Assignment 10:

**Seasonal Box-Jenkins**

Acacia Johnson

Tarleton State University

LSCM 5330 – Supply Chain Analytics

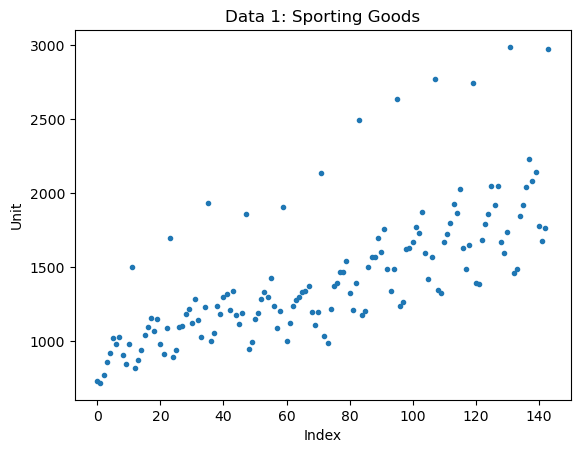
Dr. Triss Ashton

April 13th, 2025

# Assignment 10: Seasonal Box-Jenkins

The AutoRegressive Integrated Moving Average (ARIMA) forecasts future values based on patterns in past values. Seasonal ARIMA uses lagged values and cyclical trends to forecast future behavior. Lagged values are produced by regressing the target variable on itself. I will perform seasonal ARIMA on two distinct data sets and then forecast five time periods.

