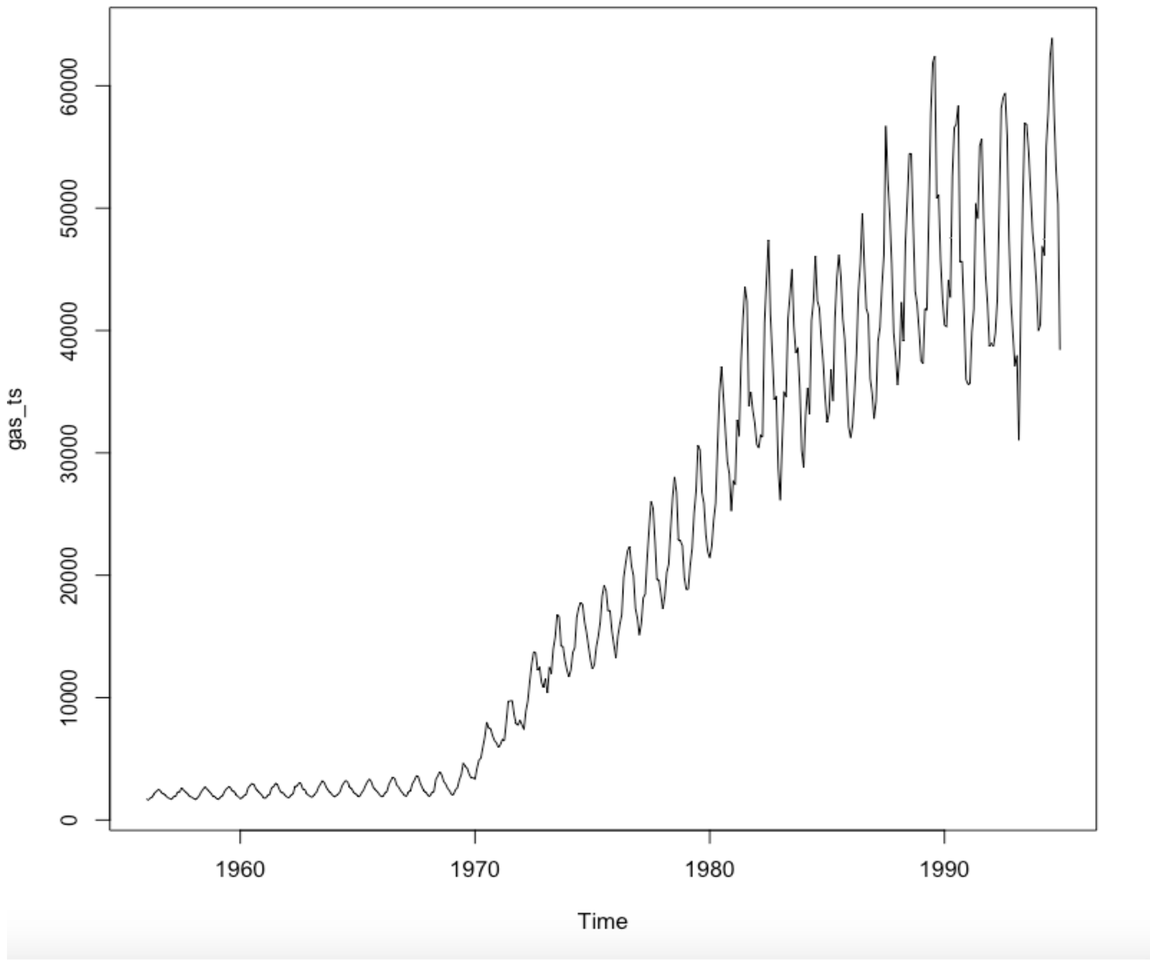
For this assignment, you are requested to download the **Forecast** package in R. The package contains methods and tools for displaying and analyzing univariate time series forecasts including exponential smoothing via state space models and automatic ARIMA modelling. Explore the **gas** (Australian monthly gas production) dataset in Forecast package to do the following:

- Read the data as a time series object in R. Plot the data **(5 marks)**
- What do you observe? Which components of the time series are present in this dataset? **(5 marks)**
- What is the periodicity of dataset? **(5 marks)**
- Is the time series Stationary? Inspect visually as well as conduct an ADF test? Write down the null and alternate hypothesis for the stationarity test? De-seasonalise the series if seasonality is present? **(20 marks)**
- Develop an ARIMA Model to forecast for next 12 periods. Use both manual and auto.arima (Show & explain all the steps) **(20 marks)**
- Report the accuracy of the model **(5 marks)**

1. Read the data as a time series object in R. Plot the data

Conversion of the Data into time series object . This is done by the function “**ts**”.

Below figure shows the plot of the data:



Fig(1)

Data Plot of the time series gas_ts

2. What do you observe? Which components of the time series are present in this dataset? What is the periodicity of dataset?

Trend : Increasing trend as the years pass by in the monthly production of gas.

Seasonality: There is a constant Seasonality

Cyclic components are present in the data set.

