LESSON 3

Box & Jenkins approach

JEN-WEN LIN, PhD, CFA

Date: 06 January, 2019

Steps to time series modeling

1. Plot the time series and check for

- Trend, seasonal and other cyclic components, any apparent sharp changes in behavior, as well as any outlying observations

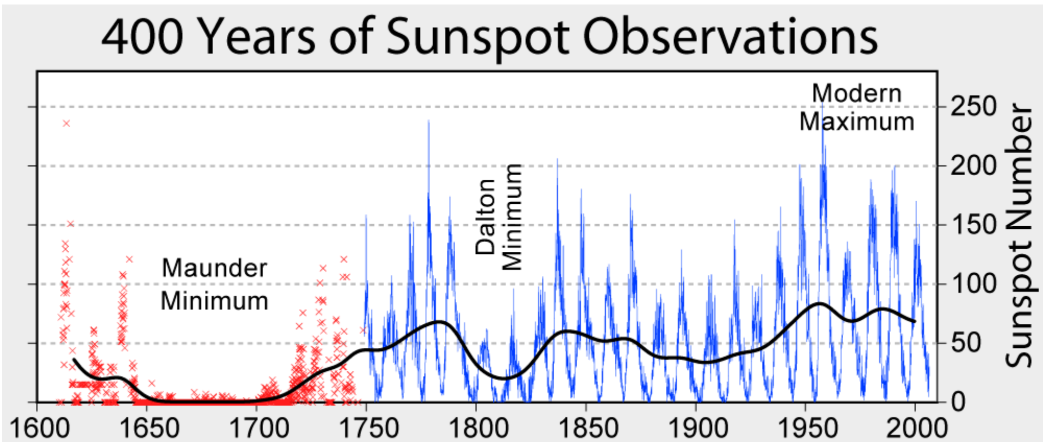
2. Remove trend and seasonal components to get residuals

3. Choose a model to fit the residuals

4. Forecasting can be carried out by forecasting residual and then inverting the transformation carried out in Step 2.

Modeling irregular component

Modeling irregular component



- The origin of ARMA models can be traced back to Yule's work on Wolfer's sunspot numbers.
- ARMA models were widely accepted by researchers and professionals after the work of Box and Jenkins (1970).

Autoregressive and moving average (ARMA) model

- A process $\{X_t\}$ is said to be an ARMA(p, q) process if $\{X_t\}$ is stationary and if for every t

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = a_t + \theta_1 a_{t-1} + \cdots + \theta_q a_{t-q}, \quad a_t \sim N(0, \sigma_a^2)$$

  - $\{X_t\}$ is said to be an $ARMA(p, q)$ process with mean μ if $\{X_t - \mu\}$ is an $ARMA(p, q)$ process.

- Express the ARMA model using compact notation:

$$\Phi(B)(X_t - \mu) = \Theta(B)a_t,$$

where

$$\Phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$$

and

$$\Theta(z) = 1 + \theta_1 z + \cdots + \theta_p z^p.$$

MA(∞) processes

- If $\{a_t\} \sim WN(0, \sigma^2)$ then we say that $\{X_t\}$ is a MA(∞) process of $\{a_t\}$ if there exists a sequence $\{\psi_j\}$ with $\sum_{j=0}^{\infty} |\psi_j| < \infty$ such that

$$X_t = \sum_{j=0}^{\infty} \psi_j a_{t-j}, \quad t = \cdots, -1, 0, 1, 2, \ldots$$

- **Theorem** (Brockwell and Davis, 1992, p.91)

  The MA(∞) process is stationary with mean zero and autocovariance function

$$\gamma(k) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|k|}$$

- We can calculate autocorrelation functions of a stationary stochastic process $\{X_t\}$ using the aforementioned theorem as long as $\{X_t\}$ can be written in the form of a MA(∞) process.

Moving average process of order q

- The moving average process of order *q*, denoted as $MA(q)$, is given by

$$X_t = a_t + \theta_1 a_{t-1} + \cdots + \theta_q a_{t-q} = \Theta(B) a_t$$

where B is the backward shift operator, $B^h X_t = X_{t-h}$,

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

and $a_t \sim WN(0, \sigma^2)$

- **Question:** What conditions do we need for $MA\ (p)$ processes to be weakly stationary?

Autocovariance of MA(2) process

$$X_t = a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2}, \qquad a_t \sim WN(0, \sigma^2)$$

$$X_t = \sum_{j=0}^{\infty} \psi_j a_{t-j},\ t = \cdots, -1,0,1,2, \ldots \qquad \gamma(k) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|k|}$$

$$\psi_j = \theta_j,\ j = 0,1,2 \text{ and } \psi_j = 0,\ j > 2$$

$$\gamma(1) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+1} = \sigma^2 (1 \cdot \theta_1 + \theta_1 \cdot \theta_2 + \theta_2 \cdot 0) = \sigma^2 (\theta_1 + \theta_1 \theta_2)$$

$$\gamma(2) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+2} = \sigma^2 (1 \cdot \theta_2 + \theta_1 \cdot 0 + \theta_2 \cdot 0) = \sigma^2 \theta_2$$

$$\gamma(3) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+3} = \sigma^2 (1 \cdot 0 + \theta_1 \cdot 0 + \theta_2 \cdot 0) = 0$$

Autocovariance of MA(q) processes

$$X_t = a_t + \theta_1 a_{t-1} + \cdots + \theta_q a_{t-q}, a_t \sim WN(0, \sigma^2)$$

$$\psi_j = \theta_j,\ j = 0, 1, \ldots, q \text{ and } \psi_j = 0\ \forall j > q$$

$$\gamma(k) = cov(X_t, X_{t+k})$$
$$= cov\left(\sum_{j=0}^{q} \theta_j a_{t-j}, \sum_{j=0}^{q} \theta_j a_{t+k-j}\right)$$

$$= \begin{cases} \sigma^2 \sum_{i=0}^{q-k} \theta_i \theta_{i+k}, & k = 0, \pm 1, \ldots, \pm q \\ 0, & otherwise \end{cases}$$

$$\gamma(k) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|k|}$$

Summary

- The maximum lag of the non-zero sample autocorrelation is a good indicator of the MA(q) processes.

