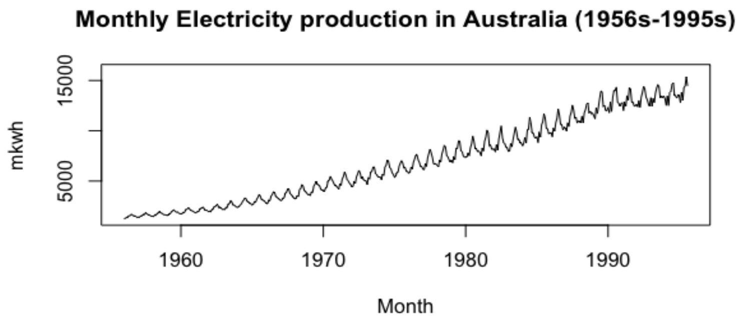
Monthly electricity production in Australia

By Matt Korzec

❑ Data Set



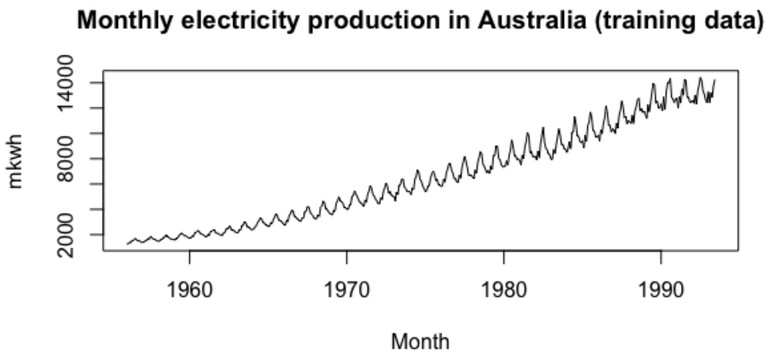
Time Range: 1956/1/1 - 1995/8/1

Sample Size: 476

Variables: Month and Million Kilowatt Hour

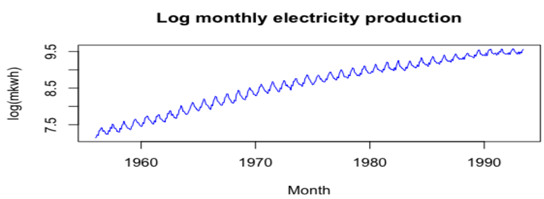
❑ Training Data

Training data: 450 & Test data: 26



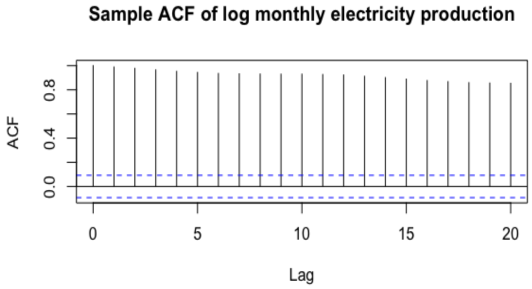
=> upward trend with increasing variance (fan shape)

❑ Log transformation



=> constant variance, increasing trend

❑ Check for weakly stationary

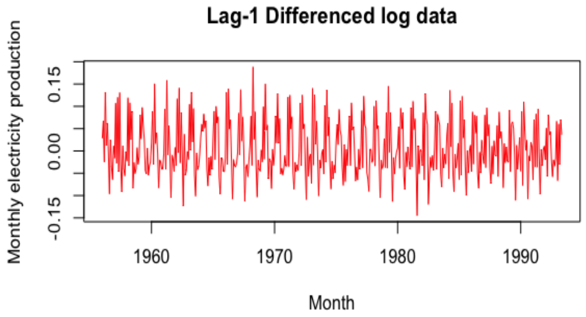


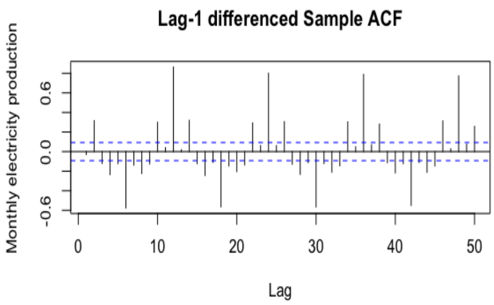

```
         Augmented Dickey-Fuller Test

data:  log_train
Dickey-Fuller = -1.6176, Lag order = 7, p-value = 0.7395
alternative hypothesis: stationary
```

