Time Series with ARIMAS

MSDS 7333

Brad Blanchard

bblanchard@smu.edu

World Changers Shaped Here

SMU®

ARIMAS

- **A**uto**R**egressive (p)
- **I**ntegrated (d)
- **M**oving **A**verage (q)

My my my, that's a lot of fancy words

- Yes
- **A**uto**R**egressive
  - Current prediction depends on the past results
  - Similar to a Markov chain, but can be longer than 1 step
- **I**ntegrated
  - Not only is the past result important, but the DIFFERENCES between the past results are important
- **M**oving **A**verage
  - The average of previous results is also important.

Where to start

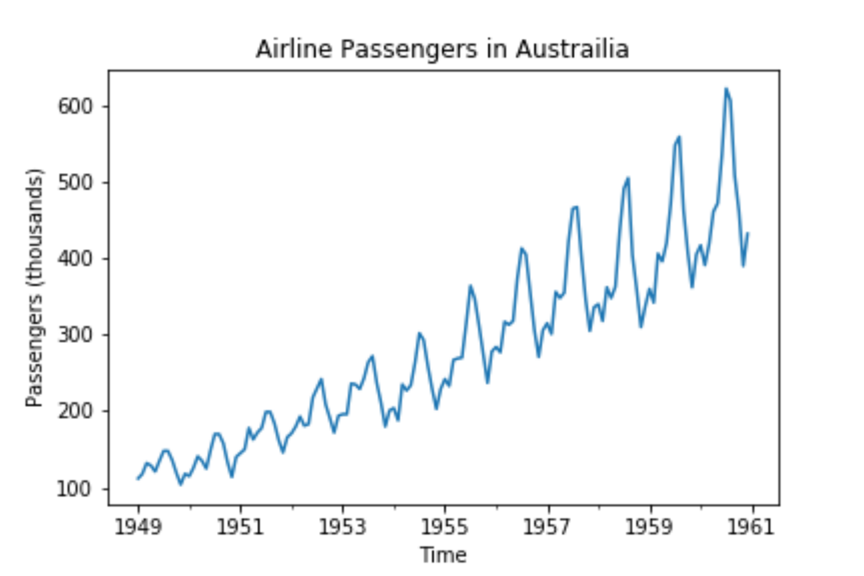
- ARIMA is really a group of models with three parameters:
- p: autoregressive part
- d: Integrated part
- q: Moving average part
- The trick is to find the right (p,d,q) combination to properly fit your model.

- Thanks Capt. Obvious!

Requirements

- Time Series
  - Linear time only (sorry Special Relativity)
- The series must be stationary
  - No trends
  - No seasonal patterns
    - Really? Yeah, really
  - This seems like a very strict requirement
    - Yep

So how do we model something like this?

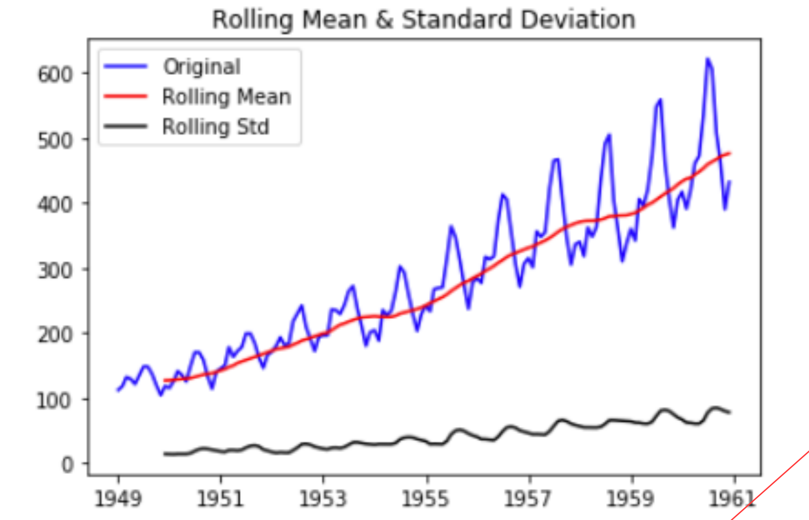


- It has a trend
- It has a seasonal Pattern
- Pretty sure <u>it is not</u> stationary

Make it stationary!