- The ACF of MA(q) processes

$$X_t = a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \cdots + \theta_q a_{t-q},$$

cut off after lag q, i.e.

$$\rho_k = \begin{cases} \dfrac{\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \cdots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \cdots + \theta_q^2} & , k = 1, \ldots, q \\ 0 & , k > q \end{cases}$$

Autoregressive model of order p

- The autoregressive process of order *p*, AR(*p*), is

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = \Phi(B) X_t = a_t,$$

where $a_t \sim WN(0, \sigma^2)$, $B^h X_t = X_{t-h}$, $h \in Z$, and

$$\Phi(B) = \left(1 - \phi_1 B - \cdots - \phi_p B^p\right)$$

Questions to think about

1. How to check the weak stationarity of $\{X_t\}$?
2. Derive the autocovariance function of $\{X_t\}$?
3. Use ACF for model identification?

AR(1) processes

- The autoregressive process of order one is given by

$$X_t - \phi X_{t-1} = a_t, \{a_t\} \sim WN(0, \sigma^2),$$

where a_t is uncorrelated with X_s for all $s < t$.

- Using recursive substitution, we may express an AR(1) process as the for of an MA(∞) process

$$X_t = \sum_{i=0}^{\infty} \phi^j a_{t-j}\,.$$

AR(1) processes

An AR(1) process with $|\phi| < 1$ is called causal or a future-independent autoregressive process. In this course, we consider only causal processes.

$$\gamma(k) = \mathrm{cov}(X_t, X_{t+k})$$

$$= \mathrm{cov}(\sum_{l=0}^{\infty} \phi^l a_{t-l}, \sum_{j=0}^{\infty} \phi^j a_{t+k-j})$$

$$= \mathrm{cov}(\sum_{l=0}^{\infty} \phi^l a_{t-l}, \sum_{s=0}^{k-1} \phi^s a_{t+k-s} + \sum_{l=0}^{\infty} \phi^{l+k} a_{t-l})$$

$$= \phi^k \gamma(0)$$

$$\gamma(0) = Var(X_t) = \sigma^2 / (1 - \phi^2)$$

finite if $|\phi| < 1$, why?

Remark and questions

- We have shown answers for an AR(1) process using definition and its MA(∞) representation.
- Can we apply the same technique to higher order AR process and answer the questions asked earlier?
  1. Check weak stationarity of an AR(p) model
  2. Derive autocorrelation functions of an AR(p) model
  3. Can we use the pattern of ACFs of an AR(p) process for model identification

General approach to checking stationarity of an AR(*p*) process

- A general way of checking the (weak) stationarity condition of autoregressive processes is that the roots of the following equation

$$\Phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p = 0$$

must lie outside the unit circle.

- *Example:* Check the stationarity of an AR(1) model

$$(1 - \phi B)X_t = a_t$$

  - $\Phi(B) = 1 - \phi B = 0$ so $B = 1/\phi$
  - $|B| = |1/\phi| > 1$ so $|\phi| < 1$

autocovariance functions of autoregressive processes

For simplicity, consider a stationary AR(1) process

$$X_t = \phi X_{t-1} + a_t,\ \ a_t \sim NID(0, \sigma^2), \qquad (\star)$$

For $k = 1$: multiply X_{t-1} on both sides of $(\star)$ and take expectation on both sides of the equation

$$X_t X_{t-1} = \phi X_{t-1}^2 + X_{t-1} a_t$$

Take Expecation:

$$E(X_t X_{t-1}) = \phi \cdot Var(X_t)$$

$$\Rightarrow \gamma(1) = \phi \cdot \gamma(0)$$

$$X_t X_{t-1} = \phi X_{t-1}^2 + X_{t-1} a_t$$

Take Expecation:

$$\boxed{E(X_t X_{t-1})} = \boxed{\phi \cdot Var(X_t)}$$

$$\Rightarrow \gamma(1) = \phi \cdot \gamma(0)$$

$$X_{t-1} = \sum_{j=0}^{\infty} \phi^j \cdot a_{t-1-j}$$

$$\text{cov}(a_t, X_{t-1}) = \text{cov}(a_t, \sum_{j=0} \phi^j a_{t-1-j}) = 0$$

For $k = 2$: multiply X_{t-2} on both sides of (*) and take expectation on both sides of the equation

$$X_t X_{t-2} = \phi X_{t-1} X_{t-2} + X_{t-2} a_t$$

Take Expecation:

$$E(X_t X_{t-2}) = \phi \cdot E(X_{t-1} X_{t-2})$$

$$\Rightarrow \gamma(2) = \phi \cdot \gamma(1)$$

Using the result that $\gamma(1) = \phi \cdot \gamma(0)$

$$\Rightarrow \gamma(2) = \phi \cdot \gamma(1) = \phi^2 \cdot \gamma(0)$$

For k ≥3, similarly we have

$$X_t X_{t-k} = \phi X_{t-1} X_{t-k} + X_{t-k} a_t$$

Take Expecation:

$$E(X_t X_{t-k}) = \phi \cdot E(X_{t-1} X_{t-k})$$

$$\Rightarrow \gamma(k) = \phi \cdot \gamma(k-1)$$

$$\Rightarrow \quad \cdots\cdots$$

$$\Rightarrow \boxed{\gamma(k) = \phi^k \gamma(0)}$$

Question: Can we apply the above technique to a general AR(p) process?

