PREDICTING TRAFFIC CONGESTION WITH THE AUTO-REGRESSIVE MOVING AVERAGE (ARMA) MODEL

A Presentation by Stephanie Boggs

Home University- Jackson State University

Program Site- Salisbury University

August 5th, 2019

Explore **E**merging **C**omputing **I**n **S**cience and **E**ngineering **REU**

Mixed Auto-Regressive Moving Average Model – ARMA(p , q)

ARMA Model consists of two closely related polynomials used for understanding and perhaps predicting future values in a series.

Before considering how the models may be combined they must be examined separately

Auto-Regression

$$X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + \varepsilon_t + \sum_{i=1}^{q} \Theta_i \varepsilon_{t-i}$$

Moving Average

AR Model (AR(p)) perspective

Predicts future behavior based on the correlation of past values and the values that succeed them.

$$X_t = c + \sum_{i=1}^{p} \phi_i X_{t-i} + \varepsilon_t$$

Specific lagged values of x_t are used as predictor variables. Lags are where results from one time period affect following periods

Variables at time (t) & order (i)

$$X_t = c + \sum_{i=1}^{p} \phi_i X_{t-i} + \varepsilon_t$$

- Current time step is dependent on previous time steps
- Model parameter

Order of AR model

$$X_t = c + \sum_{i=1}^{p} \phi_i X_{t-i} + \varepsilon_t$$

The number of immediately preceding values in the series that are used to predict the value at the present time

Determining this order may be done by plotting a **partial autocorrelation** function of data.

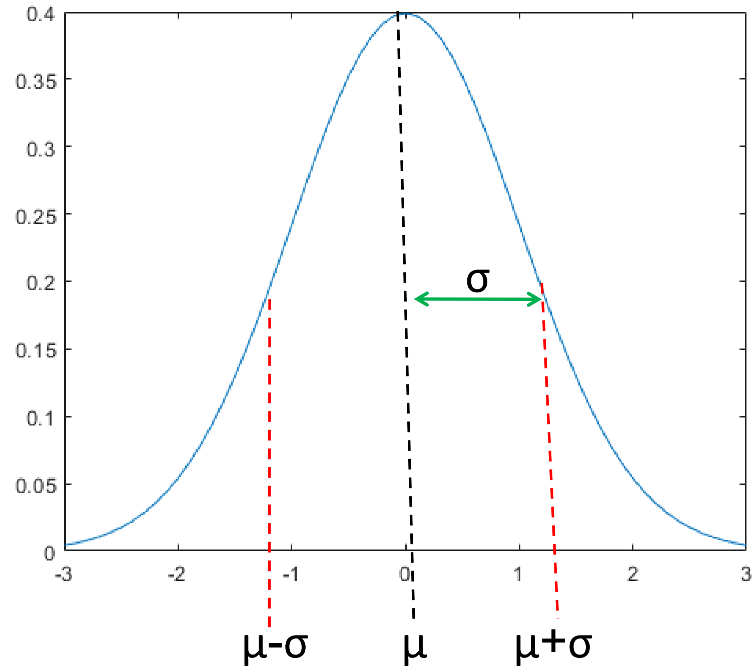
At lag *k*, this is the correlation between series values that are *k* intervals apart, accounting for the values of the intervals between.

Error (White Noise)

$$X_t = c + \sum_{i=1}^{p} \phi_i X_{t-i} + \boxed{\varepsilon_t}$$

- Randomly Determined (Stochastic)

- Characterized by independent normal random variables sampled by Gaussian Distribution (mean = 0)



MA Model (MA(q)) perspective

The error of any period, t, is linearly dependent on previous errors.

$$X_t = \sum_{i=1}^{q} \Theta_i \varepsilon_{t-i}$$

Fitting the MA estimates is more complicated than it is in autoregressive models (AR models), because the lagged error terms are not observable.

Variables at time (t) and order (i)

$$X_t = \sum_{i=1}^{q} \Theta_i \varepsilon_{t-i}$$

- This model's parameter accounts for the weights applied to prior values in the time series.
- Errors for previous periods (Residuals)

Order of MA Model

$$X_t = \sum_{i=1}^{q} \Theta_i \varepsilon_{t-i}$$

Determining the order of this model may be done by plotting an **autocorrelation** function of data.

➢At lag *k*, this is the correlation between series values that are *k* intervals apart

ARMA in Traffic Prediction

- ARMA Model is atheoretic and used for prediction of random processes.
- Traffic is difficult to model & predict because it lacks inherent symmetry
- This model has not been used in traffic prediction

Methodology & Tools

- MATLAB – Econometrics Toolbox
  - Econometric Modeler
  - Correlation functions
- Box-Jenkins Method for ARMA prediction
- Compare experimental results to real-life values

