Times Series Analysis (V) – Autoregressive Moving Average (ARMA) Models

Jianyong Sun
School of Mathematics and Statistics
Xi'an Jiaotong University

28th Sep., 2017

Recap

- ARMA(p, q):
$$\phi(L)Y_t = \theta(L)e_t$$
- Every Explosive has a Cause.
- MA(q) is AR(∞) and AR(p) is MA(∞)
- Parameter Redundancy
- Casualty, Stationarity and Invertibility.
  - Causality: a process is stationary but does not depend on future
  - roots of $\phi(z) = 0$ lie outside the unit circle.
  - Invertibility: Model uniqueness
  - roots of $\theta(z) = 0$ lie outside the unit circle.

The Ψ-weights for an ARMA model

For a casual ARMA model $\phi(L)Y_t = \theta(L)e_t$ where the zeros of $\phi(z)$ are outside the unit circle, recall that $Y_t = \sum_{j=0}^{\infty} \psi_j e_{t-j}$. To get the $\Psi-$weights, we must match the coefficients in $\phi(z)\psi(z) = \theta(z)$:

$$(1-\phi_1 z-\phi_2 z^2-\cdots)(\psi_0+\psi_1 z+\psi_2 z^2+\cdots) = (1-\theta_1 z-\theta_2 z^2-\cdots)$$

we have:

$$\begin{aligned}
\psi_0 &= 1 \\
\psi_1 - \phi_1\psi_0 &= -\theta_1 \\
\psi_2 - \phi_1\psi_1 - \phi_2\psi_0 &= -\theta_2 \\
&\vdots
\end{aligned}$$

where we would take $\phi_j = 0$ for $j > p$ and $\theta_j = 0$ for $j > q$.

The Ψ-weights for an ARMA model

Generally, we have:

$$\psi_j - \sum_{k=1}^{p} \phi_k \psi_{j-k} = 0 \text{ for } \ j \geq \max(p, q+1) \qquad (1)$$

with initial conditions

$$\psi_j - \sum_{k=1}^{j} \phi_k \psi_{j-k} = -\theta_j, 0 \leq j \leq \max(p, q+1) \qquad (2)$$

The Ψ-weights for an ARMA model

Consider the model

$$Y_t = 0.9Y_{t-1} + 0.5e_{t-1} + e_t$$

Because $\max(p, q+1) = 2$, using Eq.(2), we have

$$\psi_0 = 1, \psi_1 = 0.9 + 0.5 = 1.4$$

By Eq.(1), for $j = 2, 3, \cdots$,, the Ψ-weights satisfy

$$\psi_j - 0.9\psi_{j-1} = 0$$

So the general solution is $\psi_j = c0.9^j$. Use the initial condition $\psi_1 = 1.4$, so $1.4 = 0.9c$ and $c = 1.4/0.9$, therefore $\psi_j = 1.4(0.9)^{j-1}$ for $j \geq 1$.

Autocorrelation and Partial Autocorrelation

Consider MA(q) process: $Y_t = \theta(L)e_t$ where $\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q$. It can shown that

$$
\begin{aligned}
E(Y_t) &= 0 \\
\gamma_k &= Cov(Y_{t+k}, Y_t) = Cov\left(\sum_{j=0}^{q} \theta_j e_{t+k-j}, \sum_{i=0}^{q} \theta_i e_{t-i}\right) \\
&= \begin{cases} \sigma_e^2 \sum_{j=0}^{q-k} \theta_j \theta_{j+k} & 0 \le k \le q \\ 0 & k > q. \end{cases} \\
\rho_k &= \begin{cases} \frac{\sum_{j=0}^{q-k} \theta_j \theta_{j+k}}{1+\theta_1^2+\cdots+\theta_q^2} & 1 \le k \le q \\ 0 & k > q. \end{cases}
\end{aligned}
$$

Autocorrelation and Partial Autocorrelation

For a causal ARMA(p,q) model, $\phi(L)Y_t = \theta(L)e_t$, where the zeros of $\phi(z)$ are outside the unit circle, write:

$$Y_t = \sum_{j=0}^{\infty} \psi_j e_{t-j}$$

It follow immediately that

$$\begin{aligned} E(Y_t) &= 0 \\ \gamma_k &= cov(Y_{t+k}, Y_t) = \sigma_e^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+k}, k \geq 0 \end{aligned}$$

Again Eqs.(1) and (2) can be used to solve for the Ψ-weights.

