Description :

Source :

"Makridakis, Wheelwright and Hyndman (1998)"

Description :

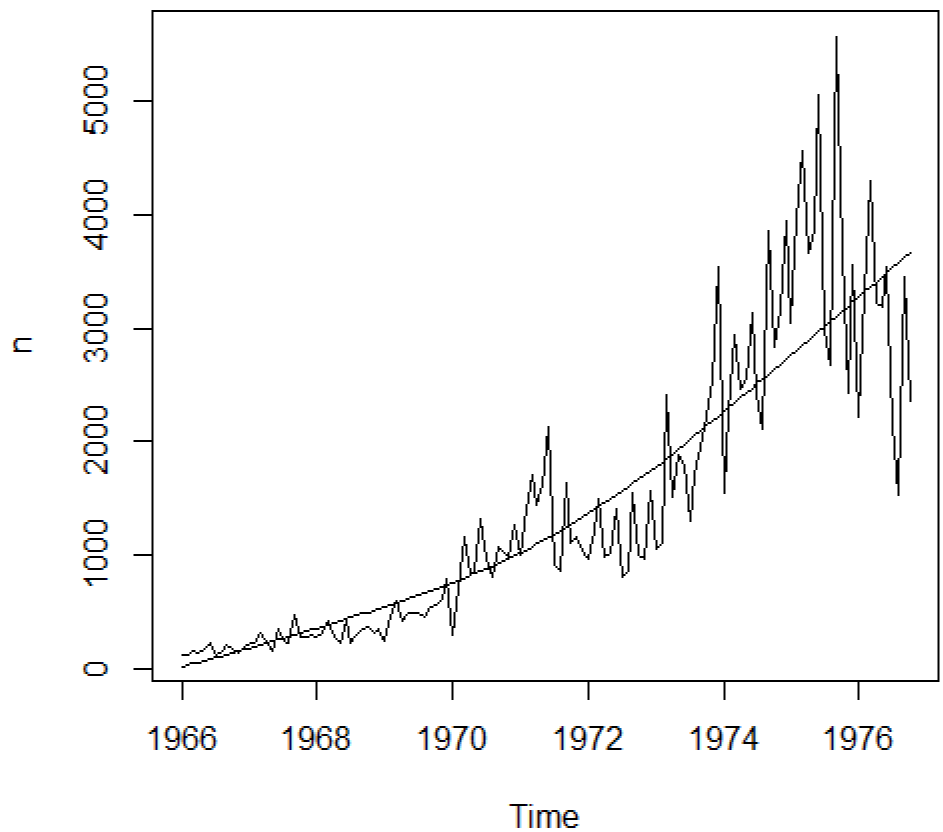
"Monthly shipment of pollution equipment (in thousands of french francs)"

Subject :

"Production"

```
> n = tsdl [[468]]
> attributes(n)
```

Plot The Time Series :




```
> plot(n)

> lines(n)

> lines(lowess(n))
```

- We plot The time series to know is there trend in our data or not and to know the seasonality .

As we see In this data it appear there is seasonal component in the set and not stationarity .

Testing Stationarity :

```
> adf.test(n)

	Augmented Dickey-Fuller Test

data:  n
Dickey-Fuller = -1.6188, Lag order = 5, p-value = 0.735
alternative hypothesis: stationary
```