ACF of AR(p) processes

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots\ldots \phi_p X_{t-p} + a_t$$

Stationarity condition of an AR(p) models are that all *p* roots of the characteristic equation outside of the unit circle

ACF

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \ldots\ldots \phi_p \rho_{k-p}$$

$$\left.\begin{aligned}
\rho_1 &= \phi_1 \rho_0 + \phi_2 \rho_1 + \ldots\ldots \phi_p \rho_{p-1} \\
\rho_2 &= \phi_1 \rho_{11} + \phi_2 \rho_0 + \ldots\ldots \phi_p \rho_{p-2} \\
&\vdots \\
\rho_p &= \phi_1 \rho_{p-1} + \phi_2 \rho_{p-2} + \ldots\ldots \phi_p \rho_0
\end{aligned}\right\}$$

System (Yule-Walker equations) to solve for the first *p* autocorrelations: *p* unknowns and *p* equations

In general, ACFs of an AR(p) model decay as mixture of exponentials and/or damped sine waves--depending on real/complex roots

Alternative measure FOR temporal dependence

Partial autocorrelation function

- The correlation between X_t and X_{t+k} after mutual linear dependency on the intervening variables, X_{t+1}, X_{t+2}, …, and X_{t+k-1} has been removed.

- The conditional correlation

$$\phi_{kk} = corr(X_t, X_{t+k} | X_{t+1}, \dots, X_{t+k-1})$$

  is usually referred to as the partial autocorrelation functions (PACF) in time series analysis.

  - PACF between X_t and X_{t+k} can be obtained as the regression coefficient associated with X_t when regressing X_{t+k} on its *k* lagged variables X_{t+k-1}, X_{t+k-2}, …, and X_t.

Calculate PACF using Yule-Walker equations

- A general method for finding the partial autocorrelation function for any stationary process with autocorrelation function ρ_k is as follows.

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

$$\rho_j = \phi_{k1}\rho_{j-1} + \phi_{k2}\rho_{j-2} + \phi_{k3}\rho_{j-3} + \cdots + \phi_{kk}\rho_{j-k},$$

for $j = 1,2, \dots, k$. That is, we regard $\rho_1, \dots, \rho_k$ as given and wish to solve for ϕ_{kk}.

Solving YW equations using Cramer's rule

- For the AR(2) process, the Yule-Walker equations may be written as

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-1}, \text{ for } k \geq 1.$$

- Given a set of ACFs, we can solve $\phi_{11}, \phi_{22}, \phi_{33}$ based on Yule-Walker equations:

$$\phi_{11} = \rho_1 = \frac{\phi_1}{1 - \phi_2} \qquad \phi_{22} = \frac{\det\begin{bmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{bmatrix}}{\det\begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} = \phi_2$$

Cramer's rule

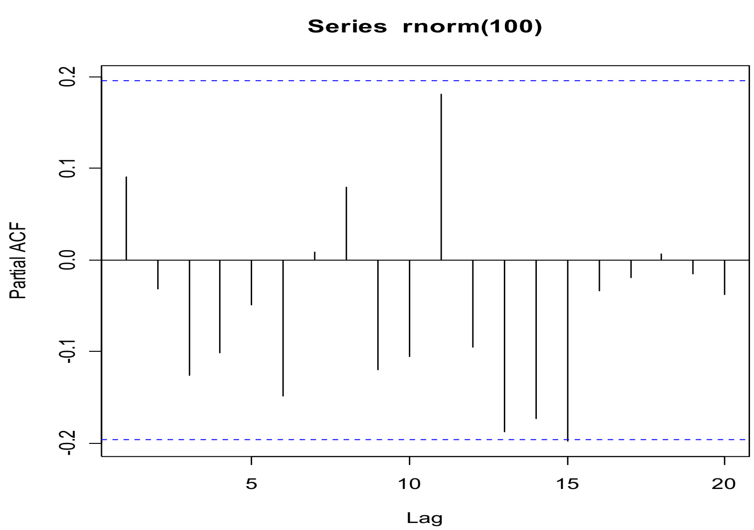
$$\phi_{33}=\frac{\det\begin{bmatrix}1 & \rho_1 & \rho_1\\ \rho_1 & 1 & \rho_2\\ \rho_2 & \rho_1 & \rho_3\end{bmatrix}}{\det\begin{bmatrix}1 & \rho_1 & \rho_2\\ \rho_1 & 1 & \rho_1\\ \rho_2 & \rho_1 & 1\end{bmatrix}}=\frac{\det\begin{bmatrix}1 & \rho_1 & \phi_1+\phi_2\rho_1\\ \rho_1 & 1 & \phi_1\rho_1+\phi_2\\ \rho_2 & \rho_1 & \phi_1\rho_2+\phi_2\rho_1\end{bmatrix}}{\det\begin{bmatrix}1 & \rho_1 & \rho_2\\ \rho_1 & 1 & \rho_1\\ \rho_2 & \rho_1 & 1\end{bmatrix}}=0 \quad \phi_{kk}=0,\ k\geq 3$$

This example show that for an AR(p) model, PACF at lag k equals zero if k is greater than p, where k and p are integers.