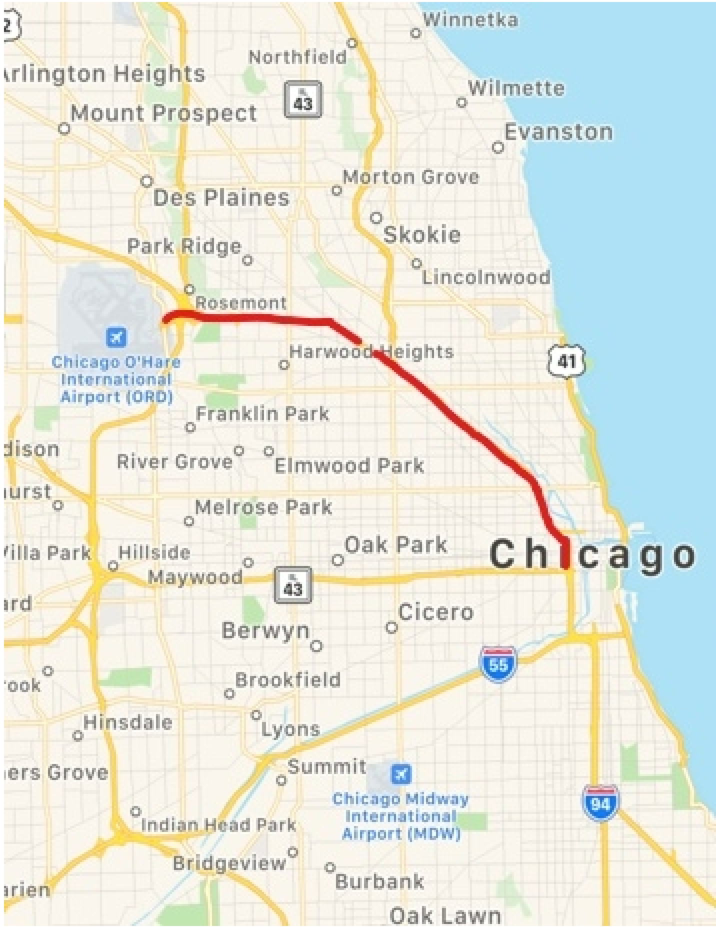
Data Overview

Chicago, Illinois : Kennedy Expressway

O'Hare Int'l Airport ⟷ I:290 J.B. Interchange

Approximately: 18 miles

Speed Limit: 70 mph

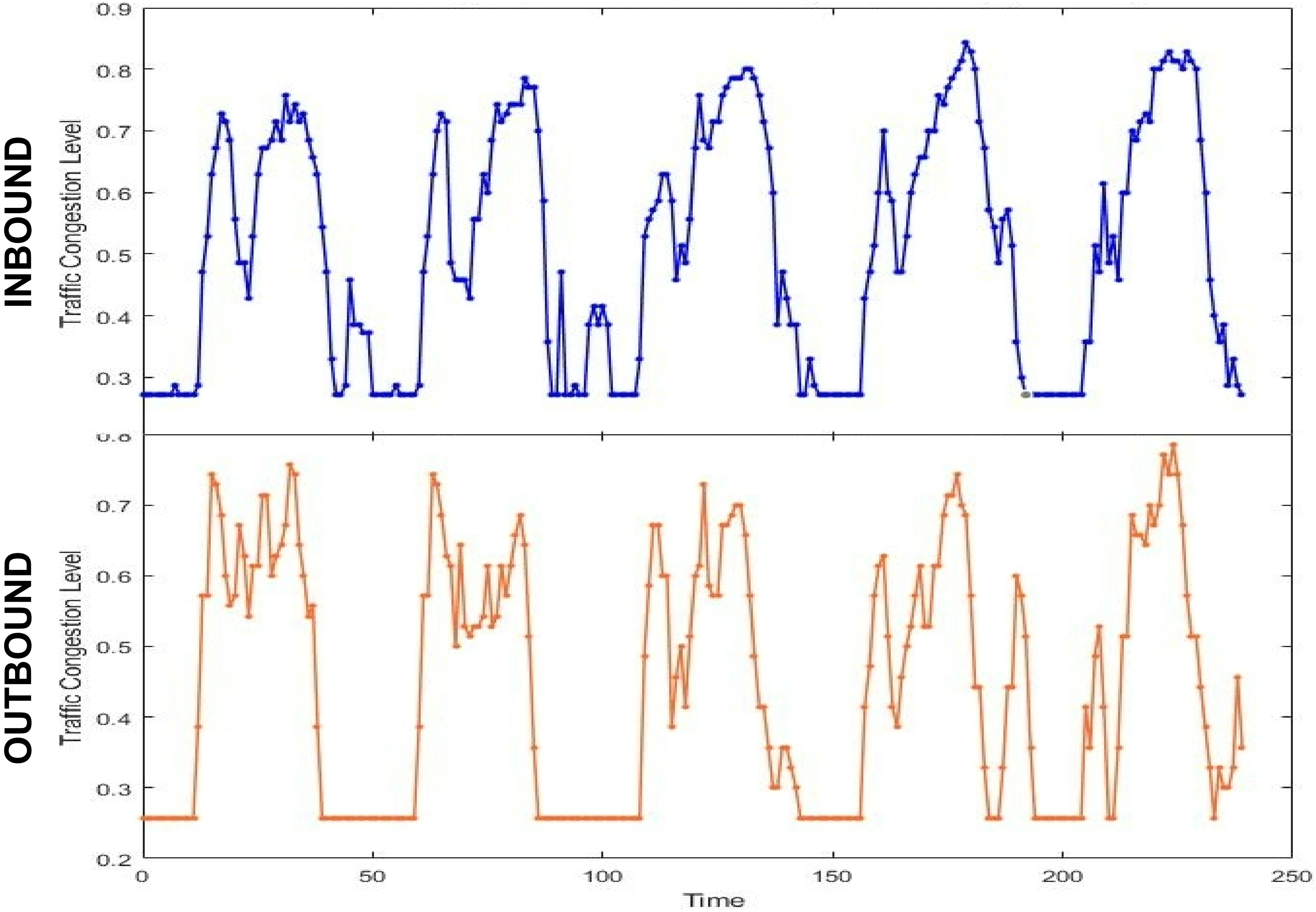


Observed Variable:

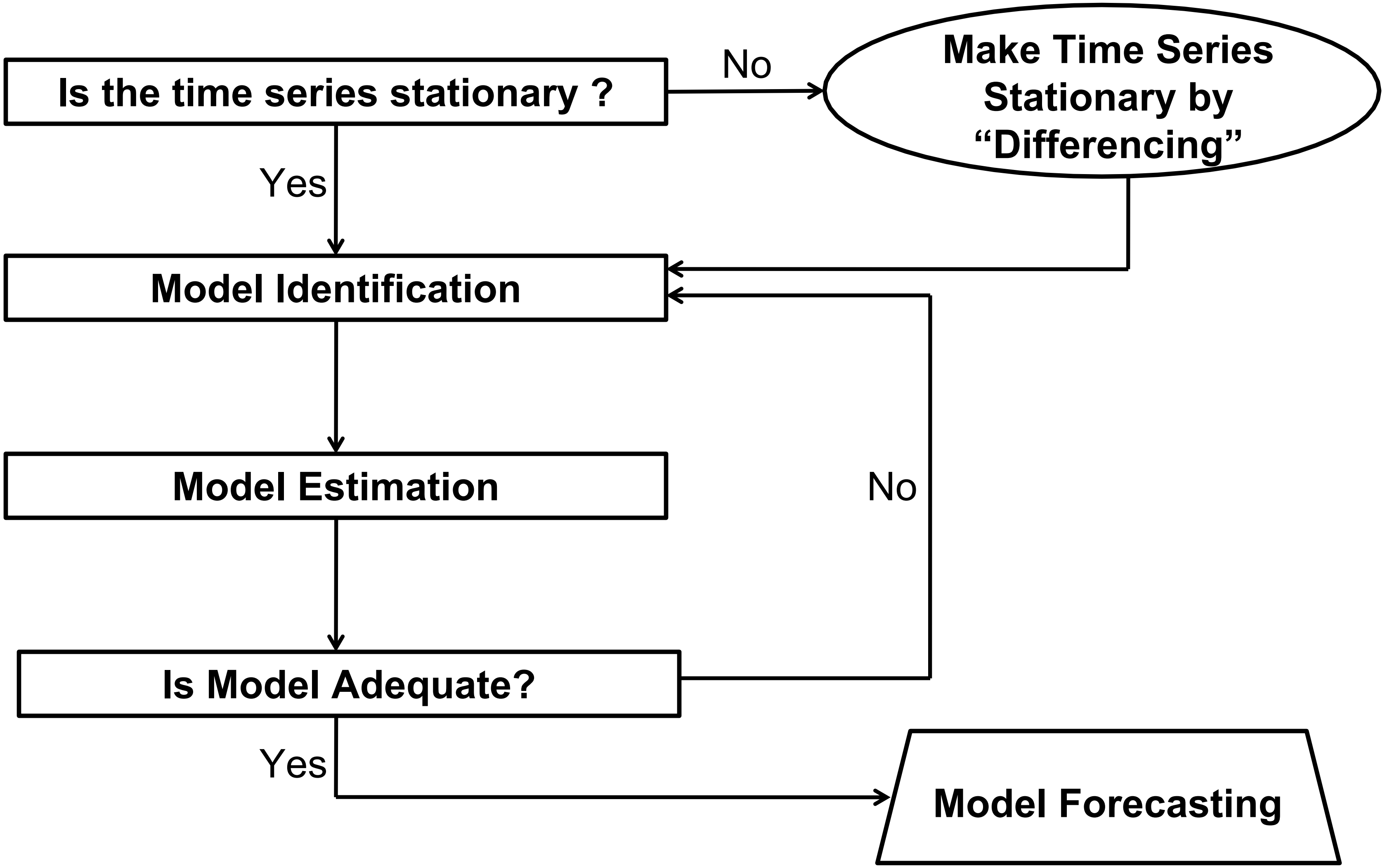
Traffic Congestion Levels:

