TIME SERIES STUDENT PROJECT

Fall 2011

AGENDA

1. Introduction
2. Initial Data Thoughts and Analysis
3. Testing for Stationarity
4. Testing for Seasonality
5. Building an initial ARIMA model with recommended differencing.
6. Comparing AR vs MA term
7. Testing more complex models to see if they improve fit
8. Analysis of Residuals
9. Forecasting proposed model

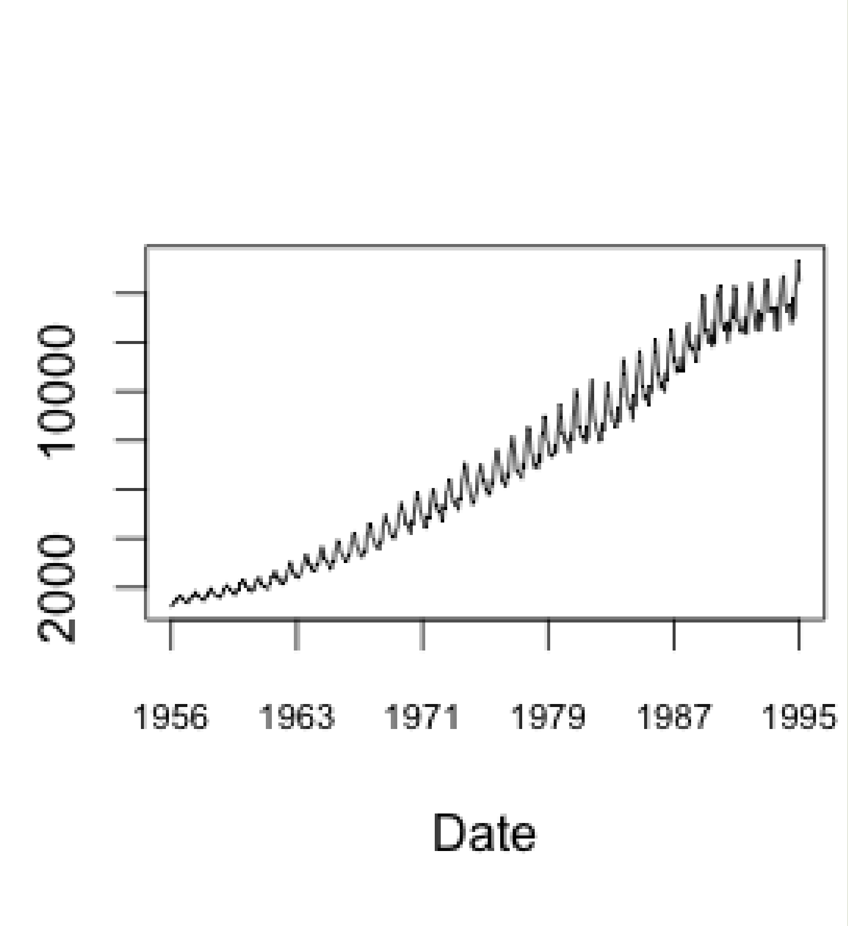
INTRODUCTION

- The purpose of this project is to demonstrate knowledge of the time series material through the analysis of an actual dataset.
- After examining some sample data sources on, I decided to choose a dataset showing the monthly electricity production in Australia.
  - https://datamarket.com/data/set/22l0/monthly-electricity-production-in-australia-million-kilowatt-hours-jan-1956-aug-1995#!ds=22l0&display=line.
  - I chose this data source, as it has a sufficient number of data points to adequately display time series. There are a few naturally intuitive trends we would expect to hold with this data series, which we will discuss in more detail

INITIAL DATA THOUGHTS

- Data Intuitions
  - The total amount of electricity consumption in a country is intuitively proportional to the population of the country.
    - Populations generally increase over time within a country. Therefore, we do not expect this process to be stationary
  - Electricity consumption is often related to outdoor temperature.
    - Therefore, we expect the process to have a seasonal component.

