FORECASTING OF MONTHLY AVERAGE SOLAR RADIATIONS

MATH1307 ASSIGNMENT 2

Subbiah Soundarapandian
MASTER OF ANALYTICS-s3825012

Contents

INTRODUCTION

- This assignment mainly aims at the time series analysis and forecasting of monthly average horizontal solar radiations reaching the ground during a period of January 1960 and December 2014.
- We also take the monthly precipitation series recorded in the same period to see if there's a direct effect or lag effect between the two variables.
- The dataset for this assignment is taken from the canvas.
- Also, in this report we study whether spurious correlation exist between the quarterly Residential Property Price Index (PPI) in Melbourne and quarterly population change over the previous quarter in Victoria between September 2003 and December 2016.
- Hence this report can be split into two parts for the afore-mentioned tasks.

METHODOLOGY

TASK 1

- We check for stationarity and seasonal components of the time series plot to apply suitable models.
- Then we apply distributed lag models, dynamic linear models, and the exponential smoothing methods along with suitable state space model to find out the best model to achieve forecasting.
- Then the best model is decided based on the error measures such as RMSE, MASE etc and penalized measures such as "aic", "bic" etc.

TASK 2

- To check the existence of Spurious Correlation between two time series, we must make sure that two time series are stationary.
- Then we prewhiten the series to remove the leftover cross correlation between the series.

ANALYSIS

TASK 1 - FORECASTING OF SOLAR RADIATIONS

VISUAL ANALYSIS AND DESCRIPTIVE STATISTICS

- We begin the task 1 by visual analysis and descriptive statistics of the features.
- The given data was analysed with the help of R code in the R-Studio. The R-code is added in appendix A.
- To proceed with visual analysis, we can plot the time series of the target variable 'solar'.

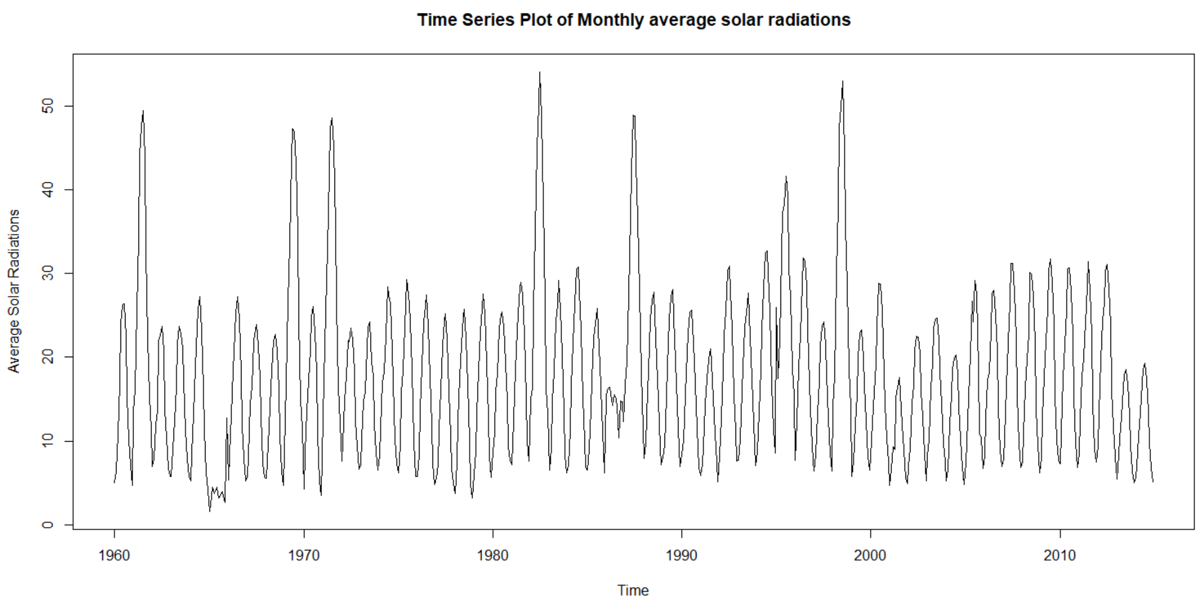

Figure 1 Time Series Plot of Monthly average solar radiations

Any time series plot can be talked on with 5 important points as follows,

- Trend
- Seasonality
- Changing Variance
- Intervention point
- Behaviour

Now let us have a look on how our time series plot is faring on the above-mentioned time series characteristics,

1. Trend – There is very less to no trend in the given time series.
2. Seasonality – There is an obvious seasonal pattern in the time series along the years.
3. Changing variance – Variance tends to change at different time points.
4. Intervention point – There is no change point or intervention point. Data point in 1965 looks more of an outlier rather than intervention point since there is no effect on the time series data after that year.
5. Behaviour – This time series data is mostly Moving Average behaviour.

Let us have a quick look at the time series plot of the predictor series as well,

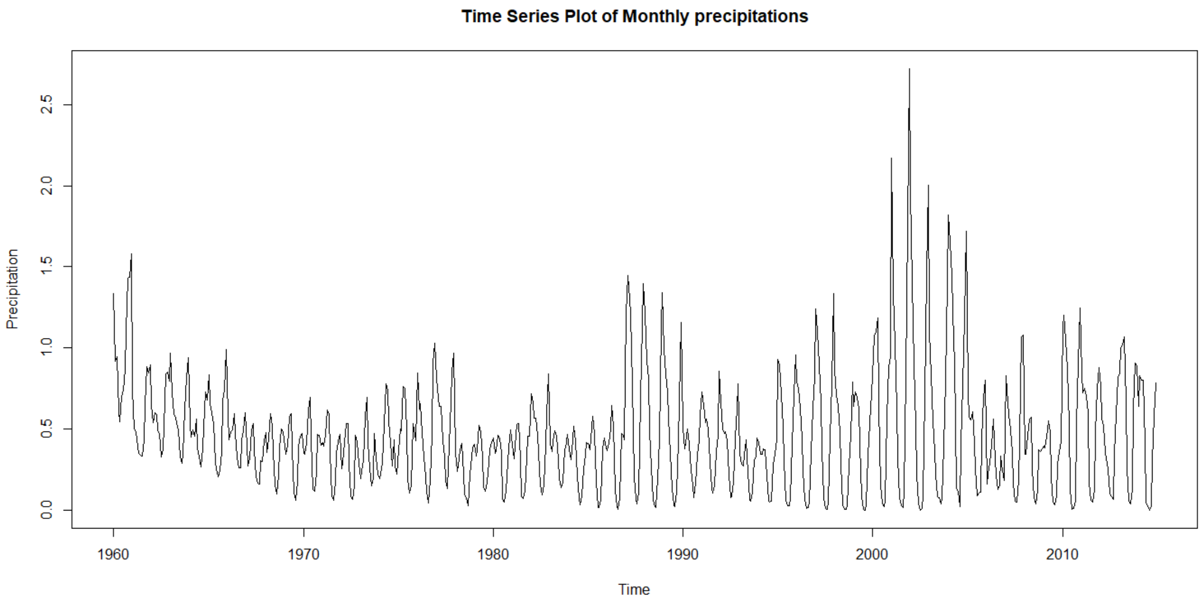

Figure 2 Time Series plot of the predictor series

The predictor series follow a similar time series pattern of the target time series variable and hence there could be a mild to good amount of correlation between them.

Now let us have a look at the descriptive statistics of the variable ‘solar’,

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.582	9.739	16.460	17.788	23.711	54.056

Figure 3 Descriptive statistics of solar radiations

- Above figure shows the descriptive statistics of the ‘solar’ variable.
- From figure 3, it is evident that 50% of average monthly solar radiations are above 16.46 w/m^2.
- 25% of the average monthly solar radiations are less than 9.739 while 25% of them are more than 23.711.
- Maximum solar radiation emitted is 54.056 w/m^2 while the minimum solar radiation emitted is 1.582 w/m^2.

Now, let us have a look at the correlation between the target time series and predictor series,

	solar	ppt
solar	1.0000000	-0.4540277
ppt	-0.4540277	1.0000000

Figure 4 Correlation Matrix - 'solar' & 'ppt' variable

From figure 4, we could safely say that both the time series are moderately negatively correlated. This information could aid us in developing the distributed lag model.

<u>AUTOCORRELATION</u>

We can check for autocorrelation in the horizontal average monthly solar radiations using ACF and PACF plot.