$$TCL = (V_{max} - V_i)/V_{max}$$

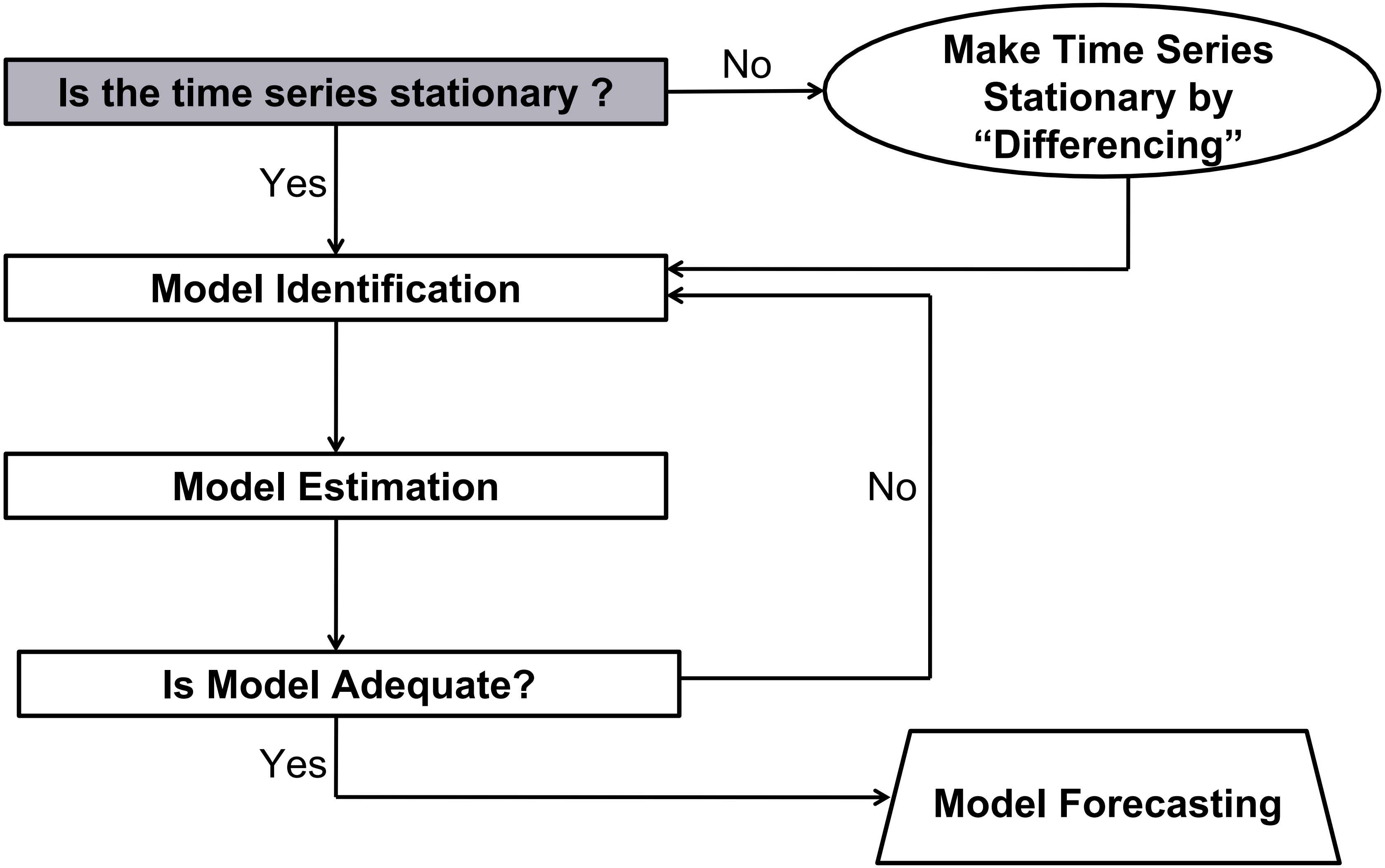
Traffic Congestion Levels for Kennedy Expressway