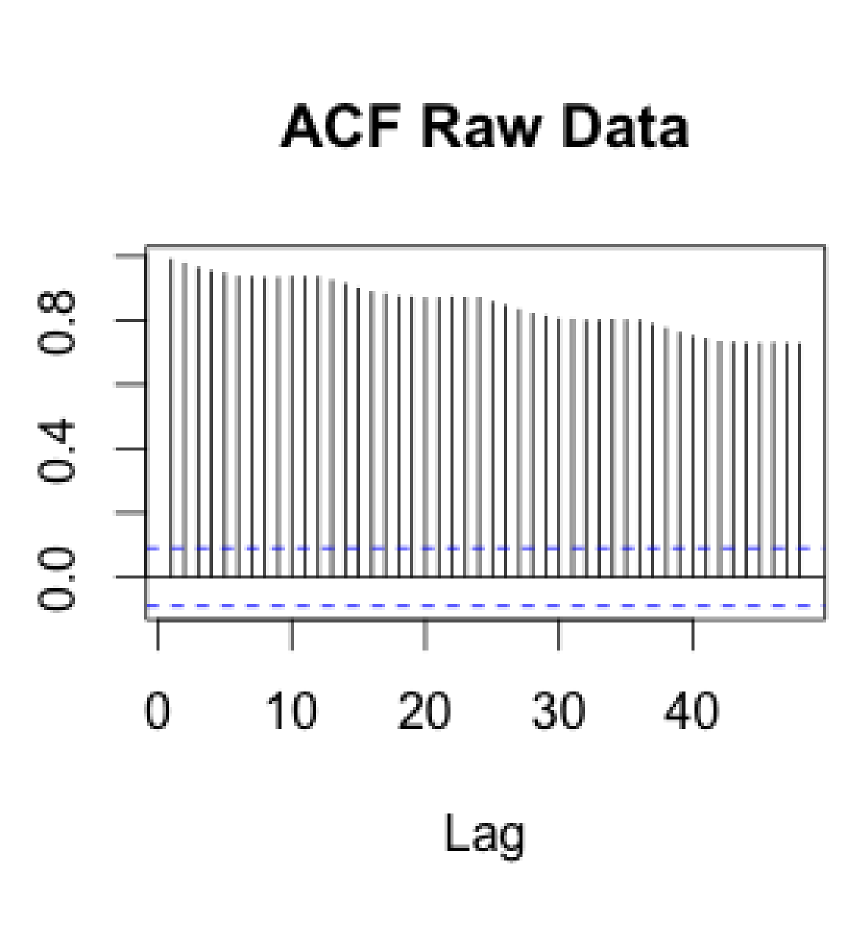
INITIAL DATA ANALYSIS



- An initial graph of the raw data verifies our intuition, and suggests we should further examine the time series for several components
  - Stationary
  - Seasonality
  - Moving Average term
  - Autoregressive term