- What does stationary mean anyway
  - Separate lecture
  - Short answer: Apply Dickey-Fuller Test
    - The test looks to see if the rolling mean increases (or decreases) – which is a measure of stationarity
  - Shorter answer: No Trends!

Look at our Example



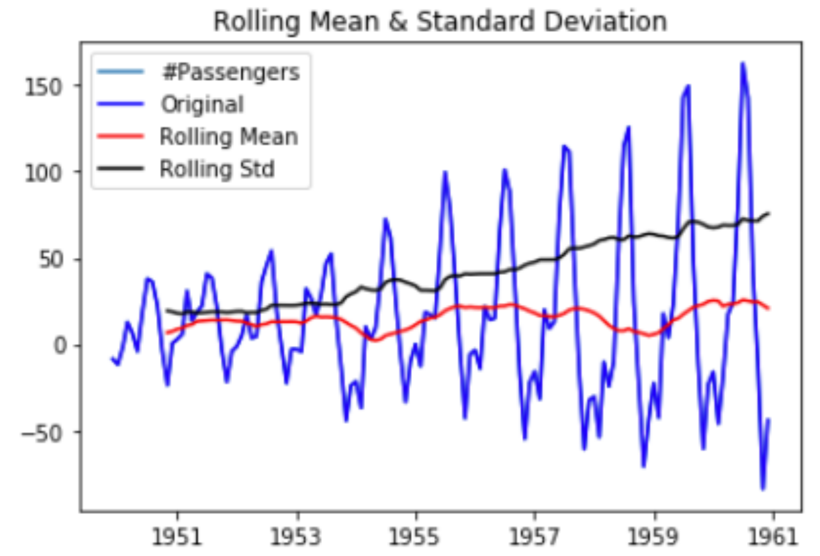

```
Results of Dickey-Fuller Test:
Test Statistic                   0.815369
p-value                          0.991880
#Lags Used                      13.000000
Number of Observations Used    130.000000
Critical Value (1%)             -3.481682
Critical Value (5%)             -2.884042
Critical Value (10%)            -2.578770
dtype: float64
```

Obvious Trend (NON STATIONARY!)

Test Statistic is greater than critical value(s)

'Lags' is just the number of values to compute rolling statistics. In this case 12(+1)

Let’s get rid of the mean

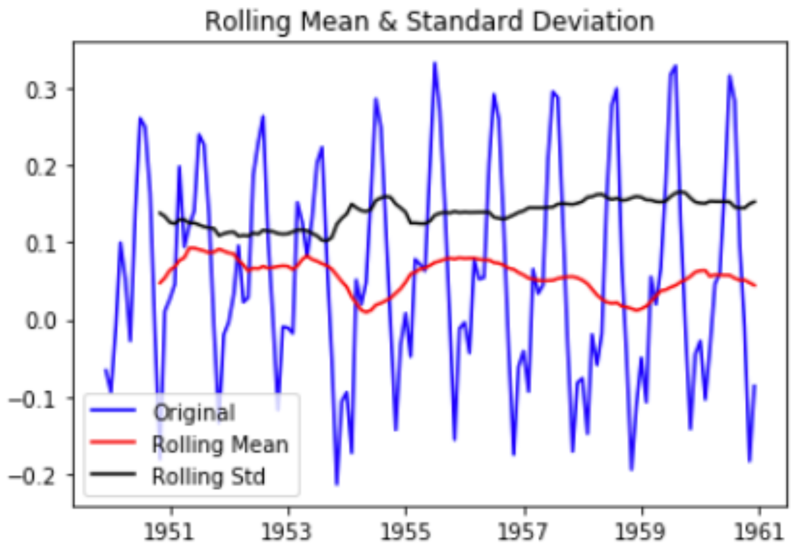



```
Results of Dickey-Fuller Test:
Test Statistic                  -3.164968
p-value                          0.022104
#Lags Used                      13.000000
Number of Observations Used    119.000000
Critical Value (1%)             -3.486535
Critical Value (5%)             -2.886151
Critical Value (10%)            -2.579896
```

Hmm there still seems to be an increasing pattern over time, but it passes the stationary test!!

