Time Series Analysis

Stationarity of Time Series - II

Contents

Case Study

Background

- Annual Sales for a specific company from year 1961 to 2017

Objective

- To assess stationarity of time series

Available Information

- Number of cases: 57
- Variables: Year, sales(in 10’s GBP)

Data Snapshot

turnover_annual data

Variables

...ons on Discrete Time Scale

Year	sales
1961	224786
1962	230034
1963	236562
1964	250960
1965	261615
1966	268316
1967	283589
1968	280160
1969	301422
1970	308018
1971	[illegible]
1974	364834
1975	392503

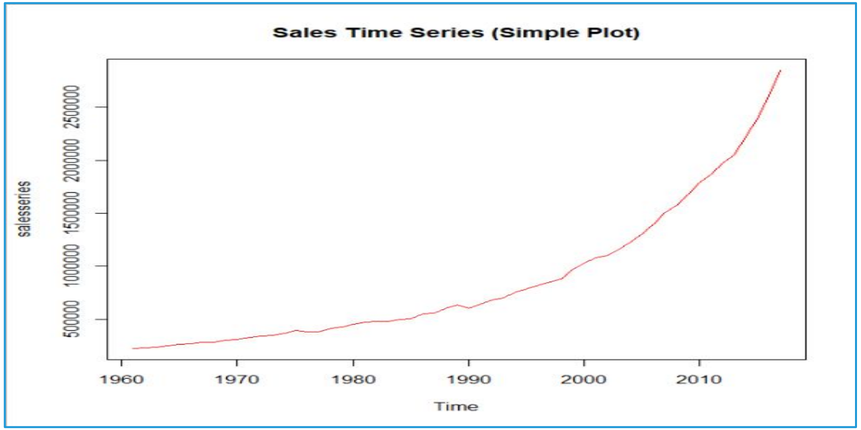
Columns	Description	Type	Measurement	Possible values
Year	Financial Year	Numeric	-	-
sales	sales(in 10’s GBP)	Numeric	In British Pound	Positive values

Creating and Plotting Time Series in R

```
#Importing the Data
turnover_annual <- read.csv("turnover_annual.csv",header=TRUE)
#Creating and Plotting a Time Series Object
salesseries<-ts(turnover_annual$sales,start=1961,end=2017)
```

ts() converts a column from a data frame to a simple time series object. The **start=** and **end=** arguments start and end period.

```
plot(salesseries,col="red",main="Sales Time Series (Simple Plot)")
```



Interpretation :

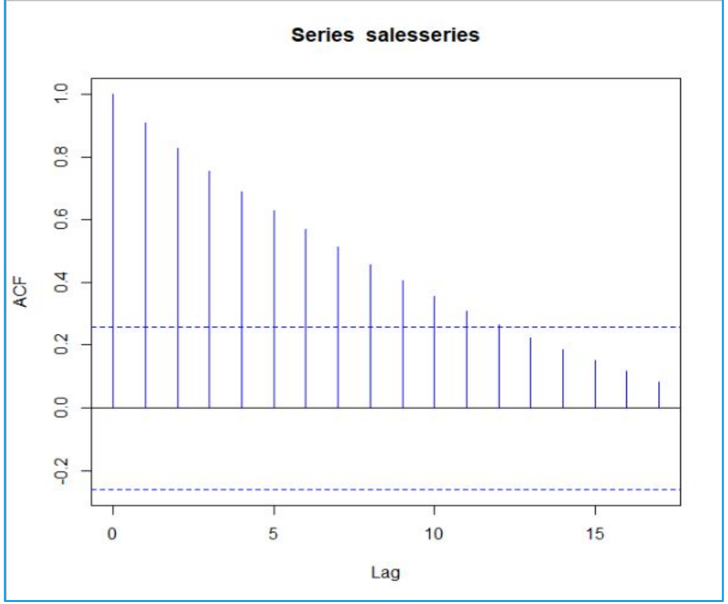
- The time-series clearly shows a positive trend.

Checking Stationarity – Correlogram

```
# ACF Plot
acf(salesseries,col="blue")
```

- **acf()** returns correlogram with X axis indicating lag number and Y axis autocorrelation value

```
# Output
```



Interpretation :

- We can observe that there is a very slow decay which is a sign of Non-stationarity.

