CS6301: R For Data Scientists

LECTURE 27: TIME SERIES II

Understanding Autocorrelation Function (ACF)

Recall Pearson Correlation Coefficient (calculated between two vectors):

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

In this case, we only have one vector, but we are looking at correlations between different points in time …

Autocorrelation Function (ACF)

$$r_k = \frac{\sum_{t=k+1}^{T} (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^{T} (y_t - \bar{y})^2},$$

Checking correlation between the y_is and themselves – *autocorrelation.*

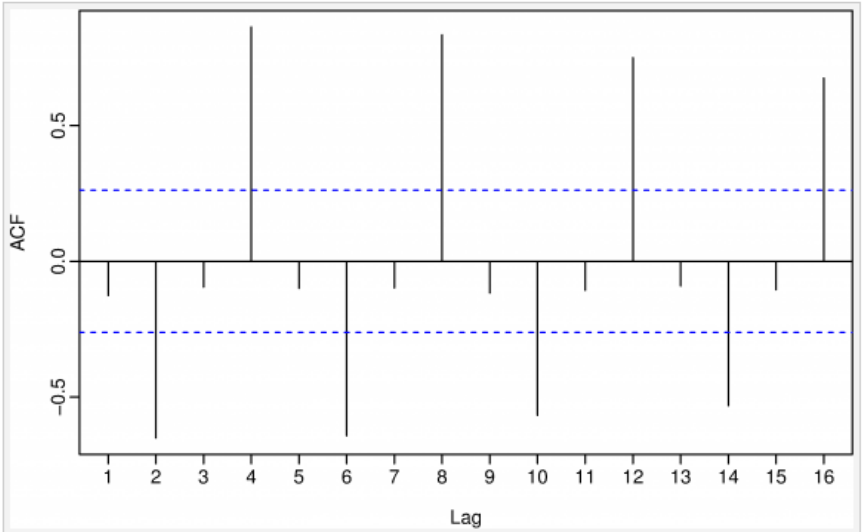


Figure 2.10: Autocorrelation function of quarterly beer production

Partial Autocorrelation Function (PACF)

If y_t and y_{t-1} are correlated, then so are y_{t-1} and y_{t-2}

But now y_t and y_{t-2} are correlated …

To overcome this problem, use partial ACF, which only measure correlation between y_t and y_{t-k}, after removing effects of other lags y_{t-1}, y_{t-2}, …, y_{t-k-1}

Stationary

A time series is **stationary** if its properties do not depend on the time it is observed

- No trend or seasonality
- Can be cyclic – cycles do not have fixed length

No predictable pattern in the long term

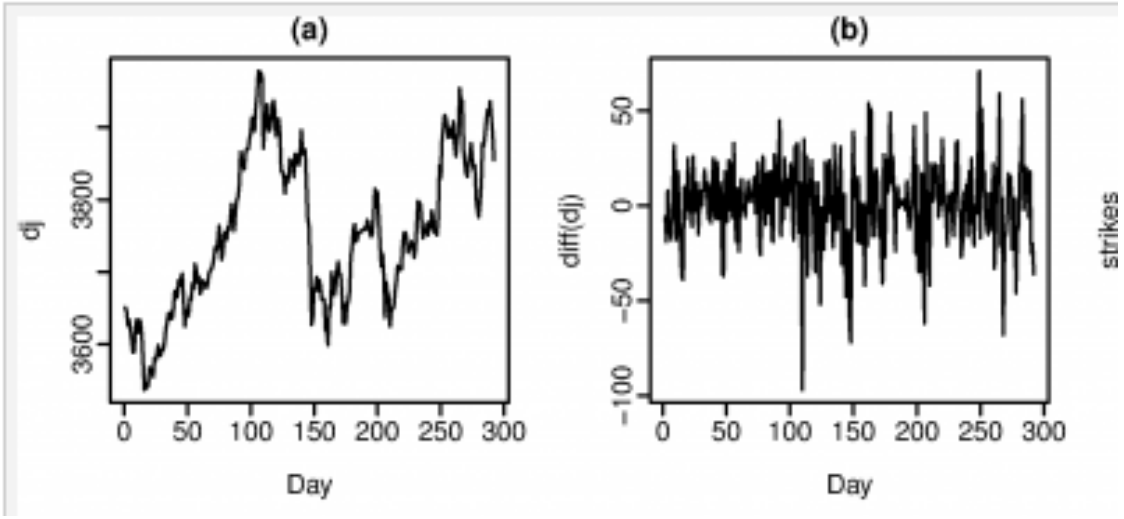
Plot will be basically horizontal, with constant variance