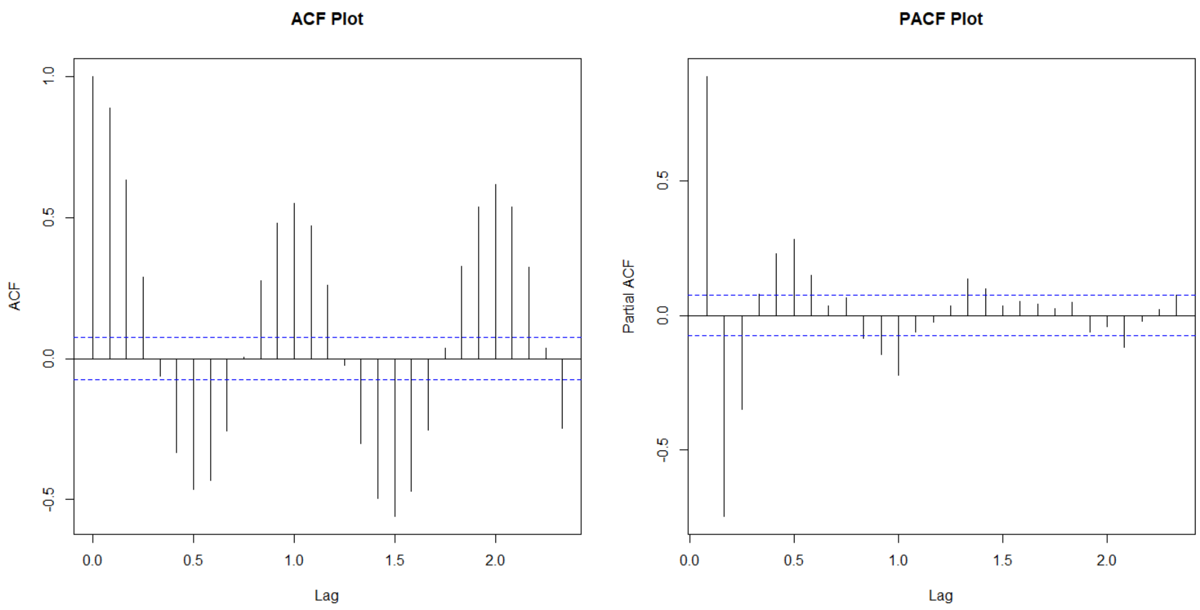

Figure 5 ACF and PACF Plot

From figure 5, wave like pattern in the ACF and PACF plot confirms the presence of seasonality in the given time series data.

<u>DECOMPOSITION</u>

- We can decompose the given time series to find out the trend and seasonal component present in it.

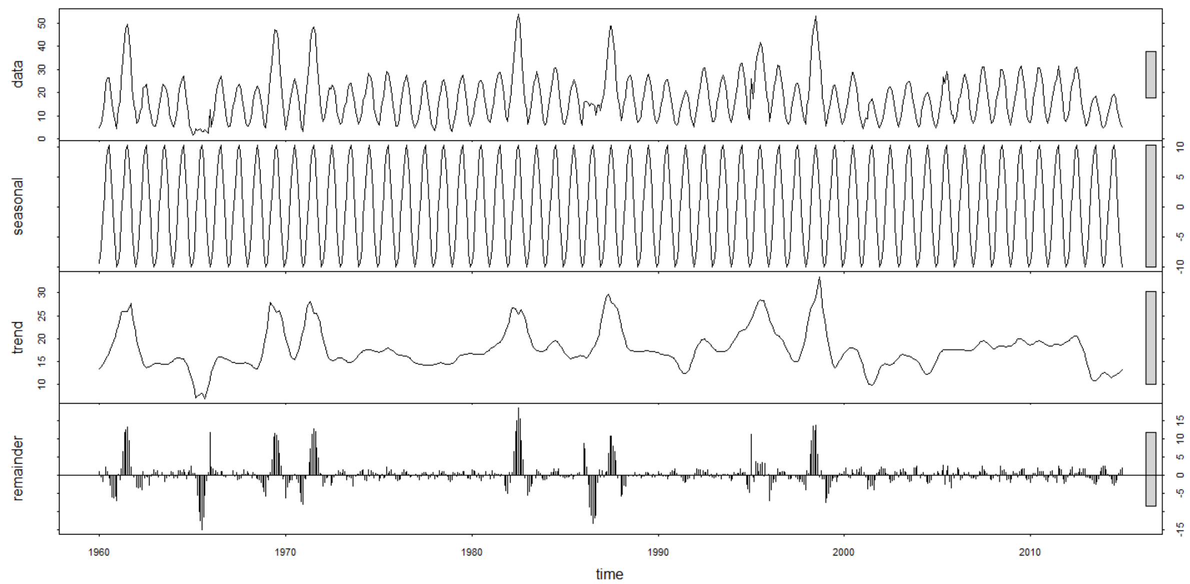

Figure 6 STL Decomposition - Solar radiations

- From the STL decomposition of the time series data above, it is clearly evident that there's a significant seasonality present.
- There is no trend, and the series tends to be stationary.

- However, we will confirm the same using the tests of non-stationarity.

TESTS OF NON-STATIONARITY:

We can confirm the stationarity of the given series using Augmented Dickey-Fuller test and Phillips-Perron test.

```
	Augmented Dickey-Fuller Test

data:  assign_ds$solar
Dickey-Fuller = -4.7661, Lag order = 8, p-value = 0.01
alternative hypothesis: stationary

	Phillips-Perron Unit Root Test

data:  assign_ds$solar
Dickey-Fuller = -8.4247, Truncation lag parameter = 6, p-value = 0.01
```

Figure 7 ADF and PP test - 'solar' variable

- Figure 7 shows that both adf and pp test rejects the null hypothesis that the series is non-stationary.

MODELLING

As mentioned in the methodology, we begin our data modelling by fitting different distributed lag models first.

Distributed Lag Models:

Default DLM:

We begin by fitting a default dlm model. The output is below,

```
Call:
lm(formula = model.formula, data = design)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.52798 -0.16830 -0.04918  0.11488  2.09202 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.656458   0.047540  13.808   <2e-16 ***
x.t          0.003542   0.004849   0.730   0.4654    
x.1         -0.012407   0.007209  -1.721   0.0857 .  
x.2         -0.005277   0.007197  -0.733   0.4637    
x.3          0.003623   0.007259   0.499   0.6179    
x.4          0.006274   0.007260   0.864   0.3878    
x.5          0.001118   0.007262   0.154   0.8777    
x.6         -0.009684   0.007279  -1.330   0.1839    
x.7          0.003709   0.007262   0.511   0.6097    
x.8          0.003945   0.007256   0.544   0.5869    
x.9          0.008526   0.007247   1.176   0.2399    
x.10        -0.001670   0.007185  -0.232   0.8163    
x.11        -0.005214   0.007196  -0.725   0.4690    
x.12        -0.007575   0.004837  -1.566   0.1179    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2853 on 634 degrees of freedom
Multiple R-squared:  0.338,	Adjusted R-squared:  0.3244 
F-statistic:  24.9 on 13 and 634 DF,  p-value: < 2.2e-16

AIC and BIC values for the model:
       AIC      BIC
1 229.4518 296.5602
```

Figure 8 DLM model - solar vs ppt

- Figure 8 shows the summary of DLM model fit to 'solar' variable against the 'ppt' independent variable.
- It is evident that no lags of the predictor series are significant. Even though p-value shows that this model could be significant, Adjusted R squared value shows that model explains only 32.44 % of the total variation of the dependent variable.
- Hence this model is not a good fit, and we can go for polynomial distributed lag model.

POLY-DLM:

Let us proceed with fitting Polynomial DL model to our target variable 'solar' against our explanatory variable 'ppt'.

```
Call:
"Y ~ (Intercept) + X.t"

Residuals:
     Min       1Q   Median       3Q      Max 
-0.54878 -0.18709 -0.03949  0.12155  2.07135 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  6.549e-01  4.501e-02   14.55   <2e-16 ***
z.t0        -6.830e-03  4.466e-04  -15.29   <2e-16 ***
z.t1         2.611e-03  1.590e-04   16.42   <2e-16 ***
z.t2        -1.937e-04  1.262e-05  -15.35   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2894 on 644 degrees of freedom
Multiple R-squared:  0.3081,	Adjusted R-squared:  0.3049 
F-statistic:  95.6 on 3 and 644 DF,  p-value: < 2.2e-16
```

Figure 9 - Poly DLM - 'solar' vs 'ppt'

- Above figure shows the summary of the polynomial distributed lag model fit to the solar radiation time series.
- Here all the z parameters are extremely significant, and the p-value is also below 0.05 to prove that this model is significant.
- But the adjusted R-Squared value is still low and hence it’s not good to proceed with forecasting using this model.

