Chapter 4 Box-Jenkins (ARIMA) Models

AMA4650
Fu Guanxing
guanxing.fu@polyu.edu.hk

Review of ACF and sample ACF (1/2)

- Let $\gamma(k)$ be the autocovariance function at lag k of a stationary time series. The autocorrelation function (ACF) $\rho(k)$ at lag k is given by

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)}, \quad k = 0, \pm 1, \pm 2, \cdots$$

Remark.

1. We have $\gamma(k) = \gamma(-k)$ and hence $\rho(k) = \rho(-k)$.
2. $|\gamma(k)| \leq \gamma(0)$ and hence $|\rho(k)| \leq 1$.
3. Sometimes, we denote by γ_k the autocovariance at lag k, i.e. $\gamma_k = \gamma(k)$. Similarly, we also have denote $\rho_k = \rho(k)$.

- To identify whether $\rho(k)$ is zero from observations, we need to use sample autocorrelation function (sample ACF)

$$r_k = \frac{\sum_{t=k+1}^{n}(x_t - \bar{x})(x_{t-k} - \bar{x})}{\sum_{t=1}^{n}(x_t - \bar{x})^2} \quad \curvearrowright$$

Review of ACF and sample ACF (2/2))

- **Question 1**: If $\rho(k) = 0$, how do we identify it from observations?
- **Answer 1**: *Use the corresponding sample ACF*. However, if ACF equals zero, in real life applications, the sample ACF will not necessarily be equal to zero, but the absolute value will be small.
- **Question 2**: how small is small?
- **Answer 2**: *large sample distribution of the sample ACF (Bartlett's lemma)*[1]
  1. Sample ACF follows normal distribution asymptotically.
  2. If we make the assumption that the true value of ACF at lag k is zero ($\rho_k = 0$), then the variance of the sample ACF is

$$\text{Var}(r_k) \approx \frac{1}{n},$$

     and the standard error is

$$se(r_k) \approx \frac{1}{\sqrt{n}},$$

     where n is the number of observations.
  3. 95% confidence interval of sample ACF is $\pm\frac{1.96}{\sqrt{n}} \approx \pm\frac{2}{\sqrt{n}}$.

[1] Refer to https://link.springer.com/content/pdf/bbm%3A978-3-319-52452-8%2F1.pdf for the statement and the proof.

Motivation

Why ARIMA models? (1/2)

- In practice, uncorrelation assumption for observations is often violated. That is, successive observations are serially dependent.
- Exponential smoothing may be inefficient and sometimes inappropriate because they do not take advantage of the serial dependence in the observations *in the most effective way*.
- To introduce dependent structure, this chapter explores a general class of models: **A**uto**r**egressive **i**ntegrated **m**oving **a**verage models (ARIMA, or Box-Jenkins).

Why ARIMA models? (2/2)

- *Wold's decomposition theorem*. Every stationary time series Y can be written as infinite moving average

$$Y_t = \mu + \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i},$$

where μ is deterministic, $\epsilon \sim WN(0, \sigma^2)$ and $\sum_{i=0}^{\infty} \psi_i^2 < \infty$.

- It is impractical to attempt to estimate infinitely many weights $\{\psi_i\}$.
- Thus, the infinite moving average representation of stationary TS is useless in practice, except for certain special cases:
  - **Case 1** - finite order MA models: only finitely many $\{\psi_i\} \neq 0$.
  - **Case 2** - finite order AR models: $\{\psi_i\}_{i=0}^{\infty}$ are generated by finite parameters.
  - **Case 3** - Mixture of finite order AR and MA models (ARMA).
- All of them are special cases of ARIMA models.