=>sample ACF dies down slowly
=>unit-root non-stationary

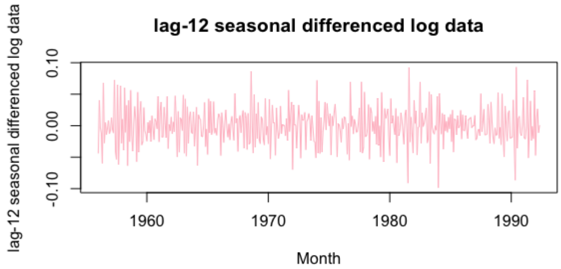
- ❑ **Lag-1 differencing**

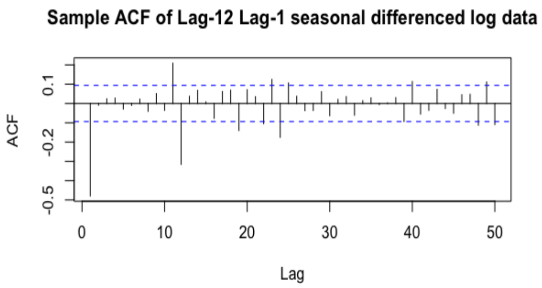




=>sample ACF dies down quickly
=>unit-root non-stationary removed
=> sample ACF dies down at lags with a period of 12
=> the period is 12

- ❑ **Lag-12 Seasonal differencing to remove seasonal non-stationary**





=> sample ACF cuts off quickly
=> seasonal non-stationary has been removed
=> weakly stationary

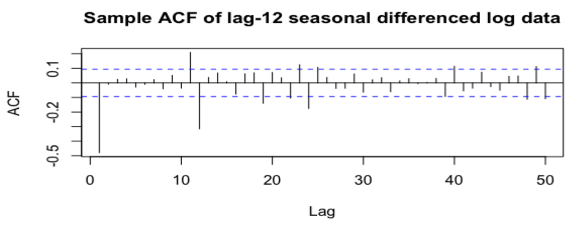
- ❑ **Test the mean of the resulting data to see if a intercept term is needed in the model**

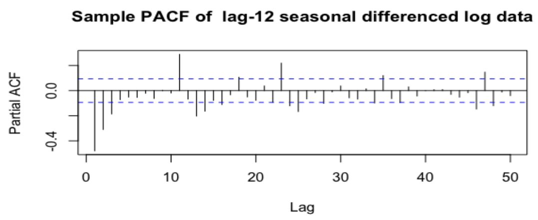
```
            One Sample t-test

data:  sdiff_train
t = -0.13563, df = 436, p-value = 0.8922
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -0.003035554  0.002643645
sample estimates:
    mean of x
-0.0001959544
```

=> fail to reject the null hypothesis
  => the mean is equal to 0
    => exclude the intercept in the model

- ❑ **Model Selection**





=> Sample ACF cuts off at lag-24;
Sample PACF dies down;
=> MA(2)_12 should be chosen

=> with in the first period, sample ACF cuts off at lag-1
=> MA(1) needed
=> we will compare several other models with SARIMA(0,1,1) x (0,1,2)_12

Model1: SARIMA(0,1,1) x $(0,1,2)_{12}$ Model

Model2: SARIMA(0,1,1) x $(0,1,1)_{12}$ Model

Model3: SARIMA(0,1,1) x $(1,1,0)_{12}$ Model

Model4: SARIMA(0,1,1) x $(1,1,1)_{12}$ Model

❑ Model Comparison

In-Sample Comparison

Model	SARIMA (0,1,1) x $(0,1,2)_{12}$	SARIMA (0,1,1) x $(0,1,1)_{12}$	SARIMA (0,1,1) x $(1,1,0)_{12}$	SARIMA (0,1,1) x $(1,1,1)_{12}$
AIC	-2126.185	-2123.582	-2051.789	-2124.926
RSE	0.02086	0.02096	0.02291	0.02089