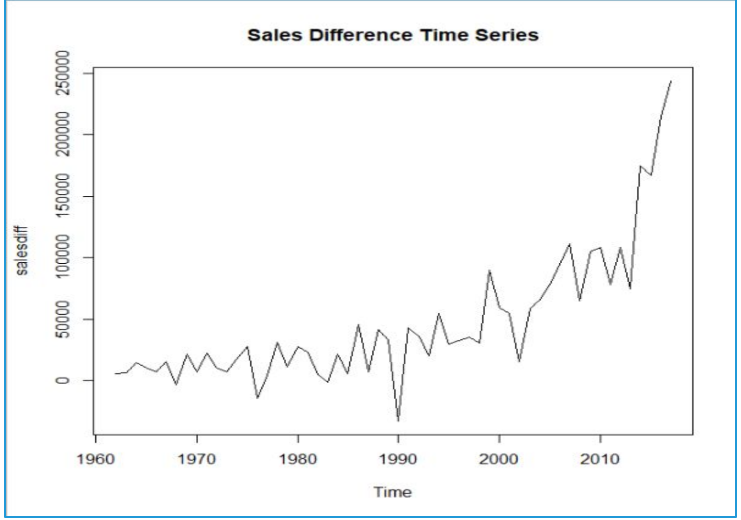
Plot of 1st Order Differenced Time Series

```
# Creating and Plotting a Differenced Series
salesdiff<-diff(salesseries,difference=1)

plot(salesdiff,main="Sales Differenced Time Series")
```

- **diff()** returns suitably lagged differences.
- **difference=** an integer indicating the order of the difference. Here order is 1.

```
# Output
```

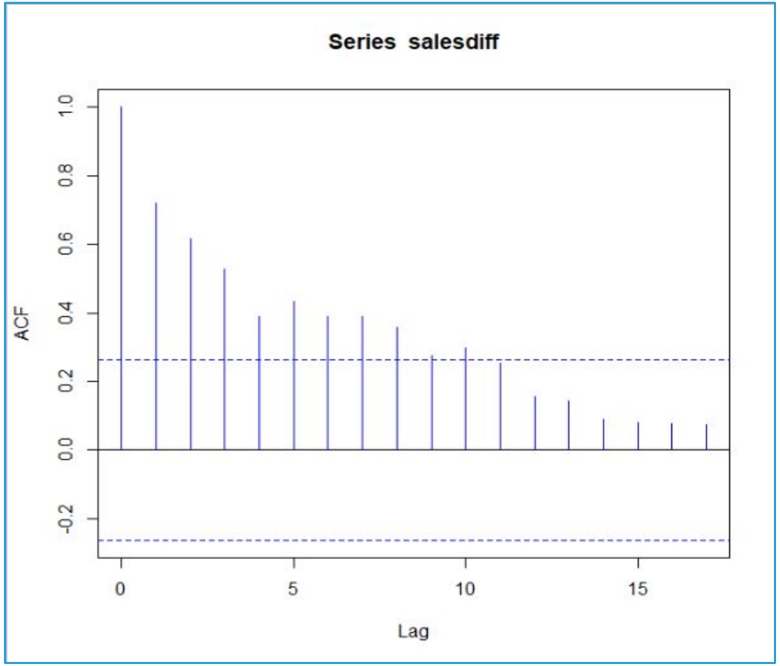


Interpretation :

- Even after first order differencing, the series looks non-stationary.

Correlogram for 1^{st} Order Differenced Time Series

```
# ACF Plot
acf(salesdiff,col="blue")
# Output
```



Interpretation :