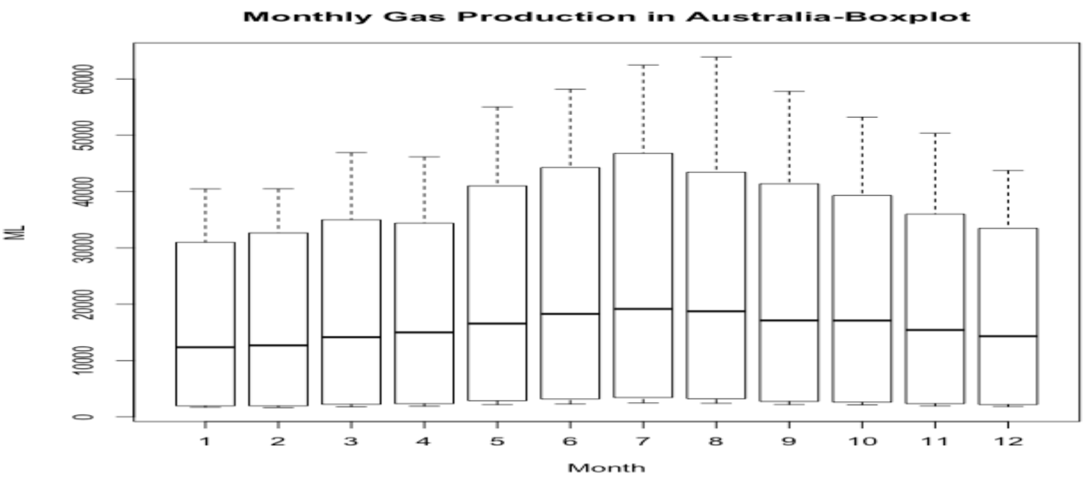
As the years pass by the monthly production of gas has been increasing gradually.

```
View(gas_ts)

lass(gas_ts)

[1] "ts"

start(gas_ts)

[1] 1956    1

end(gas_ts)

[1] 1994   12

frequency(gas_ts)

[1] 12

summary(gas_ts)

Min. 1st Qu.  Median    Mean 3rd Qu.    Max.

   1646    2659   16346   20877   37924   63896
```

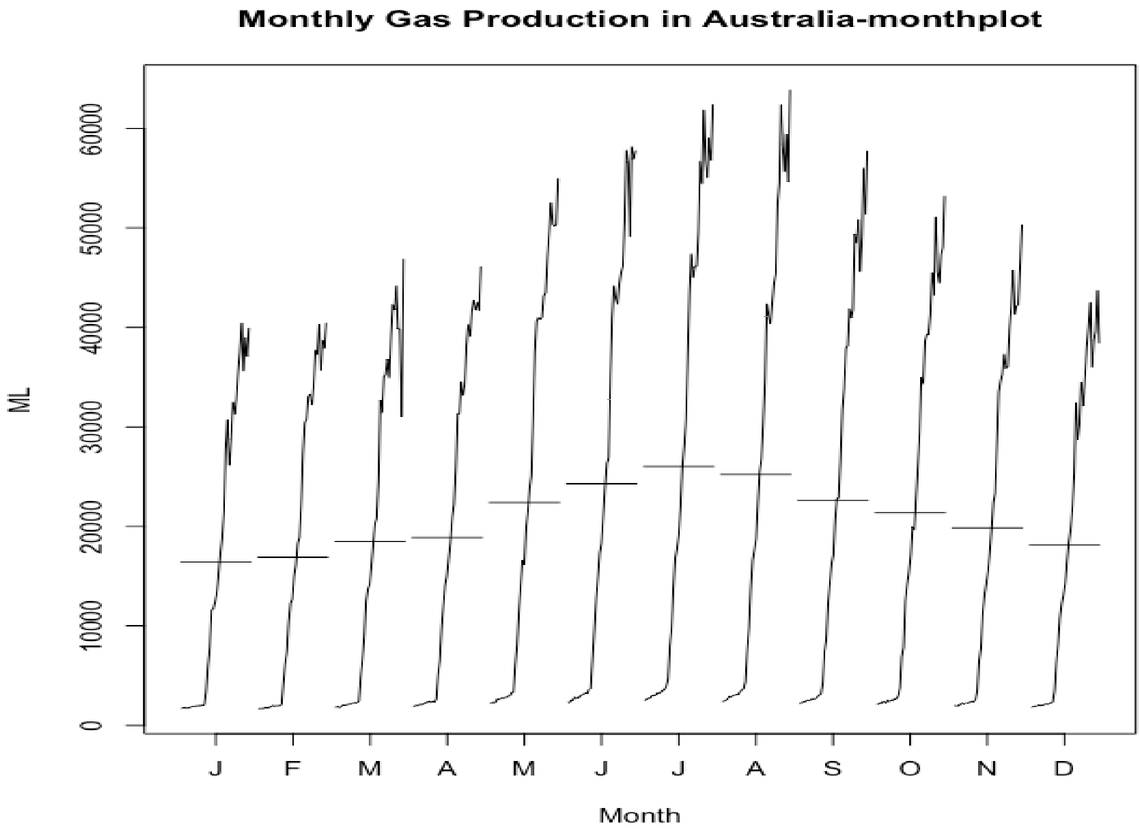
cycle(gas_ts)