Autocorrelation and Partial Autocorrelation

It is also possible to obtain a homogeneous difference equation in terms of γ_k.

$$
\begin{aligned}
\gamma_k &= Cov(Y_{t+k}, Y_t) = Cov\left(\sum_{j=1}^{p}\phi_j Y_{t+k-j} + \sum_{j=0}^{q}\theta_j e_{t+k-j}, Y_t\right) \\
&= \sum_{j=1}^{p}\phi_j\gamma_{k-j} + \sigma_e^2\sum_{j=k}^{q}\theta_j\psi_{j-k}, k \geq 0
\end{aligned}
$$

where we need the fact for $k \geq 0$

$$
Cov(e_{t+k-j}, Y_t) = Cov\left(e_{t+k-j}, \sum_{k=0}^{\infty}\psi_k e_{t-k}\right) = \psi_{j-k}\sigma_e^2 \qquad (1)
$$

From Eq.(3), we can write a general homogeneous equation for the ACF of a causal ARMA process

$$\gamma_k - \phi_1 \gamma_{k-1} - \cdots - \phi_p \gamma_{k-p} = 0, k \geq \max(p, q+1)$$

with initial conditions:

$$\gamma_k - \sum_{j=1}^{p} \phi_j \gamma_{k-j} = \sigma_e^2 \sum_{j=k}^{q} \theta_j \psi_{j-k}, 0 \leq k < \max(p, q+1)$$

Dividing the equations by γ_0 will allow us to solve for the ACF.

Example: The ACF of an ARMA(1,1)

Consider the ARMA(1,1) process $Y_t = \phi Y_{t-1} + \theta e_{t-1} + e_t$ where $|\phi| < 1$. Based the homogeneous equation, we have:

$$\gamma_k - \phi\gamma_{k-1} = 0, k = 2, 3, \cdots$$

So we get the general solution

$$\gamma_k = c\phi^k, k = 1, 2, \cdots$$

To obtain the initial conditions, we can use previous equation:

$$\gamma_0 = \phi\gamma_1 + \sigma_e^2[1 + \theta\phi + \theta^2]$$

and

$$\gamma_1 = \phi\gamma_0 + \sigma_e^2\theta$$

Example: The ACF of an ARMA(1,1)

Solving for γ_0 and γ_1, we obtain:

$$\gamma_0 = \sigma_e^2 \frac{1 + 2\theta\phi + \theta^2}{1 - \phi^2}$$
$$\gamma_1 = \sigma_e^2 \frac{(1 + \theta\phi)(\phi + \theta)}{1 - \phi^2}$$

To solve for c, note that $\gamma_1 = c\phi$, or $c = \gamma_1/\phi$. Hence the specific solution for $k \geq 1$ is

$$\gamma_k = \sigma_e^2 \frac{(1 + \theta\phi)(\phi + \theta)}{1 + 2\theta\phi + \theta^2} \phi^{k-1}, k \geq 1$$

Finally, dividing through by γ_0 yields the ACF:

$$\rho_k = \frac{(1 + \theta\phi)(\phi + \theta)}{1 + 2\theta\phi + \theta^2} \phi^{k-1}, k \geq 1$$

Example: The ACF of an ARMA(1,1)

Observations

- The general pattern of ρ_k is not different from that of an AR(1) given as follows:

$$\rho_k = \phi^k, k \geq 0$$

- It is unlikely that we will be able to tell the difference between ARMA(1,1) and AR(1) based solely on an ACF estimated from a sample.

The Partial Autocorrelation Function (PACF)

- For an MA(q) model, the ACF will be zero for lags greater than q.
- The ACF provides a considerable amount of information about the order of the dependence for MA.
- If the process is ARMA or AR, the ACF alone tells us little about the orders of dependence.
- We need a function that behave like the ACF of MA models, $\rightarrow$ Partial ACF

The Partial Autocorrelation Function (PACF)

To motivate, consider a causal AR(1) model, $Y_t = \phi Y_{t-1} + e_t$, then

$$\begin{aligned} \gamma_2 &= Cov(Y_t, Y_{t-2}) = Cov(\phi Y_{t-1} + e_t, Y_{t-2}) \\ &= cov(\phi^2 Y_{t-2} + \phi e_{t-1} + e_t, Y_{t-2}) = \phi^2 \gamma_0 \end{aligned}$$

- This result follows from causality because Y_{t-2} involves $\{e_{t-2}, e_{t-3}, \cdots\}$ which all are uncorrelated with e_t and e_{t-1}
- The correlation between Y_t and Y_{t-2} is not zero, as it would be for an MA(1), because Y_t is dependent on Y_{t-2} through Y_{t-1}
- Suppose we break this chain of dependence by removing (or partial out) the effect Y_{t-1}, i.e. to consider the correlation between

  $$Y_t - \phi Y_{t-1} \text{ and } Y_{t-2} - \phi Y_{t-1}$$

  because it is the correlation between Y_t and Y_{t-2} with the linear dependence of each on Y_{t-1} removed.