DANGER WILL ROBINSON!!!

Much better after log transform




```
Results of Dickey-Fuller Test:
Test Statistic                 -3.162908
p-value                         0.022235
#Lags Used                     13.000000
Number of Observations Used   119.000000
Critical Value (1%)            -3.486535
Critical Value (5%)            -2.886151
Critical Value (10%)           -2.579896
```

Still need to deal with our seasonality!

Notice the stats didn't actually change!

Even though we've shifted the pattern from ever-spreading to truly a seasonal cycle (be careful when you do these transforms, we do them for visual reasons and not for statistical ones)

But why 12 for moving average?

- Seasonality to Data
- Using 5 doesn’t work as well (this is highly dependent on your data and its context)
- Using the rolling average is important

So we got the series stationary now what!

- Time to find p, d, q
  - Use intuition
  - Use ~~high explosive~~ brute force

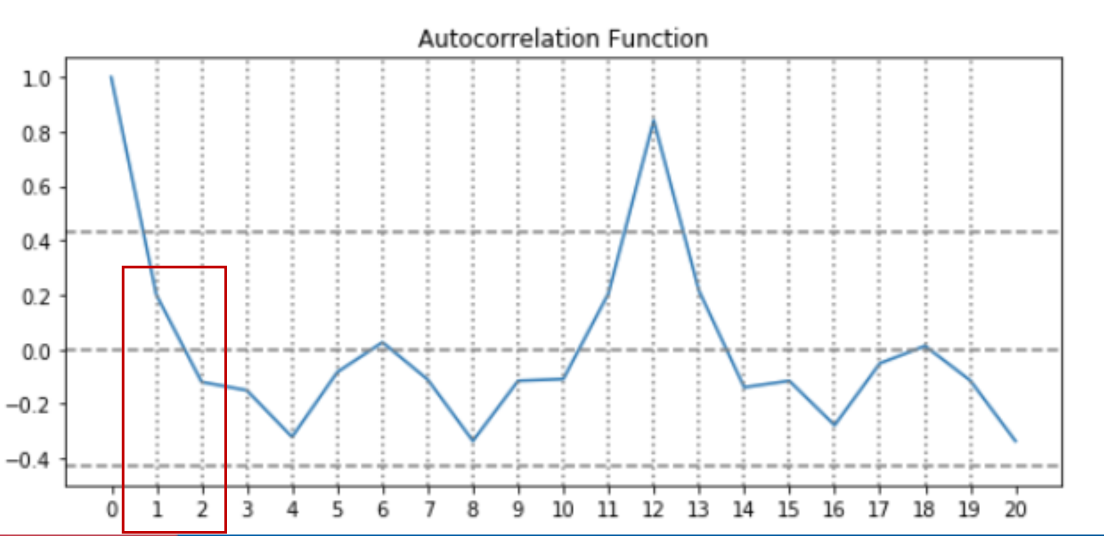
Intuition Method

- Use Autocorrelation, Partial Autocorrelation

Rules for ‘d’

- **Identifying the order of differencing and the constant:**
- Rule 1: If the series has positive autocorrelations out to a high number of lags (say, 10 or more), then it probably needs a higher order of differencing.
- Rule 2: If the lag-1 autocorrelation is zero or negative, or the autocorrelations are all small and patternless, then the series does *not* need a higher order of differencing. If the lag-1 autocorrelation is -0.5 or more negative, the series may be overdifferenced. **BEWARE OF OVERDIFFERENCING.**
- Rule 3: The optimal order of differencing is often the order of differencing at which the standard deviation is lowest. (Not always, though. Slightly too much or slightly too little differencing can also be corrected with AR or MA terms. See rules 6 and 7.)
- Rule 4: A model with <u>no</u> orders of differencing assumes that the original series is stationary (among other things, mean-reverting). A model with <u>one</u> order of differencing assumes that the original series has a constant average trend (e.g. a random walk or SES-type model, with or without growth). A model with <u>two</u> orders of total differencing assumes that the original series has a time-varying trend (e.g. a random trend or LES-type model).
- Rule 5: A model with <u>no</u> orders of differencing normally includes a constant term (which allows for a non-zero mean value). A model with <u>two</u> orders of total differencing normally does <u>not</u> include a constant term. In a model with <u>one</u> order of total differencing, a constant term should be included if the series has a non-zero average trend.