Box Jenkins Method for ARMA Prediction

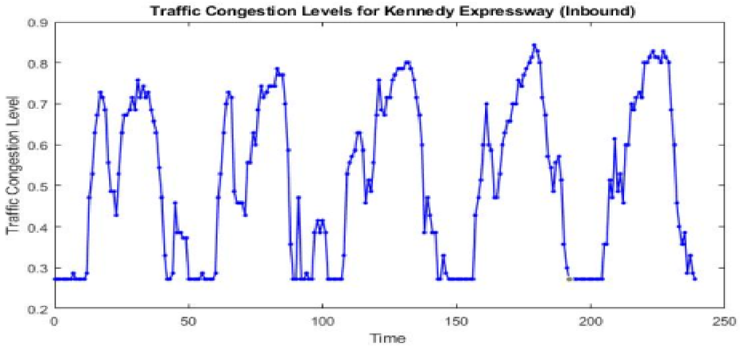


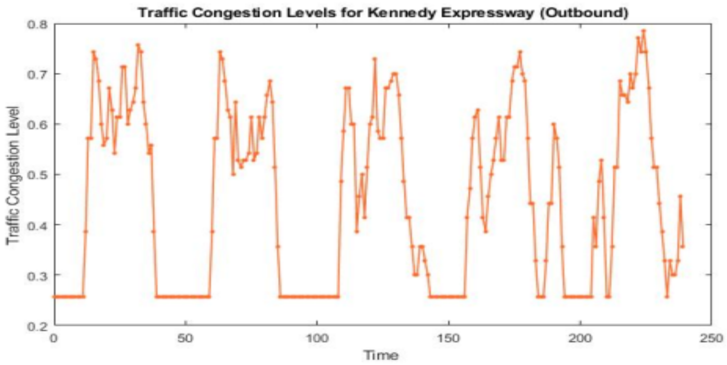
Box Jenkins Method for ARMA Prediction

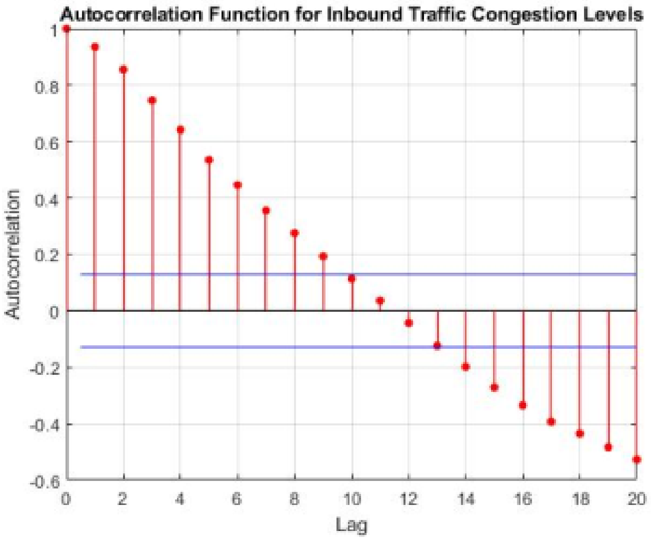


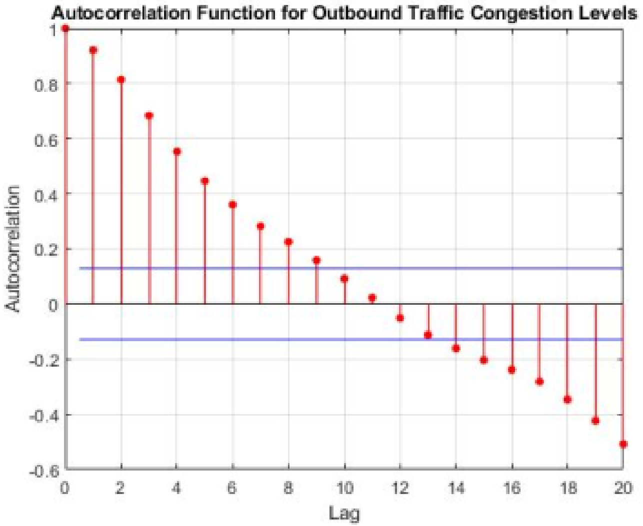
Time series is non-stationary

- Autocorrelation pattern signifies a trend.
- Dependence on time