Koyck- DLM:

- Now let us fit Koyck-DLM model to the target variable ‘solar’ against our independent variable ‘ppt’.
- Koyck model uses the lag of the target variable itself while fitting the model.

```
Call:
"Y ~ (Intercept) + Y.1 + X.t"

Residuals:
     Min       1Q   Median       3Q      Max 
-0.72134 -0.11963 -0.02830  0.09304  1.27156 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.2412388  0.0247280   9.756  < 2e-16 ***
Y.1          0.7268717  0.0239547  30.344  < 2e-16 ***
X.t         -0.0064145  0.0009641  -6.653 6.05e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.206 on 656 degrees of freedom
Multiple R-Squared: 0.6616,	Adjusted R-squared: 0.6606 
Wald test: 615.4 on 2 and 656 DF,  p-value: < 2.2e-16 

Diagnostic tests:
NULL

                              alpha         beta       phi
Geometric coefficients:  0.8832435 -0.006414528 0.7268717
```

Figure 10 Koyck DLM - solar vs ppt

- It is evident that this model performs much better than the previous two models.
- All the parameters are 99% significant and p-value of the model is very less showing that this model is significant.
- Adjusted R-Squared value of 0.66 shows that this model explains 66.06% of total variation in the target variable ‘solar’.
- Now we can have a look at the residuals of this model to check the goodness of fit of this model.

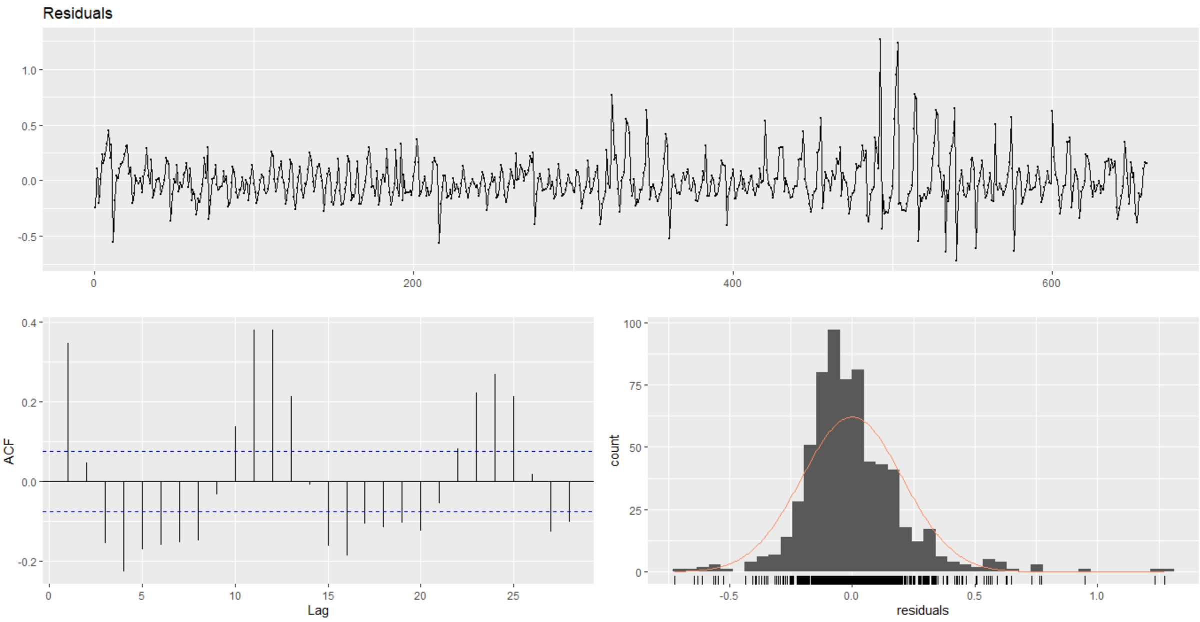

Figure 11Residuals - Koyck DLM

- Despite the residuals being stationary and normally distributed, the ACF plot shows that still some seasonal correlation is left in residuals.
- This shows that the model does not capture 100% of the correlation between the two time series data. Hence it is better not to go with this model for forecasting of solar radiations.

<u>Autoregressive Distributed Lag Model</u>:

- Now let us have a look at the last Distributed lag model.
- Since this model could take up different values of p and q parameters (autoregressive lag parameters), we use for loop and get the best one out of all the possible models.
- So, we got the model with p and q parameters 2 and 3 respectively, as the best model.
- It is reasonable we get higher MA (q) parameter since our target time series is mostly moving average.

```
Time series regression with "ts" data:
Start = 4, End = 660

Call:
dynlm(formula = as.formula(model.text), data = data, start = 1)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.83604 -0.09525 -0.01438  0.08487  1.27685 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.245930   0.025213   9.754  < 2e-16 ***
X.t         -0.003905   0.002553  -1.530  0.12661    
X.1         -0.005296   0.004297  -1.232  0.21824    
X.2          0.005081   0.002571   1.976  0.04856 *  
Y.1          1.009839   0.038709  26.088  < 2e-16 ***
Y.2         -0.269904   0.054360  -4.965 8.78e-07 ***
Y.3         -0.109669   0.038994  -2.812  0.00507 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1826 on 650 degrees of freedom
Multiple R-squared:  0.7351,	Adjusted R-squared:  0.7326 
F-statistic: 300.6 on 6 and 650 DF,  p-value: < 2.2e-16
```

Figure 12 ARDLM - solar vs ppt

- The ARDLM model selected with lowest aic and bic values among other models proved to be better than the previous models.
- The p-value of the model shows that this model is very much statistically significant.
- This model explains 73.26% of the total variation in the target variable.
- The second lag of the predictor variable and the first three lags of the target variable are 90% significant or greater.
- Now let us check the goodness of fit of this model using residuals.

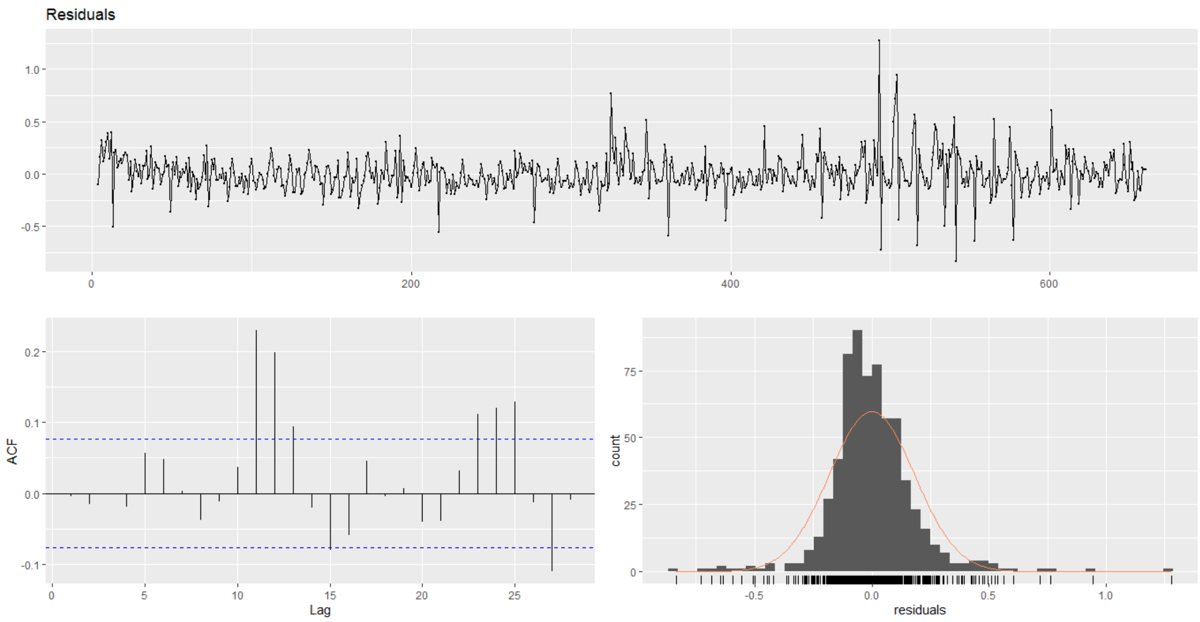

Figure 13 Residuals – ARDLM

- The residuals are much better compared to the previous model but still there is some significant correlation left in the ACF plot.
- Hence it is better to go with different set of time series regression models.