Distribution of sample PACF

- Under the hypothesis that the underlying process is white noise sequence, sample PACF are normally distributed with $var(\widehat{\phi_{kk}}) = 1/n$ asymptotically.
- Hence, $\pm {}^{2}/_{\sqrt{n}}$ can be used as critical limits (95% confidence level) to test for the hypothesis of a white noise process.

Sample autocorrelation function of a simulated NID sequences



Sample partial autocorrelation function

Causal and invertible processes (Brockwell and Davis, 1992)

Causal process:	$\{X_t\}$ can be expressed in terms of $\{a_s, s \leq t\}$. Such processes are called causal or future-independent autoregressive process.
Invertible process:	No restrictions on the $\{\theta_i\}$ are required for a (finite-order) MA process to be stationary. The imposition of the invertible condition ensures that there is a unique MA process for a given set of ACF.

Duality between AR and MA processes

- A finite-order stationary AR(p) process corresponds to a MA(∞) process, and a finite-order invertible MA(q) corresponds to an AR process of infinite-order .

Casual/stationary AR(p) ➔ MA(∞)

Invertible MA(q)➔ AR(∞)

- This dual relationship also exists in autocorrelation and partial autocorrelation functions.

Remarks on invertibility

- Consider the following first-order MA processes:

$$\text{A:}\quad X_t = a_t + \theta a_{t-1} \qquad\qquad \text{A:}\quad a_t = X_t - \theta X_{t-1} - \theta^2 X_{t-2} - \cdots,\ \text{if } |\theta| < 1$$

$$\text{B:}\quad X_t = a_t + \frac{1}{\theta} a_{t\text{-}1} \qquad\qquad \text{B:}\quad a_t = X_t - \frac{1}{\theta} X_{t\text{-}1} - \frac{1}{\theta^2} X_{t-2} - \cdots,\ \text{if } |\theta| > 1$$

$$\begin{aligned}
\gamma_A(0) &= \text{var}(a_t + \theta a_{t-1}) = (1+\theta^2)\sigma^2 \\
\gamma_A(1) &= \text{cov}(a_t + \theta a_{t-1}, a_{t+1} + \theta a_t) = \theta\sigma^2 \\
\rho_A(1) &= \gamma_A(1)/\gamma_A(0) = \theta/(1+\theta^2)
\end{aligned}$$

$$\begin{aligned}
\gamma_B(0) &= \text{var}(a_t + \frac{1}{\theta} a_{t-1}) = (1+\frac{1}{\theta^2})\sigma^2 \\
\gamma_B(1) &= \text{cov}(a_t + \frac{1}{\theta} a_{t-1}, a_{t+1} + \frac{1}{\theta} a_t) = \frac{1}{\theta}\sigma^2 \\
\rho_B(1) &= \frac{\frac{1}{\theta}\sigma^2}{(1+\frac{1}{\theta^2})\sigma^2} = \theta/(1+\theta^2)
\end{aligned}$$

Two different models possess the same autocorrelation functions

Causal ARMA processes

- An ARMA(p,q) process, defined by the equation $\Phi(B)X_t = \Theta(B)a_t$, is said to be causal if there exists a sequence of constants $\{\psi_j\}$ such that $\sum_{j=0}^{\infty}|\psi_j| < \infty$ and $X_t = \sum_{j=0}^{\infty} \psi_j a_{t-j}, t = 0, \pm 1, \pm 2, \ldots$
- In compact notation, we have

$$\Psi(z) = \sum_{j=0}^{\infty} \psi_j \cdot z^j = \frac{\Theta(z)}{\Phi(z)}, \quad |z| \leq 1$$

Show how to calculate $\{\psi_j\}$ in class

Invertible ARMA models

- An $ARMA(p,q)$ process is said to be invertible if there exists a sequence of constants $\{\pi_j\}$ such that $\sum_{j=0}^{\infty}|\pi_j| < \infty$ and

$$a_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}, t = 0, \pm 1, \pm 2, \ldots$$

$$\Pi(z) = \sum_{j=0}^{\infty} \pi_j \cdot z^j = \frac{\Phi(z)}{\Theta(z)}, \quad |z| \leq 1$$

Show how to calculate $\{\pi_j\}$ in class

Wold Decomposition

- Any zero-mean process $\{X_t\}$ which is not *deterministic* can be expressed as a sum of $X_t = U_t + V_t$, where $\{U_t\}$ denotes an $MA(\infty)$ process and $\{V_t\}$ is a deterministic process which is uncorrelated with $\{U_t\}$.
  - If the values $X_{n+j}, j \geq 1$ of the process $\{X_t, t = 0, \pm 1, \pm 2, ...\}$ were perfectly predictable in terms of $\mu_n = sp\{X_t, -\infty < t \leq n\}$. Such processes are called *deterministic*.
  - If X_n comes from a deterministic process, it can be predicted (or determined) by its past observations of the process, i.e., $X_t, t < n.$

Model identification

32

The first stage of Box-Jenkins analysis

Review theoretical ACFs and PACFs of ARMA models

Model identification using ACF and PACF with R examples

MA(1) processes

- Autocorrelation functions

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{\theta\sigma^2}{(1+\theta^2)\sigma^2} = \frac{\theta}{1+\theta^2}$$

$$\rho_j = 0 \qquad j > 1$$

- Partial autocorrelation functions

$$\phi_{11} = \rho_1 = \frac{\theta}{1+\theta^2} = \frac{\theta(1-\theta^2)}{1-\theta^4}$$

$$\phi_{22} = -\frac{\rho_1^2}{1-\rho_1^2} = -\frac{\theta^2}{1+\theta^2+\theta^4} = -\frac{\theta^2(1-\theta^2)}{1-\theta^6}$$

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

MA(q) processes

- The ACF of MA(q) processes,

$$X_t = a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \cdots + \theta_q a_{t-q},$$

cut off after lag q, i.e.

$$\rho_k = \begin{cases} \dfrac{\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \cdots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \cdots + \theta_q^2} & , k = 1, \ldots, q \\ 0 & , k > q \end{cases}$$

- The maximum lag of the non-zero sample autocorrelation is a good indicator of the MA(q) processes.

