Non-Stationary Time Series - ARIMA

STAT 1321/2320

Kiran Nihlani

Department of Statistics
University of Pittsburgh

Outline

Assessing Non-Stationarity - Graphical Inspection

- The nonstationarity can sometimes be seen from a regular plot of the time series: for example, if we can see that the mean or the variance changes over time.
- Sometimes the ACF plot can reveal nonstationarity as well.
  - With a nonstationary series, the ACF typically does not die off quickly as the lag increases.

The Dickey-Fuller Unit Root Test

- There are two versions of the test that can be used in appropriate situations.
- Version 1:

$$H_0 : \text{Difference Stationary} \quad \text{vs.} \quad H_a : \text{Stationary}$$

- Use adf.test(x, alternative="stationary") from the **tseries** package.
- Version 2:

$$H_0 : \text{Trend Stationary} \quad \text{vs.} \quad H_a : \text{Stationary}$$

- Use adfTest(x, lags, type="ct") from the **fUnitRoots** package. Specify lags as the maximum number of lags suggested for the AR model based on AIC.

Other Tests

Ljung-Box Test

- The test may be used to test for autocorrelation in the dataset.

$$H_0 : \text{White Noise/No autocorrelation} \quad \text{vs.} \quad H_a : \text{Autocorrelation}$$

Shapiro-Wilks Test

- The test may be used to test for normality of data.

$$H_0 : \text{Normality} \quad \text{vs.} \quad H_a : \text{Non-normality}$$

ARIMA (p,d,q)

- We have seen that many real time series exhibit non-stationary behavior.
- For these, ARIMA would be a better model than ARMA-type models.
- An ARIMA model has 3 orders - p, d, q
  - d is the order of differencing required to coerce the time series into stationarity.
  - The differenced series then has an ARMA(p, q) model.

$$\nabla^d X_t = ARMA(p, q)$$

- Differencing of orders 1 and 2 may be used to handle linear and quadratic trends, respectively.

ARIMA Process using Backshift Operators

- An ARIMA(p,d,q) model can be written using the backshift operators as:

$$\phi(B)(1-B)^d X_t = \phi(B)\nabla^d X_t = \theta(B)W_t$$

where

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 + \cdots \phi_p B^p \qquad \text{and} \qquad \theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots \theta_q B^q$$

auto.arima() function

- The **auto.arima** function may be used to find the best ARIMA model based on AIC.
- By default, it uses a stepwise model selection approach to make the search faster (use **trace=T** argument to print the models tried at each step).
- The function considers values of p and q up to 5 and values of d up to 2 but these values can be adjusted.

Comparisons between ARIMA and ARMA Models

- The comparison between an ARMA(p,q) model with $d = 0$ and ARIMA(p,d,q) with $d > 0$ using AIC is not recommended (why?).
- The correct amount of differencing should be chosen first, and then AIC/BIC can be used to guide the choices of p and/or q.
- You may try differencing once or twice and test the differenced series at each step for stationarity using the ADF test.
- Differencing an already stationary series will always give a stationary series. This is overdifferencing and should not be done.

Seasonal ARIMA

- In practice, many time series contain a seasonal periodic component, which repeats every s observations.
  - For example, with monthly observations, where $s = 12$, we may typically expect X_t to depend on values at annual lags, such as X_{t-12}, and perhaps X_{t-24}.
  - We may also see non-seasonal dependence on values for X_{t-1} and X_{t-2}.
- We may need a seasonal ARIMA (SARIMA) model in such cases.
- The order of a SARIMA model is specified as $\underbrace{(p, d, q)}_{\text{non-seasonal.}} \times \underbrace{(P, D, Q)}_{\text{seasonal.}}{}_s$

seas. period.

ARMA : stationary (p, q)

ARIMA

↳ Non - stationary. (difference).

(p, d, q)

trend $\rightarrow \nabla^d X_t \equiv ARMA(p, q)$.

seasonality : $X_t - X_{t-s}$

s: seasonality period.

$$X_t = \phi X_{t-1} + w_t. \qquad AR(1)$$

$$X_t = \Phi X_{t-12} + w_t. \qquad SAR(1).$$

$s = 12$. 1 seasonal lag.

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \Phi X_{t-12}$$

$$X_t = w_t + \theta w_{t-1} \qquad MA(1).$$

$$X_t = w_t + \Theta w_{t-12} \qquad SMA(1)$$

$s = 12$.

SARIMA.

Understanding SARIMA Models

- The order of a SARIMA model is specified as $(p, d, q) \times (P, D, Q)_s$
- This implies that the twice differenced time series

ARMA (p, q) x (P, Q)

$$\nabla^d \nabla_s^D X_t$$

is ARMA$(p, q) \times (P, Q)_s$.