Dynamic Linear Models:

- We will move to a more general version of time series regression models called Dynamic Linear Models (DLM).
- In this model, we can feed the trend and seasonal component of the target time series data to find out if those have any effect.
- Since there is not any intervention point, we don't use step function to use as a pulse for intervention point.

Before proceeding with the fit, let us first transform the data to reduce the changing variance. This helps to get more clearer output.

We opt for Box-Cox transformation since it reduced the changing variance along the time series.

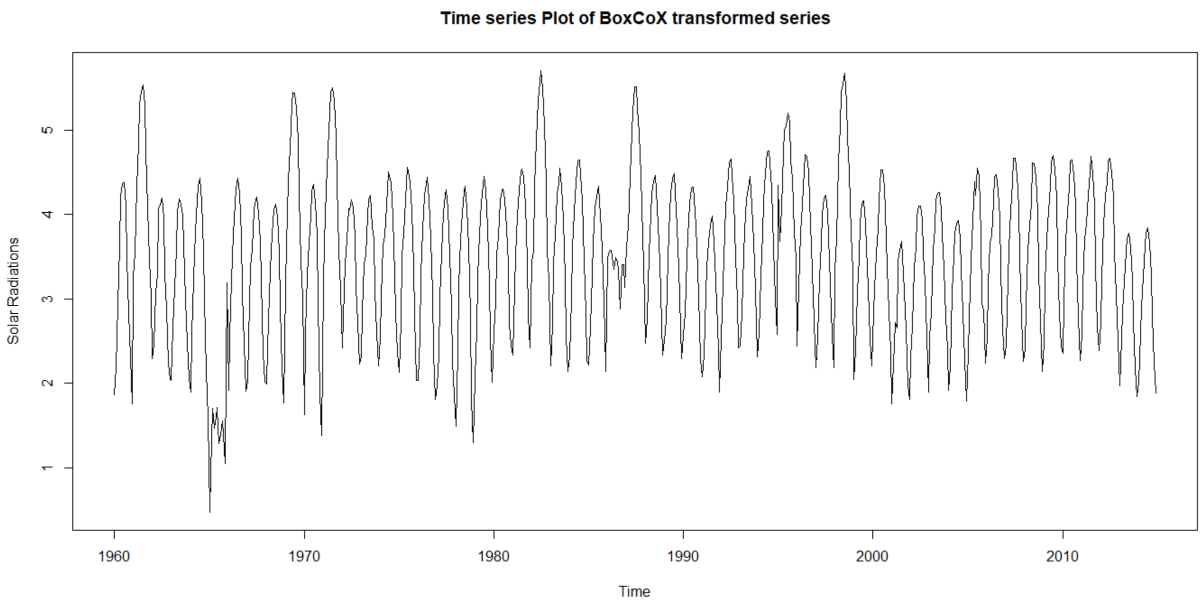

Figure 14 Box Cox transformed time series plot of solar radiations

- As we could see, the changing variance has been eliminated and hence we can proceed with dynamic linear model fit.

In the first model, we try to feed trend, seasonal component, and the first lag of the target variable ‘solar’,

```
Time series regression with "ts" data:
Start = 1960(2), End = 2014(12)

Call:
dynlm(formula = Y.t ~ L(Y.t, k = 1) + trend(Y.t) + season(Y.t))

Residuals:
     Min       1Q   Median       3Q      Max 
-1.39559 -0.07812 -0.00330  0.07887  2.32769 

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)     0.3758606  0.0586997   6.403 2.93e-10 ***
L(Y.t, k = 1)   0.8763872  0.0189699  46.199  < 2e-16 ***
trend(Y.t)      0.0002064  0.0006623   0.312 0.755437    
season(Y.t)Feb  0.3572822  0.0514735   6.941 9.50e-12 ***
season(Y.t)Mar  0.6107383  0.0524235  11.650  < 2e-16 ***
season(Y.t)Apr  0.3915906  0.0560503   6.986 7.04e-12 ***
season(Y.t)May  0.4683522  0.0589378   7.947 8.58e-15 ***
season(Y.t)Jun  0.3251289  0.0627439   5.182 2.94e-07 ***
season(Y.t)Jul  0.2121657  0.0648468   3.272 0.001126 ** 
season(Y.t)Aug -0.0167453  0.0654453  -0.256 0.798135    
season(Y.t)Sep -0.2380456  0.0633489  -3.758 0.000187 ***
season(Y.t)Oct -0.4609516  0.0594086  -7.759 3.36e-14 ***
season(Y.t)Nov -0.5971723  0.0548854 -10.880  < 2e-16 ***
season(Y.t)Dec -0.4409011  0.0519562  -8.486  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2685 on 645 degrees of freedom
Multiple R-squared:  0.9186,	Adjusted R-squared:  0.917 
F-statistic: 560.2 on 13 and 645 DF,  p-value: < 2.2e-16
```

Figure 15 Dynamic Linear Model 1

- Since there is no trend, the trend component is insignificant in this model. All the other parameters are 95% significant.
- This model seems be quite good in explaining the variation in the target variable since the adjusted R-Squared value is 0.917.
- The model is significant since the p-value is very much less than the significance value.

The residuals can be checked,

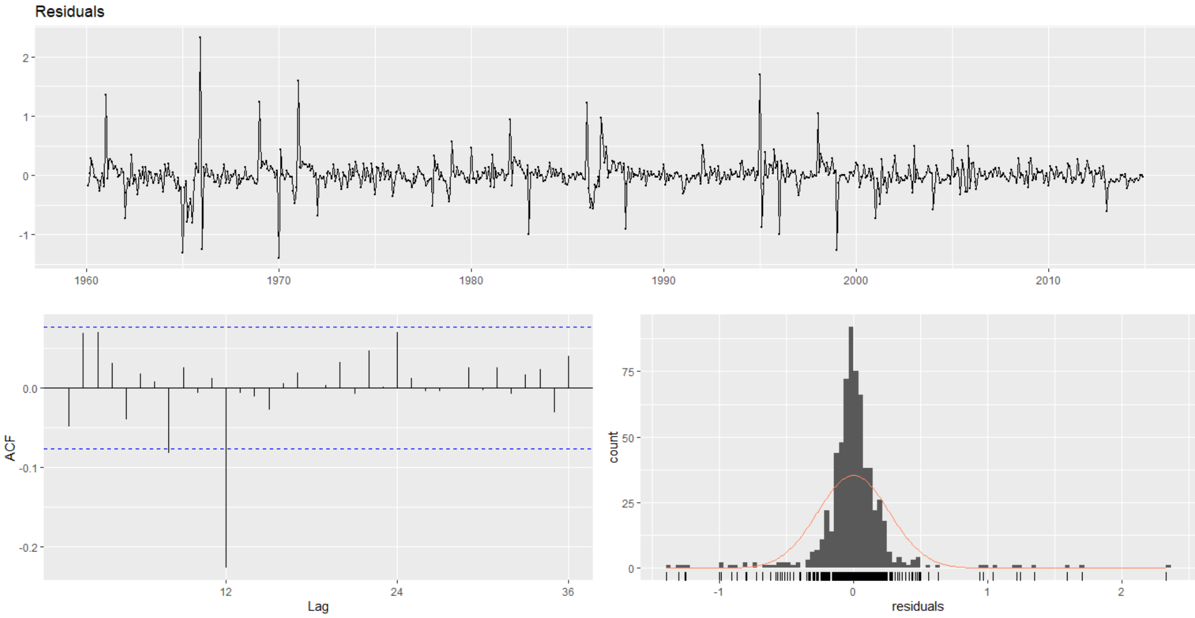

Figure 16 Residuals - dynlm-1

- The residuals give a good impression of the model but however there is still some autocorrelation left at the seasonal lags.
- Hence this model fails to capture the complete autocorrelation at the seasonal lags.

Since all the previous models fail to capture the seasonality completely, we go for exponential smoothing methods.

Exponential smoothing methods:

- Exponential smoothing methods are usually used to find point estimates while forecasting.
- Since this data has very minimal to low trend, it is better to skip smooth exponential model and the holt's linear model.
- Seasonality is obvious in our data and hence we go for holt winter model with additive & multiplicative seasonal components.
- We fit the afore mentioned model one by one and decide the best model based on the error measures and "AIC, BIC" values.