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1956	1	2	3	4	5	6	7	8	9	10	11	12
1957	1	2	3	4	5	6	7	8	9	10	11	12
1958	1	2	3	4	5	6	7	8	9	10	11	12
1959	1	2	3	4	5	6	7	8	9	10	11	12
1960	1	2	3	4	5	6	7	8	9	10	11	12
1961	1	2	3	4	5	6	7	8	9	10	11	12
1962	1	2	3	4	5	6	7	8	9	10	11	12
1963	1	2	3	4	5	6	7	8	9	10	11	12
1964	1	2	3	4	5	6	7	8	9	10	11	12
1965	1	2	3	4	5	6	7	8	9	10	11	12
1966	1	2	3	4	5	6	7	8	9	10	11	12
1967	1	2	3	4	5	6	7	8	9	10	11	12
1968	1	2	3	4	5	6	7	8	9	10	11	12
1969	1	2	3	4	5	6	7	8	9	10	11	12
1970	1	2	3	4	5	6	7	8	9	10	11	12
1971	1	2	3	4	5	6	7	8	9	10	11	12
1972	1	2	3	4	5	6	7	8	9	10	11	12
1973	1	2	3	4	5	6	7	8	9	10	11	12
1974	1	2	3	4	5	6	7	8	9	10	11	12
1975	1	2	3	4	5	6	7	8	9	10	11	12
1976	1	2	3	4	5	6	7	8	9	10	11	12
1977	1	2	3	4	5	6	7	8	9	10	11	12
1978	1	2	3	4	5	6	7	8	9	10	11	12
1979	1	2	3	4	5	6	7	8	9	10	11	12
1980	1	2	3	4	5	6	7	8	9	10	11	12
1981	1	2	3	4	5	6	7	8	9	10	11	12
1982	1	2	3	4	5	6	7	8	9	10	11	12
1983	1	2	3	4	5	6	7	8	9	10	11	12
1984	1	2	3	4	5	6	7	8	9	10	11	12
1985	1	2	3	4	5	6	7	8	9	10	11	12
1986	1	2	3	4	5	6	7	8	9	10	11	12
1987	1	2	3	4	5	6	7	8	9	10	11	12
1988	1	2	3	4	5	6	7	8	9	10	11	12
1989	1	2	3	4	5	6	7	8	9	10	11	12
1990	1	2	3	4	5	6	7	8	9	10	11	12
1991	1	2	3	4	5	6	7	8	9	10	11	12
1992	1	2	3	4	5	6	7	8	9	10	11	12
1993	1	2	3	4	5	6	7	8	9	10	11	12
1994	1	2	3	4	5	6	7	8	9	10	11	12

```
boxplot(gas_ts ~ cycle(gas_ts), xlab = "Month", ylab = "ML", main = "Monthly Gas Production in Australia-Boxplot")
```

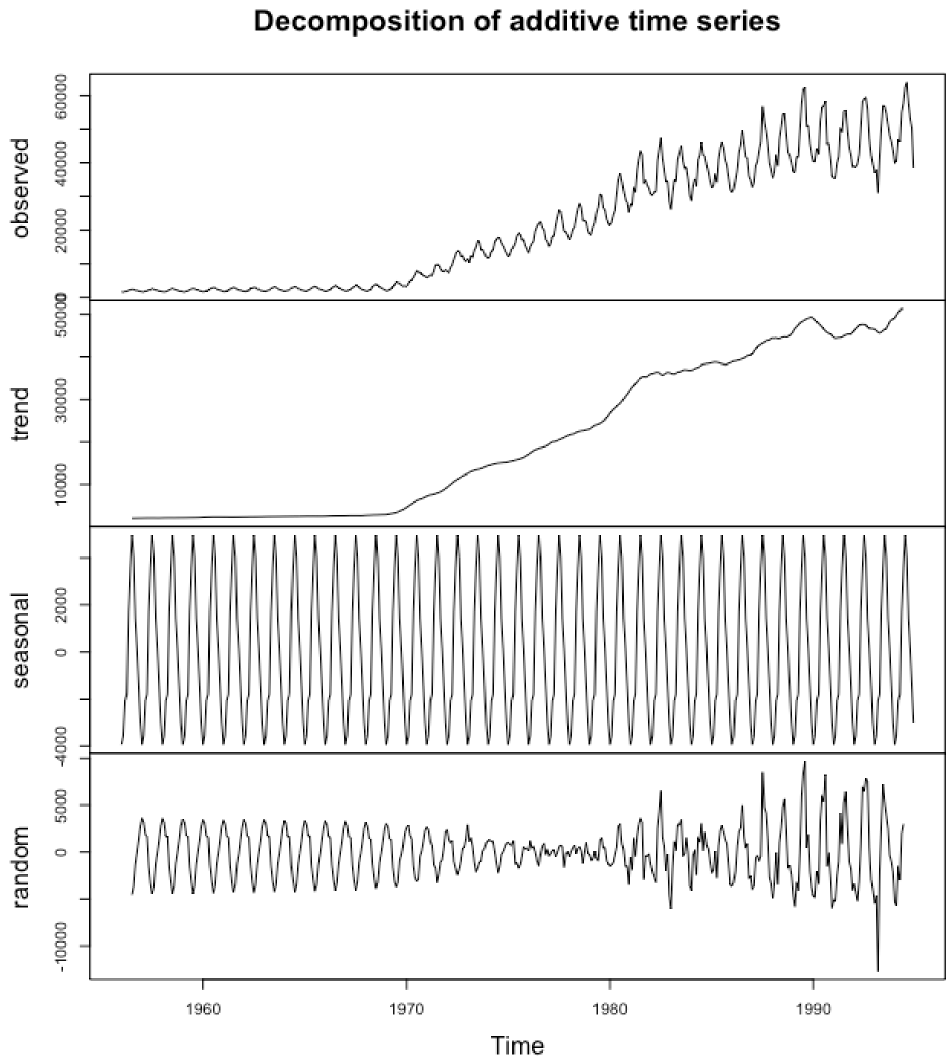



```
monthplot(gas_ts, main = "Monthly Gas Production in Australia-monthplot", xlab ="Month", ylab = "ML")
```