‘d’ examples (our differenced/stationary plot)



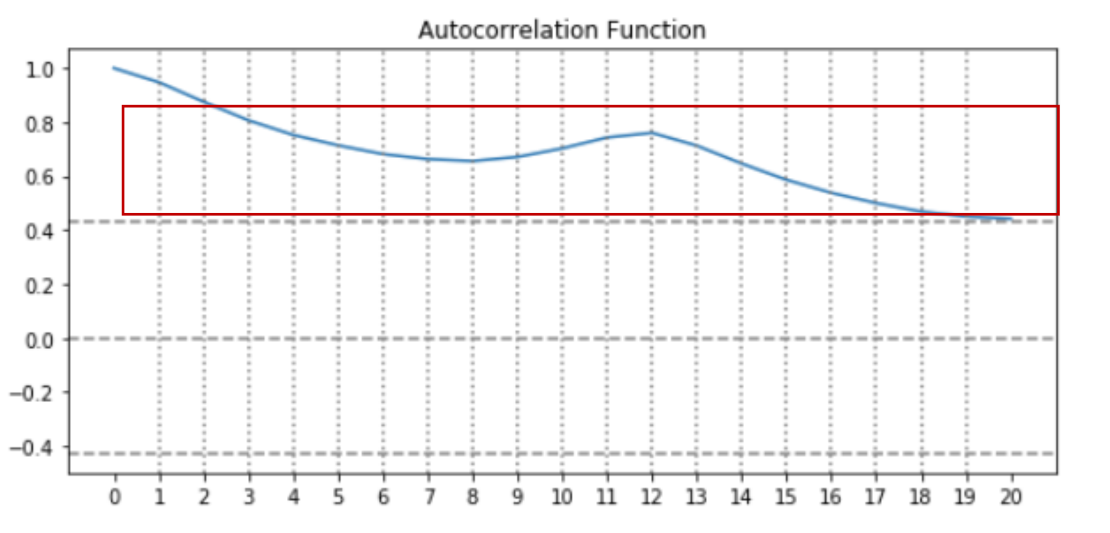
d = 0

Rule 1: No. (almost immediately drops below UCL)

Rule 2: Yes. At label 1 the value is ~ 0

- **Identifying the order of differencing and the constant:**
- Rule 1: If the series has positive autocorrelations out to a high number of lags (say, 10 or more), then it probably needs a higher order of differencing.
- **Rule 2: If the lag-1 autocorrelation is zero or negative**, or the autocorrelations are all small and patternless, then the series does *not* need a higher order of differencing. If the lag-1 autocorrelation is -0.5 or more negative, the series may be overdifferenced. **BEWARE OF OVERDIFFERENCING.**

‘d’ example (raw airline population)



Rule 1: Yes. (above UCL to 20)