4.1 Moving average processes

Moving average process

Definition 4.1 (The MA(q) process/model)

The proces $\{X_t, t \in \mathbb{Z}\}$ is called a moving average process of order q, or **MA(q)**, if

$$X_t = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}, \quad \text{where } \{Z_t\} \sim WN(0, \sigma^2).$$

1. Current observ. $=$ weighted sum of current noise and past noise of order q.
2. General MA(q) model: $X_t = \mu + \theta_0 Z_t + \theta_1 Z_{t-q} + \cdots + \theta_q Z_{t-q}$, $\theta_0 \neq 0$.
   - W.l.o.g. assume $\mu = 0$ and $\theta_0 = 1$. Else, consider

     $$\bar{X}_t = \frac{X_t - \mu}{\theta_0},$$

     which follows

     $$\bar{X}_t = Z_t + \bar{\theta}_1 Z_{t-1} + \cdots + \bar{\theta}_q Z_{t-q},$$

     where $\bar{\theta}_1 = \frac{\theta_1}{\theta_0}, \cdots \bar{\theta}_q = \frac{\theta_q}{\theta_0}$.
3. MA(q) can be rewritten as $X_t = \sum_{i=0}^{q} \theta_i Z_{t-i}$ with $\theta_0 = 1$ conventionally.

Examples of MA processes

Today's value = linear combination of past shocks/errors/noise.

Example 1: change of price of oil[2] and typhoon

$$\Delta Oilp_t = \mu + \epsilon_t + 0.5\epsilon_{t-1},$$

where $\Delta Oilp_t$ is the change of oil price at t and ϵ_t is the strength of typhoon at t.

Example 2: change of stock price and orders of noise traders

$$\Delta S_t = \mu + \sigma_0\epsilon_t + \sigma_1\epsilon_{t-1},$$

where ΔS_t is the change of price at t and ϵ_t is the volume of noise traders at t.

[2] https://oilprice.com

Statistical properties of MA(q): $X_t = \sum_{i=0}^{q} \theta_i Z_{t-i}$ (1/2)

Statistical properties of MA(q): $X_t = \sum_{i=0}^{q} \theta_i Z_{t-i}$ (2/2)

- ACF of MA(q) has **cut-off** at lag q!

Example 4.1: MA(1)

Statistical properties of MA(1): $X_t = Z_t + \theta Z_{t-1}$.

- **Expectation/mean**: $\mathbb{E}[X_t] = 0$.
- **Variance**: $Var(X_t) = (1+\theta^2)\sigma^2$.
- **Autocovariance**: $\gamma(0) = (1+\theta^2)\sigma^2, \quad \gamma(1) = \theta\sigma^2, \quad \gamma(k) = 0$ for $k > 1$.
- **ACF**: $\rho(1) = \frac{\theta}{1+\theta^2}$ and $\rho(k) = 0$ for $k > 1$.

Remark:
Do NOT try to construct an MA(1) model with ACF $\{\rho(k)\} = (1, 0.8, 0, \cdots)$

Example 4.2: MA(2) (1/2)

Statistical properties of MA(2): $X_t = Z_t + \theta Z_{t-1} + \theta_2 Z_{t-2}$.

- **Expectation/mean**: $\mathbb{E}[X_t] = 0$.
- **Autocovariance**:

  $$\gamma(0) = \sigma^2(1+\theta_1^2+\theta_2^2),\ \ \gamma(1) = \sigma^2(\theta_1+\theta_1\theta_2),\ \ \gamma(2) = \sigma^2\theta_2,\ \ \gamma(k) = 0,\ k > 2.$$