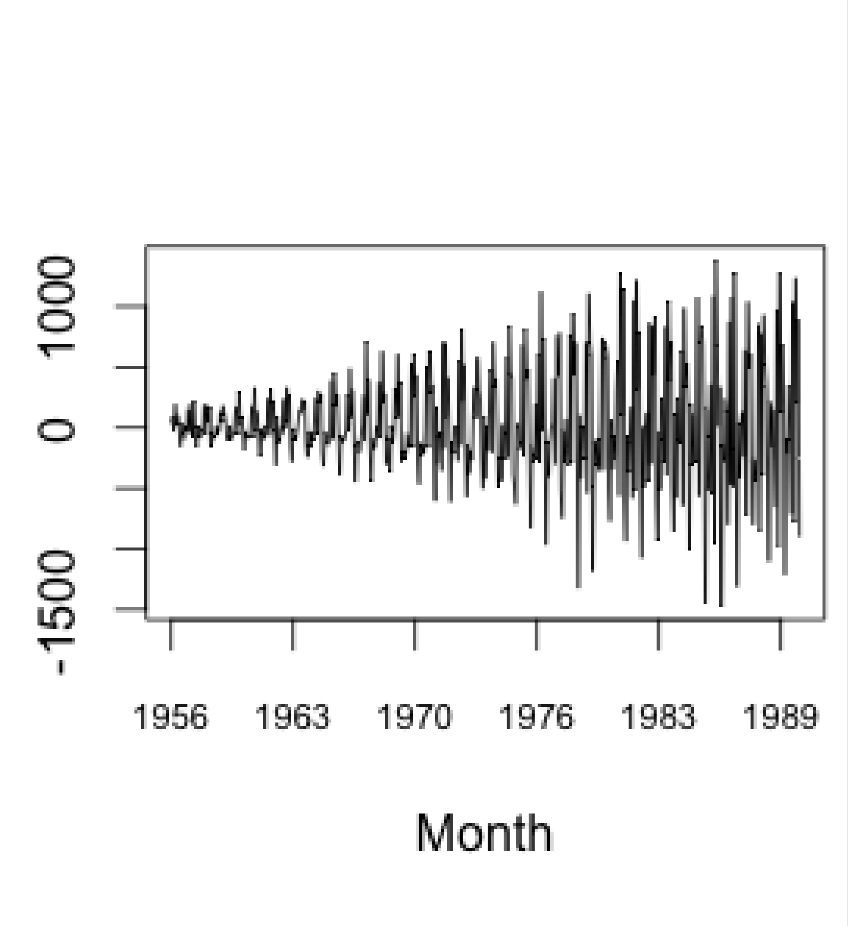
STATIONARITY AND DIFFERENCING

- A stationary time series is one whose properties such as mean, variance, autocorrelation are constant at all times
  - From the raw data on the previous slide, we see the average varies substantially over time
  - We will check if the first difference makes the process stationary



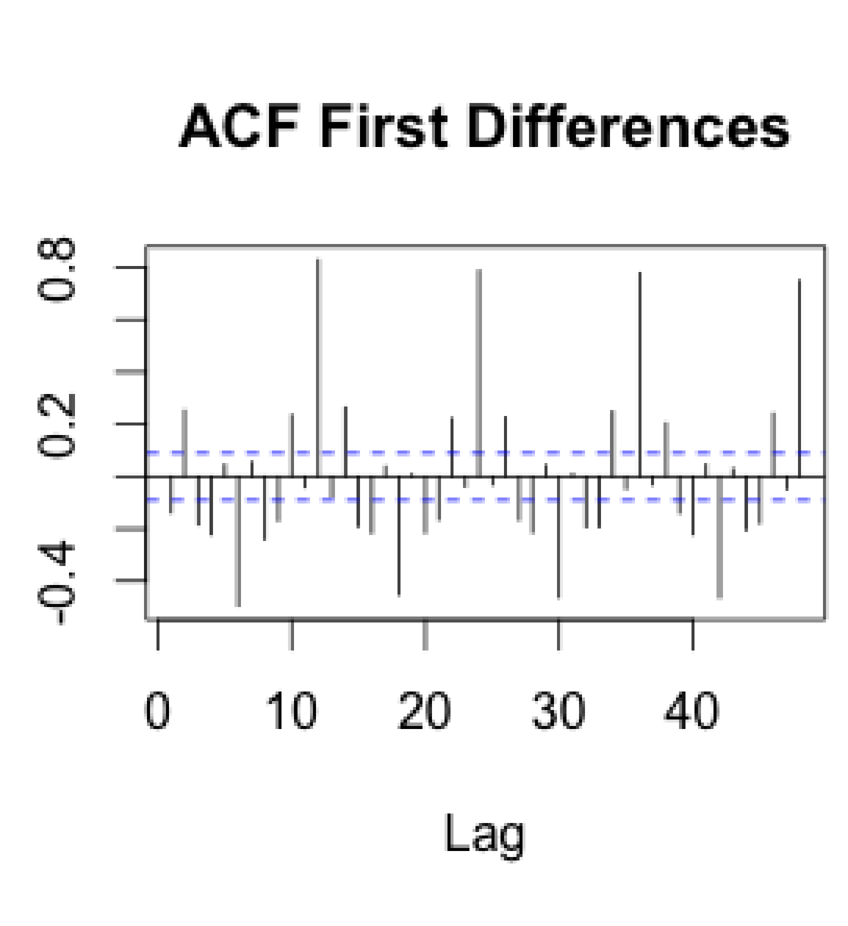
STATIONARITY AND DIFFERENCING

- Taking the first difference, we see the overall mean appears to be closer to 0, but there is still evidence of a strong seasonality trend.
- There also appears to be larger variance in later years
  - Changing Variance indicates a transformation should be ideally be investigated.