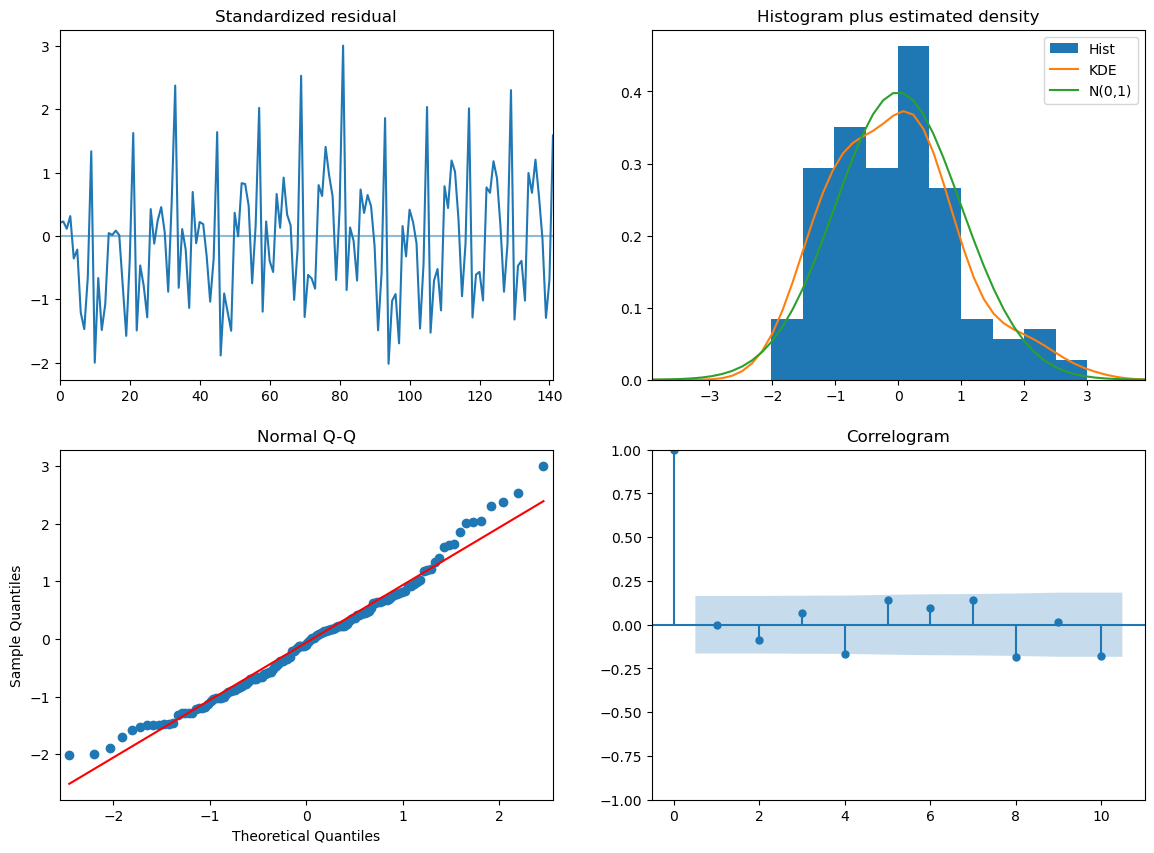
# Data 1

* 1. The original data follows a positive linear trend . There appears to be a cyclical pattern every 12 observations. The trend leads me to believe seasonality plays an important part in this data set.

## Stationary Tests

1. The stationary test used this time is an Augmented Dickey-Fuller (ADF). We fail to reject the null hypothesis because the p-value is 0.930. For clarification, the null hypothesis is that the data has a unit root, or is not stationary. This confirms Data 1 is non-stationary and requires differencing.

## Regression with ARIMA

1. I start by performing a stepwise search to find the lowest AIC (Akaike information criterion). The AIC returned is -90.157. The ADF model produces a seasonal ARIMA (4,0,5). This means there is four auto-regressive (AR), zero differences taken (I), and five moving-average (MA) component respectively. The output summary shows how each ARIMA (ar.L4, ma.L3, ma.L4, ma.L5) has significance. All other intercepts and terms have no significance.
2. Since the ADF test reveals that the data is non-stationary, we need to take a difference by setting d=1. The new model produces an AIC of -106.308, which is significantly lower that the first model. The data set is now stationary with a seasonal ARIMA (5,1,5). Some coefficients are not significant and can be removed form the model (ar.L5 and ma.L1). Despite the model still containing insignificant lags, there is a solid formation to the ARIMA(5,1,5). The following plots show no heteroskedascity, minimal auto-correlation, and normally distributed errors and residuals.
3. **Forecasting the Next Five Values**
4. The final step is to use the ARIMA model to predict future values. The python code creates each value and their respective confidence intervals. The predictions follow closely to the pre-established cycle in the original data. Each time period fits within the confidence interval set by ARIMA(5,1,5).
5. Forecast for Next 5 Time Periods:
6. 144 7.611391
7. 145 7.516177
8. 146 7.578740
9. 147 7.779646
10. 148 7.620974
11. Confidence Intervals for Next 5 Time Periods: [[7.30007695 7.92270526]
12. [7.20421633 7.82813821]
13. [7.24719958 7.91028008]
14. [7.44691756 8.11237467]
15. [7.27145165 7.97049612]]

# Data 2

## The original data follows a positive linear trend . There is no visible seasonality or outliers within the graph. The trend leads me to believe a difference will need to be taken to create stationary data..