AR(1) processes

- Autocovariance and Autocorrelation:

$$\gamma_j = \phi\gamma_{j-1} \quad j \geq 1$$

$$\rho_j = \frac{\gamma_j}{\gamma_0} = \frac{\phi\gamma_{j-1}}{\gamma_0} = \phi\rho_{j-1}, \qquad j \geq 1$$

- Partial autocorrelation functions:

$$\phi_{11} = \rho_1 = \phi$$

$$\phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1-\rho_1^2} = \frac{\phi^2 - \phi^2}{1-\rho_1^2} = 0$$

$$\boxed{\phi_{kk} = 0 \quad k \geq 2}$$

AR(P) processes

$$X_t = \mu + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots\ldots\phi_p X_{t-p} + a_t$$

Stationarity condition is that all *p* roots of the characteristic equation outside of the unit circle

ACF

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \ldots..\phi_p\rho_{k-p}$$

$$\left.\begin{aligned}
\rho_1 &= \phi_1\rho_0 + \phi_2\rho_1 + \ldots..\phi_p\rho_{p-1} \\
\rho_2 &= \phi_1\rho_{11} + \phi_2\rho_0 + \ldots..\phi_p\rho_{p-2} \\
&\vdots \\
\rho_p &= \phi_1\rho_{p-1} + \phi_2\rho_{p-2} + \ldots..\phi_p\rho_0
\end{aligned}\right\}$$

System to solve for the first *p* autocorrelations: *p* unknowns and *p* equations

ACFs decay as mixture of exponentials and/or damped sine waves--depending on real/complex roots

PACF

$$\phi_{kk} = 0 \text{ for } k > p$$

TABLE 6.1(Wei, 2005)

Characteristics of theoretical ACF and PACF for stationary processes

Process	ACF	PACF
AR(p)	Tails off as exponential decay or damped sine wave	Cuts off after lag p
MA(q)	Cuts off after lag q	Tails off as exponential decay or damped sine wave
ARMA(p,q)	Tails off after lag (q-p)	Tails off after lag (p-q)

Model adequacy

The Final stage of Box-Jenkins approach

This stage is also called model diagnostic checking which involves techniques like over-fitting, residual plots and, more importantly, checking if residuals are approximately uncorrelated.

Why residuals are uncorrelated?

- The residuals of a fitted $ARMA(p, q)$ model is

$$\hat{a}_t = X_t - \hat{\phi}_1 X_{t-1} - \cdots - \hat{\phi}_p X_{t-p} - \hat{\theta}_1 a_{t-1} - \cdots - \hat{\theta}_q a_{t-q},$$

where $\hat{\phi}_k$, $\hat{\theta}_k$ $\forall k$ are the parameter estimates obtained from the second stage, and $\{\hat{a}_t\}$ are the residuals of the fitted model

  - Residuals can be seen as the sample estimates of $\{\hat{a}_t\}$ and therefore are approximately uncorrelated (white noise) because of the estimation process.
  - Remark: residuals are not independent in the classical regression model

Autocorrelation among residuals

- Residual autocorrelation functions at lag *k*

$$\hat{\rho}_k = \frac{\sum_{t=1}^{n-k} \hat{a}_t \hat{a}_{t+k}}{\sum_{t=1}^{n} \hat{a}_t^2}$$

- How many lags are enough??
- The overall tests that check an entire group of residual autocorrelation functions (assuming that the model is adequate) are called portmanteau tests.
- In spirits, portmanteau tests may be seen as a variant of the goodness of fit tests.

Popular portmanteau tests

Box and Pierce (1970)

$$Q_{BP} = n \cdot \sum_{1}^{m} \hat{\rho}_k^2 \sim \chi^2_{m-(p+q)}$$

Ljung and Box (1978)

$$Q_{LB} = \sum_{1}^{m} \frac{n \cdot (n+2)}{(n-k)} \hat{\rho}_k^2 \sim \chi^2_{m-(p+q)}$$

Example: Li (2004), page 11

k	1	2	3	4	5	6	7	8	9	10
	$\hat{\rho}_k$ .4	.15	.07	.06	.09	.03	.05	.06	.5	.01

- $X_t = (1-0.4B)a_t$ was fitted to a series of 80 observations.

$$Q_{BP} = 80(.4^2 + .15^2 + \dots + .01^2) = 16.696$$

$$Q_{LB} = 80(82)(.4^2/79 + .15^2/78 + \dots + .01^2/70) = 17.488$$

- The upper *5%* critical value from the chi-squared distribution with *9*(=*10*-*1*) degrees of freedom is 16.92.

More about portmanteau tests

- Pros:
  - Practical purposes
  - Minimal requirement for using the fitted model
- Cons:
  - Lack power if comparing with traditional statistical tests, such as likelihood ratio tests
- Possible improvements and other applications:
  - 1. Finite sample adjustments
  - 2. Complicated functional of residual autocorrelations
  - 3. Monte Carlo test: See Lin & McLeod(2006)
  - 4. Other applications: portmanteau tests for randomness and ARMA models with infinite variance innovations

Model selection

In time series analysis, several models may adequately represent a given data set.