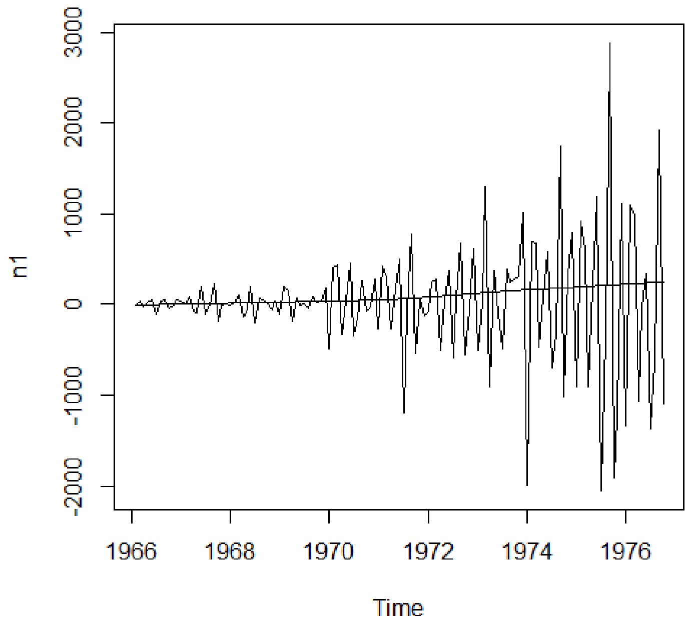
We can test the stationary of the series from this command > adf.test(n) and the Hypothesis will be :

$$H_0 : The\ series\ is\ not\ stationary\ .$$

$$H_1 : The\ series\ is\ stationary\ .$$

So since the $\mathbf{P\ .Value > \alpha\ , P\ .Value = 0.735}$ and $\boldsymbol{\alpha} = \mathbf{0.05}$ so we Accept H_0 that's mean the series is not stationary

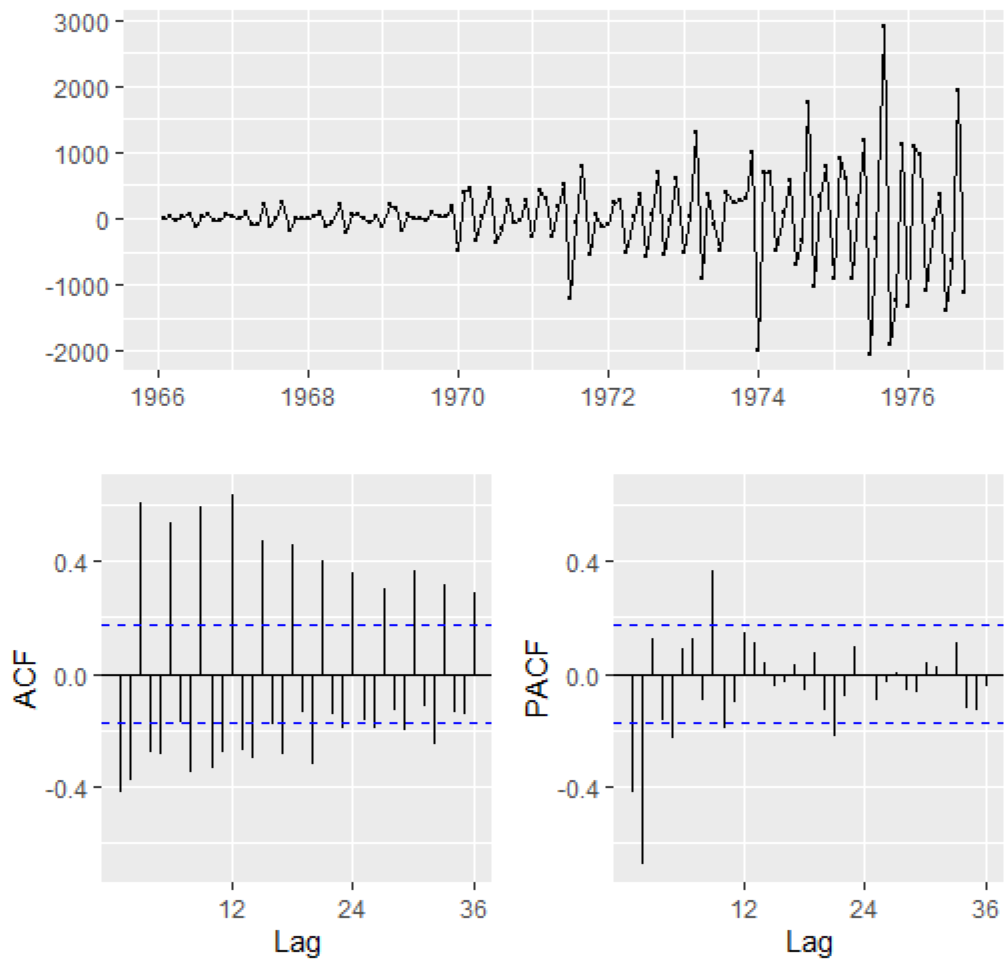
- Now we should take the difference for series to make it stationary




```
> n1=diff(n,diffrence=1)
> plot(n1)
> lines(n1)
> lines(lowess(n1))
> adf.test(n1))
Augmented Dickey-Fuller Test data:  n1
Dickey-Fuller = -5.7566, Lag order = 5, p-value = 0.01
alternative hypothesis: stationary
```

