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

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

26th Sep., 2017

Recap

- General Linear Process

$$Y_t = e_t + \psi_1 e_{t-1} + \psi_2 e_{t-2} + \cdots$$

  s.t. $\sum_{i=1}^{\infty} \psi_i^2 < \infty$

- MA(q):

$$Y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_p e_{t-q}$$

  with $\{e_t\}$ the stochastic component and $E(X_t) = 0$

- AR(p):

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + e_t$$

Autoregressive Model (AR(p))

A p-th order autoregressive process, $\{Y_t\}$, satisfy the equation:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_p Y_{t-p} + e_t$$

where Y_t is stationary, e_t is a Gaussian white noise with mean zero and variance σ_e^2. The mean of Y_t is zero. If the mean μ of Y_t is not zero, replace Y_t by $Y_t - \mu$, we have:

$$Y_t - \mu = \phi_1(Y_{t-1} - \mu) + \phi_2(Y_{t-2} - \mu) + \phi_p(Y_{t-p} - \mu) + e_t$$

or

$$Y_t = \alpha + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_p Y_{t-p} + e_t$$

where $\alpha = \mu(1 - \phi_1 - \cdots - \phi_p)$

Autoregressive Moving Average Models

Apply Lag-operator (also called backshift order) to write AR(p):

$$\begin{aligned} Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_p Y_{t-p} &+ e_t \\ Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2} - \phi_p Y_{t-p} &= e_t \\ (1 - \phi_1 L^1 - \phi_2 L^2 - \cdots - \phi_p L^p) Y_t &= e_t \end{aligned}$$

or more concisely:

$$\phi(L) Y_t = e_t$$

$\phi(L)$ is called autoregressive characteristic polynomial (or operator)

Express AR(1) as a General Linear Process

$$\begin{aligned} Y_t &= \phi Y_{t-1} + e_t; \quad Y_{t-1} = \phi Y_{t-2} + e_{t-1} \\ Y_t &= \phi(\phi Y_{t-2} + e_{t-1}) + e_t \Rightarrow Y_t = e_t + \phi e_{t-1} + \phi^2 Y_{t-2} \end{aligned}$$

Repeat the substitution $k-1$ times, we get

$$Y_t = e_t + \phi e_{t-1} + \phi^2 e_{t-2} + \cdots + \phi^{k-1} e_{t-k+1} + \phi^k Y_{t-k}$$

or let $k \to \infty$, $Y_t = \sum_{j=0}^{\infty} \phi^j e_{t-j}$. Note that

$$\lim_{k\to\infty} E\left(Y_t - \sum_{j=0}^{k-1} \phi^j e_{t-j}\right)^2 = \lim_{k\to\infty} \phi^{2k} E(Y_{t-k}^2) = 0$$

$$\begin{aligned}
\gamma_k &= Cov(Y_{t+k}, Y_t) = E\left[\left(\sum_{j=0}^{\infty} \phi^j e_{t+k-j}\right)\left(\sum_{i=0}^{\infty} \phi^k e_{t-i}\right)\right] \\
&= E[(e_{t+k} + \cdots + \phi^k e_t + \phi^{t+1} e_{t-1} + \cdots)(e_t + \phi e_{t-1} + \cdots)] \\
&= \sigma_e^2 \sum_{j=0}^{\infty} \phi^{k+j} \phi^j = \sigma_e^2 \phi^k \sum_{j=0}^{\infty} \phi^{2j} = \frac{\sigma_e^2 \phi^k}{1-\phi^2}
\end{aligned}$$

Thus

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \phi^k \text{ for } \ k = 1, 2, 3, \cdots$$