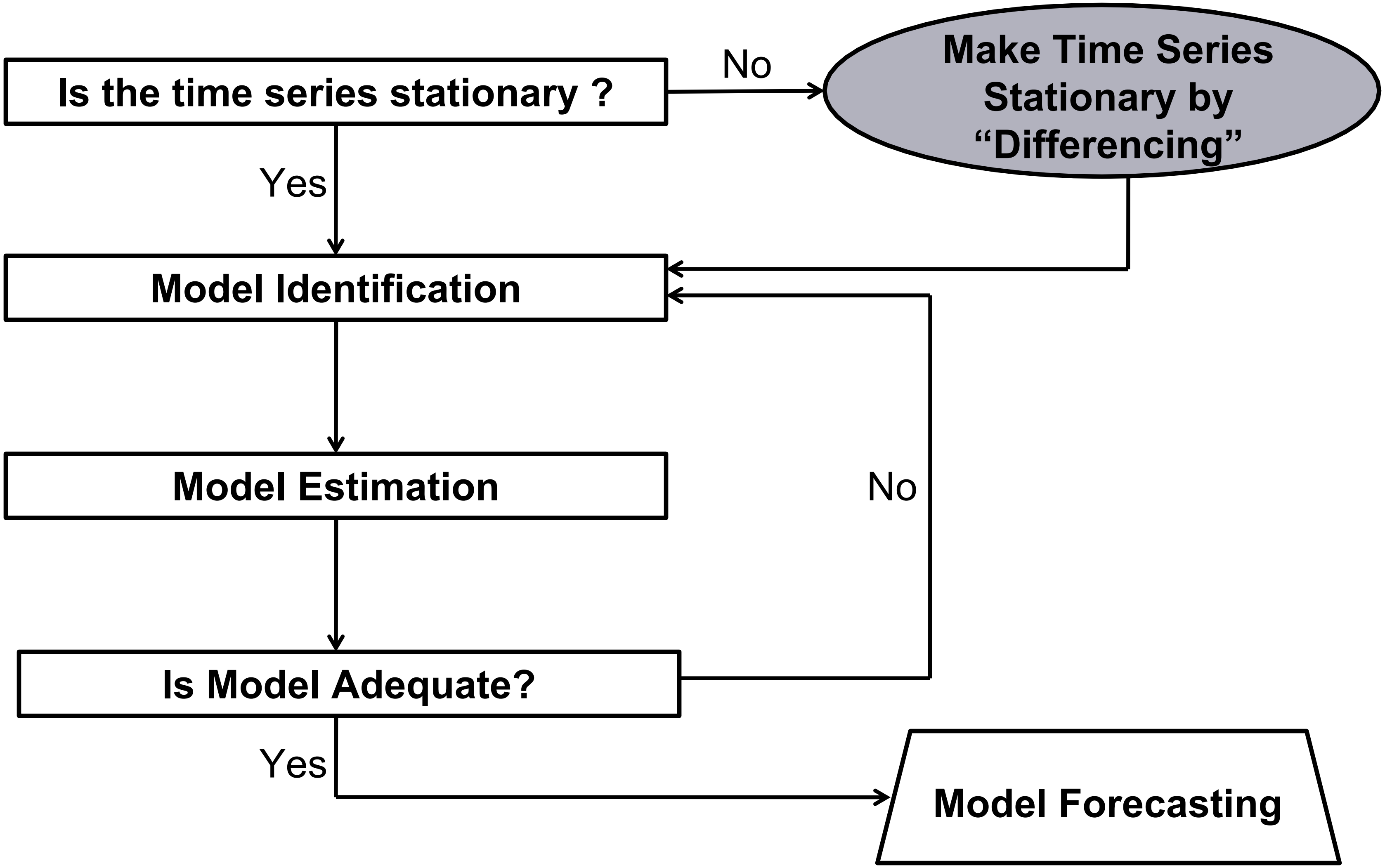








Box Jenkins Method for ARMA Prediction

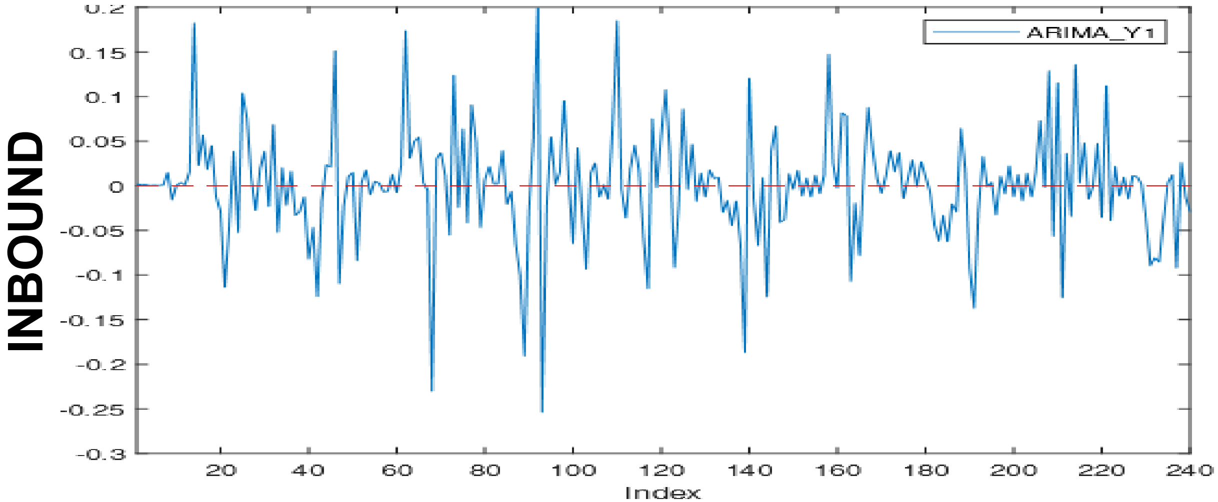


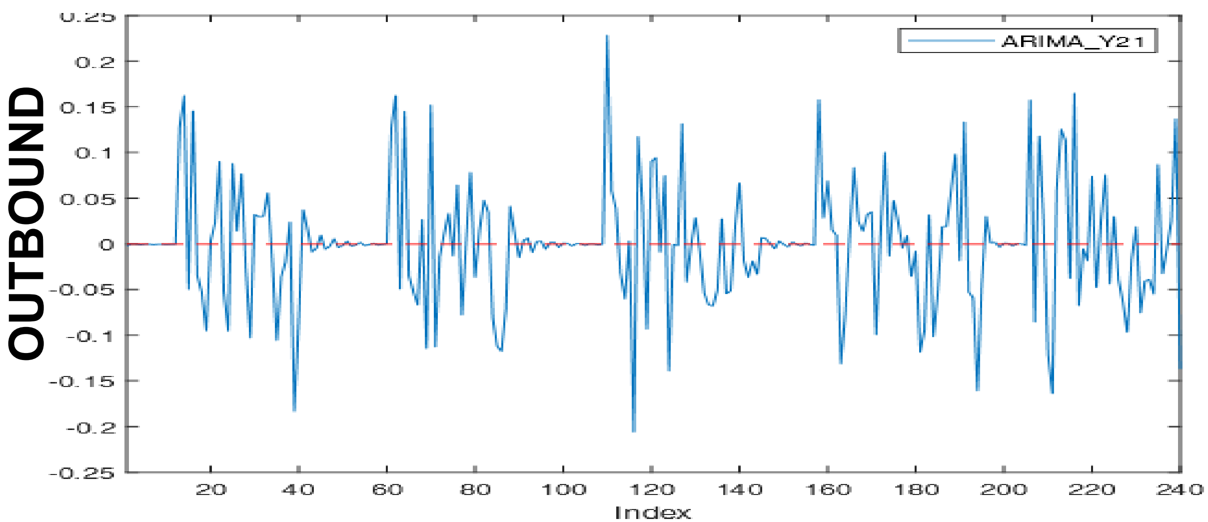
ARMA become AR(I)MA

- Notation - ARIMA(p,d,q)
- (I) – Integrate

  Eliminated non-stationarity by a process known as "differencing"
- Differencing removes random trends by correlating raw values with a gradient