So now we see the series become stationary and since the $\mathbf{P\,.\,Value} < \boldsymbol{\alpha}\,,\mathbf{P\,.\,Value} = \mathbf{0.01}$ and $\boldsymbol{\alpha} = \mathbf{0.05}$ so we Reject H_0 that's mean the series is stationary

Model Type :



From the AutoCorrelation (ACF) graph and the Parial AutoCorrelation (PACF) graph we can nominate two models are suggested :

SARIMA(0,1,1)(0,0,2)12 , SARIMA(0,1,1)(0,0,1) .

Testing Coeffiicients :

For model SARIMA(0,1,1)(0,0,2)12 :

```
> fit1 = arima(n,order=c(0,1,1),seasonal=list(order=c(0,0,2),period=12))
> coeftest(fit1)
z test of coefficients:

     Estimate Std. Error z value  Pr(>|z|)
ma1  -0.603531   0.062874 -9.5990 < 2.2e-16 ***
sma1  0.798564   0.098913  8.0734 6.839e-16 ***
sma2  0.338155   0.102954  3.2845  0.001022 **
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The Hypothesis will be :

$$H_0 : \theta_1 = 0 \quad vs \quad H_1 : \theta_1 \neq 0$$

$$H_0 : \theta_2 = 0 \quad vs \quad H_1 : \theta_2 \neq 0$$

$$H_0 : \vartheta_1 = 0 \quad vs \quad H_1 : \vartheta_1 \neq 0$$

Since the $\boldsymbol{P.value's\ in\ all\ parameters < \ \alpha = 0.05}$

So we reject H_0 , that's mean the value is significantly different from zero .

For model SARIMA(0,1,1)(0,0,1)

```
> fit2 = arima(n,order=c(0,1,1),seasonal=list(order=c(0,0,1),period=12))
> coeftest(fit2)
z test of coefficients:

     Estimate Std. Error  z value  Pr(>|z|)
ma1  -0.617118   0.061525 -10.0303 < 2.2e-16 ***
sma1  0.631052   0.073322   8.6066 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The Hypothesis will be :

$$H_0 : \theta_1 = 0 \quad vs \quad H_1 : \theta_1 \neq 0$$

$$H_0 : \vartheta_1 = 0 \quad vs \quad H_1 : \vartheta_1 \neq 0$$

Since the $\boldsymbol{P.value's\ in\ all\ parameters < \ \alpha = 0.05}$

So we reject H_0 , that's mean the value is significantly different from zero .

Testing Residuals :

For Model SARIMA(0,1,1)(0,0,2)12 :

```
> tsdiag(fit1)
Box.test(res1,lag=12,type=c("Ljung-Box"))
  Box-Ljung test
data:  res1
X-squared = 2.542, df = 12, p-value = 0.5244
```

- We test the residuals are correlated or not :

$$\boldsymbol{H_0 : \rho_1 = \rho_2 = \cdots = \rho_k = 0}$$
$$\boldsymbol{H_1 : at\ least\ two\ \neq 0}$$

From the **grarph and p.value for Ljung-Box statistic we** notice all **the p.value's = 0.5244 > α = 0.05** , so we Accept H_0 , that's mean all the residuals are not correlated .