Autoregressive Moving Average Models

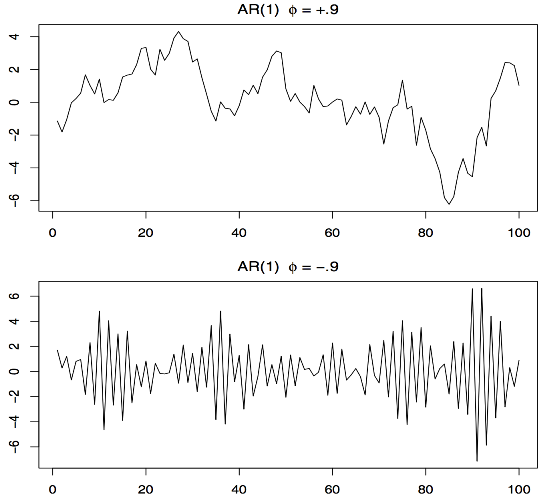

Figure: Simulated AR(1) models: $\phi = 0.9$(top); $\phi = -0.9$ (bottom)

Explosive AR Models and Causality

- AR(1): $Y_t = \phi Y_{t-1} + e_t$ with $|\phi| > 1$, **explosive process**
- because the values of the time series quickly become large in magnitude.
- $|\phi|^j \to \infty$ as $j \to \infty$, $\sum_{j=0}^{k-1} \phi^j e_{t-j}$ will not converge (in mean square) as $k \to \infty$
- So to get $Y_t = \sum_{j=0}^{\infty} \phi^j e_{t-j}$ will not work directly
- However, ...

We can modify that argument to obtain a stationary model!!!
$Y_{t+1} = \phi Y_t + e_{t+1}$, then

$$\begin{aligned} Y_t &= \phi^{-1}(Y_{t+1} - e_{t+1}) = \phi^{-1}(\phi^{-1} Y_{t+2} - \phi^{-1} e_{t+2}) - \phi^{-1} e_{t+1} \\ &\vdots \\ &= \phi^{-k} Y_{t+k} - \sum_{j=1}^{k-1} \phi^{-j} e_{t+j} \end{aligned}$$

by iterating forward k steps.

Because $|\phi|^{-1} < 1$, it suggests that the **future** dependent AR(1) model

$$Y_t = -\sum_{j=1}^{\infty} \phi^{-j} e_{t+j}$$

is stationary.
Useless because it requires us to know the future to be able to predict the future.

Causality

When a process does not depend on the future, such as AR(1) with $|\phi| < 1$, we will say the process is causal.

Every Explosion Has a Cause

With a explosive model, such as $Y_t = \phi Y_{t-1} + e_t$ with $|\phi| > 1$, we have its **non-causal** stationary counterpart:

$$Y_t = -\sum_{j=1}^{\infty} \phi^{-j} e_{t+j}$$

with $E(Y_t) = 0$ and

$$\begin{aligned}\gamma_k &= Cov(Y_{t+k}, Y_t) = Cov\left(-\sum_{j=1}^{\infty} \phi^{-j} e_{t+k+j}, -\sum_{i=1}^{\infty} \phi^{-i} e_{t+i}\right) \\ &= \sigma_e^2 \phi^{-2} \phi^{-k} / (1 - \phi^{-2})\end{aligned}$$

Every Explosion Has a Cause

Consider the causal process

$$X_t = \phi^{-1} X_{t-1} + v_t$$

where $v_t \backsim \mathcal{N}(0, \sigma_e^2 \phi^{-2})$
It is stochastically equal to the $Y_t = \phi Y_{t-1} + e_t$ process. Note that

$$\rho_k^x = \left(\sigma_e^2 \phi^{-2}\right) \frac{\phi^{-k}}{1 - \phi^{-2}}$$

Definition

The moving average model of order q, or MA(q) model, is defined to be

$$Y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q}$$

where there are q lags in the moving average. Although it is not necessary yet, we assume that e_t is a Gaussian white noise series with mean zero and variance σ_e^2, unless otherwise stated.

Moving Average Model

Definition

The **moving average operator** is

$$\theta(L) = 1 - \theta_1 L^1 - \theta_2 L^2 - \cdots - \theta_q L^q$$

Unlike the autoregressive process, the moving average process is stationary for any values of the parameters $\theta_1, \cdots, \theta_q$.