Dickey Fuller test: A hypothesis test on whether the time series is stationary

Differencing

Can transform a nonstationary time series to a stationary one by differencing

- This means taking the difference between successive observations



AR and MA Models

We will now look at a different way of modeling time series

We will be interested in seeing how shocks or errors impact future values of our variable

For this approach, we will assume that the data has been differenced to make it stationary

Auto Regression (AR) and Moving Average (MA) are two common approaches

- Note “Moving Average” is a poorly chosen name – this is different than the SMA we saw earlier

Autoregression

Regression forecasts a value for the variable of interest using the observed predictor values

In autoregression, we forecast the value of the variable using past values of the variable itself (regress the variable onto itself)

This is just standard linear regression like we did before – use previous values of *y* as predictors

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + e_t,$$

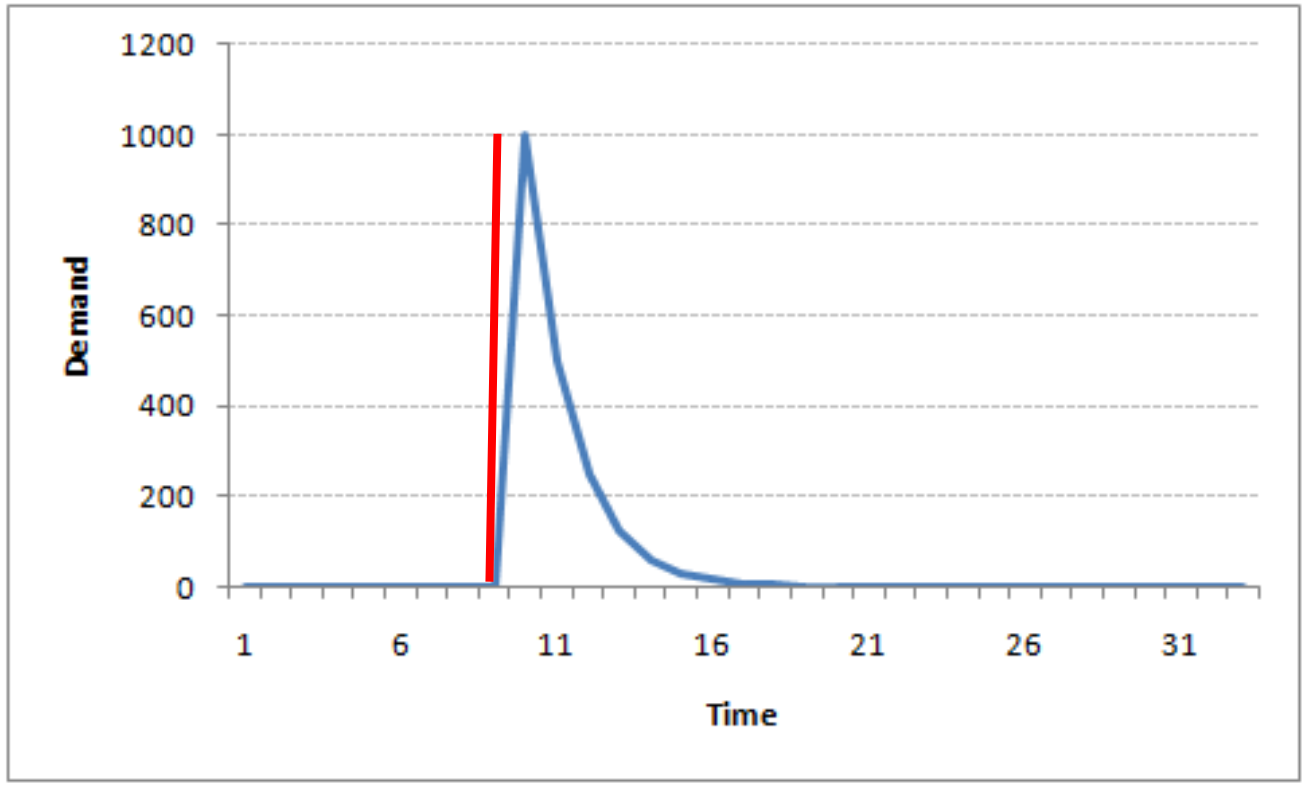
AR Models

The order of the model is a measure of how far in the past the last term is …

In the previous example, this is *p*

Key property: Shocks to the system (major changes in y_t) decay more slowly

- This shock could be caused by big e_t at some moment t_0
- Suppose we have an AR(1) model, i.e. $y_{t+1} = 0.5 * y_t + e_t$
- Then a big change at $t = t_0$ will die off exponentially in future y_t, but in theory never completely die