Rule 2: No. At label 1 the value is ~2x UCL

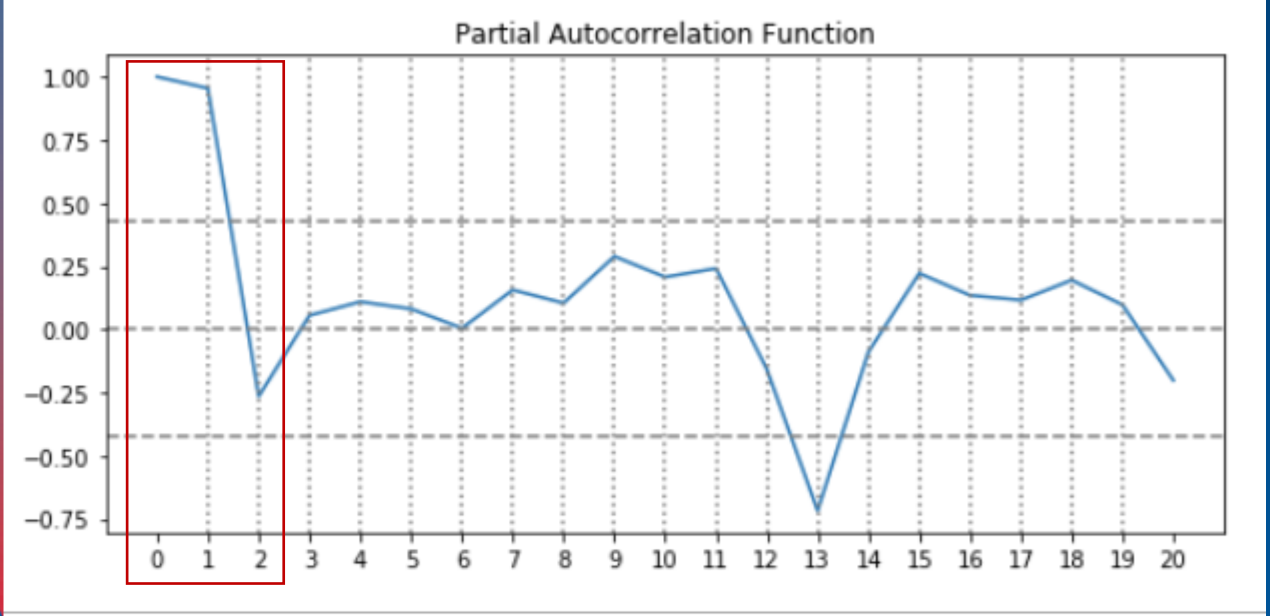
- **Identifying the order of differencing and the constant:**
- **Rule 1: If the series has positive autocorrelations out to a high number of lags** (say, 10 or more), then it probably needs a higher order of differencing.
- Rule 2: If the lag-1 autocorrelation is zero or negative, or the autocorrelations are all small and patternless, then the series does *not* need a higher order of differencing. If the lag-1 autocorrelation is -0.5 or more negative, the series may be overdifferenced. **BEWARE OF OVERDIFFERENCING.**

d >= 1

Rules for ‘p’, ‘q’

- **Identifying the numbers of AR and MA terms:**
- Rule 6: If the partial autocorrelation function (PACF) of the differenced series displays a sharp cutoff and/or the lag-1 autocorrelation is positive--i.e., if the series appears slightly "underdifferenced"--then consider adding one or more AR terms to the model. The lag beyond which the PACF cuts off is the indicated number of AR terms.
- Rule 7: If the autocorrelation function (ACF) of the differenced series displays a sharp cutoff and/or the lag-1 autocorrelation is negative--i.e., if the series appears slightly "overdifferenced"--then consider adding an MA term to the model. The lag beyond which the ACF cuts off is the indicated number of MA terms.
- Rule 8: It is possible for an AR term and an MA term to cancel each other's effects, so if a mixed AR-MA model seems to fit the data, also try a model with one fewer AR term and one fewer MA term--particularly if the parameter estimates in the original model require more than 10 iterations to converge. **BEWARE OF USING MULTIPLE AR TERMS AND MULTIPLE MA TERMS IN THE SAME MODEL.**
- Rule 9: If there is a unit root in the AR part of the model--i.e., if the sum of the AR coefficients is almost exactly 1--you should reduce the number of AR terms by one and increase the order of differencing by one.
- Rule 10: If there is a unit root in the MA part of the model--i.e., if the sum of the MA coefficients is almost exactly 1--you should reduce the number of MA terms by one and reduce the order of differencing by one.
- Rule 11: If the long-term forecasts* appear erratic or unstable, there may be a unit root in the AR or MA coefficients.

Example



+1 AR term (rule 6)

Rule 6: If the partial autocorrelation function (PACF) of the differenced series displays a sharp cutoff and/or the lag-1 autocorrelation is positive--i.e., if the series appears slightly "underdifferenced"--then consider adding one or more AR terms to the model. The lag beyond which the PACF cuts off is the indicated number of AR terms.