Is the time series Stationary? Inspect visually as well as conduct an ADF test? Write down the null and alternate hypothesis for the stationarity test? De-seasonalise the series if seasonality is present?

Additive and Multiplicative observations

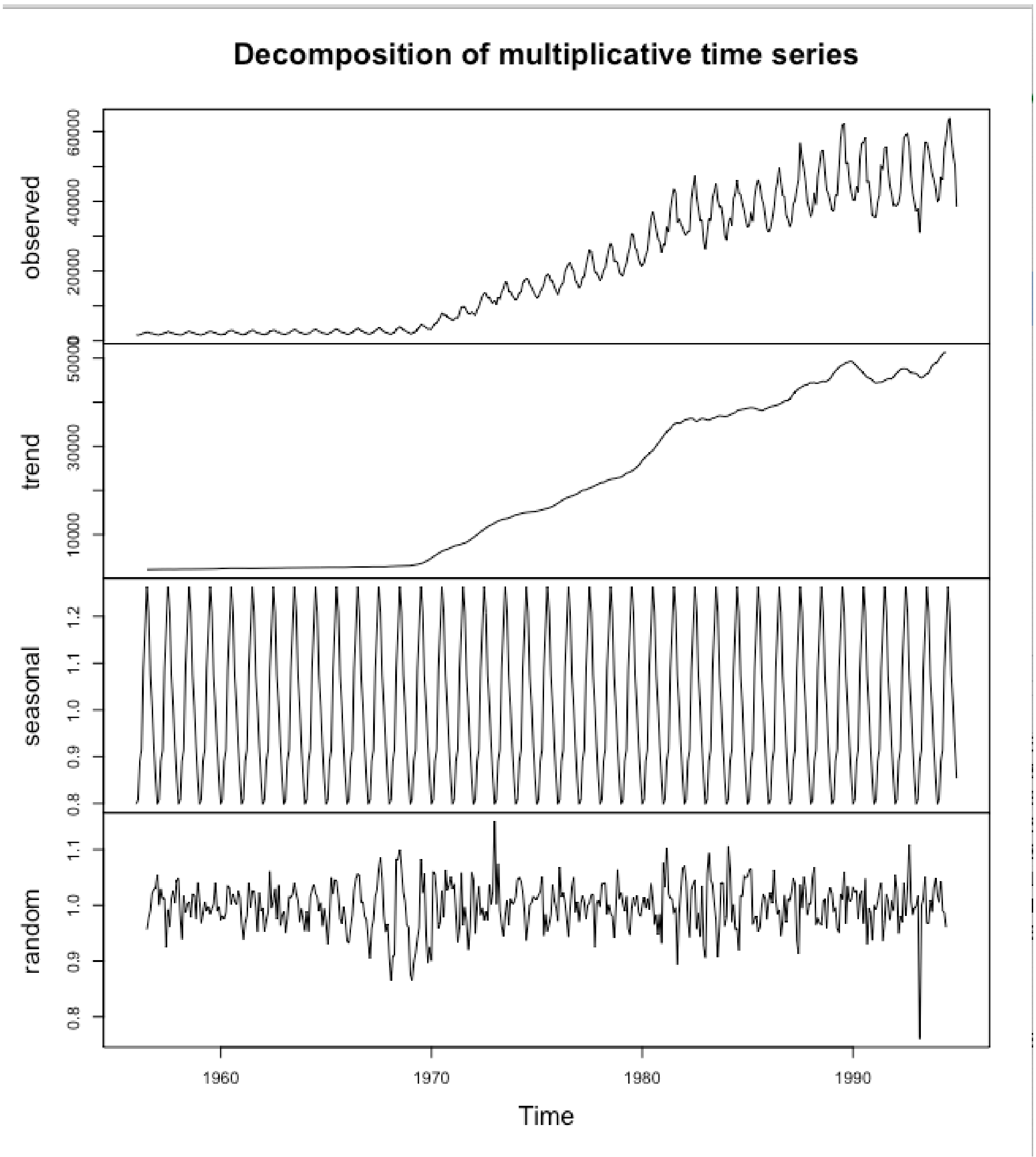


$type
[1] "additive"

attr(,"class")
[1] "decomposed.ts"