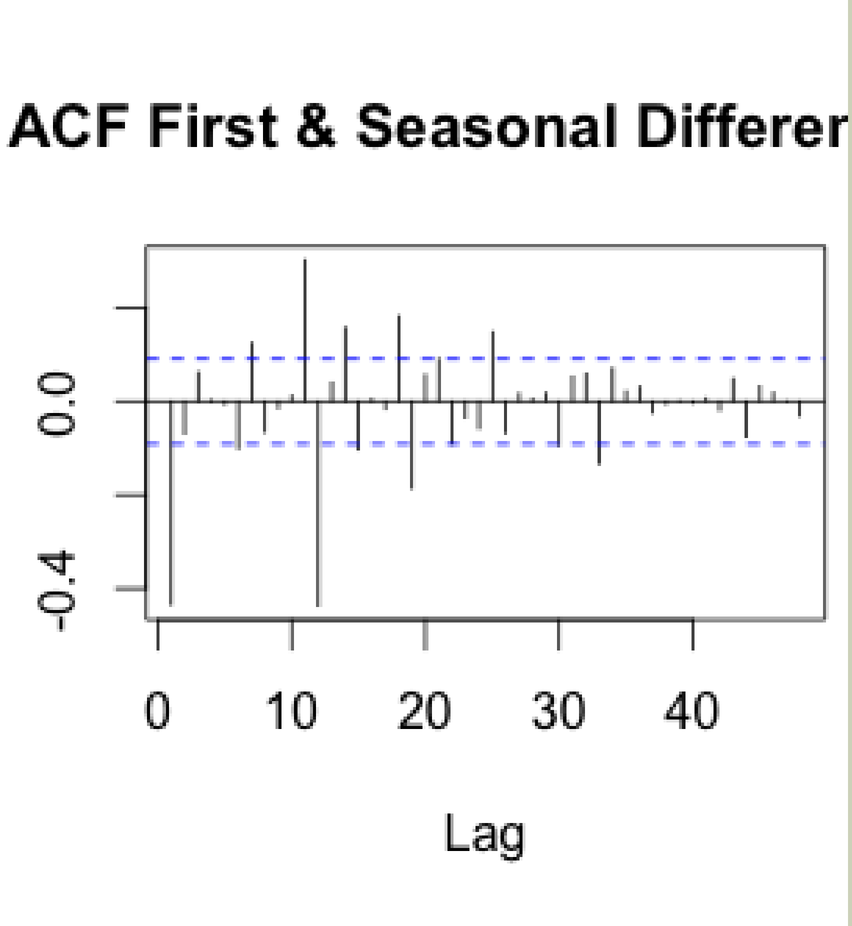
STATIONARITY AND DIFFERENCING

- Next, we plot the Autocorrelation Function of the First Difference
  - We see the sharp spikes at lags 12, 24, 36, 48, indicating a seasonal component is necessary



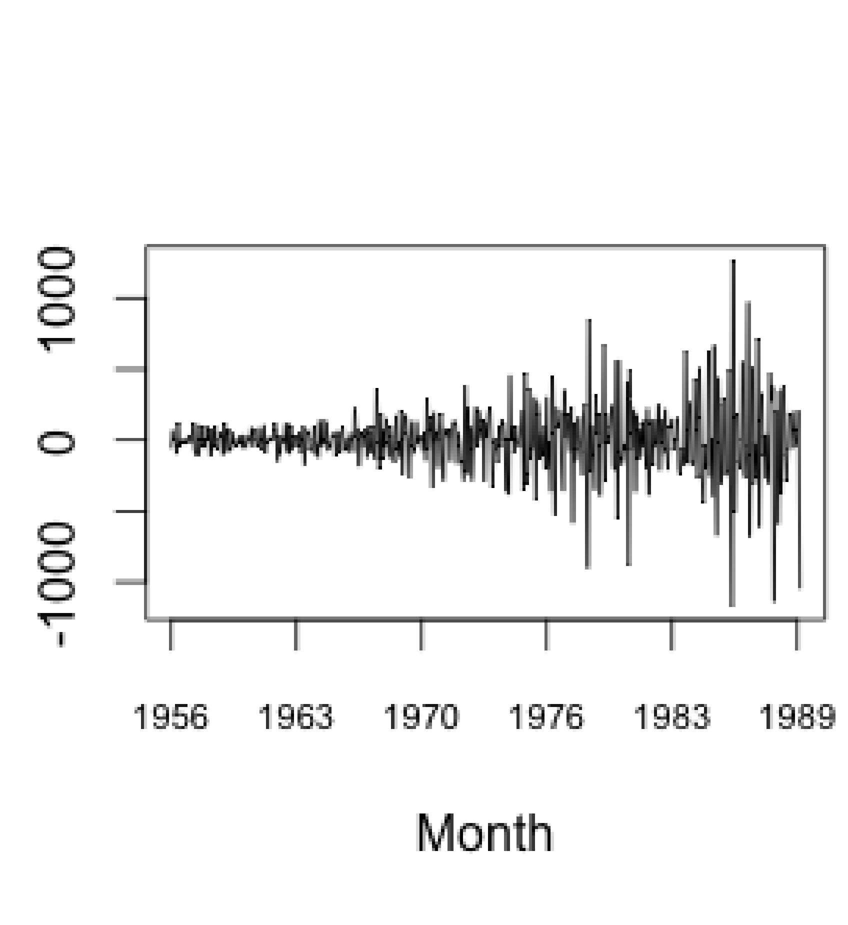
ADJUSTING FOR SEASONALITY

- Per the graph in the previous slide, we adjust for seasonality.
- To the right is the Autocorrelation function of the Seasonal difference