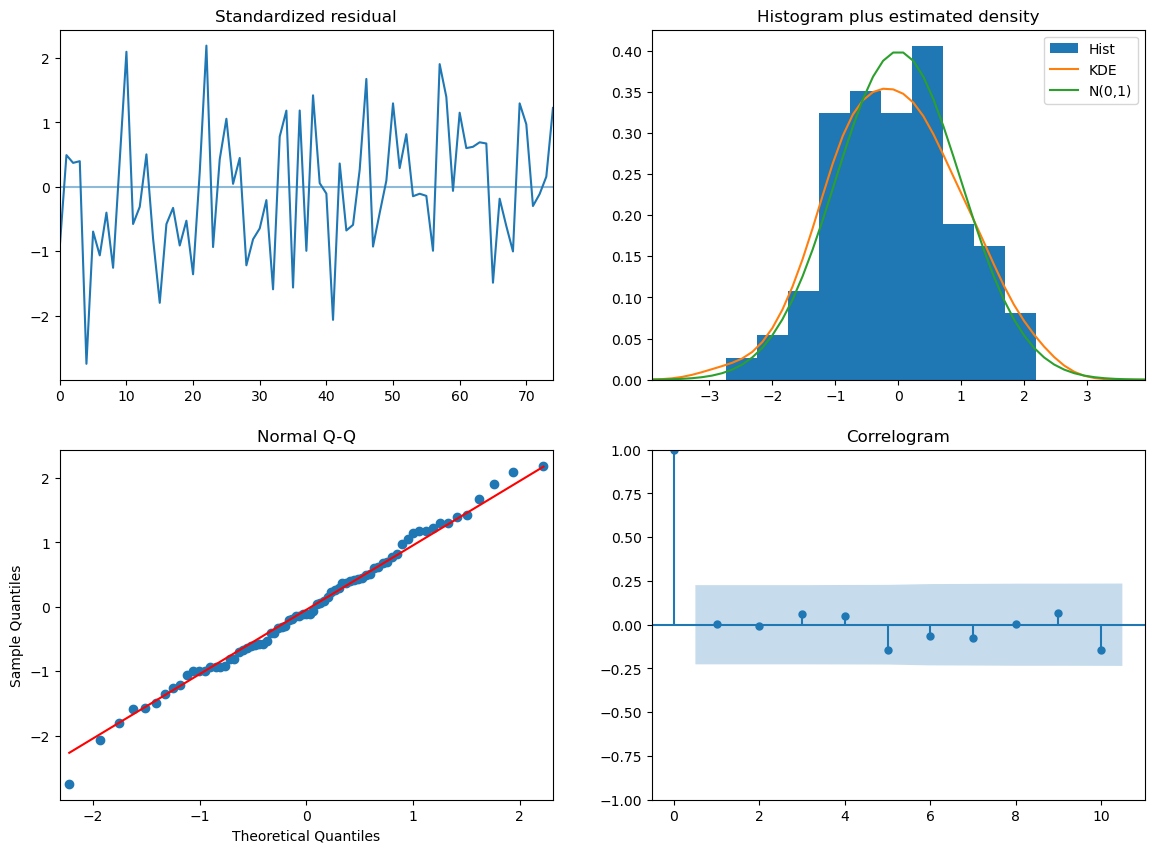
## Stationary Tests

The first stationary test used this time is an Augmented Dickey-Fuller (ADF). We fail to reject the null hypothesis because the p-value is 1.0. For clarification, the null hypothesis is that the data has a unit root, or is not stationary. This confirms Data 2 is non-stationary and may require differencing.

## Regression with ARIMA

I start by performing a stepwise search to find the lowest AIC (Akaike information criterion). The AIC returned is -247.216. The ADF model produces a seasonal ARIMA (4,0,2). This means there is four auto-regressive (AR), zero differences taken (I), and two moving-average (MA) component respectively. The output summary shows how each coefficient (ar.L3, ma.L1, ma.L2) has significance. We can also tell there is no auto-correlation (Ljung-Box = 0.52) and the residuals are normally distributed (Jarque-Bera = 0.67).

Since the ADF test reveals that the data is non-stationary, we need to take a difference by setting d=1. The new model produces an AIC of -268.94, which is significantly lower that the first model. The data set is now stationary with a seasonal ARIMA (0,1,2). The reduce number of AR terms increase the significance and strength of all variables. Ljung-Box now equals 0.98 and Jarque-Bera equals 0.93. The following plots show no heteroskedascity, no auto-correlation, and normally distributed errors and residuals.

**Foreca****sting the Next Five Values**

The final step is to use the ARIMA model to predict future values. The python code creates each value and their respective confidence intervals. The predictions follow closely to the pre-established cycle in the original data. Each time period fits within the confidence interval set by ARIMA(0,1,2).

Forecast for Next 5 Time Periods:

76 6.517314  
77 6.502228  
78 6.511137  
79 6.520047  
80 6.528956  
  
Confidence Intervals for Next 5 Time Periods:

[[6.44301109 6.59161654]  
 [6.41114471 6.59331083]  
 [6.41893762 6.60333672]  
 [6.42674389 6.61334924]  
 [6.43456305 6.62334888]]

# Python Code