Multiplicative :

```
decompgas = decompose(gas_ts, type = "multiplicative")
plot(decompgas)
decompgas
```

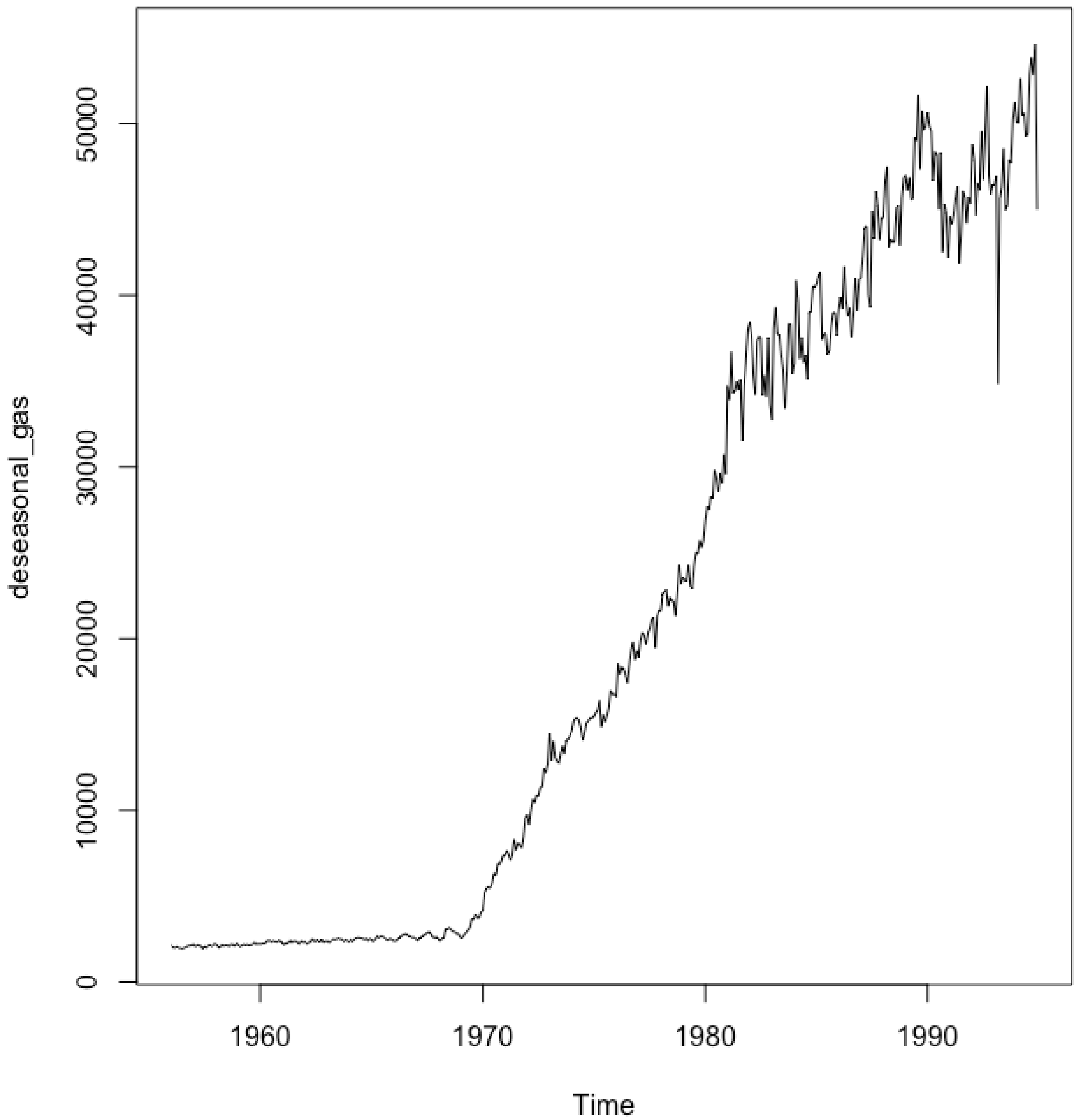


$type
[1] "multiplicative"

attr(,"class")
[1] "decomposed.ts"