Shock e_t

Value y_t

More On AR Models

For these types of processes, we would to see a decaying ACF …

Current values are correlated to past values, but with a decaying level

Looking at the ACF may give an idea of how far back we need to go (what order to use)

Software can analyze and give suggestions

Moving Average Process

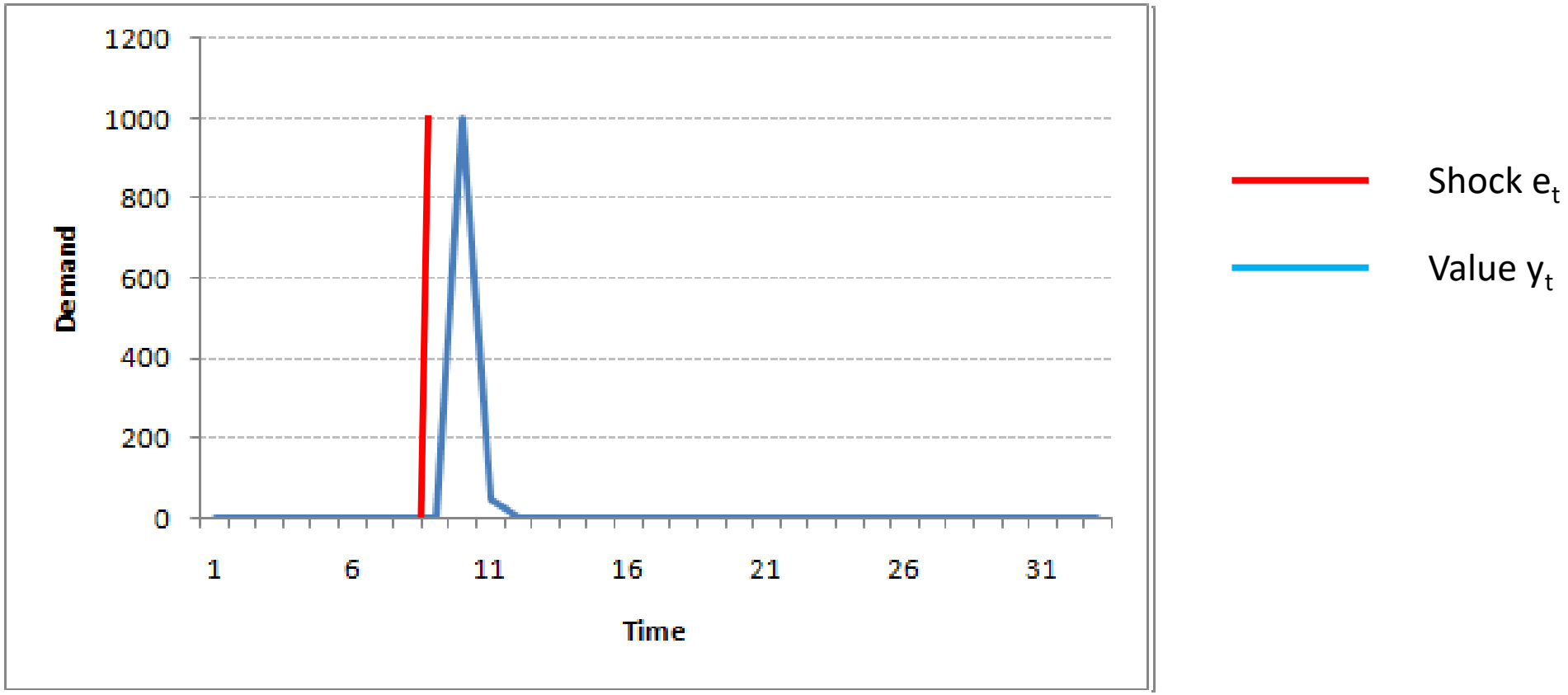
Moving Average models forecast the future value of the variable using past forecast errors/shocks:

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

Recall we are assuming we have difference the series to be stationary

- If we do enough differencing, *c* is zero

For these models, shocks tend to die much more quickly – once outside the “window”


1200
1000
800
600
400
200
0
Demand
1
6
11
16
21
26
31
Time
Shock e_t
Value y_t

AR Versus MA

AR: A hot new service is introduced, and people rush to sign up. Eventually the buzz dies down, and the number of people using the service (which is a function of the number using the service the previous time periods) dies down

MA: A hot new product is introduced and people rush to buy it, leaving a shortage – many people who want the product don't get it (call this the error). People who want the product must get it later, but eventually the demand dies down as the product loses popularity

AR Versus MA – Take Two

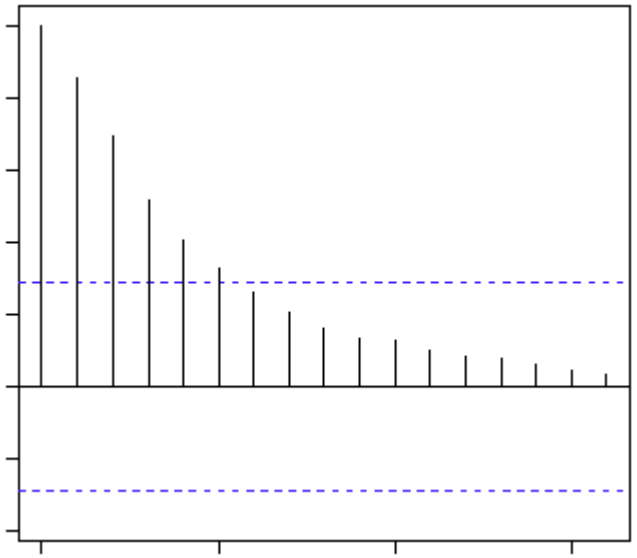
AR: Look at correlation of y_t to previous y's, should see gradual decline after *p* (order of model), but not sudden drop off

- Correlation "propagates back"
- Also, PACF should only show a short spike

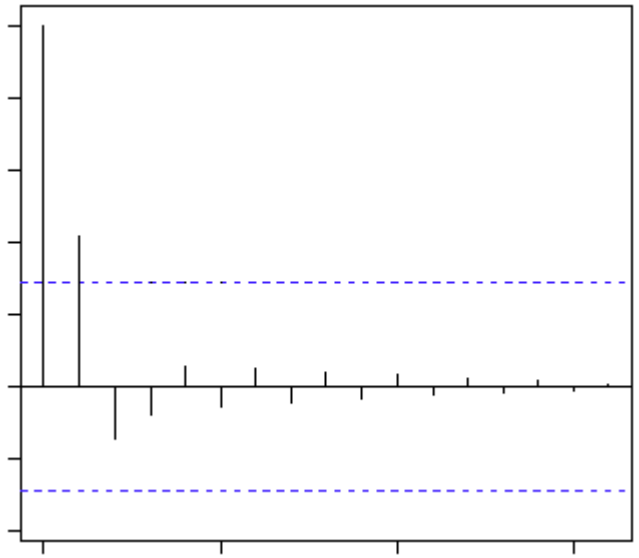
MA: No correlation beyond *p* (order of the model) – sudden drop off in ACF