Non-Stationary to Stationary Time Series

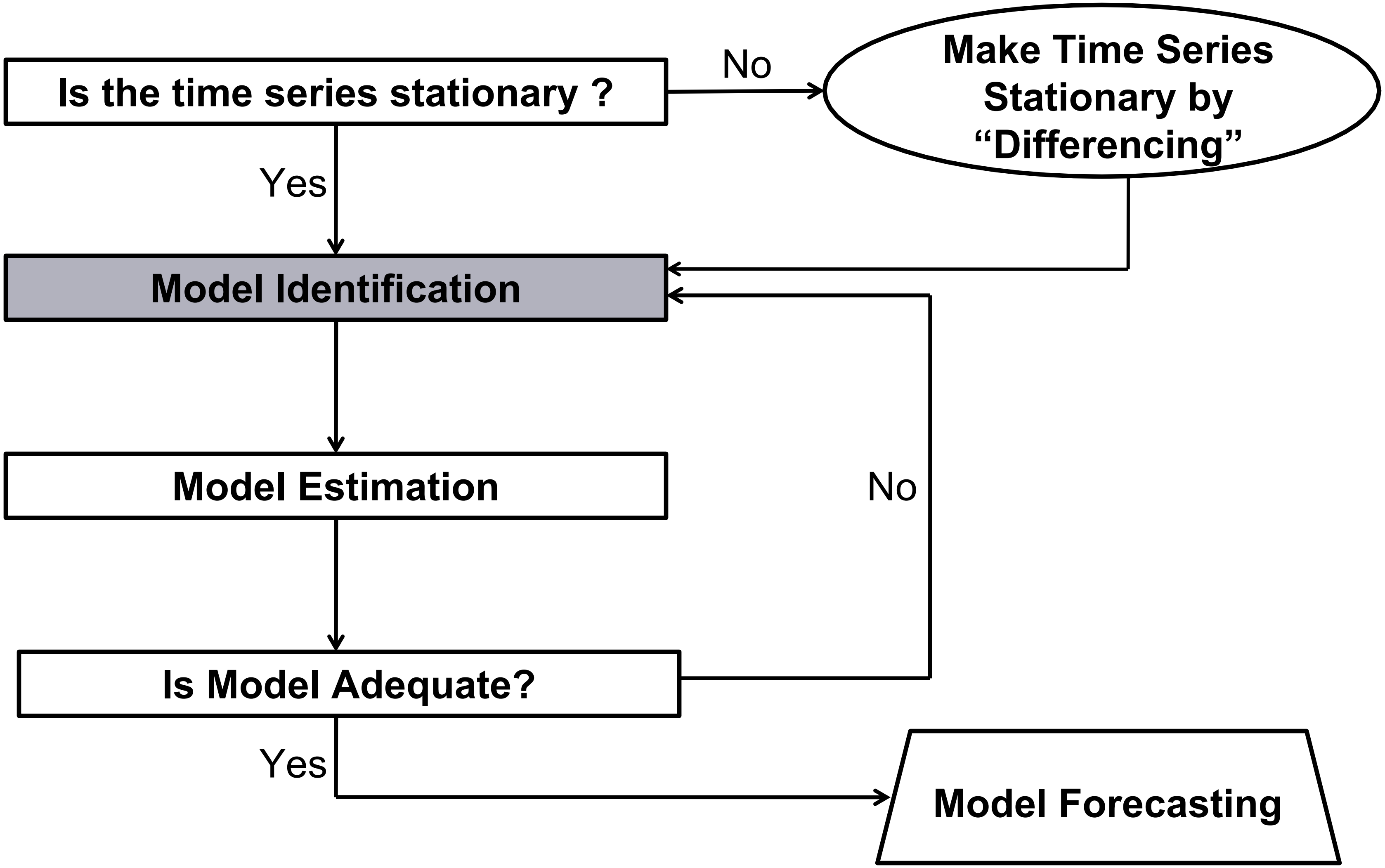




Mean = 0

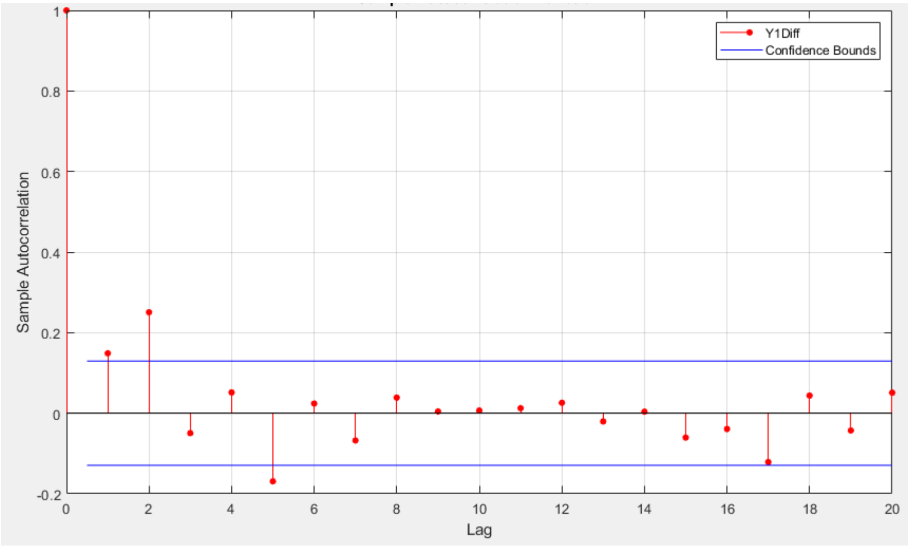
Constant Variance

Box Jenkins Method for ARMA Prediction

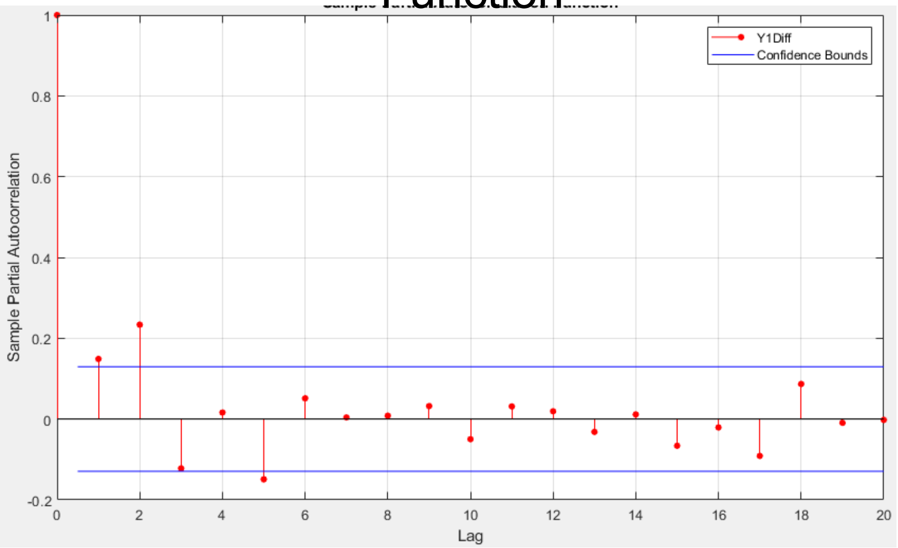


Correlograms

Inbound Autocorrelation Function