Moving Average Model

Consider the MA(1) model $Y_t = e_t - \theta e_{t-1}$. Then $E(Y_t) = 0$ and

$$\begin{aligned}
\gamma_0 &= (1+\theta^2)\sigma_e^2 \\
\gamma_1 &= -\theta\sigma_e^2 \\
\gamma_k &= 0 \text{ for } \; k > 1 \\
\rho_1 &= \frac{-\theta}{1+\theta^2}
\end{aligned}$$

Non-uniqueness of MA Models and Invertibility:

- For an MA(1) model, ρ_k is the same for θ and $\frac{1}{\theta}$
- The pair $\sigma_e^2 = 1$ and $\theta = 5$ yield the same autocovariance function as the pair $\sigma_e^2 = 25$ and $\theta = 1/5$
- Thus the MA(1) processes

$$
\begin{aligned}
Y_t &= e_t + \frac{1}{5} e_{t-1}, e_t \backsim \mathcal{N}(0, 25) \\
X_t &= v_t + 5 v_{t-1}, v_t \backsim \mathcal{N}(0, 1)
\end{aligned}
$$

  are the same because of normality.
- We can only observe the time series Y_t and X_t, so we cannot distinguish between the models.

Moving Average Model

To discover which model is the invertible model, consider MA(1): $Y_t = e_t - \theta e_{t-1}$, or

$$e_t = Y_t + \theta e_{t-1}$$

Iterating k times, we have:

$$e_t = \sum_{j=0}^{k-1} \theta^j Y_{t-j} + \theta^k e_{t-k}$$

If $|\theta| < 1$, we have:

$$e_t = \sum_{j=0}^{\infty} \theta^j Y_{t-j}$$

Given the two models, which one will you choose?

Moving Average Model

$$e_t = \sum_{j=0}^{\infty} \theta^j Y_{t-j}$$

or

$$e_t = Y_t + \theta Y_{t-1} + \theta^2 Y_{t-2} + \cdots$$

or

$$Y_t = (-\theta Y_{t-1} - \theta^2 Y_{t-2} - \cdots) + e_t$$

That is, MA(1) is AR(∞) if and only if $|\theta| < 1$.

Moving Average Model

- An MA(1) can be written as

$$Y_t = \theta(L)e_t$$

  with $\theta(L) = 1 - \theta L$.
- The inversion of $\theta(L)$ exists if and only if $|\theta| < 1$
- Let $\theta(z) = 1 - \theta z$, if $|\theta| < 1$, $\frac{1}{1-\theta z} = \sum_{j=0}^{\infty} \theta^j z^j$

For an MA(q), we define the **MA characteristic polynomial**:

$$\theta(z) = 1 - \theta_1 z - \theta_2 z^2 - \cdots \theta_q z^q$$

and the corresponding **MA characteristic equation**:

$$1 - \theta_1 z - \theta_2 z^2 - \cdots \theta_q z^q = 0$$

It can be shown that MA(q) model is invertible if and only if the roots of the MA characteristic equation exceed 1 in modulus. Consider the two MA(1)s

$$\begin{aligned} Y_t &= e_t + \frac{1}{5} e_{t-1} \\ X_t &= v_t + 5 v_{t-1} \end{aligned}$$

Their respective roots are -5 and -0.2.