The Partial Autocorrelation Function (PACF)

In this way, we have broken the dependence chain between Y_t and Y_{t-2}. In fact

$$Cov(Y_t - \phi Y_{t-1}, Y_{t-2} - \phi Y_{t-1}) = Cov(e_t, Y_{t-2} - \phi Y_{t-1}) = 0$$

Here the tool is partial autocorrelation, which is the correlation between Y_s and Y_t with the linear effect of everything 'in the middle' removed.

The Partial Autocorrelation Function (PACF)

To formally define the PACF for mean-zero stationary time series, let $\hat{Y}_{t+k}$ for $k \geq 2$, denote the regression of Y_{t+k} on $\{Y_{t+k-1}, Y_{t+k-2}, \cdots, Y_{t-1}\}$, that is:

$$\hat{Y}_{t+k} = \beta_1 Y_{t+k-1} + \beta_2 Y_{t+k-2} + \cdots + \beta_{k-1} Y_{t+1}$$

No intercept is needed since Y_t is zero-mean. In addition, let $\hat{Y}_t$ denotes the regression of Y_t on $\{Y_{t+1}, Y_{t+2}, \cdots, Y_{t+k-1}\}$, then

$$\hat{Y}_t = \beta_1 Y_{t+1} + \beta_2 Y_{t+2} + \cdots + \beta_{k-1} Y_{t+k-1}$$

Because of stationarity, the coefficients $\beta_1, \cdots, \beta_{k-1}$ are the same.

Regression in the Population Sense

Note that the term regression here refers to regression in the population sense. That is $\hat{Y}_{t+k}$ is the linear combination of $\{Y_{t+k-1}, Y_{t+k-2}, \cdots, Y_{t-1}\}$ that minimizes the mean squared error, i.e. $E(Y_{t+k} - \sum_{j=1}^{k-1} \alpha_j Y_{t+j})^2$

The Partial Autocorrelation Function (PACF)

Definition

The **partial autocorrelation function (PACF)** of a stationary process Y_t, denoted ϕ_{kk}, for $k = 1, 2, \cdots$, is

$$\phi_{11} = corr(Y_{t+1}, Y_t) = \rho_1$$

and

$$\phi_{kk} = corr(Y_{t+k} - \hat{Y}_{t+k}, Y_t - \hat{Y}_t), k \geq 2$$

where

$$\hat{Y}_{t+k} = \beta_1 Y_{t+k-1} + \beta_2 Y_{t+k-2} + \cdots + \beta_{k-1} Y_{t+1}$$

Note both $Y_{t+k} - \hat{Y}_{t+k}$ and $Y_t - \hat{Y}_t$ are uncorrelated with $\{Y_{t+1}, \cdots, Y_{t+k-1}\}$

The Partial Autocorrelation Function (PACF)

Definition

If Y_t is a normally distributed time series, we can let

$$\phi_{kk} = Corr(Y_t, Y_{t-k} | Y_{t-1}, Y_{t-2}, \cdots, Y_{t-k+1})$$

That is, ϕ_{kk} is the correlation in the bivariate distribution of Y_t and Y_{t-k} conditional on $Y_{t-1}, Y_{t-2}, \cdots, Y_{t-k+1}$.
For normally distributed series, the two definitions coincide. By convention $\phi_{11} = 1$.