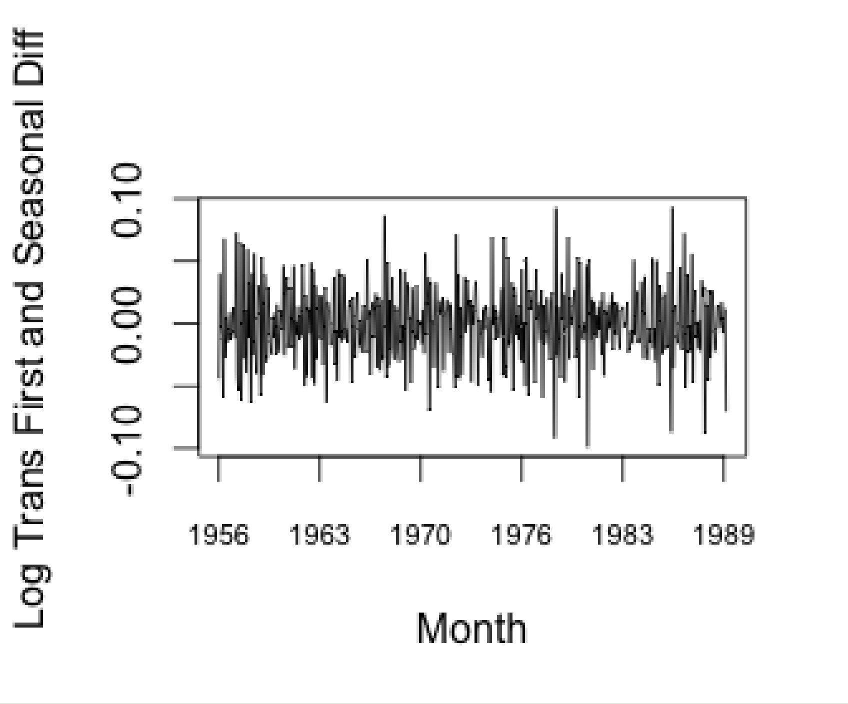
ADJUSTING FOR SEASONALITY

- We still see increasing Variance in later years, so we will investigate a simple natural log transform