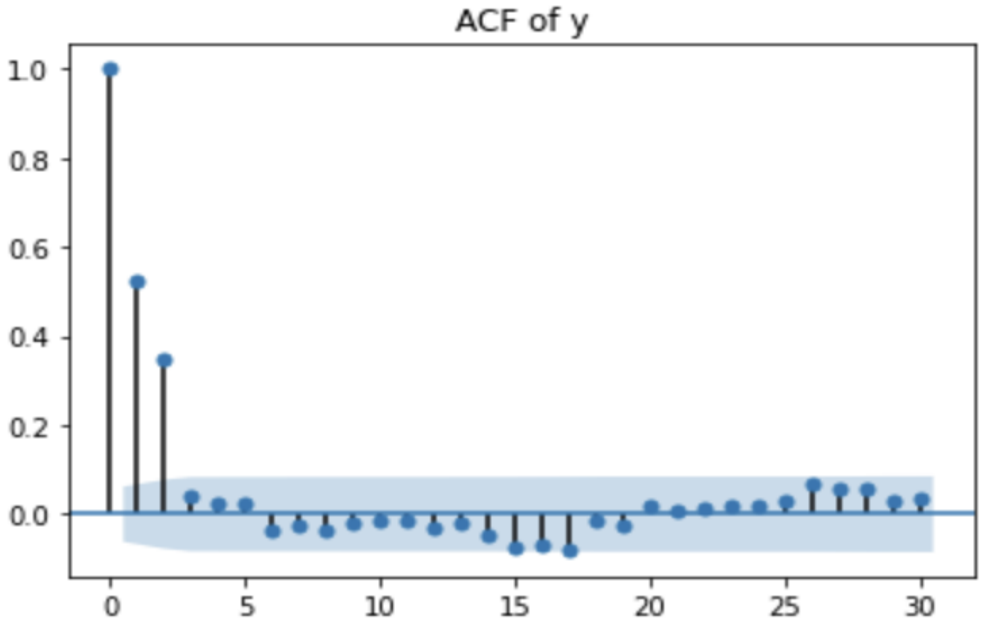
- **ACF**:

  $$\rho(1) = \frac{\theta_1 + \theta_1\theta_2}{1 + \theta_1^2 + \theta_2^2}$$
  $$\rho(2) = \frac{\theta_2}{1 + \theta_1^2 + \theta_2^2}$$

  and $\rho(k) = 0$ for all $k > 2$.

Example 4.2 (2/2)

Sample ACF of 1000 simulations from $X_t = Z_t + 0.5Z_{t-1} + 0.5Z_{t-2}$.



Remark. $\frac{2}{\sqrt{1000}} \approx 0.0632$.

Backward shift operator and difference operator (1/2)

- In order to extend the discussion on general MA process, recall

1. Backward shift (or Backshift) operator: $BX_t = X_{t-1}$
2. Difference operator: $\nabla X_t = X_t - X_{t-1}$

- Relation $\nabla = 1 - B$.
- Both operators obey algebraic properties.
  - Example 1: $B^2 X_t$.

$$B^2 X_t = B(BX_t) = BX_{t-1} = X_{t-2}.$$

  - Example 2: $(1-B)^2 X_t$.

$$\begin{aligned}(1-B)^2 X_t &= (1-B)[(1-B)X_t] = (1-B)(X_t - X_{t-1}) \\ &= (X_t - X_{t-1}) - B(X_t - X_{t-1}) \\ &= X_t - 2X_{t-1} + X_{t-2} \\ &= (1 - 2B + B^2)X_t.\end{aligned}$$

As a result, we get $\nabla^2 = (1-B)^2 = 1 - 2B + B^2$.

Backward shift operator and difference operator (2/2)

- The expression like $(1-\alpha B)^{-1}X_t$ can be regarded as

$$(1+\alpha B+\alpha^2 B^2+\cdots)X_t.$$

Examples.

1. $X_t = \alpha X_{t-1} + Z_t$ is rewritten as $(1-\alpha B)X_t = Z_t$.
2. $X_t = X_{t-1} + Z_t$ is rewritten as $\nabla X_t = Z_t$.
3. $X_t = \sum_{i=0}^{q} \theta_i Z_{t-i} \sim X_t = \Theta(B)Z_t$, where

$$\Theta(y) := \sum_{i=0}^{q} \theta_i y^i.$$

It is called MA polynomial.

Invertibility of MA processes

- Recall that MA(q) can be rewritten as

$$X_t = \Theta(B)Z_t = (\theta_0 + \theta_1 B + \cdots + \theta_q B^q)Z_t = \sum_{i=0}^{q} \theta_i Z_{t-i}.$$

- An important concept related to MA processes is *invertibility*.