Example

Recall from previous class, that in minimum mean square error sense, the best linear predictor of Y_t based on Y_{t-1} alone is just $\rho_1 Y_{t-1}$. Thus, for any stationary process,

$$Cov(Y_t - \rho_1 Y_{t-1}, Y_{t-2} - \rho_1 Y_{t-1}) = \gamma_0(\rho_2 - \rho_1^2 - \rho_1^2 + \rho_1^2) = \gamma_0(\rho_2 - \rho_1^2)$$

Since

$$Var(Y_t - \rho_1 Y_{t-1}) = Var(Y_{t-2} - \rho_1 Y_{t-1}) = \gamma_0(1 + \rho_1^2 - 2\rho_1^2) = \gamma_0(1 - \rho_1^2)$$

Then, the lag-2 partial ACF can be expressed as

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

The PACF of an AR(1)

Consider an AR(1) process given by

$$Y_t = \phi Y_{t-1} + e_t$$

with $|\phi| < 1$. By definition $\phi_{11} = \rho_1 = 1$. To calculate ϕ_{22}, consider the regression of Y_{t+2} on Y_{t+1}, say $\hat{Y}_{t+1} = \beta Y_{t+1}$. We choose β to minimize

$$E(Y_{t+2} - \hat{Y}_{t+2})^2 = E(Y_{t+2} - \beta Y_{t+1}) = \gamma_0 - 2\beta\gamma_1 + \beta^2\gamma_0$$

Taking derivatives w.r.t. β and zeroing it, we have

$$\beta = \gamma_1/\gamma_0 = \phi$$

The PACF of an AR(1)

Next consider the regression of Y_t on Y_{t+1}, say $\hat{Y}_t = \beta Y_{t+1}$, we choose β to minimize

$$E(Y_t - \hat{Y}_t)^2 = E(Y_t - \beta Y_{t+1}) = \gamma_0 - 2\beta\gamma_1 + \beta^2\gamma_0$$

This is the same equation as before, so $\beta = \phi$. Hence

$$\begin{aligned} \phi_{22} &= Corr(Y_{t+2} - \hat{Y}_{t+2}, Y_t - \hat{Y}_t) \\ &= Corr(Y_{t+2} - \beta Y_{t+1}, Y_t - \beta Y_{t+1}) \\ &= Corr(e_{t+2}, Y_t - \phi Y_{t+1}) = 0 \end{aligned}$$

by causality. Thus $\phi_{22} = 0$

The PACF of an AR(p)

The model implies that $Y_{t+k} = \sum_{j=1}^{p} \phi_j Y_{t+k-j} + e_{t+k}$ where the roots of $\phi(z)$ are outside the unit circle. When $k > p$, the regression of Y_{t+k} on $\{Y_{t+1}, \cdots, Y_{t+k-1}\}$, is

$$\hat{Y}_{t+k} = \sum_{j=1}^{p} \phi_j Y_{t+k-j}$$

(we will prove it later). Thus when $k > p$,

$$\phi_{kk} = Corr(Y_{t+k} - \hat{Y}_{t+k}, Y_t - \hat{Y}_t) = Corr(e_{t+k}, Y_t - \hat{Y}_t) = 0$$

because by causality, $Y_t - \hat{Y}_t$ depends only on $\{e_{t+k-1}, e_{t+k-2}, \cdots\}$

When $k \leq p$, ϕ_{pp} is not zero, and $\phi_{11}, \cdots, \phi_{p-1,p-1}$ are not necessary zero.
We will prove later that, $\phi_{pp} = \phi_p$.

The PACF of an AR(p)

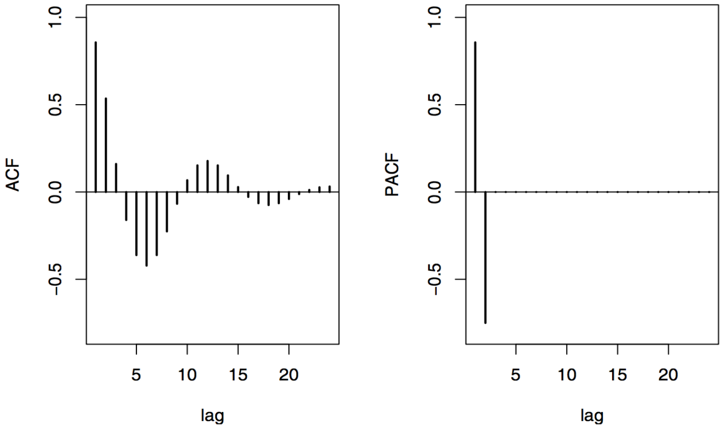

Figure: The ACF and PACF of an AR(2) model with $\phi_1 = 1.5$ and $\phi_2 = -0.75$.

The PACF of an Invertible MA(q)

For an invertible MA(q), we can write

$$Y_t = \sum_{j=1}^{\infty} \pi_j Y_{t-j} + e_t$$

From this result, it should be apparent that the PACF will never cut off, as in the case of an AR(p).