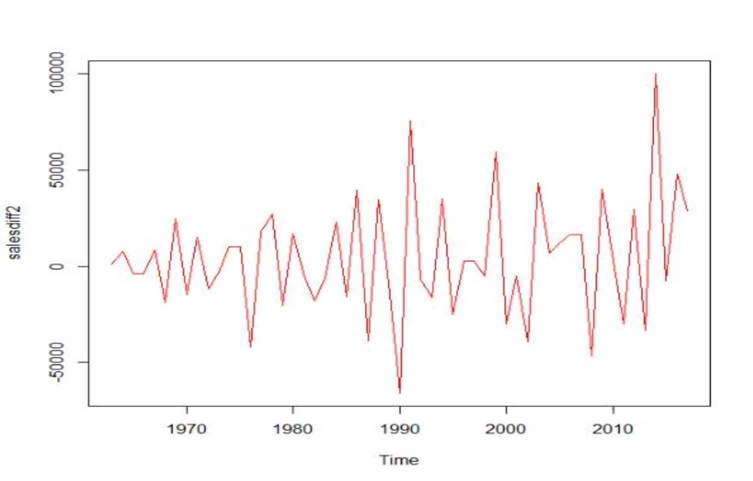
- Intuitive conclusion about the non-stationarity even after differencing is confirmed in this ACF plot as the decay is still slow.
- Stationarity is not achieved with first difference.

Determining Order of Differencing

```
# Install & load package "forecast"
install.packages("forecast")
library(forecast)
ndiffs(salesseries)
[1] 2

salesdiff2<-diff(salesseries,diff
plot(salesdiff2,col="red")
```

- **ndiffs()** by default uses a Unit Root test to estimate the number of differences required.



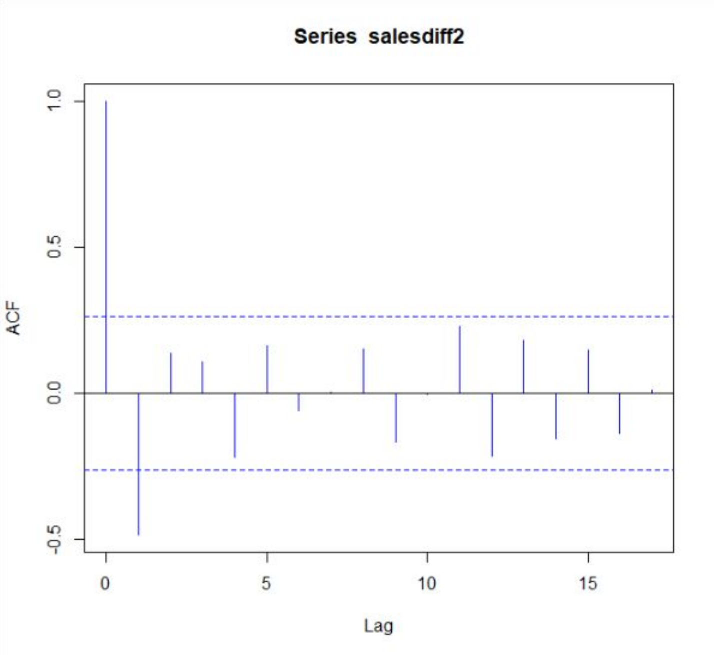
Interpretation :

- After 2nd order differencing, the series looks **stationary**.

Correlogram for 2nd Order Differenced Time Series

```
# ACF Plot
acf(salesdiff2,col="blue")
```

```
# Output
```



Interpretation :

- Intuitive conclusion about stationarity is confirmed in this ACF plot.
- Stationarity is achieved with 2^{nd} order difference.

Analytical Method – Dickey Fuller (DF) Test

- A linear stochastic process has a unit root if "1" is the root of the process's characteristic equation. Such a process is non-stationary.
- Dickey and Fuller pioneered idea of testing for unit roots for stationarity checking.

Consider X_t (t=1,2,3,...) is a time series of the form

$$\boxed{X_t=\rho X_{t-1}+U_t \quad \dots\dots\dots\dots\dots\dots(1)}$$

If $\rho=1$ then X_t becomes a random walk

- We assume that $U_t \sim IID(0, \sigma^2)$, i.e U_t is a white noise
- Therefore, we are interested in testing for $\rho=1$

* IID means independent and identically distributed

Dickey Fuller (DF) Unit Root Test

Objective	To test the **null hypothesis** that **time series is not stationary**

Null Hypothesis H_0: $\rho=1$

Alternate Hypothesis H_1: $\rho< 1$

$(X_t - X_{t-1}) = \Delta X_t = (\rho-1)\, X_{t-1} + U_t$..............from (1)