- PACF tends to decay slowly

We can use the ACF and PACF graphs to see which model should be used

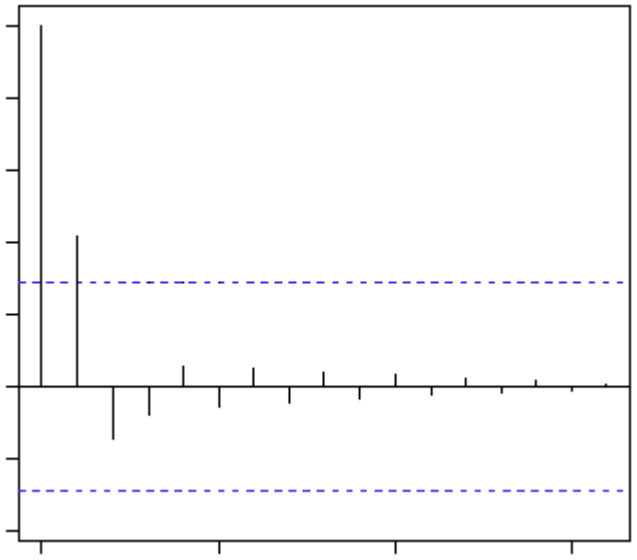


ACF

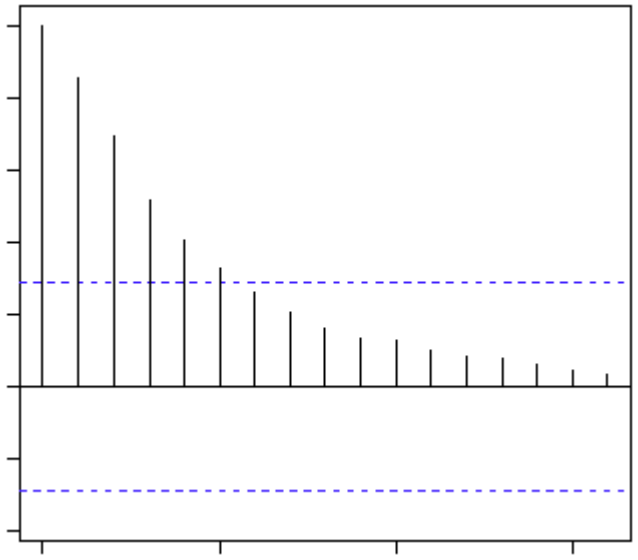


PACF

Probably an AR(2) process



ACF



PACF

Probably an MA(2) process

ARIMA

ARIMA stands for AutoRegressive Integrated Moving Average

Also called ARMA

Combine AR and MA models with differencing, get nonseasonal ARIMA models

y_t' = differenced values, may have been differenced more than once

Refer to as ARIMA(p,d,q), d is number of times differencing is applied

$$y_t' = c + \phi_1 y_{t-1}' + \cdots + \phi_p y_{t-p}' + \theta_1 e_{t-1} + \cdots + \theta_q e_{t-q} + e_t,$$

ARIMA – Simple Example

Let y_t represent sales of an item, differenced so the mean is zero

Let e_t represent the introduction of a sales coupon at time t

A simple ARIMA model could be

$$y_t = 0.9y_{t-1} + e_t + 0.2e_{t-1}$$

The AR part represents brand loyalty, the MA part the effect of a “shock”

If $y_0 = e_0 = 0$ and $e_1=1$, then

$Y_1 = 1$, $y_2 = 0.9 + 0.2 = 1.1$, $y_3 = .99$, $y_4 = .891$, etc.

Effect of shock is temporary, but lingers on in the AR part

ARIMA in R

R will do model selection for you, or you can input your own model parameters

Using the auto selection can be dangerous – should consider how to select the parameters p,d,q

- AIC and BiC can be used to help determine the order of the model (determined by the above parameters)

Once these are determined, R uses MLE to find the coefficients in the model

ARIMA - Configuring

Can use ACF and PACF to determine if AR(p) or MA(q) model is appropriate

Data may be ARIMA(p,d,0) if

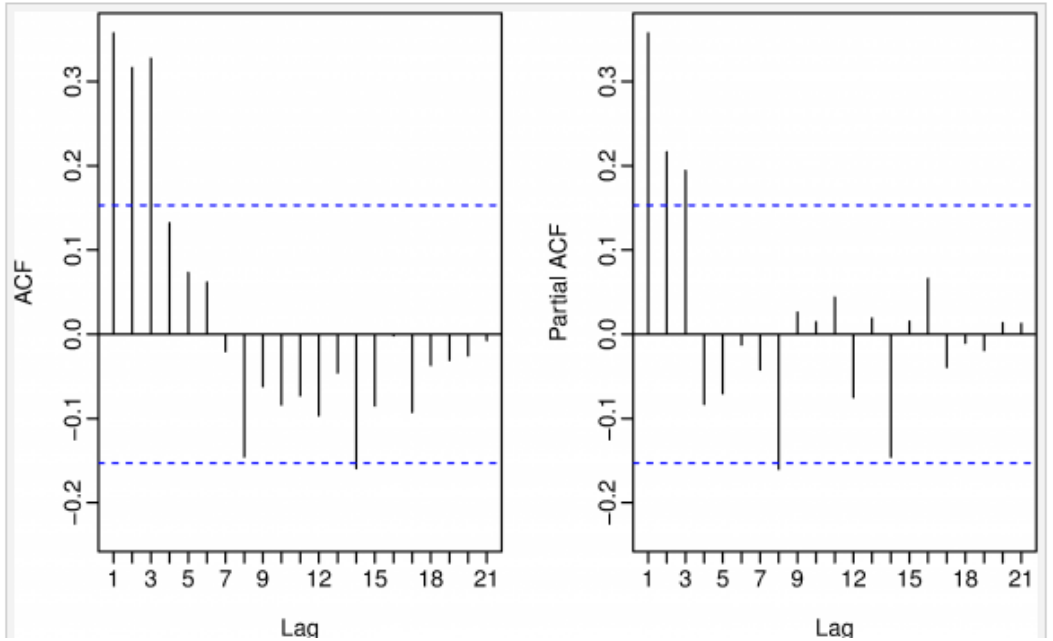
- ACF is exponentially decreasing or sinusoidal
- there is a significant spike at lag p in PACF, but none beyond lag p

