Time Series Analysis

Kichun Lee

Hanyang University, Fall, 2013

Partial Correlation

- The number of churches and the number of crimes are highly and positively correlated. The reason? It is because of population.
- For random variables X, Y, correlation $\rho_{X,Y} = Corr(X, Y)$ captures a degree of linear dependence between the two.
- Partial correlation of X and Y excluding Z is denoted by

$$\rho_{X,Y|Z} = corr(X, Y|Z)$$

and computed as follows. We will regress X on Z and Y on Z, trying to remove the influence of Z and to take the residual of each. Then compute the correlation of the residuals.

$$X = Z\alpha + e_X$$
$$Y = Z\beta + e_Y,$$

Then we can show that

$$\rho_{X,Y|Z} = \frac{\rho_{X,Y} - \rho_{X,Z}\rho_{Y,Z}}{\sqrt{1-\rho^2_{X,Z}}\sqrt{1-\rho^2_{Y,Z}}}.$$

Partial Correlation cont.

- Recall the regression: regress Y on X after centering X and Y [Example 3.14 in the book]

$$\min_{\alpha} E[(Y - X\alpha)^2]$$
$$= \min_{\alpha} Var[Y] - 2\alpha Cov(X, Y) + \alpha^2 Var(X).$$

The first order condition leads to

$$\hat{\alpha} = \frac{Cov(X, Y)}{Var(X)} \backsimeq corr(X, Y).$$

The approximation was made under the condition $Var(X) = Var(Y)$.

- Partial Correlation for AR(1) $x_t = \phi x_{t-1} + a_t$?
$\phi_{11} = Corr(x_t, x_{t-1}) = \rho(1) = \phi$.

Partial Correlation cont.

$$\begin{aligned}\phi_{22} &= Corr(x_t, x_{t-2}|x_{t-1}) \\ &= Corr(x_t - \alpha x_{t-1}, x_{t-2} - \beta x_{t-1}) \\ &= Corr(a_t, a_{t-1}) = 0\end{aligned}$$

$$\begin{aligned}\phi_{33} &= Corr(x_t, x_{t-3}|x_{t-1}, x_{t-2}) \\ &= Corr(x_t - \alpha_1 x_{t-1} - \alpha_2 x_{t-2}, x_{t-2} - \beta_1 x_{t-1} - \beta_2 x_{t-2}) \\ &= Corr(a_t, x_{t-2} - \beta_1 x_{t-1} - \beta_2 x_{t-2}) = 0\end{aligned}$$

- Partial Correlation for AR(2) $x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + a_t$?
  The same procedure reveals that
  $\phi_{11} = \rho(1) = \frac{\phi_1}{1-\phi_2}, \phi_{22} = \ldots$
  $\phi_{33} = Corr(x_t - \phi_1 x_{t-1} - \phi_2 x_{t-2}, x_{t-3} - \beta_1 x_{t-2} - \beta_2 x_{t-1}) =$
  $Corr(a_t, \ldots) = 0$
  $\phi_{44} = 0, \cdots.$

Partial Correlation cont.

- In general, PACF (partial autocorrelation function) is computed via

$$\begin{aligned}
\rho_1 &= \phi_{h1} + \phi_{h2}\rho_1 + \cdots + \phi_{hh}\rho_{h-1} \\
\rho_2 &= \phi_{h1}\rho_1 + \phi_{h2} + \cdots + \phi_{hh}\rho_{h-2} \\
&\cdots \\
\rho_h &= \phi_{h1}\rho_{h-1} + \phi_{h2}\rho_{h-2} + \cdots + \phi_{hh}\cdot\rho_{h-2}
\end{aligned}$$

  Solve for ϕ_{hh} from the system of equations.

- [Example] For AR(2),

$$\begin{bmatrix} \rho(1) \\ \rho(2) \end{bmatrix} = \begin{bmatrix} 1 & \rho(1) \\ \rho(1) & 1 \end{bmatrix} \begin{bmatrix} \phi_{21} \\ \phi_{22} \end{bmatrix}$$

Behavior of ACF and PACF

	AR(p)	MA(q)	ARMA(p, q)
ACF	Tails off	Cuts off after lag q	Tails off
PACF	Cuts off after lag p	Tails off	Tails off

Figure : Behavior of the ACF and PACF for ARMA Models

- By the behavior of ACF and PACF, we can empirically determine an appropriate ARMA model. Refer to [Example 3.17 in the book]

Behavior of ACF and PACF (cont.)