- p, q are known as non-seasonal AR and MA lags, respectively.
- P, Q are known as seasonal AR and MA lags, respectively.
- s is the period of the data.
- d and D are the order of non-seasonal and seasonal differencing, respectively.

SARIMA - Examples

$X_t = X_{t-1} + W_t$: random walk.

- Consider a seasonal lag model

∇_{12}^{1}

$$X_t = X_{t-12} + W_t \implies (1 - B^{12})X_t = W_t$$

seasonal random walk.

seasonal differencing.

This is a SARIMA $(0,0,0) \times (0,1,0)_{12}$ model.

- Consider a seasonal $AR(1)_{12}$ model

$$X_t = \Phi X_{t-12} + W_t$$

$s = 12$

$D = 1$

This is a SARIMA $(0,0,0) \times (1,0,0)_{12}$ model

SAR(1)

SARIMA - Examples

- Consider a seasonal $MA(1)_{12}$ model

seasonal lag

seasonal char. polynomial

$$X_t = W_t + \Theta W_{t-12} \implies X_t = \underbrace{(1 + \Theta B^{12})}W_t$$

$\Theta(B^{12})$

This is an SARIMA $(0,0,0) \times (0,0,1)_{12}$ model.

D Θ - ∅ -

- Consider an SARIMA $(0,0,0) \times (0,1,1)_{12}$ model

$$\nabla_{12}^{1} X_t \rightarrow X_t - X_{t-12} = \underbrace{W_t + \Theta W_{t-12}}_{SMA(1)}$$

$$SARIMA(1,0,0)\times(0,1,1)_4.$$

$$s = 4, \quad D = 1$$

$$\underbrace{(1-\phi B)}_{AR(1)}\underbrace{(1-B^4)^1}_{\text{seasonal diff}} X_t = \underbrace{w_t + \Theta w_{t-4}}_{SMA(1)}$$

AR(1).

$$\underbrace{(1 - \phi B)}_{\Phi(B)} X_t = w_t .$$

SAR(1)

$$(1 - \Phi B^{12}) X_t = w_t .$$

$\Phi(B^{12})$

1.

$$\Phi(z) = 1 - \Phi z = 0 .$$

SARIMA using Backshift Operators

diff, AR : left.
MA : right.

- A SARIMA model can be written as:

y_t

$$\phi(B)\Phi_P(B^s)(1-\overset{\text{lag-}s}{B^s})^D(1-B)^d X_t = \Theta_Q(B^s)\theta(B)W_t$$

$$\phi(B)\Phi_P(B^s)\nabla_s^D\nabla^d X_t = \Theta_Q(B^s)\theta(B)W_t$$

- Example: An SARIMA $(1,1,1)\times(1,1,1)_4$ model can be written as

$$\underset{\text{SAR(1)}}{(1-\Phi B^4)}\underset{\text{AR(1)}}{(1-\phi B)}\nabla_4^1\nabla X_t = \underset{\text{MA(1)}}{(1+\theta B)}\underset{\text{SMA(1)}}{(1+\Theta B^4)}W_t$$

$$(1-\Phi B^4)(1-\phi B)(1-B^4)^1(1-B)X_t = (1+\theta B)(1+\Theta B^4)W_t$$

$$\phi(B)\Phi_P(B^s)\,y_t = \Theta_Q(B^s)\,\theta(B)\,W_t.$$

$$X_t = \Phi_1 X_{t-4} + \Phi_2 X_{t-8} + \Phi_3 X_{t-12} + w_t.$$

$$s = 4.$$

SAR(3) with $s = 4$.

Example

Interpret the SARIMA$(0,1,1) \times (0,1,1)_{12}$ model.

$$\nabla^{1}_{12} \underbrace{\nabla X_t}_{} = \Theta_1(B^{12}) \theta(B) w_t .$$

$$\nabla^{1}_{12} (X_t - X_{t-1})$$

$$X_t - X_{t-12} - X_{t-1} + X_{t-13} = (1 - \Theta B^{12})(1 - \theta B) w_t$$

$$= w_t - \theta w_{t-1} - \Theta w_{t-12} + \theta \Theta w_{t-13} .$$

R Codes

You may use

- auto.arima() to find the order of the model (forecast package).
- sarima.sim() to simulate a series (astsa package).
- arima(x, order=c(p,d,q), seasonal = list(order = c(P,D,Q), period = s)) to fit a model (stats package).
- sarima.for() to make forecasts (astsa package).