Definition 4.2

An MA(q) process

$$X_t = \Theta(B)Z_t, \qquad \text{where } \{Z_t\} \sim WN(0, \sigma^2)$$

is said to be *invertible* if there exists constants $\{\pi_j\}$ s.t. $\sum_{j=0}^{\infty} |\pi_j| < \infty$ and

$$Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}, \quad t \in \mathbb{Z}.$$

Why invertiblity? (1/2)

- Consider an MA(1) process

$$X_t = Z_t + \beta Z_{t-1}, \quad |\beta| < 1. \qquad (*)$$

- Recall Example 4.1: ACF of (*) follows

$$\rho(0) = 1, \quad \rho(1) = \frac{\beta}{1+\beta^2}, \quad \rho(k) = 0 \text{ for all } k > 1.$$

- Consider another MA(1) process

$$X_t = Z_t + \frac{1}{\beta} Z_{t-1}, \quad |\beta| < 1. \qquad (**)$$

- ACF of (**) follows

$$\rho(0) = 1, \quad \rho(1) = \frac{\frac{1}{\beta}}{1+\frac{1}{\beta^2}} = \frac{\beta}{\beta^2+1}, \quad \rho(k) = 0 \text{ for all } k > 1.$$

- They have the same ACF. So there is an indeterminacy problem.

Why invertibility? (2/2)

The first MA(1) (*)

- Note that $\frac{1}{1+\beta B} = \sum_{n=0}^{\infty}(-\beta)^n B^n$.
- For (*), we have

$$X_t = (1+\beta B)Z_t \Rightarrow Z_t = (1+\beta B)^{-1}X_t = X_t - \beta X_{t-1} + \beta^2 X_{t-2} - \cdots,$$

where the coefficients on the r.h.s. are decaying and convergent.

The second MA(1) (**)

- Note that $\frac{1}{1+\frac{1}{\beta}B} = \sum_{n=0}^{\infty}(-\frac{1}{\beta})^n B^n$.
- For (**),

$$X_t = (1+\frac{1}{\beta}B)Z_t \Rightarrow Z_t = (1+\frac{1}{\beta}B)^{-1}X_t = X_t - \frac{1}{\beta}X_{t-1} + \frac{1}{\beta^2}X_{t-2} - \cdots,$$

where the coefficients on r.h.s. are **not** decaying or convergent $\Rightarrow$ it does not make sense as it suggests the error's dependence on X_t increases as we go back in time.

- In general, for MA process $X_t = \Theta(B)X_t$, if the representation $Z_t = \Theta(B)^{-1}X_t$ has this behavior, then the time series X is not invertible.

Invertibility of MA processes

Theorem 4.1

A necessary and sufficient condition for an MA(q) process $X_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \cdots + \theta_q Z_{t-q} = \Theta(B) Z_t$ to be invertible is that all the roots of the equation

$$\Theta(y) = 0$$

lie *outside* the unit circle.

Remark. The roots can be either real or complex.

1. For a real root y, outside the unit circle means $|y| > 1$.
2. For a complex root y, outside unit circle means the radius $|y| > 1$.

Example. For the MA(1) process $X_t = Z_t + \beta Z_{t-1} = \Theta(B) Z_t$, the root of $\Theta(y) = 1 + \beta y = 0$ is $y = -\frac{1}{\beta}$. Thus,

$$\text{invertibility} \Leftrightarrow \left| -\frac{1}{\beta} \right| > 1 \Leftrightarrow |\beta| < 1.$$

Proof of Theorem 4.1 (1/4)

- By *fundamental theorem of algebra*[3], every univariate polynomial of positive degree n with complex coefficients can be factorized as

$$f(y) = c(y - y_1)(y - y_2) \cdots (y - y_n),$$

where $y_1, y_2, \cdots, y_n$ are roots of f.