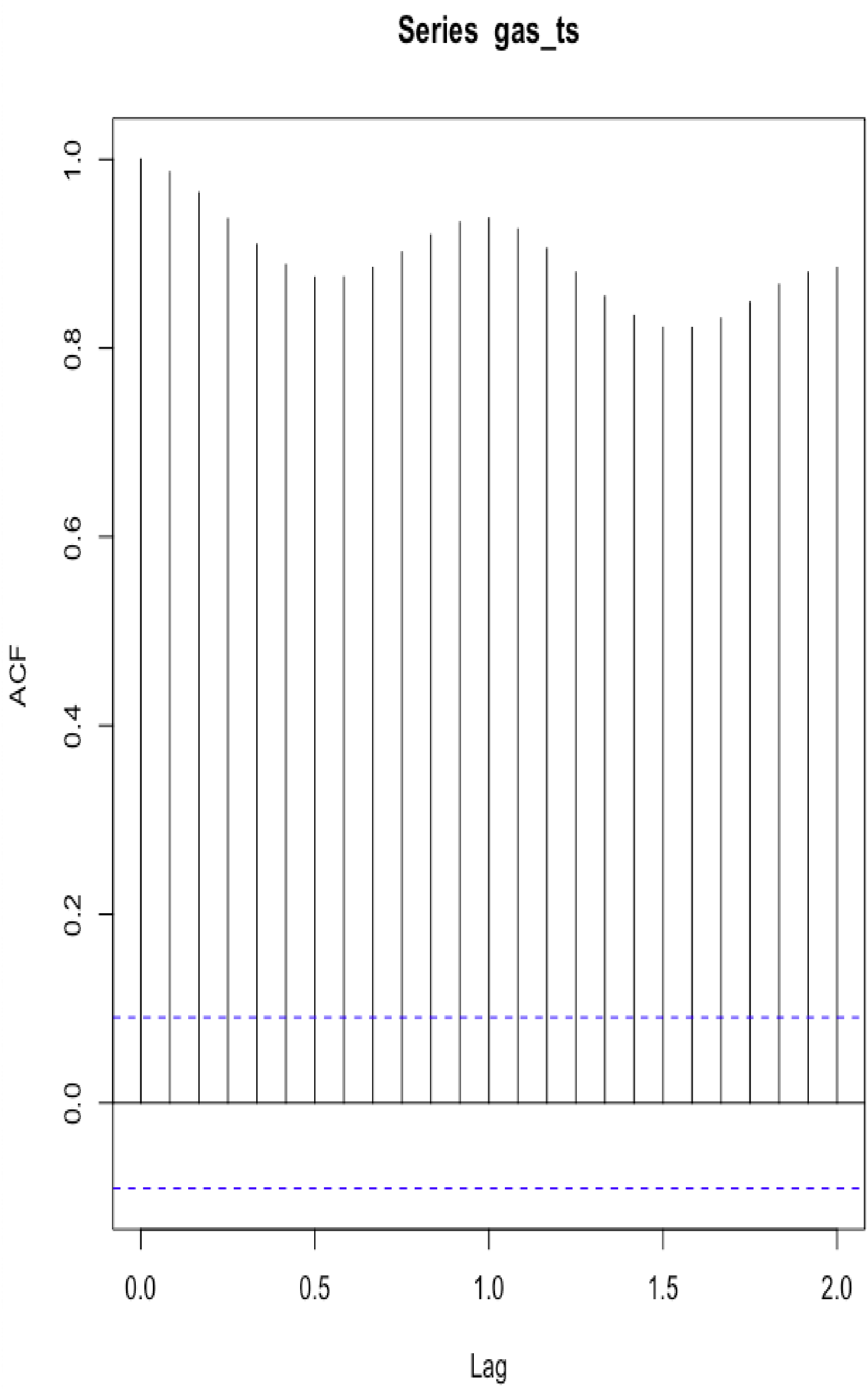
deseasonal_gas = seasadj(decompgas)

> plot(deseasonal_gas)