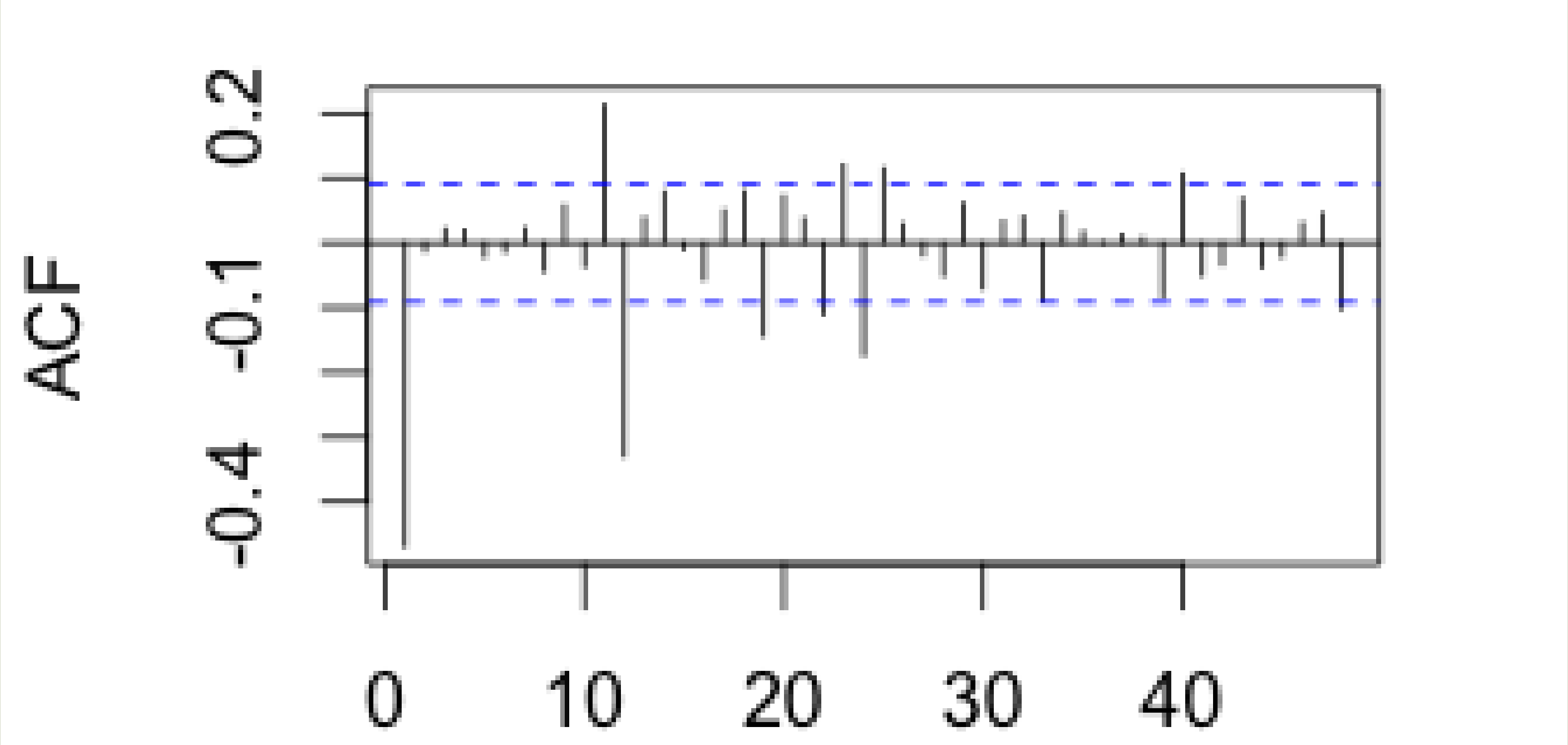


TRANSFORMATION

- Log-transformed – variance is a lot more consistent between older years and newer years



LOG TRANSFORMED ACF FIRST AND SEASONAL DIFFERENCES



- We will now replot the ACF for the log transformed variables

FITTING INITIAL ARIMA MODEL

- Based on previous graphs, we will begin to fit an ARIMA(0,1,1)x$(0,1,1)_{12}$ Model, or

$$\nabla 12\nabla Yt=et-\theta et-1-\Theta et-12+\theta\Theta et-13$$

- The resulting parameter estimates are displayed below

```
Coefficients:
          ma1      sma1
      -0.6711   -0.6761
s.e.   0.0373    0.0348

sigma^2 estimated as 0.0004375:  log likelihood
= 1129.58,  aic = -2255.15
```

INTERPRETATION OF ARIMA OUTPUT

- From the Output, we notice that the parameter estimates of the Moving Average term and seasonal Moving Average term are large with small standard errors.
- This indicates these terms are important and should be included in the ARIMA model.
- We also note that the Variance is very low, due to the log transformation.

ALTERNATE MODELS INVESTIGATED FOR COMPLETENESS - RULED OUT

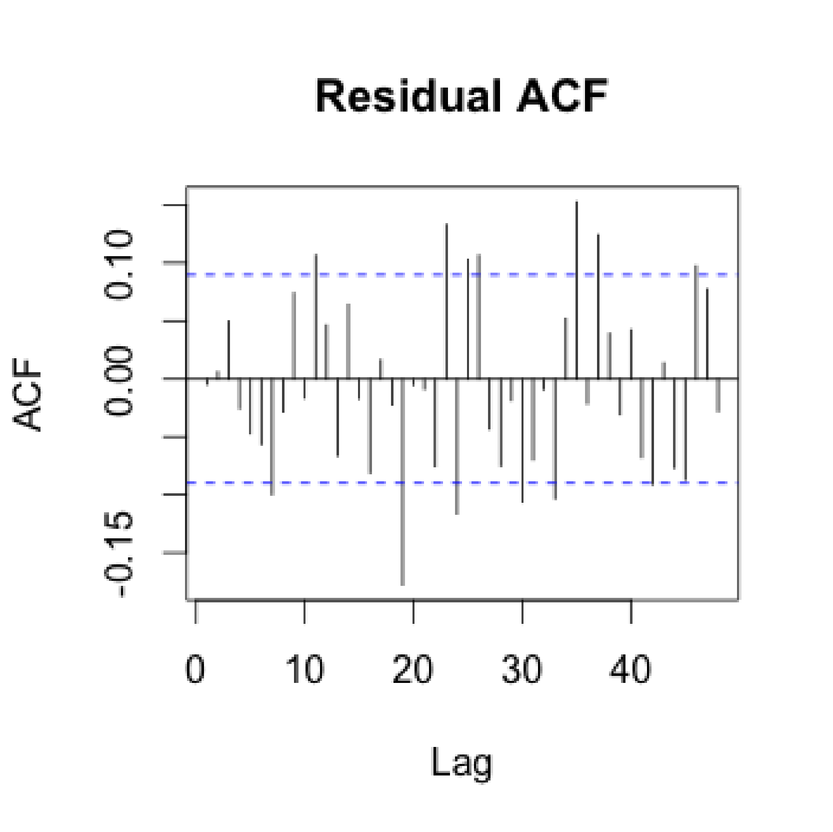
- Below, we include an AR term.
- Note the very small coefficient and large standard error.
- Therefore, we will not include this term