How to select the best model among these candidates is called model selection or order selection.

Methods for model selection

- Two of the most popular methods are Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC):

$$AIC = -2 \log ML + 2\,k, \qquad \text{(a)}$$

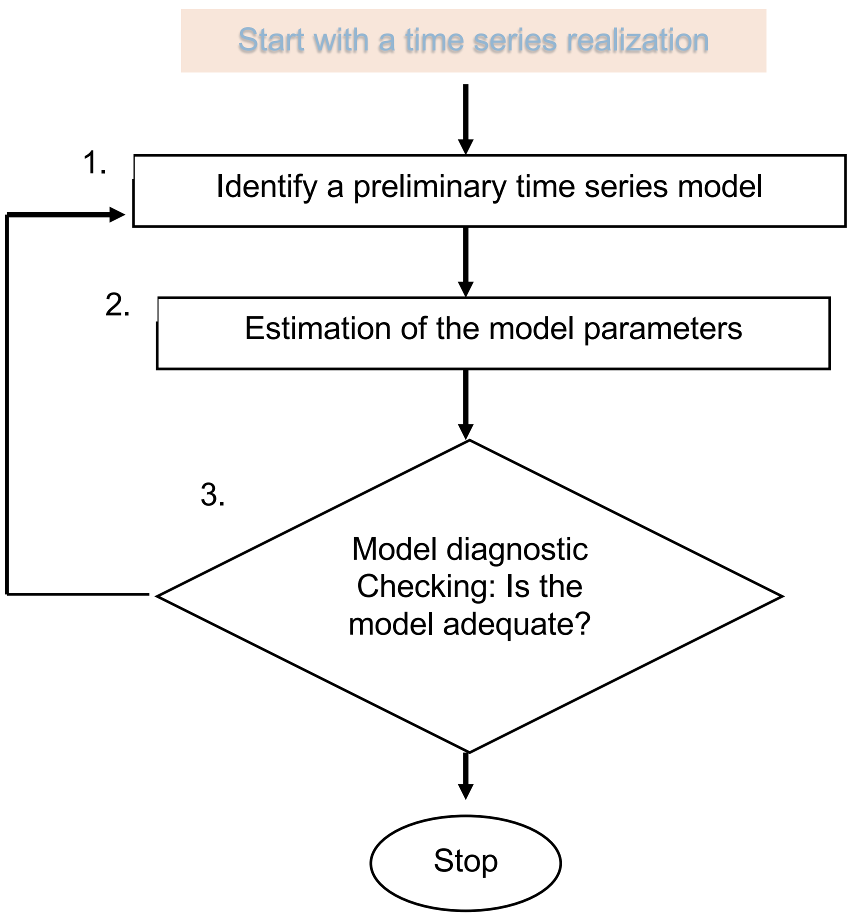
$$BIC = -2 \log ML + k \log(n), \qquad \text{(b)}$$

  where ML denotes maximum likelihood, log ML is the value of maximized log-likelihood function for a model fitted to a give data set, and k is the number of independently adjusted parameters within the model.

- *Remark*: BIC puts more penalties on the number of parameters used by fitted models, and some empirical studies indicate that the model selected by BIC performs better in the post-sample analysis, such as forecasting.