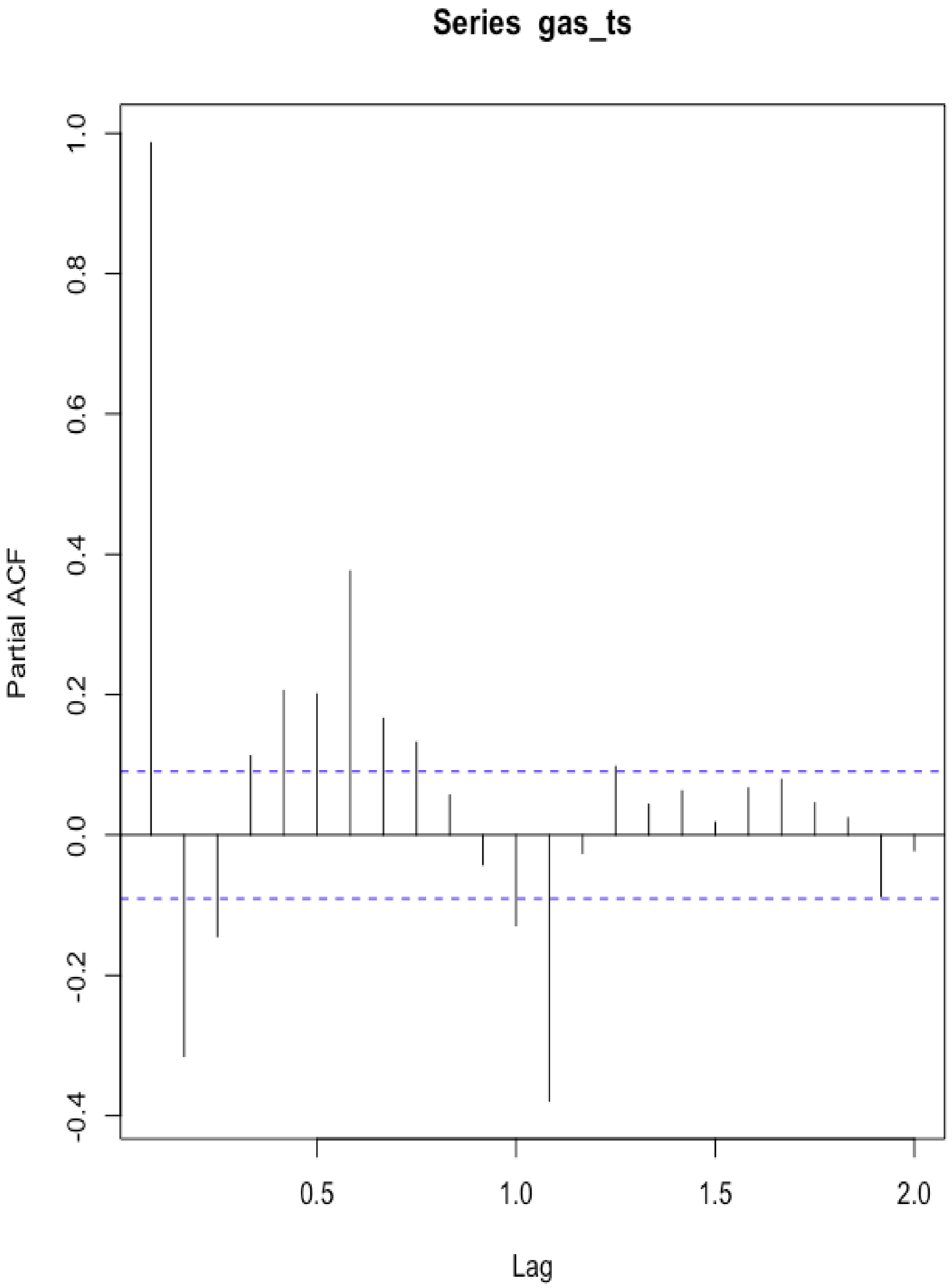


Augmented Dickey-Fuller Test
data: gas_ts
Dickey-Fuller = -2.6996, Lag order = 7, p-value = 0.282
alternative hypothesis: stationary

acf(gas_ts, lag.max = 24)