H_0: $\rho^*=0$, H_1: $\rho^*<0$, $\rho^*=(\rho-1)$

Test Statistic	$(\rho^*/SE(\rho^*))$ Test statistic follows DF distribution under null
Decision Criteria	Reject the null hypothesis **tcal < DF table value**

Dickey Fuller Test

```
# Install & load package "urca" for executing Dickey Fuller Test
install.packages("urca")
library(urca)
df<-ur.df(salesseries, lag=0)
summary(df)
```

ur.df() performs a Dickey Fuller unit root test on time series data.

```
# Output
###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression none


Call:
lm(formula = z.diff ~ z.lag.1 - 1)

Residuals:
   Min      1Q Median     3Q    Max
-73514 -19075  -8490   3412  76065

Coefficients:
        Estimate Std. Error t value Pr(>|t|)
z.lag.1 0.064334   0.003338   19.27   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 25950 on 55 degrees of freedom
Multiple R-squared:  0.871,     Adjusted R-squared:  0.8687
F-statistic: 371.5 on 1 and 55 DF,  p-value: < 2.2e-16


Value of test-statistic is: 19.2745

Critical values for test statistics:
      1pct  5pct 10pct
tau1  -2.6 -1.95 -1.61
```

Interpretation :

- Time series is non-stationary as value of test statistic is greater than 5% critical value.

Dickey Fuller Test

```
# checking stationarity for series with difference of order 2
df2 <- ur.df(salesdiff2,lags = 0)
summary(df2)
```

```
# Output
```

```
###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression none


Call:
lm(formula = z.diff ~ z.lag.1 - 1)

Residuals:
   Min      1Q Median      3Q     Max
-69843   -8676   7079   22229   84580

Coefficients:
        Estimate Std. Error t value Pr(>|t|)
z.lag.1  -1.4639     0.1229  -11.91   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 27740 on 53 degrees of freedom
Multiple R-squared:  0.7279,    Adjusted R-squared:  0.7228
F-statistic: 141.8 on 1 and 53 DF,  p-value: < 2.2e-16


Value of test-statistic is: -11.9083

Critical values for test statistics:
      1pct  5pct 10pct
tau1 -2.6 -1.95 -1.61
```

Interpretation :

- Time series is stationary as value of test statistic is less than 5% critical value.

Dickey Fuller Test – Drawbacks

- If a simple AR(1) model is used when in fact X_t follows an AR(p) process then the error term will be autocorrelated.
- Autocorrelated errors will invalidate the use of DF distributions, as it assumes errors is white noise.

Augmented Dickey-Fuller Test is a measure to fix this problem.

- The most widely offered criticism of the DF-test, however, is its tendency to over-reject the null hypothesis when its true and under-reject the null hypothesis when its false

* AR(1) is autoregressive process is the first-order process

Augmented Dickey Fuller Test

$$X_t=\alpha_0+\rho_1 X_{t-1}+\rho_2 X_{t-2}+\rho_3 X_{t-3}+\ldots+\rho_p X_{t-p}+U_t.$$

$$\Delta X_t=\alpha_0+\alpha^* X_{t-1}+\phi_1\,\Delta X_{t-1}+\phi_2\,\Delta X_{t-2}+\ldots.+\phi_p\,\Delta X_{t-p+1}+U_t.$$

$$\phi_1=-(\rho_2+\rho_3+\ldots.+\rho_p),\ \alpha^*=(\alpha_1+\alpha_2+\ldots+\alpha_p)-1$$

Null Hypothesis H_0: $\rho^*=0$
Alternate Hypothesis H_1: $\rho^*<0$
$\rho^*=(\rho-1)$

Test Statistic	$(\rho^*/SE(\rho^*))$ Test statistic follows DF distribution under null
Decision Criteria	Reject the null hypothesis if **tcal < DF table value**

Quick Recap

Correlograms	• **`acf()`** and **`pacf()`** functions in base R generate Correlograms
Differencing a Time Series	• Simple numeric function **`diff()`** can be used to difference a time series • **`ndiffs()`** function generates order of differencing by applying the Dickey Fuller test
Dickey Fuller Test	• **`ur.df()`** function from the package **`urca`** performs a Dickey Fuller test • The output gives test statistic and critical values for the test statistic