Autoregressive Moving Average Models

Definition

A time series $\{Y_t\}$ is **ARMA**(p, q) if it is stationary and

$$Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \cdots - \theta_q e_{t-q}$$

with $\phi_p, \theta_q \neq 0$ and $\sigma_e^2 > 0$. The parameters p and q are called the autoregressive and moving average orders, respectively. If Y_t has a nonzero mean μ, we can set $\alpha = \mu(1 - \phi_1 - \cdots - \phi_p)$, and write the model as

$$Y_t = \alpha + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \cdots - \theta_q e_{t-q}$$

Autoregressive Moving Average Models

ARMA(p, q) can be written in a lag-operator form:

$$
\begin{aligned}
Y_t - \phi_1 Y_{t-1} - \cdots - \phi_p Y_{t-p} &= e_t - \theta_1 e_{t-1} - \cdots - \theta_q e_{t-q} \\
(1 - \phi_1 L^1 - \phi_2 L^2 - \cdots - \phi_p L^p) Y_t &= (1 - \theta_1 L - \theta_2 L^2 - \cdots \theta_q L^q) e_t \\
\phi(L) Y_t &= \theta(L) e_t
\end{aligned}
$$

We need to discuss the **causality, stationarity** and **invertibility** of the process

ARMA Parameter Redundancy

Consider a white noise process $Y_t = e_t$. Equivalently, we write this as $.5Y_{t-1} = .5e_{t-1}$ by shifting back one unit of time. Now subtract, we have:

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

This looks like ARMA(1,1) model. Here we have hidden the fact that Y_t is white noise because of parameter redundancy or over-parameterization.

ARMA Parameter Redundancy

We can write the parameter redundant model in lag-operator form as

$$(1 - 0.5L)Y_t = (1 - 0.5L)e_t \quad (1)$$

It is clear to see that $Y_t = e_t$ which is the original model.

Note

- The consideration of parameter redundancy will be crucial when we discuss estimation for general ARMA models.
- As in the example, we might fit an ARMA(1,1) model to white noise data and find the parameters are significant
- If we are unaware of parameter redundancy, we might claim the data are correlated.

ARMA Parameter Redundancy

Problems:

- parameter redundant models
- stationary AR models that depend on the future, and
- MA models that are not unique.

AR and MA polynomials

Definition

The **AR and MA polynomials** are defined as

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p, \phi_p \neq 0$$

and

$$\theta(z) = 1 - \theta_1 z - \theta_2 z^2 - \cdots \theta_q^q, \theta_q \neq 0$$

respectively, where z is complex number.

ARMA Causality

To address the problem of future-dependent models, we need to introduce the concept of Causality.

Definition

An ARMA(p, q) model is said to be **causal**, if the time series Y_t can be written as a one-sided linear process

$$Y_t = \sum_{j=0}^{\infty} \psi_j e_{t-j} = \psi(L) e_t$$

where $\phi(L) = \sum_j \psi_j L^j$ and $\sum_{j=0}^{\infty} |\psi_j| < \infty$. Without loss of generality, we set $\psi_0 = 1$

ARMA Causality

To address the problem of future-dependent models, we need to introduce the concept of Causality.

Property

An ARMA(p, q) model is causal if and only if $\phi(z) \neq 0$ for $|z| \leq 1$. The coefficients of the linear process can be determined by solving

$$\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{\theta(z)}{\phi(z)}, |z| \leq 1$$

or an ARMA process is casual only when the roots of $\phi(z)$ lie outside the unit circle.

Model Uniqueness

Invertibility

An ARMA(p, q) model is said to be **inveritble**, if the time series $\{Y_t\}$ can be written as:

$$\pi(L)Y_t = \sum_{j=0}^{\infty} \pi_j Y_{t-j} = e_t$$

where $\pi(L) = \sum_{j=0}^{\infty} \pi_j L^j$ and $\sum_j |\pi_j| < \infty$, we set $\pi_0 = 1$.

Model Uniqueness

An ARMA(p, q) model is invertible if and only if $\theta(z) \neq 0$ for $|z| \leq 1$. The coefficients π_j of $\pi(L)$ given in previous slide can be determined by solving

$$\pi(z) = \sum_{j=0}^{\infty} \pi_j z^j = \frac{\phi(z)}{\theta(z)}, |z_| \leq 1$$

or we see that an ARMA process is invertible only when the roots of $\theta(z)$ lies outside the unit circle.