For Model SARIMA(0,1,1)(0,0,1)12 :

```
> tsdiag(fit2)
> Box.test(res2,lag=12,type=c("Ljung-Box"))
  Box-Ljung test
data:  res2
X-squared = 3.232, df = 12, p-value = 0.3302
```

- We test the residuals are correlated or not :

$$\boldsymbol{H_0 : \rho_1 = \rho_2 = \cdots = \rho_k = 0}$$
$$\boldsymbol{H_1 : at\ least\ two\ \neq 0}$$

From the **grarph p.value for Ljung-Box statistic** we notice all **the p.value's = 0.3302 > α = 0.05** , so we Accept H_0 , that's mean all the residuals are not correlated .

- Now we test are the Residuals random or not :

For model SARIMA(0,1,1)(0,0,2)12 :

$$H_0 : Residuals\ are\ random$$

$$H_1 : Residuals\ are\ not\ random$$

```
> runs.test(fit1$res)

        Runs Test

data:  fit1$res
statistic = -0.70438, runs = 62, n1 = 65, n2 = 65, n = 130, p-value =
0.4812
alternative hypothesis: nonrandomness
```

Since the $\boldsymbol{p.value = 0.4812 > \alpha = 0.05}$, so we Accept H_0 , that's mean the residuals are random .

For model SARIMA(0,1,1)(0,0,1)12 :

$$H_0 : Residuals\ are\ random$$

$$H_1 : Residuals\ are\ not\ random$$

```
> runs.test(fit2$res)

      Runs Test

data:  fit2$res
statistic = -0.1761, runs = 65, n1 = 65, n2 = 65, n = 130, p-value =
0.8602
```

Since the $\boldsymbol{p.value = 0.8602 > \alpha = 0.05}$, so we Accept H_0 , that's mean the residuals are random .

- Now we test the mean of the ε_t equal zero or not :

For model SARIMA(0,1,1)(0,0,2)12 :

$$H_0 : E(\varepsilon_t) = 0$$

$$H_1 : E(\varepsilon_t) \neq 0$$

```
> t.test(fit1$res)

        One Sample t-test

data:  fit1$res
t = 0.13093, df = 129, p-value = 0.896
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -65.36488  74.62923
sample estimates:
mean of x
 4.632172
```

Since the $\boldsymbol{p.value = 0.896 > \ \alpha = 0.05}$, so we Accept H_0 , that's mean the expection of ε_t equal zero .

For model SARIMA(0,1,1)(0,0,1)12 :

$$H_0 : E(\varepsilon_t) = 0$$

$$H_1 : E(\varepsilon_t) \neq 0$$

```
> t.test(fit2$res)

	One Sample t-test

data:  fit2$res
t = 0.43557, df = 129, p-value = 0.6639
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -57.37843  89.77344
sample estimates:
mean of x
  16.1975
```

Since the $\boldsymbol{p.value} = \mathbf{0.6639} > \boldsymbol{\alpha} = \mathbf{0.05}$, so we Accept H_0 , that's mean the expection of ε_t equal zero .

- Now we do invertibilty analysis :

$$\theta_1 = -0.6035$$

$$\theta_2 = 0.7986$$

$$\vartheta_1 = 0.3382$$

$$|\theta_2| < 1 ✓$$

$$\theta_1 + \theta_2 < 1 ✓$$

$$\theta_1 - \theta_2 < 1 ✓$$

And the same manner for another model.