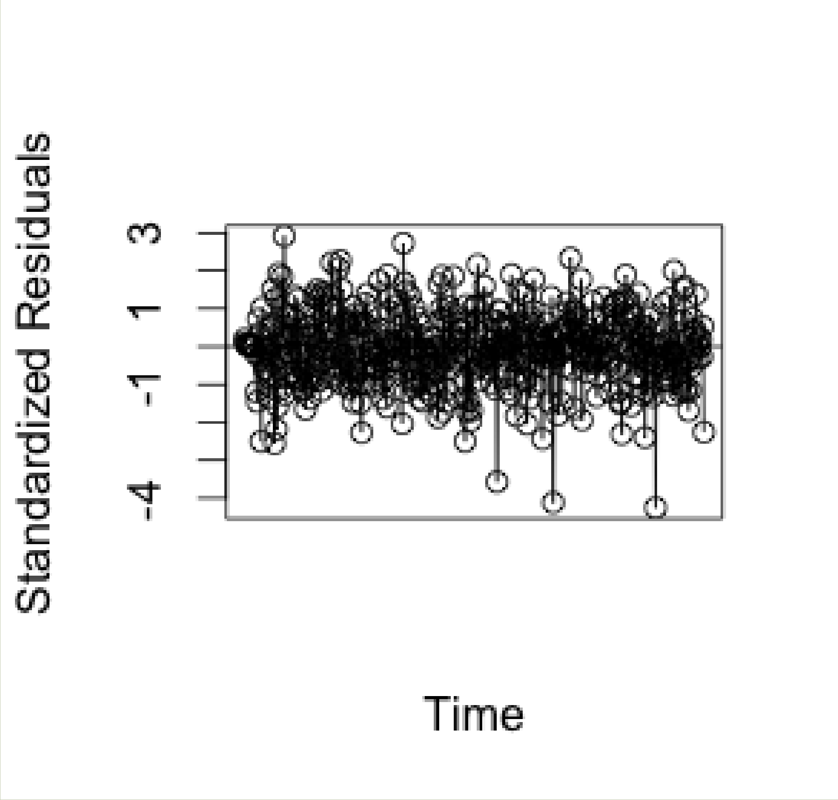
```
Coefficients:
          ar1      ma1     sma1
      -0.0173  -0.6607  -0.6743
s.e.   0.0779   0.0606   0.0358

sigma^2 estimated as 0.0004375:  log lik
elihood = 1129.6,  aic = -2253.2
```

- Below, we include a 2 period MA term (instead of 1 period).
- Note the very small coefficient and large standard error.
- Therefore, we will not include this term

```
Coefficients:
          ma1     ma2     sma1
      -0.6776  0.0109  -0.6744
s.e.   0.0484  0.0507   0.0358

sigma^2 estimated as 0.0004375:  log lik
elihood = 1129.6,  aic = -2253.2
```

