^2}$, $\rho(k) = 0$ $(k \geq 2)$
For kth lag PACF, we solve

$$\begin{pmatrix} 1 & \rho(1) & & & & \\ \rho(1) & 1 & \rho(1) & & & \\ & \rho(1) & 1 & \rho(1) & & \\ & \cdots & & & & \\ & & \rho(1) & 1 & \rho(1) \\ & & & \rho(1) & 1 \end{pmatrix} \begin{pmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{pmatrix} = \begin{pmatrix} \rho(1) \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Advanced techniques in difference equations and tedious algebras show that $\phi_{kk} = -\frac{(-\theta)^k(1-\theta^2)}{1-\theta^{2(k+1)}}$ (roughly exponential decaying).

Sample PACF

$$\begin{pmatrix} \hat{\phi}_{k1} \\ \vdots \\ \hat{\phi}_{kk} \end{pmatrix} = \begin{pmatrix} r_0 & \cdots & r_{k-1} \\ \vdots & \ddots & \vdots \\ r_{k-1} & \cdots & r_0 \end{pmatrix}^{-1} \begin{pmatrix} r_1 \\ \vdots \\ r_k \end{pmatrix}$$

Example:

Data= $(-0.63, -1.8, -0.98, -0.67, -1.14, -1.67, -2.35, -1.70)$.

- `x=c(-0.63, -1.8, -0.98, -0.67, -1.14, -1.67, -2.35, -1.70)`
- `r=acf(x)$ acf[1:1];solve(toeplitz(r),acf(x)$ acf[2:2])`
  `# lag 1 pacf`
- `r=acf(x)$ acf[1:2];solve(toeplitz(r),acf(x)$ acf[2:3])`
  `# lag 2 pacf`
- `r=acf(x)$ acf[1:3];solve(toeplitz(r),acf(x)$ acf[2:4])`
  `# lag 3 pacf`
- `pacf(x)$ acf              # all pacf`

Theorem of PACF for AR Models

For an AR(p) Model,

$$\sqrt{n}\hat{\phi}_{kk} \sim N(0,1) \text{ for } k > p$$

To test whether ϕ_{kk} is significant,

- Significant if $|\hat{\phi}_{kk}| > \frac{2}{\sqrt{n}}$
  $\Rightarrow$ should model the time series by AR(p), $p \geq k$
- NOT Significant otherwise
  $\Rightarrow$ should model the time series by AR(p), $p < k$

Example:

- Data$= (-0.63, -1.8, -0.98, -0.67, -1.14, -1.67, -2.35, -1.70)$.
- Are the PACFs significant?

Using ACF and PACF to determine the order of AR and MA model

AR(p) Model -

- ACF plot: No clear pattern except AR(1) shows exponential decay pattern
- PACF plot: Cut-off at lag p

MA(q) Model -

- ACF plot: Cut-off at lag q
- PACF plot: No clear pattern except MA(1) shows exponential decay pattern

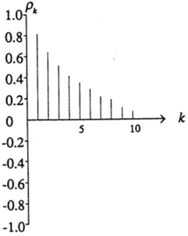
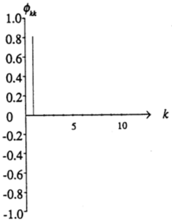
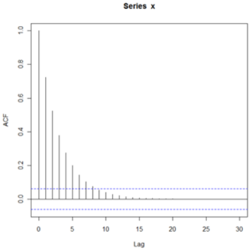
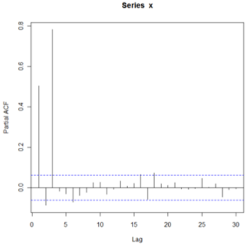
Remark: No clear pattern in ACF/PACF plots for ARMA(p, q) models.

Examples on PACF

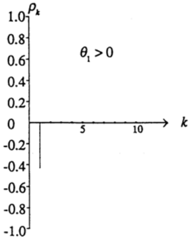
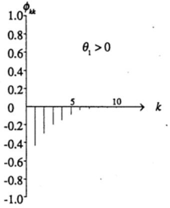
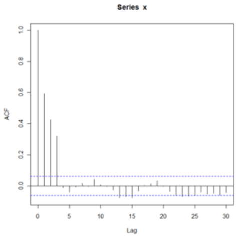
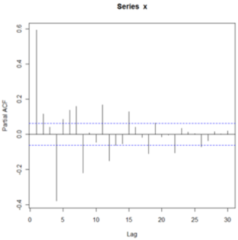
Try:

- `x=arima.sim(1000,model=list(ar=c(0.2,0.4,0.3)))`
- `pacf(x)`
- Which model will you decide?