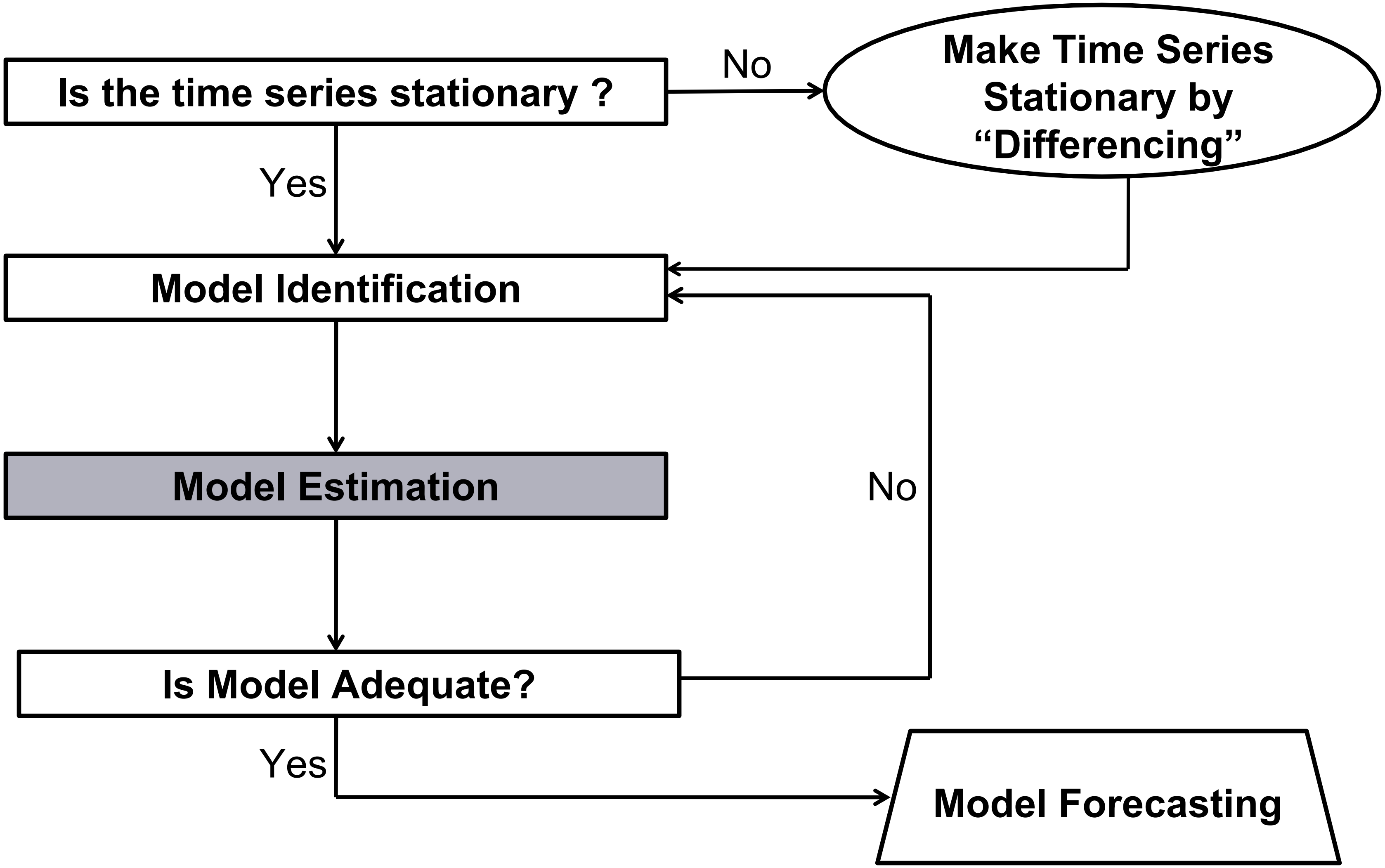
Inbound Partial Autocorrelation Function



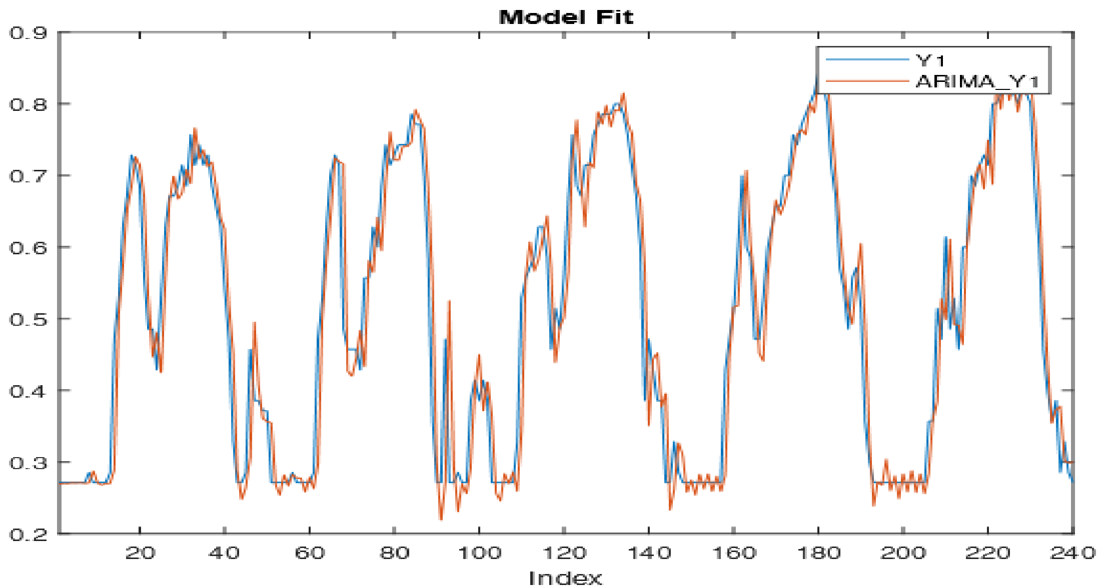
- 95% Confidence interval
- Significant correlations at the first couple of lags, followed by correlations that are not significant.
- 3 significant correlations

ARIMA (3,1,3)