Example: Parameter Redundancy, Causality and Invertibility

Consider the process

$$Y_t = 0.4Y_{t-1} + 0.45Y_{t-2} + e_t + e_{t-1} + 0.25e_{t-2}$$

or in lag-operator form,

$$(1 - 0.4L - 0.45L^2)Y_t = (1 + L + 0.25L^2)e_t$$

This appears to be an ARMA(2,2) process.

Example: Parameter Redundancy, Causality and Invertibility

The associated polynomials have:

$$\begin{aligned} \phi(z) = 1 - 0.4z - 0.45z^2 &= (1 + 0.5z)(1 - 0.9z) \\ \theta(z) = 1 + z + 0.25z^2 &= (1 + 0.5z)^2 \end{aligned}$$

a common factor that can be canceled. Then polynomial becomes

$$\phi(z) = 1 - 0.9z; \theta(z) = 1 + 0.5z$$

So the model is an ARMA(1,1) model:

$$(1 - 0.9L)Y_t = (1 + 0.5L)e_t$$

or

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

Example: Parameter Redundancy, Causality and Invertibility

The model can be written as a linear process, we can obtain the $\psi-$weights using:

$$\phi(z)\psi(z) = \theta(z)$$

or

$$(1-0.9z)(\psi_0 + \psi_1 z + \psi_2 z^2 + \cdots) = 1 + 0.5z$$

Matching coefficients, we get

$$\psi_0 = 1, \psi_1 = 1.4$$

and $\psi_j = 0.9\psi_{j-1}$ for $j > 1$. Thus $\psi_j = 1.4(0.9)^{j-1}$ for $j \geq 1$. So we have:

$$Y_t = e_t + 1.4\sum_{j=1}^{\infty} .9^{j-1} e_{t-j}$$

Similarly, the invertible representation is

$$Y_t = 1.4\sum_{j=1}^{\infty}(-0.5)^{j-1} Y_{t-j} + e_t$$

Conditional Expectation

If X and Y have joint PDF $f(x, y)$, and we denote the marginal pdf of X by $f(x)$, then the **conditional pdf** of Y given $X = x$ is given by

$$f(y|x) = \frac{f(x, y)}{f(x)}$$

For a given value of x, the conditional pdf has all of the usual properties of a pdf.

Conditional Expectation

The **conditional expectation** of Y given $X = x$ is defined as

$$E(Y|X = x) = \int_{-\infty}^{\infty} yf(y|x)dy$$

As an expected value, or mean, the conditional expectation of Y given $X = x$ has all of the usual properties. For examples:

$$E(aY + bZ + c|X = x) = aE(Y|X = x) + bE(Z|X = x) + c$$

and

$$E(h(Y)|X = x) = \int_{-\infty}^{\infty} h(y)f(y|x)dy$$

In addition, several new peroperties arise:

$$E(h(X)|X = x] = h(x)$$

That is, given $X = x$, the random variable $h(X)$ can be treated like a constant $h(x)$.

Conditional Expectation

More generally,

$$E[h(X,Y)|X=x) = E(h(x,Y)|X=x)$$

If we set $E(Y|X=x) = g(x)$, then $g(X)$ is a random variable and we can consider $E(g(X))$. It can be shown that

$$E[g(X)] = E(Y)$$

which is often written as

$$E[E(Y|X)] = E(Y)$$

If Y and X are independent, then

$$E(Y|X) = E(Y)$$

Conditional Expectation

In the Gaussian case, conditional expectation has an explicit form. Suppose $y \in \mathbb{R}^m$ and $x \in \mathbb{R}^n$, and

$$\begin{pmatrix} y \\ x \end{pmatrix} \backsim \mathcal{N}\left[\begin{pmatrix} \mu_y \\ \mu_x \end{pmatrix}, \begin{pmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{pmatrix}\right]$$

Then $y|x$ is also normal with

$$\begin{aligned} \mu_{y|x} &= \mu_y + \Sigma_{yx}\Sigma_{xx}^{-1}(x - \mu_x) \\ \Sigma_{y|x} &= \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy} \end{aligned}$$

Minimum Mean Square Error Prediction

Suppose Y is a random variable with mean μ_Y and variance σ_Y^2. If our object is to predict Y using only a constant c, what is the *best* choice for c?