First, we fit the holt winter model with additive seasonal component.

```
Forecast method: Holt-Winters' additive method

Model Information:
Holt-Winters' additive method

Call:
 hw(y = BC.Solar, h = 24, seasonal = "additive")

  Smoothing parameters:
    alpha = 0.9157
    beta  = 1e-04
    gamma = 8e-04

  Initial states:
    l = 2.734
    b = 0.0057
    s = -1.2113 -0.8556 -0.2325 0.3467 0.731 0.9435
           0.9018 0.7435 0.3468 0.026 -0.6488 -1.091

  sigma:  0.277

     AIC     AICc      BIC
2608.292 2609.245 2684.660

Error measures:
                       ME      RMSE       MAE      MPE     MAPE      MASE        ACF1
Training set -0.005519953 0.2736518 0.1427283 -1.08965 5.426877 0.2560386 -0.02891377
```

Figure 17 Holt winter additive method

- The above figure shows the summary of the Holt's-winter model fit with additive seasonal component.
- Here the beta component seems to very less and close to zero. Hence this model becomes Holt-winter's model with a drift.
- This model's error measures seem to be very much better than all the previously fit models.

Before looking at the goodness of fit, let us look at the other models and then look at the residuals of the best model.

Next, we fit the Holt's-winter model with additive seasonal component and damped trend.

```
Forecast method: Damped Holt-Winters' additive method

Model Information:
Damped Holt-Winters' additive method

Call:
 hw(y = BC.Solar, h = 24, seasonal = "additive", damped = TRUE)

  Smoothing parameters:
    alpha = 0.8944
    beta  = 0.0174
    gamma = 1e-04
    phi   = 0.8

  Initial states:
    l = 2.7287
    b = -0.039
    s = -1.1733 -0.7801 -0.1742 0.3746 0.7593 0.9102
           0.859 0.6762 0.3213 -0.0288 -0.6567 -1.0875

  sigma:  0.277

     AIC     AICc      BIC
2608.888 2609.955 2689.748

Error measures:
                       ME      RMSE       MAE        MPE     MAPE      MASE        ACF1
Training set 0.0007738631 0.2733609 0.1420258 -0.7567698 5.341186 0.2547784 -0.01788672
```

Figure 18 Damped Holt-Winters' additive method

Next, we fit the Holt-Winters’ multiplicative seasonal model,

```
Model Information:
Holt-Winters' multiplicative method

Call:
 hw(y = BC.Solar, h = 24, seasonal = "multiplicative")

  Smoothing parameters:
    alpha = 0.7386
    beta  = 0.0223
    gamma = 1e-04

  Initial states:
    l = 2.6837
    b = 0.1147
    s = 0.6703 0.7813 0.9525 1.1061 1.2112 1.2643
           1.2482 1.1954 1.089 0.9892 0.8106 0.6818

  sigma:  0.1523

     AIC     AICc      BIC
3420.201 3421.155 3496.569

Error measures:
                       ME      RMSE       MAE      MPE     MAPE      MASE       ACF1
Training set -0.009545197 0.2806843 0.1563161 -1.10757 5.730384 0.2804136 0.08292415
```

Figure 19 Holt winters' multiplicative method

Finally, we fit the multiplicative exponential Holt winters’ method and get the error measures.

```
Model Information:
Holt-Winters' multiplicative method with exponential trend

Call:
 hw(y = BC.Solar, h = 24, seasonal = "multiplicative", exponential = TRUE)

  Smoothing parameters:
    alpha = 0.5745
    beta  = 1e-04
    gamma = 0.0488

  Initial states:
    l = 2.7768
    b = 0.9958
    s = 0.5832 0.8074 0.9696 1.1502 1.2614 1.3019
           1.2699 1.2313 1.088 0.9706 0.7755 0.5909

  sigma:  0.1682

     AIC     AICc      BIC
3543.276 3544.230 3619.645

Error measures:
                      ME      RMSE       MAE        MPE     MAPE      MASE      ACF1
Training set 0.008795318 0.2866884 0.1611848 -0.5008632 5.858259 0.2891475 0.2073562
```

Figure 20 Holt-winters' multiplicative method with exponential trend

Now that we have fit all the possible Holt-winters' models, we could tabulate the error measures and proceed with which model is the best to do forecasting with.

Model/Error Measures	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Default-DLM	-	-	-	-	-	1.215	-
Poly-DLM	-	-	-	-	-	1.272	-
Koyck-DLM	-	-	-	-	-	0.881	-
ARDLM	-	-	-	-	-	0.761	-
Dynamic Linear model	3.86E-18	0.265675	0.143361	-1.1736	5.34166	0.348886	-0.04874
Holt winter - Seasonal	-0.00552	0.273652	0.142728	-1.08965	5.426877	0.256039	-0.02891
Damped Holt-winters' with additive seasonal	7.74E-04	0.273361	0.142026	-0.75677	5.341186	0.254778	-0.01789
Holt-winters' multiplicative	-0.00955	0.280684	0.156316	-1.10757	5.730384	0.280414	0.082924
Holt-winters' multiplicative method with exponential trend	0.008795	0.286688	0.161185	-0.50086	5.858259	0.289148	0.207356

Table 1 – Tabulation of error measures of all the models fit

- From the table 1, it is clearly evident that Damped Holt-winters' method with additive seasonal component performs better in terms of all the error measures except the Mean percentage error.
- Hence, it is safe to say that Holt-winter's additive seasonal model with damped trend is the model with which the forecasting can be done.

Let us have a look at the residuals of this model,

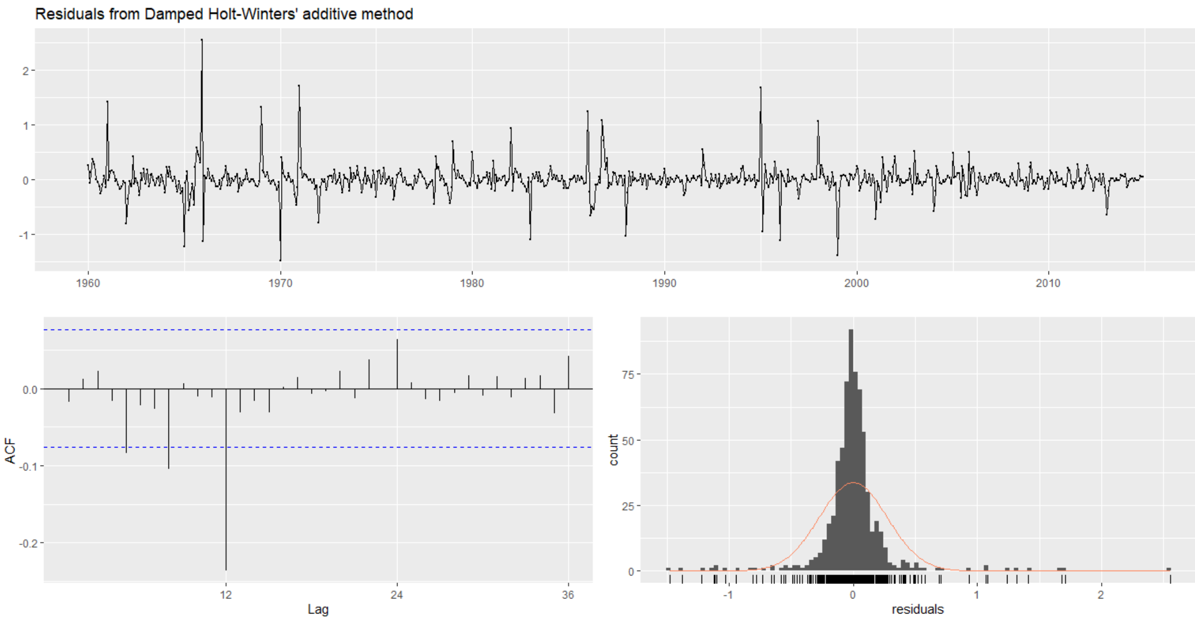

Figure 21 Residuals - Holt-Winters' additive method

- Residuals are normally distributed and stationary in nature but still some autocorrelation left in seasonal lags.
- But we will stick to this model due to the lowest of error measures among all the other models.
- However, let's explore the state space models of the Holt-winter's model and see if there is a better model than the current one.