- Now we test the Normality of the Residuals :

For model SARIMA(0,1,1)(0,0,2)12 :

H_0 : *the Residuals follow Normal Distribution*

H_1 : *the Residuals Don't follow Normal Distribution*

```
> shapiro.test(res1)


	Shapiro-Wilk normality test


data:  res1

W = 0.98315, p-value = 0.1018
```

Since the $\boldsymbol{p.value = 0.1018 > \ \alpha = 0.05}$, so we Accept H_0 , that's mean the Residual are follow Normal Distribution .

For model SARIMA(0,1,1)(0,0,1)12 :

H_0 : $the\ Residuals\ follow\ Normal\ Distribution$

H_1 : $the\ Residuals\ Don't\ follow\ Normal\ Distribution$

```
> shapiro.test(fit2$res)

	Shapiro-Wilk normality test

data:  fit2$res
W = 0.87541, p-value = 4.691e-09
```

Since the $\boldsymbol{p.value} = \mathbf{4.691e-09} < \boldsymbol{\alpha} = \mathbf{0.05}$, so we Reject H_0 , that's mean the Residual don't follow Normal Distribution , so we eliminate this model

AIC or BIC for Models:

Now we have Model SARIMA(0,1,1)(0,0,2)12 :and we calculate AIC :

```
> fit1$aic
[1] 1930.716
> fit2$aic
[1] 1939.677
```

We choose the lowest **AIC** between models , so SARIMA(0,1,1)(0,0,2)12 more appropriate and just to make sure of that we compare it with AIC of SARIMA(0,1,1)(0,0,1)12 .

Mathematical form of SARIMA(0,1,1)(0,0,2)12 :

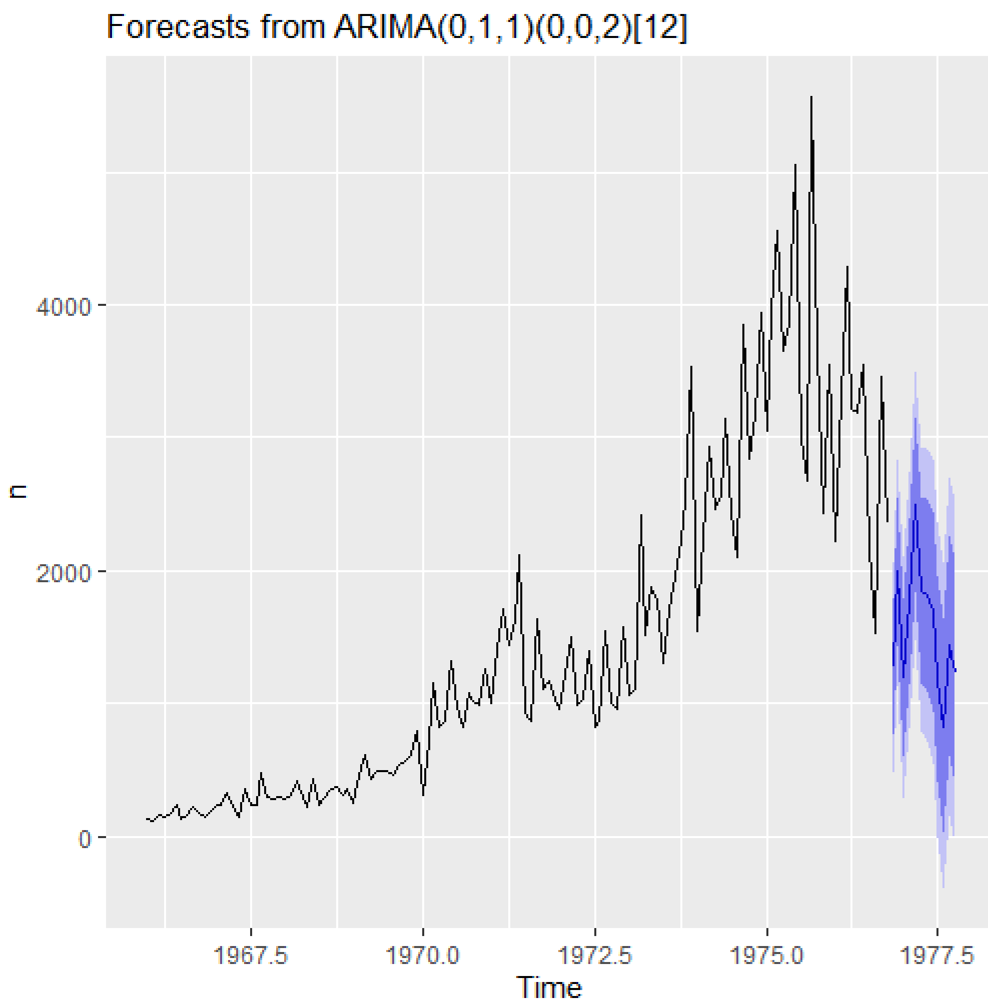
$$y_t = (1-\theta_1 B)(1-\theta_1 B^{12} - \theta_2 B^{24})(1-B)\varepsilon_t$$

```
> auto.arima(n)
Series: n
ARIMA(0,1,1)(0,0,2)[12]


Coefficients:
          ma1    sma1    sma2
      -0.6035  0.7986  0.3382
s.e.   0.0629  0.0989  0.1030
```