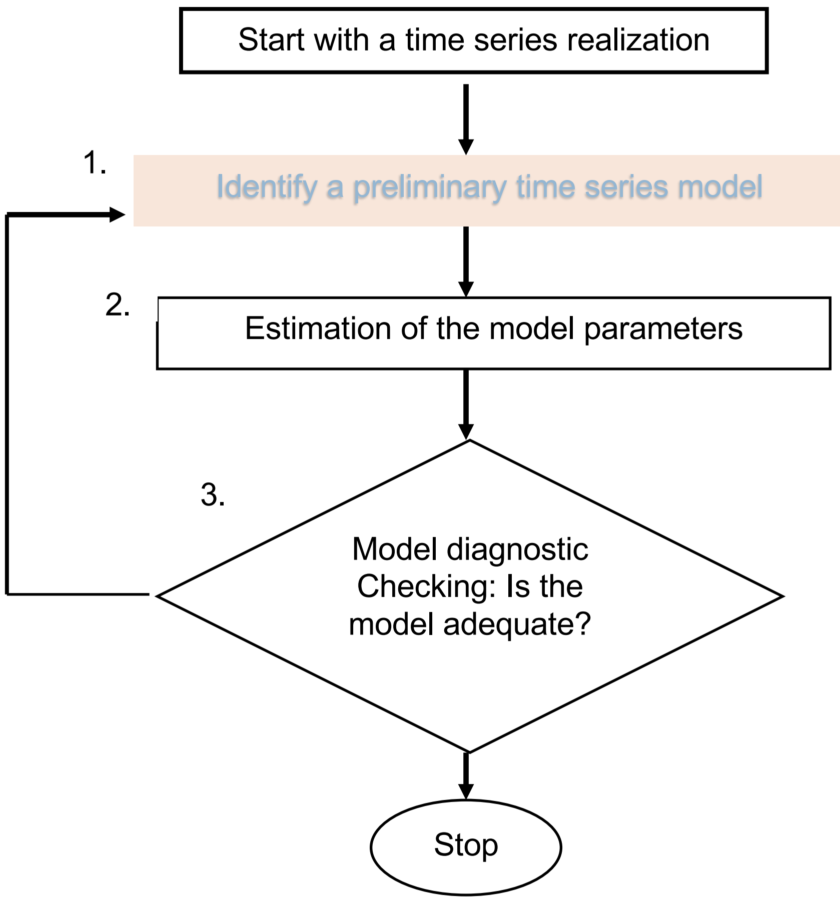
Three stage of box-Jenkins approach

Three stages of Box-Jenkins Approach



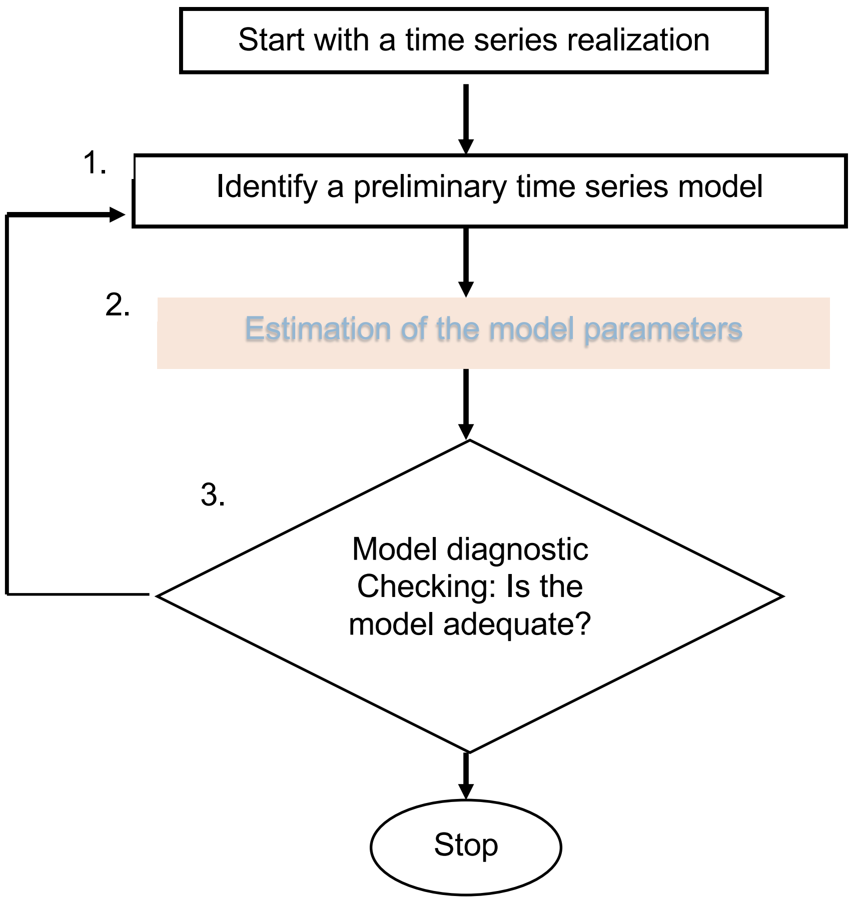
- Understand the problem of interest
- Collect data
- Plot time series data

First stage—model identification



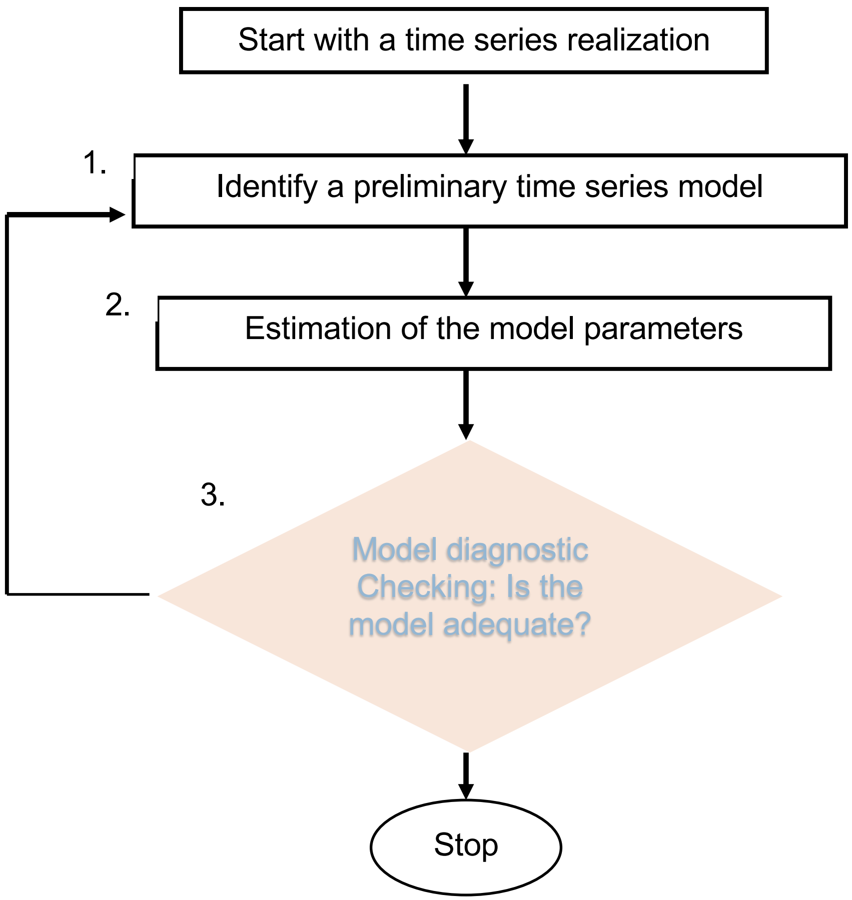
- Perform differencing and transformations to transform the data into stationarity.
- Identify preliminary ARMA(p,q) models using sample autocorrelations and sample partial autocorrelations