- Let $n = q$ and $f = \Theta$, where Θ is the MA polynomial

$$\Theta(y) = 1 + \theta_1 y + \cdots + \theta_q y^q.$$

Let all solutions of $\Theta(y) = 0$ be $\frac{1}{\lambda_1}, \cdots, \frac{1}{\lambda_q}$. Then the polynomial can be rewritten as

$$\Theta(y) = (1 - \lambda_1 y) \cdots (1 - \lambda_q y) = \prod_{i=1}^{q} (1 - \lambda_i y).$$

[3] Refer to https://en.wikipedia.org/wiki/Fundamental_theorem_of_algebra

Proof of Theorem 4.1 (2/4)

Sufficiency: outside unit circle $\Rightarrow$ invertibility.

Goal: find absolutely summable $\{\pi_j\}$ s.t. $Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}$.

Outside unit circle implies

1. $\left|\frac{1}{\lambda_i}\right| > 1$ for all $i = 1, \cdots, q$
2. $\Theta(y) \neq 0$ for all $|y| < 1$. Thus, we can define

$$\Phi(y) := \frac{1}{\Theta(y)}, \qquad |y| < 1.$$

Taking $\Phi(B)$ on both sides of $X_t = \Theta(B)Z_t$, one has $\Phi(B)X_t = Z_t$, which implies that for some constants $a_1, \cdots, a_q$,

$$Z_t = \frac{1}{\Pi_{i=1}^{q}(1-\lambda_i B)} X_t = \sum_{i=1}^{q} \frac{a_i}{1-\lambda_i B} X_t = \sum_{i=1}^{q} a_i \sum_{j=0}^{\infty} \lambda_i^j B^j X_t$$

$$= \sum_{i=1}^{q} a_i \sum_{j=0}^{\infty} \lambda_i^j X_{t-j} = \sum_{j=0}^{\infty} \left(\sum_{i=1}^{q} a_i \lambda_i^j \right) X_{t-j} := \sum_{j=0}^{\infty} \pi_j X_{t-j},$$

where, $\pi_j = \sum_{i=1}^{q} a_i \lambda_i^j$.

Proof of Theorem 4.1 (3/4)

It remains to prove that $\{\pi_j\}$ is absolutely summable. Indeed,

$$\sum_{j=0}^{\infty} |\pi_j| = \sum_{j=0}^{\infty} \left| \sum_{i=1}^{q} a_i \lambda_i^j \right| \leq q \max_{1 \leq i \leq q} |a_i| \sum_{j=0}^{\infty} \left(\max_{1 \leq i \leq q} |\lambda_i| \right)^j < \infty,$$

since $\left|\frac{1}{\lambda_i}\right| > 1 \Leftrightarrow |\lambda_i| < 1$ by assumption.

Proof of Theorem 4.1 (4/4)

Necessity: invertibility $\Rightarrow$ outside unit circle.

Goal: $|\lambda_i| < 1$ for all $i = 1, \cdots, q$.

Invertibility implies that $\frac{1}{\Theta(B)}$ is well defined because

$$X_t = \Theta(B)Z_t \Leftrightarrow Z_t = \frac{1}{\Theta(B)}X_t := \Phi(B)X_t.$$

Note that from the *sufficient proof*, once $\frac{1}{\Theta(B)}$ is well defined, we have

$$Z_t = \sum_{j=1}^{q} a_i \sum_{j=0}^{\infty} \lambda_i^j X_{t-j}.$$

Since for each $i = 1, \cdots, q$, $\sum_{j=0}^{\infty} \lambda_i^j X_{t-j}$ is well defined. It implies $|\lambda_i| < 1$, $i = 1, \cdots, q$.

Example 4.3

For an invertible MA(2) process

$$Y_t = \epsilon_t - 0.2\epsilon_{t-1} + \theta_2\epsilon_{t-2},$$

given $\rho_1 = -0.1$, determine ρ_2.