ANALYSIS OF RESIDUALS

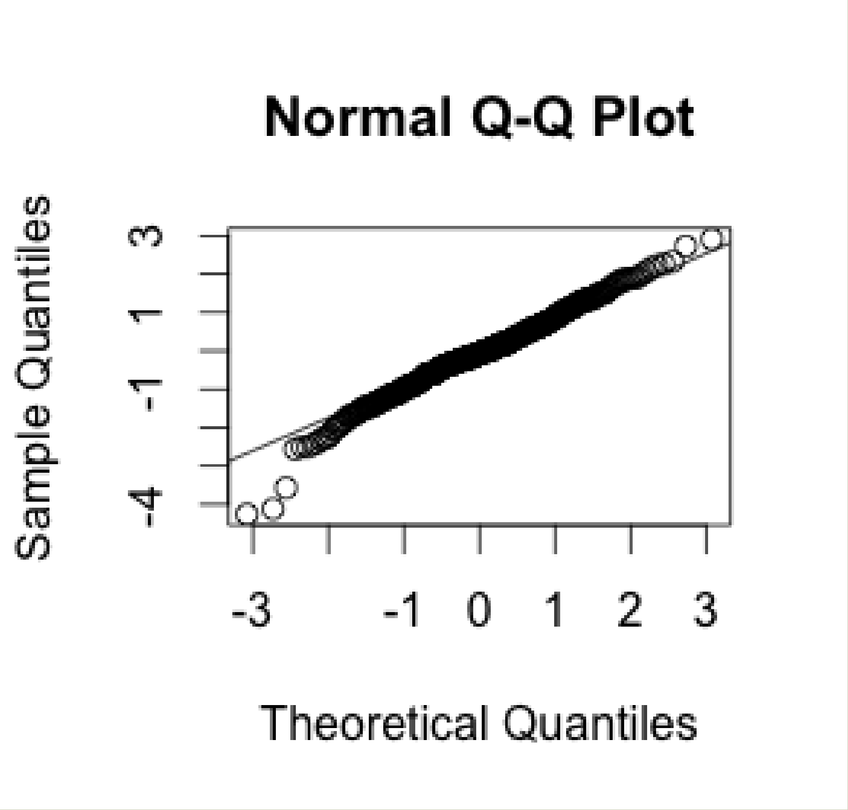




- For the most part, the autocorrelations appear insignificant
- Given the granularity of the data (monthly) the standardized residuals appear to be within reason.

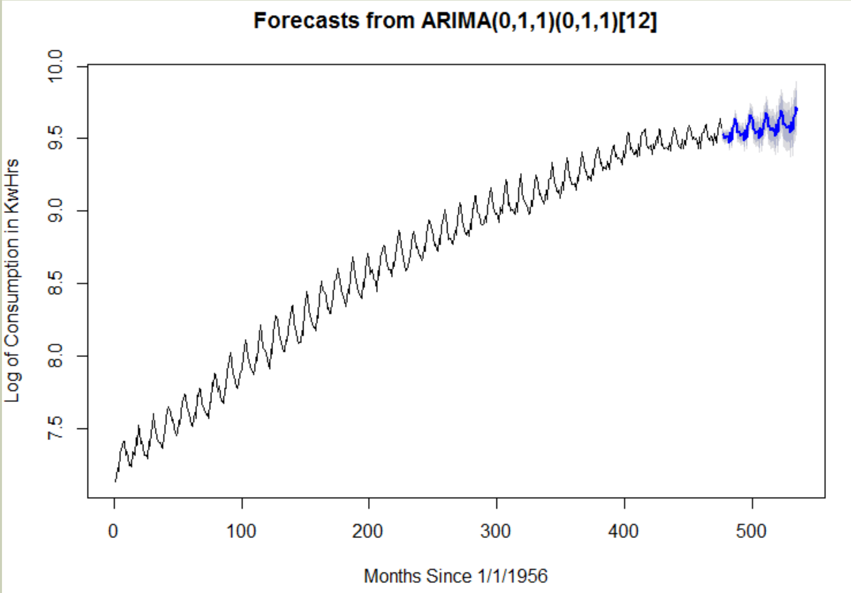
TESTING NORMALITY OF RESIDUALS

- To the right, we display the q-q plot of the residuals.
- We notice the residuals appear to be approximately Normal, deviating slightly on the left hand side indicating our data is slightly left skewed.



FORECASTING

- To the right, we see the result of our selected time series.
- We forecasted out 5 years (60 periods).
- The forecasted periods are in blue
- The gray shadow surrounding the blue show the 95% confidence values associated with the prediction



CONCLUSION

- Throughout this project, we have examined the time series data of electricity consumption in Australia between 1956 and 1995.
- We have tested for both stationarity and seasonality.
- Using our findings, we have constructed an ARIMA model (with seasonality)
- We analyzed the fit of the ARIMA Model, including an analysis of the residuals
- Using our final proposed model, we forecasted electricity consumption out 24 periods.