Brute Force (Grid Search)

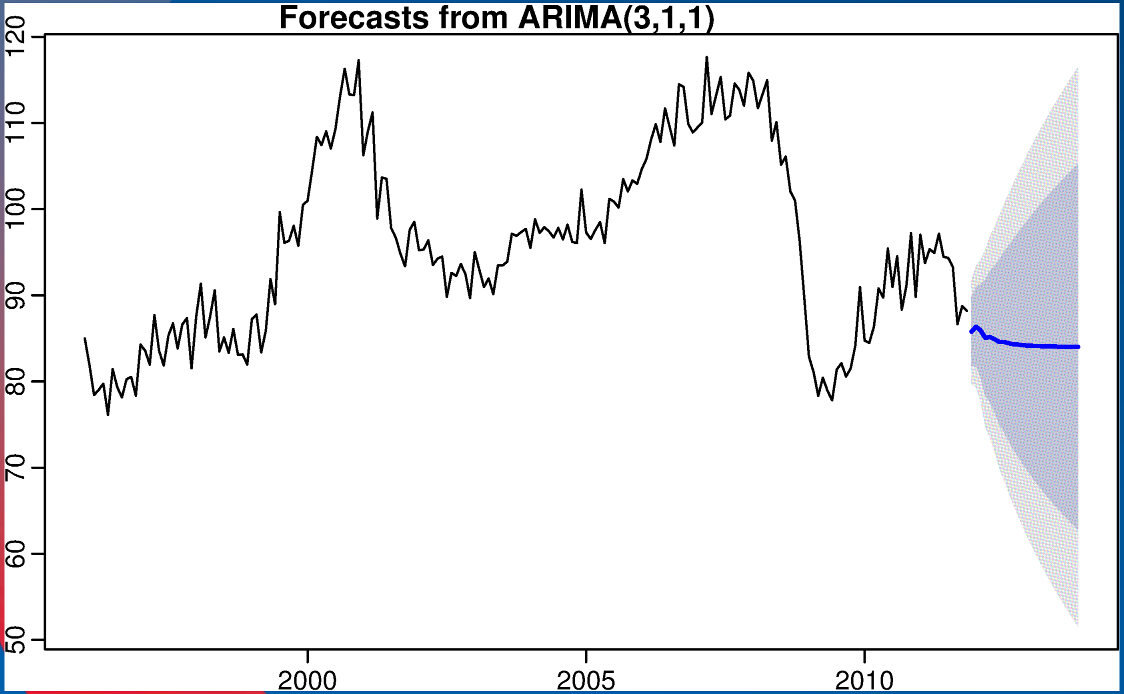
- Generally the parameters p,d,q are between 0 and 3
- So we only need 64 runs to get the best fit (4 x 4 x 4)*
- Use them all in a for loop.

- *48 runs if you're strictly adhering to Rule 8: It is possible for an AR term and an MA term to cancel each other's effects, so if a mixed AR-MA model seems to fit the data, also try a model with one fewer AR term and one fewer MA term--particularly if the parameter estimates in the original model require more than 10 iterations to converge. **BEWARE OF USING MULTIPLE AR TERMS AND MULTIPLE MA TERMS IN THE SAME MODEL.**

But wait

- ARIMA is nice as there are only 3 parameters that take on discrete values (i.e., only integers, no floats)
- Linear regression has 2 parameters we can vary over a huge range (continuous)
  - N * M values (N, M are the parameters)
  - ~ n^2 pick 5 each, 25 runs
- XGBOOST has 5-6 tuning parameters (actually more)
  - N^5 pick 5 each 3125 runs….
- It pays to know what your parameters do!!!!

We didn't talk about prediction!!!



Predict 1 point at a time
Horrible confidence intervals

Usefulness for predicting the future is suspect!

There is a huge random component to these models (otherwise we'd all be rich)

References

- https://www.analyticsvidhya.com/blog/2016/02/time-series-forecasting-codes-python/
- https://people.duke.edu/~rnau/411arim.htm
- https://machinelearningmastery.com/arima-for-time-series-forecasting-with-python/
- https://www.machinelearningplus.com/time-series/arima-model-time-series-forecasting-python/

Assignment

- Pick a stock
- Get at least 4 years worth of data
- Estimate p, d, q using techniques discussed in class
- Do a grid search for parameters
- What is your final decision on parameters and WHY (REMEMBER RULE 8! Advise against (3,1,3) models).
- *If you'd like* – try splitting your data into a training and testing set; apply your trained model on the test data