Box Jenkins Method for ARMA Prediction

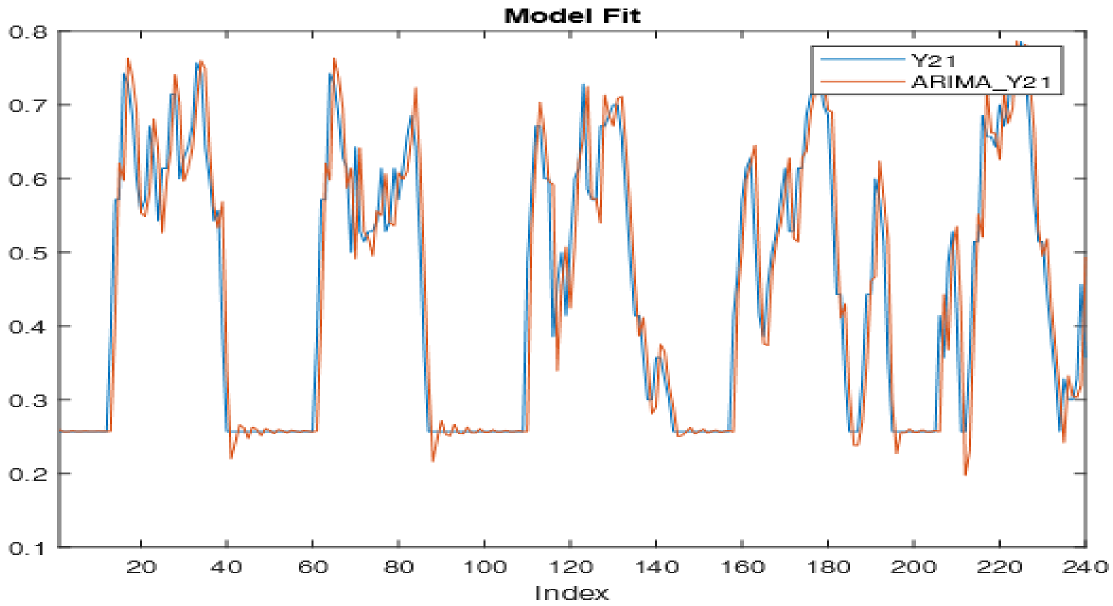


Inbound TCL Model Estimation



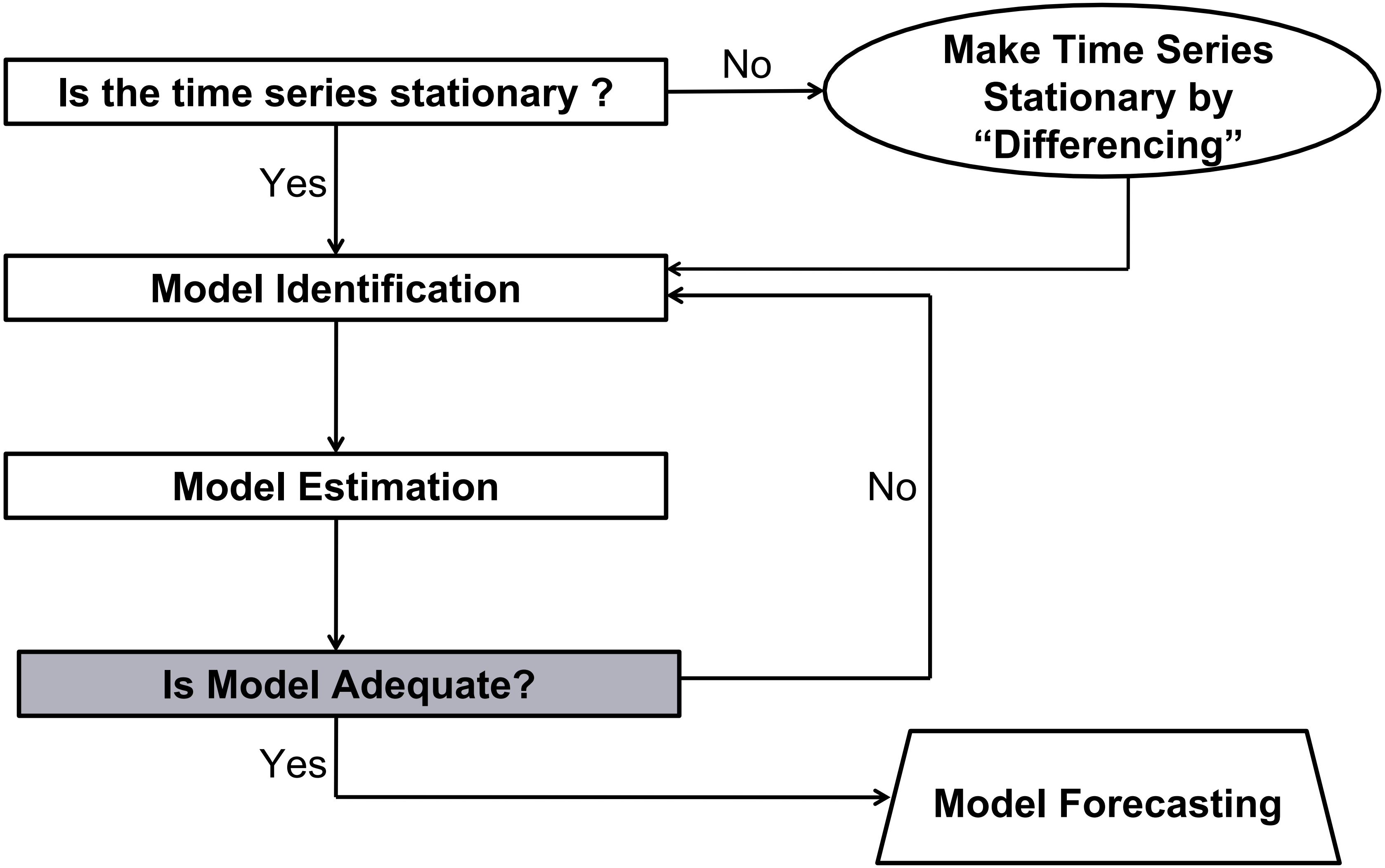
Parameter	Value	Standard Error	t Statistic	P-Value
Constant	-0.0010351	0.0071638	-0.14448	0.88512
AR{1}	-0.091047	0.34733	-0.26213	0.79322
AR{2}	0.39037	0.21629	1.8048	0.071104
AR{3}	-0.51859	0.22857	-2.2688	0.023281
MA{1}	0.26441	0.34906	0.75747	0.44877
MA{2}	-0.16467	0.2933	-0.56145	0.57449
MA{3}	0.53425	0.19214	2.7805	0.0054269
Variance	0.0038219	0.0002559	14.9353	1.9427e-50

Outbound TCL Model Estimation



Parameter	Value	Standard Error	t Statistic	P-Value
Constant	0.0010851	0.011272	0.09626	0.92331
AR{1}	-0.70176	0.25896	-2.7099	0.006731
AR{2}	-0.43793	0.20949	-2.0905	0.036577
AR{3}	0.25024	0.24757	1.0108	0.31213
MA{1}	0.88162	0.2619	3.3663	0.00076193
MA{2}	0.72256	0.20824	3.4699	0.00052071
MA{3}	0.0072267	0.26196	0.027587	0.97799
Variance	0.004298	0.00032998	13.0253	8.7895e-39

Box Jenkins Method for ARMA Prediction



Future Work

- Model Forecasting

  Forecast function

- Alternate method for determining p & q orders

  Akaike Information Criterion (AIC)- It associates the number of model parameters and the goodness of fit. It also associates a penalty factor to avoid over-fitting

Reflection:

What did I learn this summer?

MATLAB
Statistical Analysis
Wallops Flight Facility
Planning
Organization
Volleyball

References

- Kumar, R., & Soodan, B. S. (2019). Transient Analysis of a Single-Server Queuing System with Correlated Inputs and Reneging. *Reliability: Theory & Applications*, *14*(1), 102–106. Retrieved from http://ecnhts-proxy.jsums.edu:2069/login.aspx?direct=true&db=aph&AN=136703607&site=ehost-live
- The Agency for Science, Technology and Research (A*STAR). (2016, June 17). A mathematical analysis of urban traffic models clarifies dispute over which approach is best. *ScienceDaily*. Retrieved June 13, 2019 from www.sciencedaily.com/releases/2016/06/160617114007.htm
- Freud, R.J., Wilson, W.J. ((1998) Regression Analysis: Statistical Modeling of a Response Variable. San Diego,CA: Academic Press
- The Box-Jenkins Method. Chapter 470. NCSS Statistical Software. https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/Procedures/NCSS/The_Box-Jenkins_Method.pdf

Questions ?