Examples on ACF and PACF

	ACF	PACF
AR(1)		
AR(3)		

Examples on ACF and PACF

	ACF	PACF
MA(1)		
MA(4)		

Agenda

Order Selection

- ACF and PACF are graphical methods to determine the order of AR and MA model
- It is more desirable to have a systematic order selection criterion for a general ARMA model
- Some common model selection criterion:
  - FPE (Final Prediction Error)
  - AIC (Akaike's Information Criterion)
  - BIC (Bayesian Information Criterion)

Final Prediction Error (FPE)

- $FPE = \widehat{\sigma}^2 \left(\frac{n+p}{n-p}\right)$ where $\widehat{\sigma}^2$ is the MLE of σ^2
- **Idea of FPE:**
  - For an AR(p) Model, the Mean Square Error of parameter estimate is

$$MSE = E(\widehat{\phi} - \phi)'(\widehat{\phi} - \phi) \approx \sigma^2 \left(\frac{n+p}{n}\right) \tag{2}$$

  - The best unbiased estimator for σ^2 is $\widehat{\sigma}^2 \frac{n}{n-p}$
  - Substituting σ^2 by $\widehat{\sigma}^2 \frac{n}{n-p}$ in (2) gives the best estimator of MSE (define as **FPE**).
  - We find the model (AR(1), AR(3), etc) that minimizes FPE.
  - Note the trade-off between goodness of fit and the model complexity

Akaike's Information Criterion (AIC)

AIC -

$$-2\log L\left(\widehat{\beta}, \frac{S_x(\widehat{\beta})}{n}\right) + 2(p+q+1)$$

AICC (AIC corrected) -

$$-2\log L\left(\widehat{\beta}, \frac{S_x(\widehat{\beta})}{n}\right) + \frac{2(p+q+1)n}{n-p-q-2}$$

where

$$\begin{aligned} S_x(\widehat{\beta}) &= \sum_{t=1}^{n}(X_t - \widehat{\phi}_1 X_{t-1} - \cdots - \widehat{\phi}_p X_{t-p} - \widehat{\theta}_1 Z_{t-1} - \cdots - \widehat{\theta}_q Z_{t-q})^2 \\ \widehat{\beta} &= (\widehat{\phi}_1, \ldots, \widehat{\phi}_p, \widehat{\theta}_1, \ldots, \widehat{\theta}_q) \qquad \text{is the MLE} \\ L(\widehat{\beta}, \widehat{\sigma}^2) &= \left(\frac{1}{2\pi\widehat{\sigma}^2}\right)^{\frac{n}{2}} \exp\left[-\frac{1}{2\pi\widehat{\sigma}^2} S_x(\widehat{\beta})\right] \end{aligned}$$

Order Selection - Akaike's Information Criterion

Idea -

- AIC and AICC estimates the expected likelihood function
- $E[L_Y(\widehat{\beta}, \widehat{\sigma}^2)]$ where
  - $\widehat{\beta}$, $\widehat{\sigma}^2$ are MLE from the observation $\mathbf{X} = \{X_1, \ldots, X_n\}$
  - Y is a 'new' observation independent of $\mathbf{X}$
- $\mathbf{X}$ and Y are different to avoid the problem of overfitting
  - $\widehat{\beta}$, $\widehat{\sigma}^2$ were chosen to maximize $L_{\mathbf{X}}(\beta, \sigma^2)$
- We find the model (AR(1), ARMA(1,1), etc) that minimizes AIC.
- Note the trade-off between model complexity and goodness of fit

Order Selection - Bayesian Information Criterion

Bayesian Information Criterion (BIC)

$$BIC = (n-p-q)\log\left[\frac{n\widehat{\sigma}^2}{n-p-1}\right] + n(1+\log\sqrt{2\pi})$$
$$+(p+q)\log\left[\frac{\sum_{i=1}^{n} X_i^2 - n\widehat{\sigma}^2}{p+q}\right]$$

- Motivated by Bayesian argument:
  - BIC $\approx$ the posterior probability of a particular model given data
- We find the model (AR(1), ARMA(1,1), etc) that minimizes BIC.
- Consistent order selection procedure
  - As $n \to \infty$, the order selected by BIC will be equal to the true order with probability going to 1
  - AIC/FPE are not consistent (i.e. get a wrong model), but achieve minimum prediction error

Example

- $FPE = \widehat{\sigma}^2 \left(\frac{n+p}{n-p}\right)$
- $AIC = -2 \log L\left(\widehat{\beta}, \frac{S_x(\widehat{\beta})}{n}\right) + 2(p+q+1)$
- $BIC = (n-p-q)\log\left[\frac{n\widehat{\sigma}^2}{n-p-1}\right] + n(1+\log\sqrt{2\pi}) + (p+q)\log\left[\frac{\sum_{i=1}^{n} X_i^2 - n\widehat{\sigma}^2}{p+q}\right]$

```
IC=function(x,order.input=c(1,0,1)){
fit=arima(x,order=order.input);
n=length(x);p=order.input[1];q=order.input[3];sig=fit$ sigma2
FPE=sig*(n+p)/(n-p);AIC=fit$ aic
BIC=(n-p-q)*log(n*sig/(n-p-1))+n*(1+log(sqrt(2*pi)))+
(p+q)*log((sum(x∧2)-n*sig)/(p+q)); return(c(FPE,AIC,BIC)) }
```