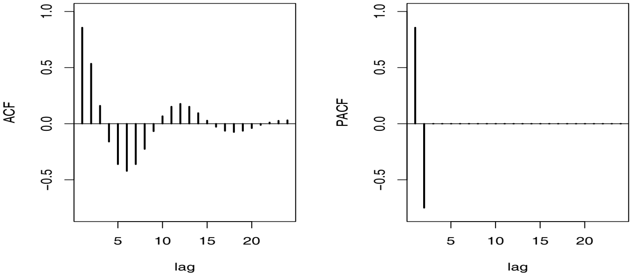


Figure : Behavior of the ACF and PACF in Example 3.17

- AR(2) seems to be appropriate.
- Also use [Property 1.1]: for white noise

$$\hat{\rho}_x(h) \sim N(0, 1/\sqrt{n}) \text{ for large } n.$$

Model building

- Given $x_1, x_2, \cdots$,, let us try to fit it into ARMA(p,q).
    1. We need to decide $p, q \rightarrow$ Model Identification
    2. Estimate unknown parameters $\rightarrow$ Model Estimation
    3. Verify that it is a reasonable model $\rightarrow$ Diagnostic Checking
    4. Then, we predict
- Often, the steps of Model Identification and Diagnostic Checking are hard to separate, so they are considered together.
- For the above steps, we use
    1. ACF and PACF
    2. Asymptotic (large-n) tests
        1. Box-Ljung test
        2. Sign test
        3. Rank test
        4. Q-Q plot
    3. AIC, BIC, FPE, $\cdots$

Testing if ACF follows WN

- We want to test if $\hat{\rho}(h)$ is significant:

$$\begin{cases} H_0 : \hat{\rho}(h) \text{ is the same as that of WN} \\ H_a : \text{ not } H_0. \end{cases}$$

  Recall that in [Property 1.1]

$$\hat{\rho_x}(h) \sim N(0, 1/\sqrt{n}) \text{ for large } n.$$

  To understand it, use $\hat{\rho}(h) = \frac{1}{N-h} \sum_{t=1}^{N-h} a_t a_{t-h}$
  $E[\hat{\rho}(h)] = 0$ and
  $Var[\hat{\rho}(h)] = \frac{1}{N-h}$.

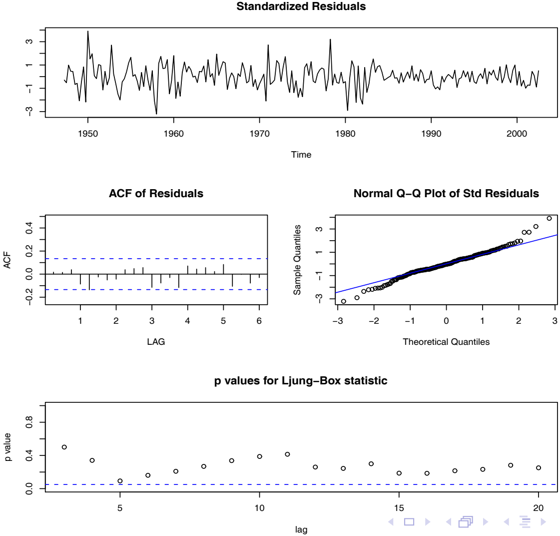
- Usually, we have the CI of $\hat{\rho}(h)$ to be $2/\sqrt{N}$.

Checking residuals

- After a model is fit, the residual $\hat{a}_t$ should behave just like white noise as was assumed.
- Another way to test if $\hat{a}_t$ follows white noise is Ljung-Box-Pierce Q-statistics. It is a kind of χ^2 test.
  - For autocorrelation function, $\hat{r}_h$, of residuals after fitting ARMA(p,q)
    $\hat{r}_h \sim N(0, 1/n)$
    $n\hat{r}_h^2 \sim \chi_1^2$.
  - Thus, Box-Pierce statistics for ARMA(p,q)
    $Q = n\sum_{h=1}^{k} \hat{r}_h^2 \sim \chi_{k-p-q}^2$
  - large $Q \approx$ the assumption not satisfied $\approx$ p-value small $\approx$ there can be a better model than the used one
  - Sometimes a modified version is used
    $Q = n(n+2)\sum_{h=1}^{k} \hat{r}_h^2/(n-k) \sim \chi_{k-p-q}^2$
  - Practically, k is chosen to be around 20
- The sum of the squared residuals (SSR) tells us how much the model fits.
- The Box-Pierce test involves how the residuals as a group behave like white noise.

Checking residuals (cont.)

- We can also draw a normal-probability plot or a q-q plot to see if the residuals follow a normal distribution.



Model selection by AIC, BIC

Then we will look at AIC and BIC values for model selection.

- While adding more parameters makes residuals small, it might worsen the ability of prediction.
- Thus, for theoretical prediction performance, we use AIC, Akaike, criterion from ARMA(p,q)
  $\mathrm{AIC} = -2\log(\hat{L}) + \frac{2(p+q+1)n}{n-p-q-2}$.
  The term $\hat{L}$ is the likelihood value after fitting the ARMA(p,q) model.
  The second term is a penalty factor for large p or q.
  So we consider the compromise between model fitting and the number of parameters.
- We would like to find the model which minimizes AIC values.
- Another criterion is BIC, Bayesian Information Criterion:
  $\mathrm{BIC} = -2\log(\hat{L}) + 2(p+q+1)\log n$.
- Refer to [Example 2.2 in the book]