Mean square error of prediction

A common criterion is to choose c to minimize the mean square error of prediction, that is, to minimize

$$g(c) = E[(Y - c)^2]$$

$g(c)$ is quadratic in c, so we get

$$c = E(Y) = \mu.$$

and the minimum value of $g(c)$ is just σ_Y^2.

Minimum Mean Square Error Prediction

Now consider the situation where a second random variable X is available and we wish to use the observed value X to help predict Y. Let $\rho = Corr(X, Y)$.
We first suppose that only linear functions $a + bX$ can be used for the prediction. The mean square error is then given by:

$$g(a, b) = E(Y - a - bX)^2$$

Expand it, we have:

$$g(a, b) = E(Y^2) + a^2 + b^2 E(X^2) - 2aE(Y) + 2abE(X) - 2bE(XY)$$

To obtain a, b, take derivatives and zeroing, we get

$$\begin{aligned}\frac{\partial g(a,b)}{\partial a} &= 2a - 2E(Y) + 2bE(X) = 0\\ \frac{\partial g(a,b)}{\partial b} &= 2bE(X^2) + 2aE(X) - 2E(XY) = 0\end{aligned}$$

or

$$\begin{aligned}a + E(x)b &= E(Y)\\ E(X)a + E(X^2)b &= E(XY)\end{aligned}$$

Minimum Mean Square Error Prediction

Solve the equations, we obtain:

$$\begin{aligned} b &= \frac{E(XY) - E(X)E(Y)}{E(X^2) - E(X)^2} = \frac{Cov(X,Y)}{Var(X)} = \rho\frac{\sigma_Y}{\sigma_X} \\ a &= E(Y) - bE(X) = \mu_Y - \rho\frac{\sigma_Y}{\sigma_X}\mu_X \end{aligned}$$

If we let $\hat{Y}$, then we can write

$$\hat{Y} = \left[\mu_Y - \rho\frac{\sigma_Y}{\sigma_X}\mu_X\right] + \rho\frac{\sigma_Y}{\sigma_X}X$$

or

$$\left[\frac{\hat{Y} - \mu_Y}{\sigma_Y}\right] = \rho\left[\frac{X - \mu_X}{\sigma_X}\right]$$

Observations

- we see

$$\min g(a,b) = \sigma_Y^2(1-\rho^2)$$

  which provides that $-1 \leq \rho \leq 1$ since $g(a,b) \geq 0$.

- The minimum mean square error obtained when we use a linear function of X to predict Y is reduced by a factor of $1-\rho^2$ compared with that obtained by ignoring X and simply using the constant μ_Y for our prediction

Minimum Mean Square Error Prediction

Consider the more general problem of predicting Y with an arbitrary function of X. The criterion is again to minimize the mean square error of prediction, that is to choose a function $h(X)$, that minimize

$$E[Y - h(X)]^2 = E(E\{[Y - h(X)]^2 | X\})$$

Thus,

$$h(x) = E(Y|X = x)$$

Since the choice of $h(x)$ minimizes the inner expectation, it must also provide the overall minimum. Thus

$$h(X) = E(Y|X)$$

is the best predictor of Y of all functions of X.

If X and Y have a bivariate normal distribution, it is well-known that

$$E(Y|X) = \mu_Y + \rho \frac{\sigma_Y}{\sigma_X}(X - \mu_X)$$

In this case, the linear predictor is the best of all functions.
More generally, if Y is to be predicted by a function of $X_1, X_2, \cdots, X_n$, then it can be argued that the minimum square error predictor is given by

$$E(Y|X_1, Y_2, \cdots, X_n)$$

Questions?