Second stage—model estimation



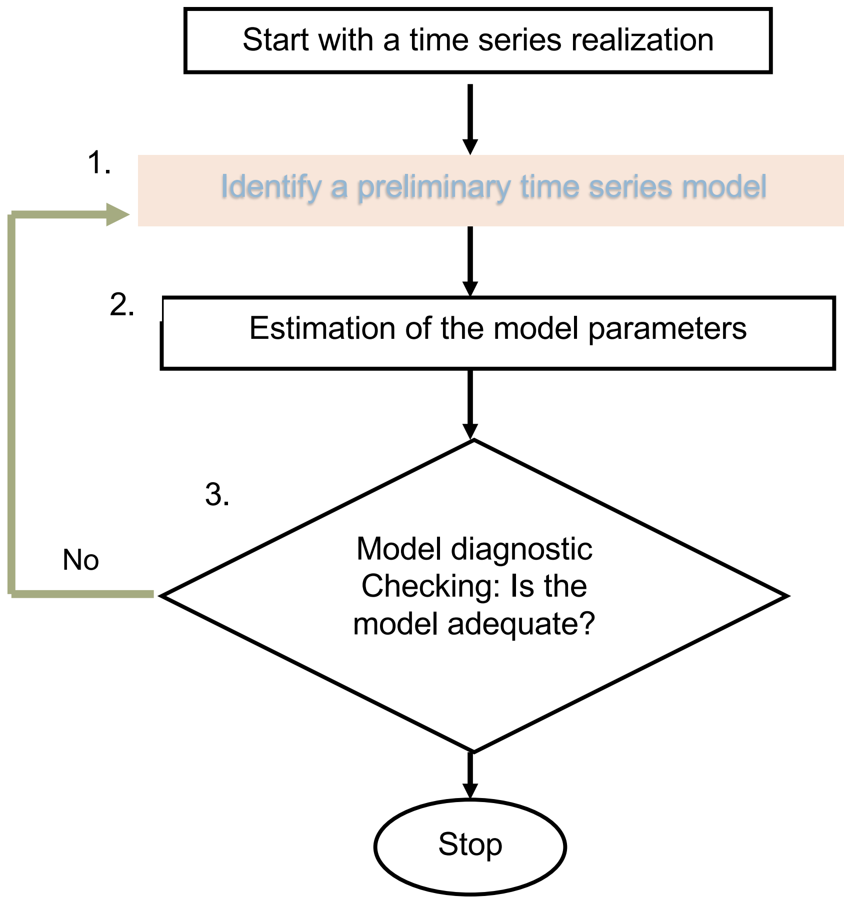
- Method of moments
- Maximum likelihood Estimation
- Kalman Filter
- Others
- To discuss latter in this course

Third stage—model evaluation



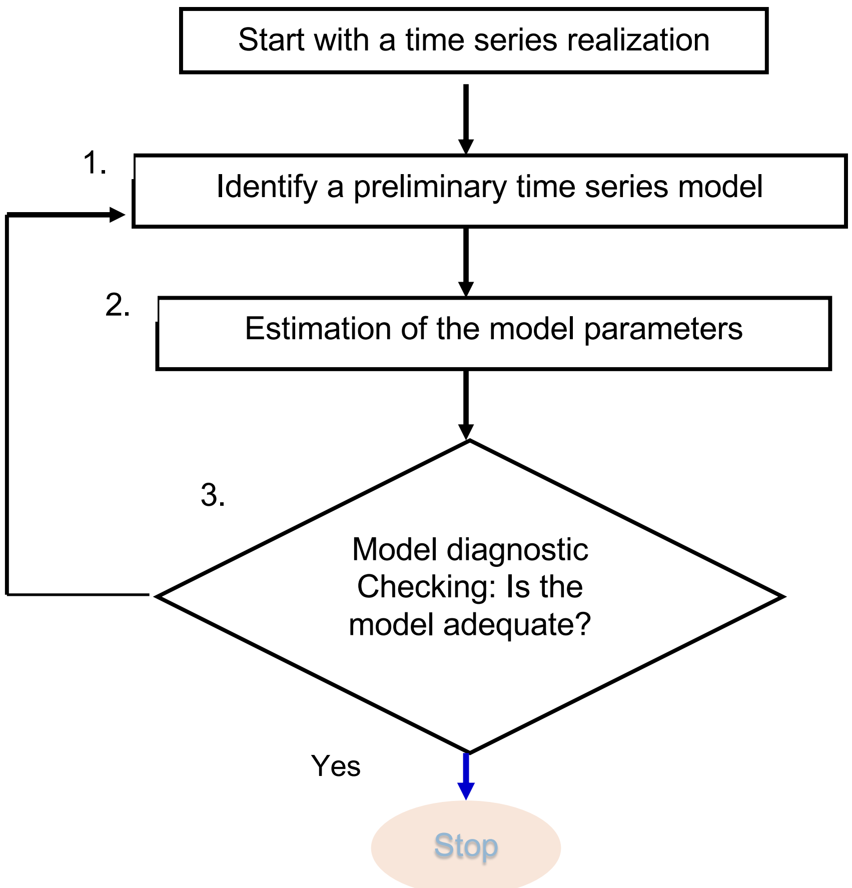
- Model adequacy is checked by examining if the residuals of the fitted model are approximately uncorrelated (after taking into account the effect of estimation)

Model is not adequate



- The fitted model fails diagnostic checks.
- Return to the first stage and identify another time series model.

Model is adequate



- If the fitted model passes diagnostic checks, we may the use the model for our analysis.