The PACF of an Invertible MA(q)

For an MA(1), $Y_t = e_t - \theta e_{t-1}$, with $|\theta| < 1$, calculations similar to previous example, we have

$$\phi_{22} = \frac{-\theta^2}{1 + \theta^2 + \theta^4}$$

In general, we show that

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

Please refer to page 19.

The PACF of an Invertible MA(q)

- The partial correlation of an MA(1) model never equals zero, but essentially decay to zero exponentially fast as the lag increases
- it is like the autocorrelation function of the AR(1) process

Partial ACF for Stationary Process

A general method for finding the partial ACF for any stationary process with ACF ρ_k (see Anderson 1971)

For a given lag k, it can be shown that the ϕ_{kk} satisfy the Yule-Walker equations:

$$\rho_j = \phi_{k1}\rho_{j-1} + \phi_{k2}\rho_{j-2} + \cdots + \phi_{kk}\rho_{j-k} \text{ for } j = 1, 2, \cdots, k$$

More explicitly, we can write these k linear equations as:

$$\begin{aligned}
\phi_{k1} + \rho_1\phi_{k2} + \rho_2\phi_{k3} + \cdots + \rho_{k-1}\phi_{kk} &= \rho_1 \\
\rho_1\phi_{k1} + \phi_{k2} + \rho_1\phi_{k3} + \cdots + \rho_{k-2}\phi_{kk} &= \rho_2 \\
&\vdots \\
\rho_{k-1}\phi_{k1} + \rho_{k-2}\phi_{k2} + \rho_{k-3}\phi_{k3} + \cdots + \phi_{kk} &= \rho_k
\end{aligned}$$

Partial ACF for Stationary Process

- The solutions to this linear equation system yield ϕ_{kk} for any stationary process.
- If the process is AR(p), then since for $k = p$ are just the Yule-Walker equations, which the AR(p) model is known to satisfy, we must have $\phi_{pp} = \phi_p$
- We have already seen $\phi_{kk} = 0$ for $k > p$.

Partial ACF for Stationary Process

Observations

- The PACF for MA models behaves much like the ACF for AR models
- The PACF for AR models behaves much like the ACF for MA models.
- Because an invertible ARMA model has an infinite AR representation, the PACF will not cut off.

Table: Behavior of the ACF and PACF for ARMA Models

	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

The Sample Partial ACF

- For an observed time series, we need to be able to estimate the partial ACF.
- According to the Yule-Walker equation, we can estimate ρ_k's with sample autocorrelation, then solve it to obtain ϕ_{kk}
- It is called the **sample partial autocorrelation function (sample ACF)** , denoted it by $\hat{\phi}_{kk}$.

The Sample Partial ACF

Levinson (1947) and Durbin (1960) gave an efficient method for obtaining the solutions to the Yule-Walker equation for either theoretical or sample partial autocorrelations. It is shown that the equations can be solved recursively as follows:

$$\phi_{kk} = \frac{\rho_k - \sum_{j=1}^{k-1} \phi_{k-1,j}\rho_{k-j}}{1 - \sum_{j=1}^{k-1} \phi_{k-1,j}\rho_j}$$

where

$$\phi_{k,j} = \phi_{k-1,j} - \phi_{k,k}\phi_{k-1,k-j}$$

The Sample Partial ACF

Example: using $\phi_{11} = \rho_1$ to get started, we have:

$$\phi_{22} = \frac{\rho_2 - \phi_{11}\rho_1}{1 - \phi_{11}\rho_1} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$$

with $\phi_{21} = \phi_{11} - \phi_{22}\phi_{11}$, then

$$\phi_{33} = \frac{\rho_3 - \phi_{21}\rho_2 - \phi_{22}\rho_1}{1 - \phi_{21}\rho_1 - \phi_{22}\rho_2}$$

The Sample Partial ACF

- we can calculate numerically as many values for ϕ_{kk} as desired.
- these recursive equations give us theoretical partial ACF
- by replacing ρ's with r's, we obtained the estimated or sampled partial ACF.

To assess the possible magnitude of the sample ACF, Quenoulle (1949) proved that

Hypothesis Test

Under the hypothesis that an AR(p) model is correct, the sample partial ACF at lags greater than p are approximately normally distributed with zero means and variances $1/n$. Thus for $k > p$, $\pm 2/\sqrt{n}$ can be used as critical limits on $\hat{\phi}_{kk}$ to test the null hypothesis that an AR(p) model is correct.

Questions?