Out-Of-Sample Comparison

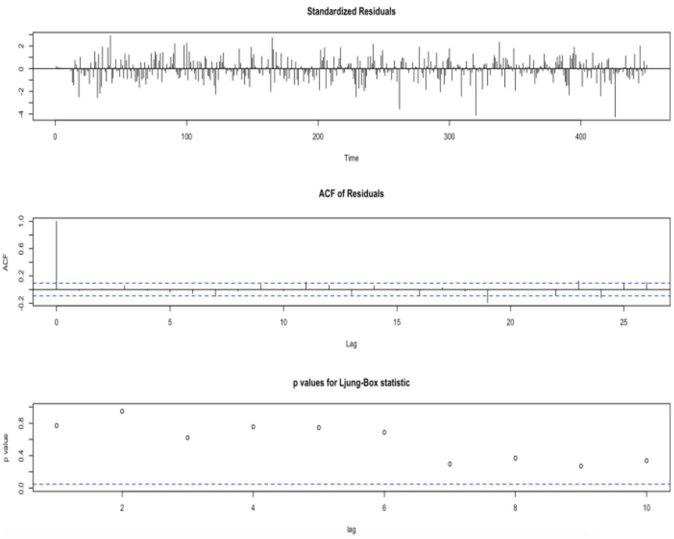
Model	SARIMA (0,1,1) x $(0,1,2)_{12}$	SARIMA (0,1,1) x $(0,1,1)_{12}$	SARIMA (0,1,1) x $(1,1,0)_{12}$	SARIMA (0,1,1) x $(1,1,1)_{12}$
RMSE	0.0161	0.0155	0.0176	0.0159
Mean absolute error	0.0123	0.0119	0.0134	0.0121

=> The four models are very close to each other. To keep our final model simple, we finally chose SARIMA(0, 1, 1) x (0, 1, 1)_12 model.

❑ Final Model: SARIMA(0,1,1) x $(0,1,1)_{12}$

$$(1-B)(1-B^{12})x_t = (1-0.6723B)(1-0.6811B^{12})a_t$$
$$with\ a_t \sim (0, 0.0004396)$$

❑ Check Adequacy of the model



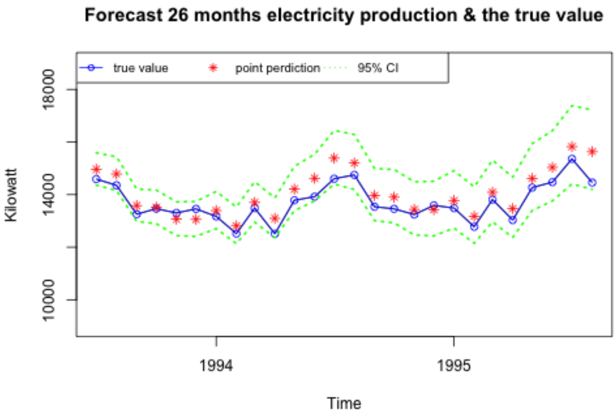

```
        Box-Ljung test

data:  model1$residual
X-squared = 11.246, df = 8, p-value = 0.1881
```

=> The model is adequate.

❑ Forecasting

Lower 95%CI	Upper 95% CI	Point Prediction
14359.52	15589.58	14961.91
14156.73	15435.68	14782.38
12980	14210.75	13581.44
12893.88	14171.91	13517.8
12439.23	13723.76	13065.72
12409.15	13740.28	13057.76
12706.71	14119.1	13394.3
12133.52	13527.95	12811.78
12953.2	14489.36	13699.77
12361.99	13872.24	13095.36
13391.62	15074.4	14208.12
13758.95	15534.83	14619.95
14400.77	16441.37	15387.28
14194.72	16282.14	15202.64
13012.05	14993.24	13967.56
12922.71	14955.74	13902.11
12463.97	14486.37	13437.18
12430.62	14507.55	13428.99
12725.32	14911.48	13775.1
12147.99	14291.05	13176.02
12965.05	15310.92	14089.25
12369.84	14662.92	13467.66
13396.33	15938.11	14612.06
13759.87	16429.59	15035.6
14406.66	17382.39	15824.74
14197.45	17217.78	15634.85



=> all testing data fall in the prediction interval with 95% confidence interval

=> the model works fine.