State Space Models

- This could be done with the help of ets() function in R. However, it's better to find the auto best fit to the data by selecting the model with lowest AIC and BIC values.
- When tried the above method in R, we got ets(A,A,A) as the best model with minimum AIC and BIC values.

```
ETS(A,A,A)

Call:
 ets(y = BC.Solar, model = "ZZZ", ic = "aic")

  Smoothing parameters:
    alpha = 0.9157
    beta  = 1e-04
    gamma = 8e-04

  Initial states:
    l = 2.734
    b = 0.0057
    s = -1.2113 -0.8556 -0.2325 0.3467 0.731 0.9435
           0.9018 0.7435 0.3468 0.026 -0.6488 -1.091

  sigma:  0.277

     AIC     AICc      BIC
2608.292 2609.245 2684.660

Training set error measures:
                       ME      RMSE       MAE      MPE     MAPE      MASE        ACF1
Training set -0.005519953 0.2736518 0.1427283 -1.08965 5.426877 0.2560386 -0.02891377
```

Figure 22 Automatic fit of state space model of exponential smoothing model

- Even though this model has the lowest AIC or BIC values, damped holt-winters' additive method had better error measures.
- Hence it is better to proceed with forecasting using the damped Holt-winters' additive model.

FORECASTING

Now that we have got our best model, we can proceed with forecasting of horizontal solar radiations falling on earth for the next two years.

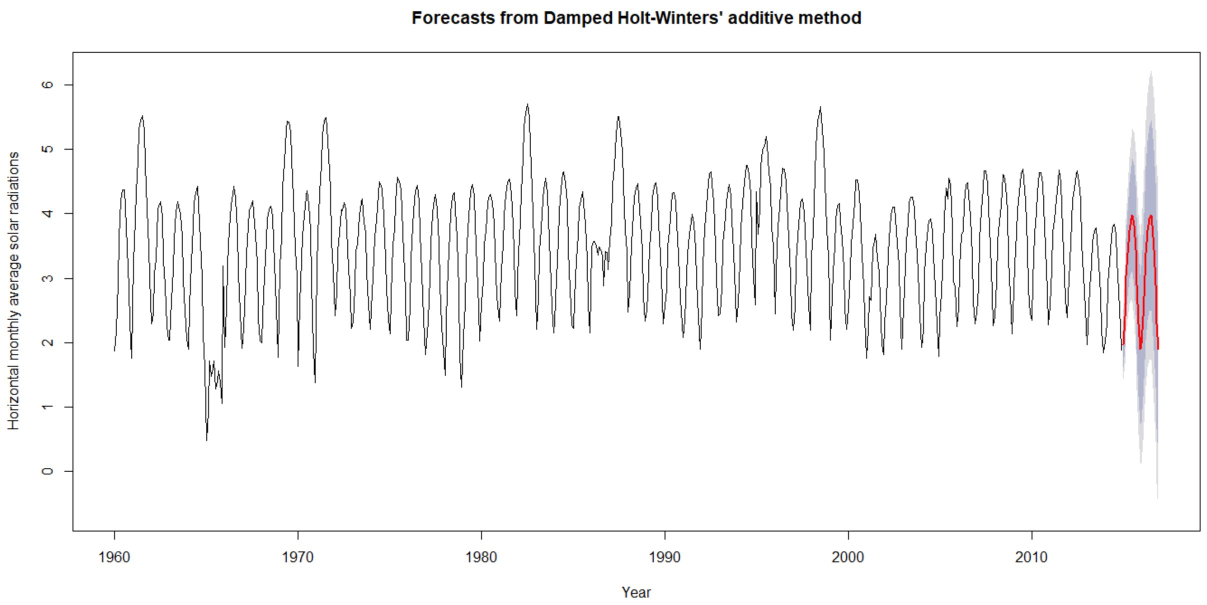

Figure 23 Two Year forecasts of solar radiations from Damped Holt-Winters' additive method

Forecasts:

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2015	1.971580	1.6166530	2.326507	1.4287661	2.514394
Feb 2015	2.403611	1.9241281	2.883095	1.6703050	3.136918
Mar 2015	3.032471	2.4524704	3.612472	2.1454365	3.919506
Apr 2015	3.383288	2.7162253	4.050351	2.3631034	4.403473
May 2015	3.738971	2.9938409	4.484102	2.5993926	4.878550
Jun 2015	3.922139	3.1055515	4.738726	2.6732763	5.171001
Jul 2015	3.973779	3.0908892	4.856669	2.6235155	5.324043
Aug 2015	3.822976	2.8779641	4.767987	2.3777053	5.268246
Sep 2015	3.438681	2.4350387	4.442324	1.9037424	4.973620
Oct 2015	2.890072	1.8307744	3.949370	1.2700161	4.510128
Nov 2015	2.284320	1.1719465	3.396694	0.5830915	3.985549
Dec 2015	1.891345	0.7281601	3.054530	0.1124073	3.670283
Jan 2016	1.977363	0.7653657	3.189360	0.1237732	3.830952
Feb 2016	2.408238	1.1492377	3.667238	0.4827633	4.333712
Mar 2016	3.036172	1.7317910	4.340554	1.0412932	5.031052
Apr 2016	3.386249	2.0379579	4.734540	1.3242157	5.448282
May 2016	3.741340	2.3504820	5.132198	1.6142063	5.868474
Jun 2016	3.924034	2.4918402	5.356227	1.7336828	6.114384
Jul 2016	3.975295	2.5029006	5.447690	1.7234620	6.227128
Aug 2016	3.824188	2.3126412	5.335736	1.5124765	6.135900
Sep 2016	3.439652	1.8899249	4.989378	1.0695492	5.809754
Oct 2016	2.890848	1.3038483	4.477848	0.4637412	5.317955
Nov 2016	2.284941	0.6615135	3.908369	-0.1978771	4.767759
Dec 2016	1.891842	0.2327788	3.550905	-0.6454760	4.429159

Figure 24 Point forecasts with confidence intervals

- Figure 22 shows the plot of 24-month forecast of solar radiations while figure 23 shows the point forecasts with 80% and 95% confidence intervals.

TASK 2 - DEMONSTRATION OF SPURIOUS CORRELATION

- In this task, we will look at the time series plot of the quarterly Residential Property Price Index (PPI) in Melbourne and time series plot of the quarterly population change over the previous quarter in Victoria between September 2003 and December 2016.

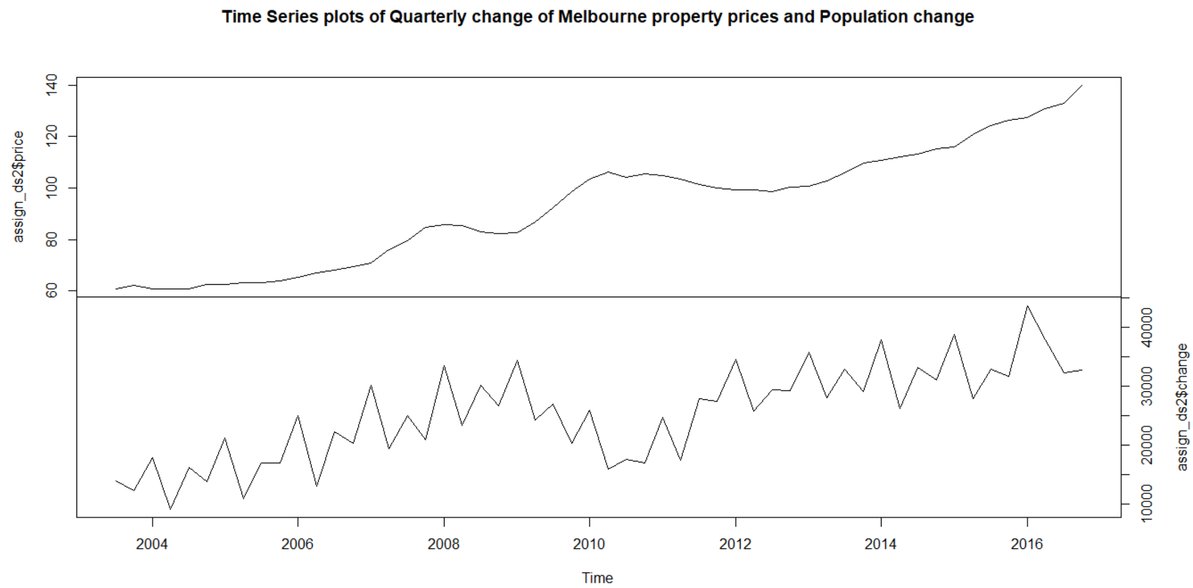

Figure 25 Intersection of property price index and population change time series plots

- From figure 25, it is evident that both the time series have a linear positive trend.
- Hence there might be some spurious correlation between the two.