Data may be ARIMA(0,d,q) if

- the PACF is exponentially decaying or sinusoidal
- there is a significant spike at lag q in ACF, but none beyond lag q

ARIMA - Configuring

Example: This would suggest an ARIMA (0,d,3) model, since PACF tends to decay exponentially and there are three significant spikes in the ACF and then no significant spikes thereafter



ARIMA Procedure

You can select your own model:

fit =Arima(usconsumption[,1], order=c(0,0,3))

Modelling procedure

When fitting an ARIMA model to a set of time series data, the following procedure provides a useful general approach.

1. Plot the data. Identify any unusual observations.
2. If necessary, transform the data (using a Box-Cox transformation) to stabilize the variance.
3. If the data are non-stationary: take first differences of the data until the data are stationary.
4. Examine the ACF/PACF: Is an AR(p) or MA(q) model appropriate?
5. Try your chosen model(s), and use the AICc to search for a better model.
6. Check the residuals from your chosen model by plotting the ACF of the residuals, and doing a portmanteau test of the residuals. If they do not look like white noise, try a modified model.
7. Once the residuals look like white noise, calculate forecasts.

ARIMA Procedure

Note the automated procedure will do steps 3-5

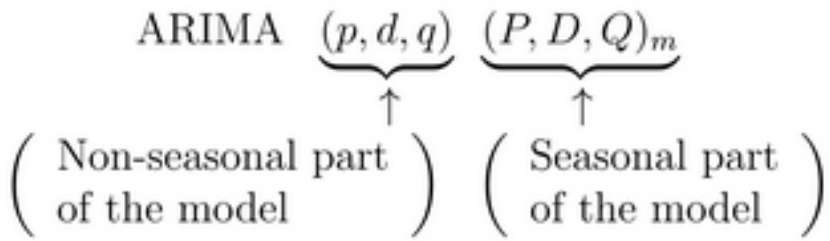
Even if you use automated selection, still recommended you do the remaining steps

How to tell is series is stationary?

- Dickey-Fuller Test
  - Assumption is $y_t = \rho y_{t-1} + noise$, or $y_t - y_{t-1} = (\rho - 1) y_{t-1} + noise$
  - DF does hypothesis test on $\rho = 1$

Seasonal ARIMA

Seasonality: Just add another ARIMA part for the seasonal model:



Still difference to make stationary

Evaluating Models

Akaike information criterion (AIC) is a popular metric

Used to estimate the quality of the model

Formal definition: Given a family of models with a number of estimated parameters *k* and a Maximum Value for the Likelihood function $\hat{L}$, we define

$$AiC = 2k - 2\ln(\hat{L})$$

Smaller numbers are better.

Links ...

http://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/

https://www.youtube.com/watch?v=lUhtcP2SUsg (MA)

https://www.youtube.com/watch?v=AN0a58F6cxA (AR)

https://www.youtube.com/watch?v=-xPDcd7WzhU (MA vs AR I)

https://www.youtube.com/watch?v=LHDqNaXZn9Q (MA vs AR II)

https://www.youtube.com/watch?v=R-oWTWdS1Jg (ACF vs PACF)

https://www.youtube.com/watch?v=Pg0RnP1uLVc (ARMA)