Forecasting :

We ready now to forecast with our model SARIMA(0,1,1)(0,0,1)12 :



```
> f=forecast(fit1, h=12)
>  autoplot(f)
> f
         Point Forecast     Lo 80    Hi 80      Lo 95    Hi 95
Nov 1976      1272.2026  755.22060 1789.185  481.546871 2062.858
Dec 1976      1993.8976 1437.76631 2550.029 1143.368183 2844.427
Jan 1977      1192.7796  600.07925 1785.480  286.322649 2099.237
Feb 1977      1768.2914 1141.15079 2395.432  809.162603 2727.420
Mar 1977      2498.9638 1839.17825 3158.749 1489.908855 3508.019
Apr 1977      1851.1821 1160.29228 2542.072  794.557327 2907.807
May 1977      1812.2115 1091.55879 2532.864  710.068279 2914.355
Jun 1977      1692.7338  943.49949 2441.968  546.878815 2838.589
Jul 1977      1186.6784  409.91351 1963.443   -1.280979 2374.638
Aug 1977       829.0518   25.69921 1632.404 -399.569954 2057.673
Sep 1977      1433.9166  604.82854 2263.005  165.935834 2701.897
Oct 1977      1221.9339  367.88552 2075.982  -84.220390 2528.088
```

We Forecast the next 10 values .

Attachments

The Packages and libraries are used :

```
install.packages("zoo")
install.packages("devtools")
devtools::install_github("FinYang/tsdl")
install.packages("nortest")
install.packages("fBasics")
install.packages("forecast")
install.packages("tseries")
install.packages("randtests")
install.packages("astsa")
install.packages("lmtest")
library("zoo")
library(tsdl)
library(nortest)
library(fBasics)
library(forecast)
library(tseries)
library(randtests)
library(astsa)
library(lmtest)
```

Codes Are used :

Code	Meaning
>attributes(m)	To describe the data for what
> plot.ts(m)	To plot your series
> adf.test(m)	To test the stationarity of series
>fit1 = arima(m,order=c(0,0,0),include.mean=T)	To identify your model
>coeftest(fit1)	To test the coefficients of model
>Box.test(res,lag=12,type=c("Ljung-Box"))	To test the residuals are correlated or not
> tsdiag(fit)	To show the graph and test residuals are correlated or not
> runs.test(res1)	To test the residuals are random or not
> t.test(res1)	To test the expection of residuals are zero or not
>Shapiro.test(res1)	To resdiuals are follow normal distribution
>aic$fit	To know what model have less AIC
> f=forecast(fit1, h=12) > autoplot(f)	To predict next values To plot the forecast
> n1=diff(n,diffrence=1)	To take difference fot the series

INDEX