We can confirm the same using the cross-correlation plot,

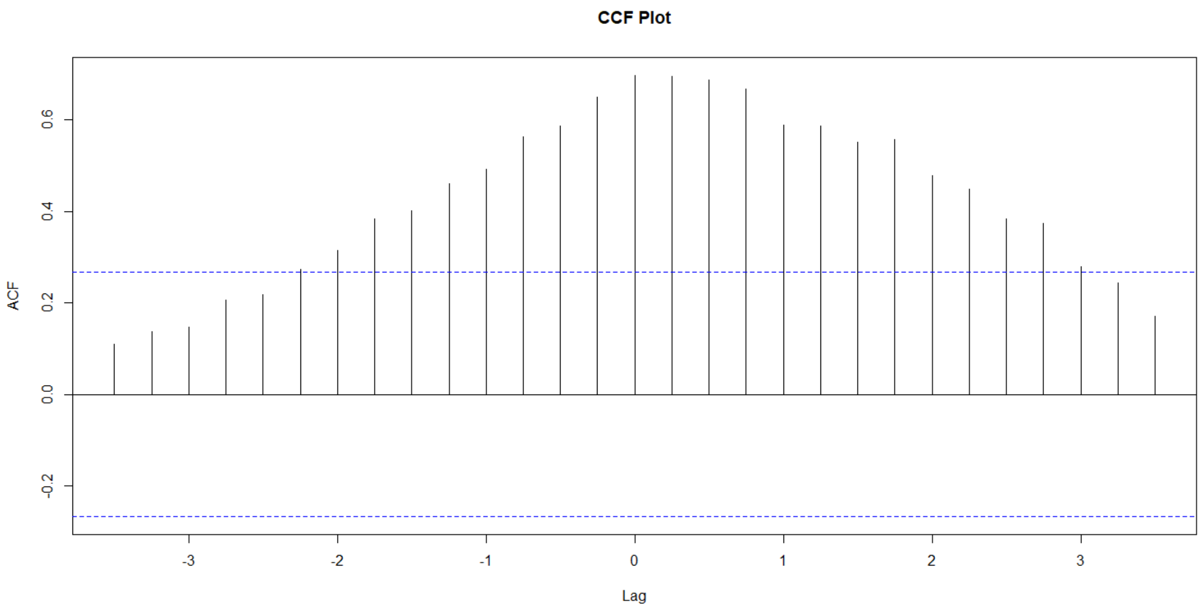

Figure 26 Cross correlation plot

- Figure 26 shows that at lag zero, there is a cross correlation of around 0.7.
- This is a spurious correlation since it makes no sense that there is so much correlation between property price indices and population change.
- This can be dealt by pre-whitening the series.

To pre-whiten, both the series should be stationary. Hence both the series are differenced up to certain levels until they become completely stationary.

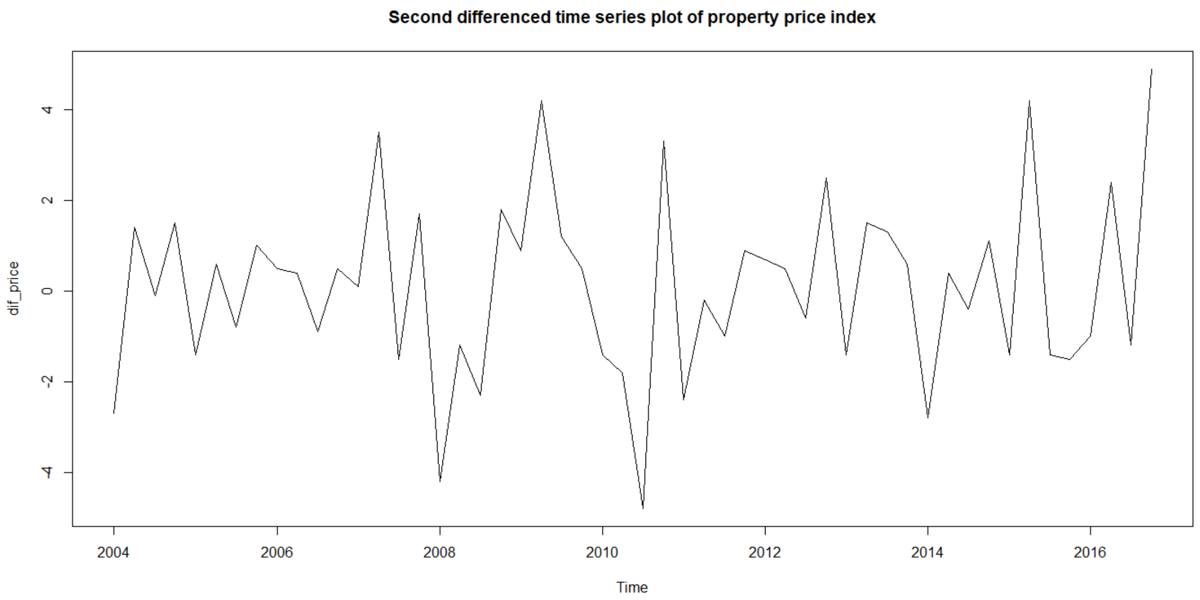

Figure 27 Second differenced time series plot of PPI

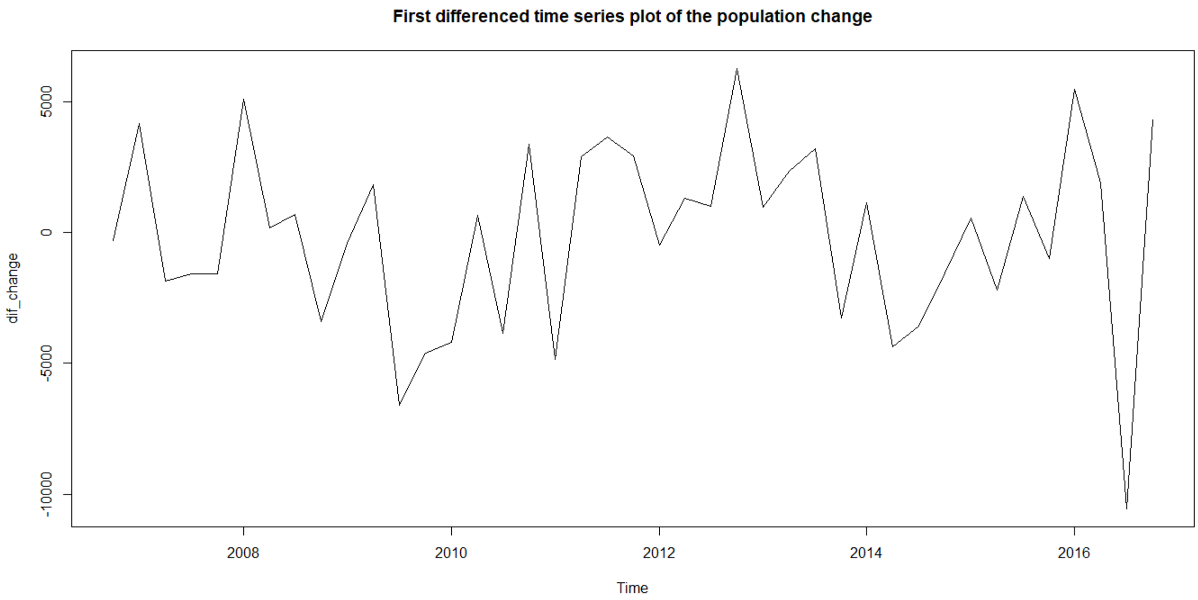

Figure 28 First differenced time series plot of population change

- Figure 27 and 28 confirms that both the time series are stationary after second differencing and first differencing of the Melbourne Quarterly PPI data and Population change data, respectively.
- Now we can prewhiten the two plots and look at the cross-correlation of the same.