Example:

- Data$=(-0.63, -1.8, -0.98, -0.67, -1.14, -1.67, -2.35, -1.70)$.
- Is ARMA(1,1) or AR(2) better?
  - `x=c(-0.63,-1.8,-0.98,-0.67,-1.14,-1.67,-2.35,-1.70)`
  - `IC(x,order=c(1,0,1))`
  - `IC(x,order=c(2,0,0))`

Agenda

Residual Analysis

Fitted values and residuals:

- Once we obtained the estimates, say $\hat{\phi}$ and $\hat{\theta}$ in ARMA(1,1) model $Y_t = \phi Y_{t-1} + Z_t + \theta Z_{t-1}$, we can compute the *fitted values*, $\{\widehat{Y}_t\}$, and *residuals*, $\{\widehat{Z}_t\}$, by the following recursions:
  1. $\hat{Y}_1 = 0$; $\quad\widehat{Z}_t = Y_1 - \widehat{Y}_1$
  2. $\hat{Y}_2 = \hat{\phi} Y_1 + \hat{\theta}\widehat{Z}_1$; $\quad\widehat{Z}_2 = Y_2 - \widehat{Y}_2$
  3. $\hat{Y}_3 = \hat{\phi} Y_2 + \hat{\theta}\widehat{Z}_2$; $\quad\widehat{Z}_3 = Y_3 - \widehat{Y}_3$
  4. $\ldots\ldots$
  5. $\hat{Y}_n = \hat{\phi} Y_{n-1} + \hat{\theta}\widehat{Z}_{n-1}$; $\quad\widehat{Z}_n = Y_n - \widehat{Y}_n$
- If the model fits the data well, then the fitted values $\{\widehat{Y}_t\}$ should be closed to the observed time series $\{Y_t\}$, i.e., the *residuals*

$$\widehat{Z}_t = Y_t - \widehat{Y}_t$$

is small. More importantly, $\{\widehat{Z}_t\}$ should be similar to the *white noise* sequence $\{Z_t\}$.

Residual Analysis

- Check the goodness of fit of the model by studying the residual

$$\widehat{Z}_t = Y_t - \widehat{Y}_t$$

- **STEPS** -

1. Time series plot of $\widehat{Z}_t$ (should look like white noises)
2. ACF plot of the ACF of $\widehat{Z}_t$, $\widehat{r}_Z(j)$.
   Good fit if most $\widehat{r}_Z(j)$ are within the C.I. $\left(-\frac{2}{\sqrt{n}}, \frac{2}{\sqrt{n}}\right)$, $j \neq 0$.
3. Portmanteau Statistics

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

  - A common choice of h is between 10 to 30
  - $Q(h) \to \chi^2(h-p-q)$. If $Q(h)$ is bigger than the 95% percentile of a $\chi^2(h-p-q)$ distribution, we reject the null hypothesis that $\widehat{Z}_t \sim$ WN.
  - i.e., good fit if the null hypothesis is not rejected.

Residual Analysis

- **STEPS** -

1. Time series plot of $\widehat{Z}_t$ (should look like white noises)
2. ACF plot of $\widehat{Z}_t$, $\widehat{r}_Z(j)$.
   Good fit if all $\widehat{r}_Z(j)$ are within the C.I. $\left(-\frac{2}{\sqrt{n}}, \frac{2}{\sqrt{n}}\right)$, $j \neq 0$.
3. Portmanteau Statistics $Q(h) = n(n+2)\sum_{j=1}^{h} \frac{\widehat{r}_Z^2(j)}{n-j} \sim \chi^2(h-p-q)$.

Example ARMA(1,1): $Y_t = \phi Y_{t-1} + Z_t + \theta Z_{t-1}$.

```
set.seed(6104)
x=arima.sim(1000,model=list(ar=c(0.2), ma=c(0.6)))
n=length(x)
fit=arima(x,order=c(1,0,1))
par(mfrow=c(2,1))
ts.plot(fit$ res)
r.z=as.numeric(acf(fit$ res,12)$ acf) ## h=12
portmanteau.stat=n*(n+2)*sum((r.z[-1]∧2)/(n-(1:12)))
portmanteau.stat>qchisq(0.95,12-1-1) ## FALSE: not reject H0
```

or `tsdiag(fit)`

Agenda

Three Stages -

1. Model Specification
   - Trend, seasonal effect, choosing ARIMA model by FPE/AIC/BIC
2. Model Identification (estimating coefficient)
   - MM, YW, LSE, CLS, MLE,
3. Model Checking(diagnostic)
   - Residual analysis (ts.plot/acf of $\{\widehat{Z}_t\}$)
   - Portmanteau test