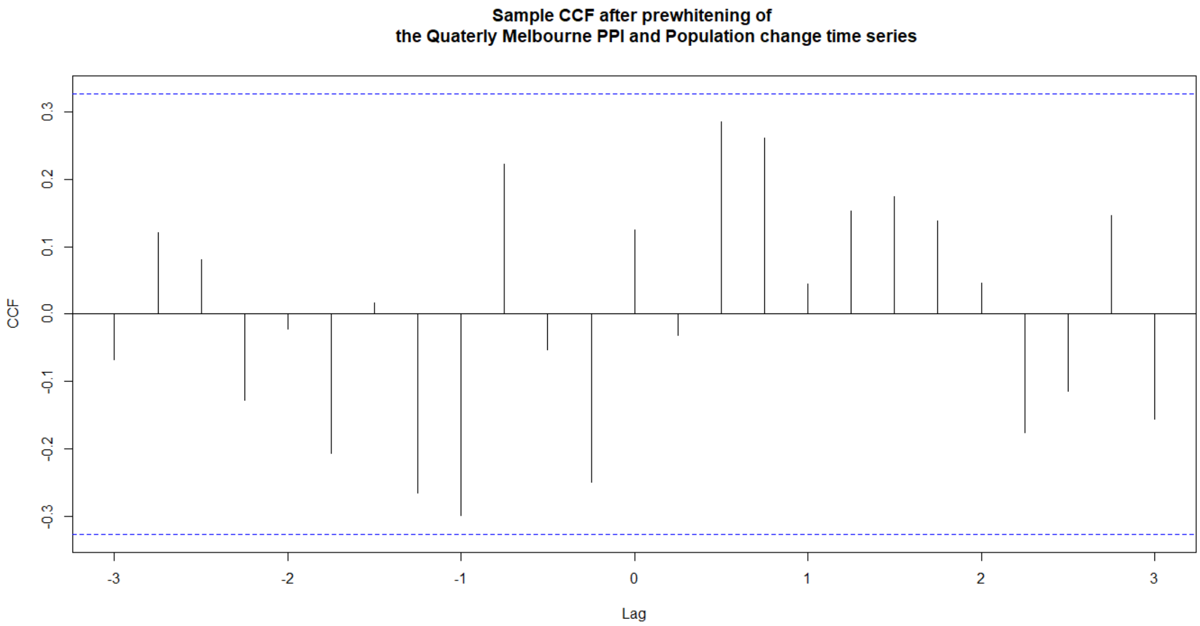

Figure 29 CCF Plot after prewhitening

- Figure 29 shows the sample cross correlation plot of the two given time series.
- Hence it is safe to say that after prewhitening, the spurious correlation between the two series has been eliminated.

CONCLUSION

This report studied the efficiency of different time series regression models on the emission of monthly average solar radiation data and its predictor variable precipitations recorded on the same time points. We applied distributed lag models, dynamic linear models, and the exponential smoothing methods with relevant state space models. At the end, Holt-winters' damped additive model with very less Root Mean Square value of 0.25 proved to be a good model. Hence forecasting is done with the Holt-winters' damped additive model for the next two years i.e., Jan 2015 to Dec 2016.

The Second part of the report demonstrated the existence of spurious correlation between the quarterly Melbourne property price indices and the quarterly Melbourne population change over the same period. Spurious correlation was eliminated by pre-whitening of the two time series data.

APPENDIX A

```
#IMPORT LIBRARIES


library(readr)

library(x12)

# install.packages("trend")

library(trend)

library(TSA)

library(tseries)

library(dLagM)

library(forecast)

library(dynlm)

setwd("F:/Subbu/RMIT/sem 4/Forecasting/assign 2")

source("MATH1307_utilityFunctions.R")


assign_ds <- read_csv("data1.csv")

head(assign_ds)

str(assign_ds)


#CONVERTING TO TIMESERIES

assign_ds$solar <- ts(assign_ds$solar,start = c(1960,1), end = c(2014,12), frequency = 12)

assign_ds$ppt <- ts(assign_ds$ppt,start = c(1960,1), end = c(2014,12), frequency = 12)

summary(assign_ds$solar)

summary(assign_ds$ppt)


#TIME SERIES PLOT

plot(assign_ds$solar, ylab = "Average Solar Radiations", main = "Time Series Plot of Monthly average solar radiations")

plot(assign_ds$ppt, ylab = "Precipitation", main = "Time Series Plot of Monthly precipitations")


# CORRELATION BETWEEN TWO TIME SERIES
```

```
cor_data = ts(assign_ds)
cor(cor_data)

# TEST OF STATIONARITY
adf.test(assign_ds$solar)
PP.test(assign_ds$solar)

#DECOMPOSITION

fit.solarstl <- stl(assign_ds$solar, t.window=15, s.window="periodic", robust=TRUE)
plot(fit.solarstl)

par(mfrow = c(1,2))
acf(assign_ds$solar, main = 'ACF Plot')
pacf(assign_ds$solar, main = 'PACF Plot')
par(mfrow = c(1,1))

# DLMs

model_1 = dlm(x = as.vector(assign_ds$solar), y = as.vector(assign_ds$ppt), q = 12)
summary(model_1)
checkresiduals(model_1)

model_2 = polyDlm(x = as.vector(assign_ds$solar), y = as.vector(assign_ds$ppt), q = 12,
          k = 2, show.beta = FALSE)
summary(model_2)
checkresiduals(model_2)

model_3 = koyckDlm(x = as.vector(assign_ds$solar), y = as.vector(assign_ds$ppt))
summary(model_3)
checkresiduals(model_3$model)
```

```
#ARDLM

for (i in 1:5){
  for(j in 1:5){
    model_4 = ardlDlm(x = as.vector(assign_ds$solar), y = as.vector(assign_ds$ppt),
                   p = i, q = j)
    cat("p=",i,"q=",j, "AIC =", AIC(model_4$model), "BIC = ", BIC(model_4$model), "\n")
  }
}

model4.1.1 = ardlDlm(x = as.vector(assign_ds$solar), y = as.vector(assign_ds$ppt),
                  p = 2, q = 3)
summary(model4.1.1)
checkresiduals(model4.1.1)

MASE(model_1, model_2, model_3, model4.1.1)

########################

# DYNAMIC LINEAR MODELS

# TRANSFORMATION

BC = BoxCox.lambda(assign_ds$solar)
BC

BC.Solar = ((assign_ds$solar^(BC)) - 1) / BC
plot(BC.Solar, main = "Time series Plot of BoxCoX transformed series", ylab = "Solar Radiations")

Y.t = BC.Solar
```

```
dynlm_1 = dynlm(Y.t ~ L(Y.t , k = 1 ) + trend(Y.t) + season(Y.t))

summary(dynlm_1)


checkresiduals(dynlm_1)


AIC(dynlm_1)




plot(BC.Solar, main = "Time series Plot of BoxCoX transformed series", ylab = "ASX Price", type = "l",
col = 'Green')

lines(dynlm_1$fitted.values)


# MASE

accuracy(dynlm_1)

accuracy(dynlm_2)


#####################################

# SIMPLE EXPONENTIAL SMOOTHING

# ses_fit1 <- ses(BC.Solar,initial = 'simple', h=24)

# summary(ses_fit1)

#

# ses_fit2 <- ses(BC.Solar,initial = 'optimal', h=24)

# summary(ses_fit2)


# HOLT WINTER METHOD


hw_fit1 <- hw(BC.Solar,seasonal="additive", h=24)

summary(hw_fit1)

hw_fit2 <- hw(BC.Solar,seasonal="additive",damped = TRUE, h=24)
```

```r
summary(hw_fit2)
checkresiduals(hw_fit2)
hw_fit3 <- hw(BC.Solar,seasonal="multiplicative", h=24)
summary(hw_fit3)
hw_fit4 <- hw(BC.Solar,seasonal="multiplicative",exponential = TRUE,h=24)
summary(hw_fit4)

plot(hw_fit2,ylab="Horizontal monthly average solar radiations",
    plot.conf=FALSE, type="l", fcol="RED", xlab="Year")



# ETS - HOLT WINTER METHOD

ETS_HW1 <- ets(BC.Solar,model ="ZZZ",  ic = "aic")
summary(ETS_HW1)

ETS_HW2 <- ets(BC.Solar,model ="ZZZ", ic = "bic")
summary(ETS_HW2)

# frc.AAdA = forecast(fit.AAdA, h =5)
# AAA_forecast = forecast(ETS_HW1, h = 24)
# plot(AAA_forecast)

###############

# TASK 2

assign_ds2 <- read_csv("data2.csv")
assign_ds2$price <- ts(assign_ds2$price, start = c(2003,3), frequency = 4)
assign_ds2$change <- ts(assign_ds2$change, start = c(2003,3), frequency = 4)
# plot(assign_ds2$price)
```

```
plot(ts.intersect(assign_ds2$price, assign_ds2$change),yax.flip = TRUE, main = "Time Series plots of Quarterly change of Melbourne property prices and Population change")


ccf(assign_ds2$price, assign_ds2$change, main = 'CCF Plot')


# dif_ts=ts.intersect(diff(diff(assign_ds2$price,12)),diff(diff(assign_ds2$change,12)))
dif_price = diff(assign_ds2$price, differences = 2)
plot(dif_price, main = 'Second differenced time series plot of property price index')
dif_change = diff(diff(assign_ds2$change,12))
plot(dif_change, main = 'First differenced time series plot of the population change')
prewhiten(dif_price, dif_change,ylab='CCF', main="Sample CCF after prewhitening of
    the Quaterly Melbourne PPI and Population change time series")
```