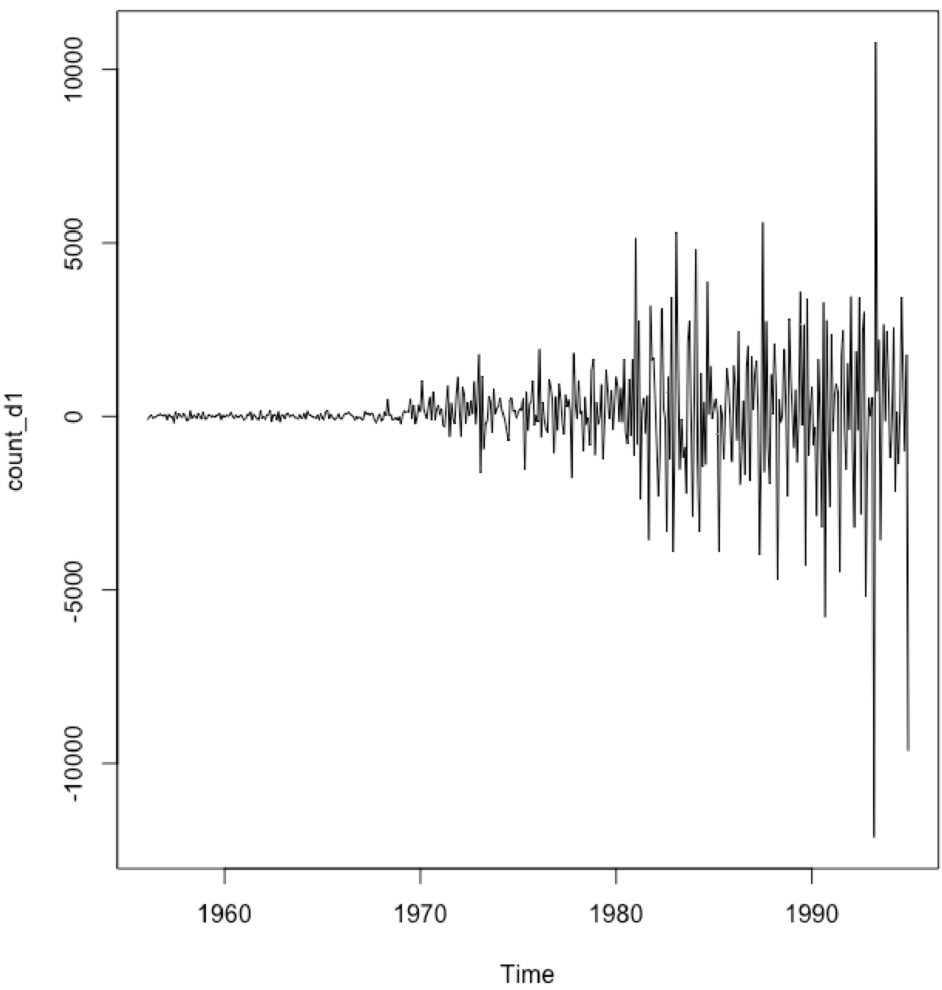


pacf(gas_ts, lag.max = 24)

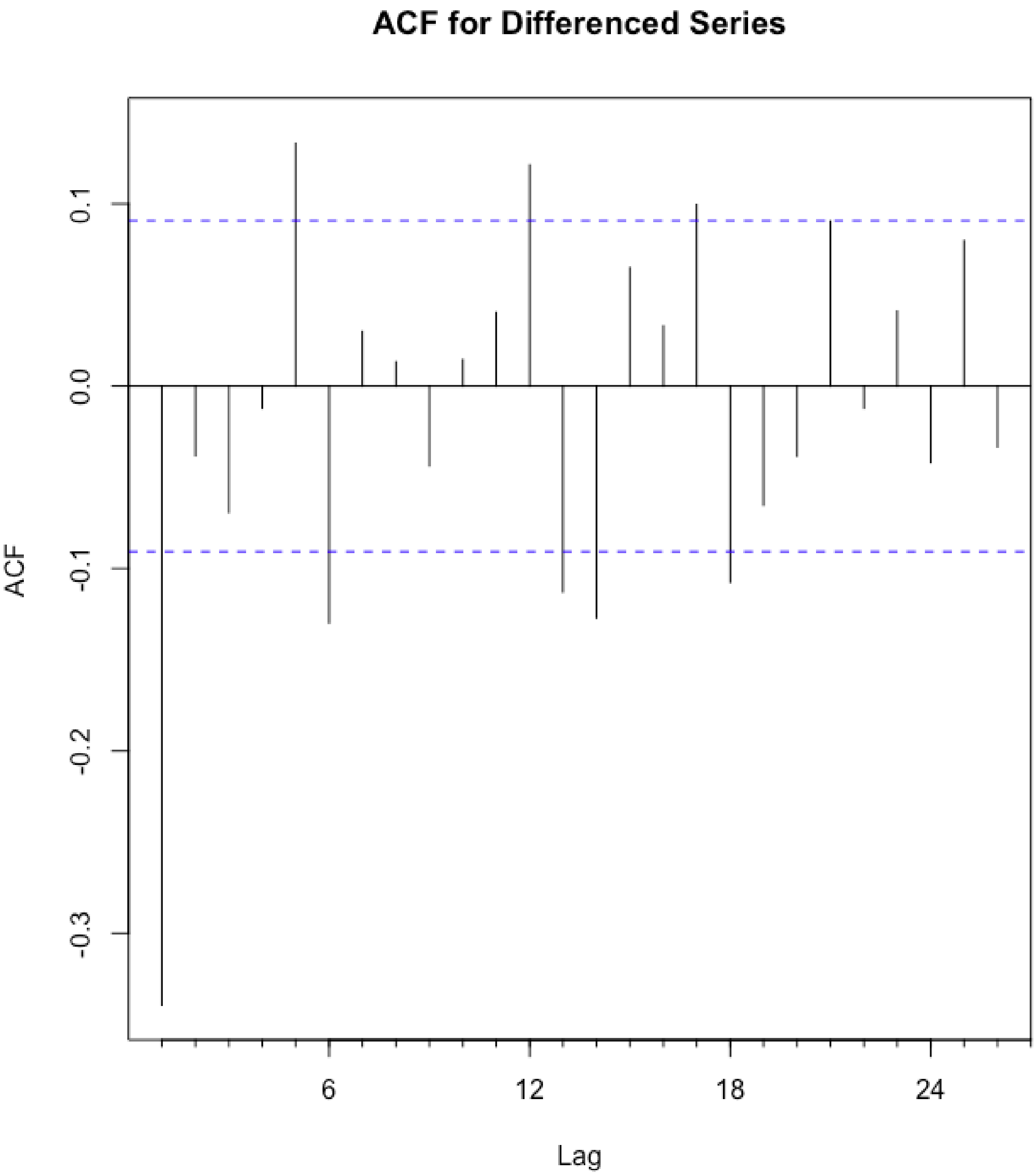


Differencing the Time Series Data :

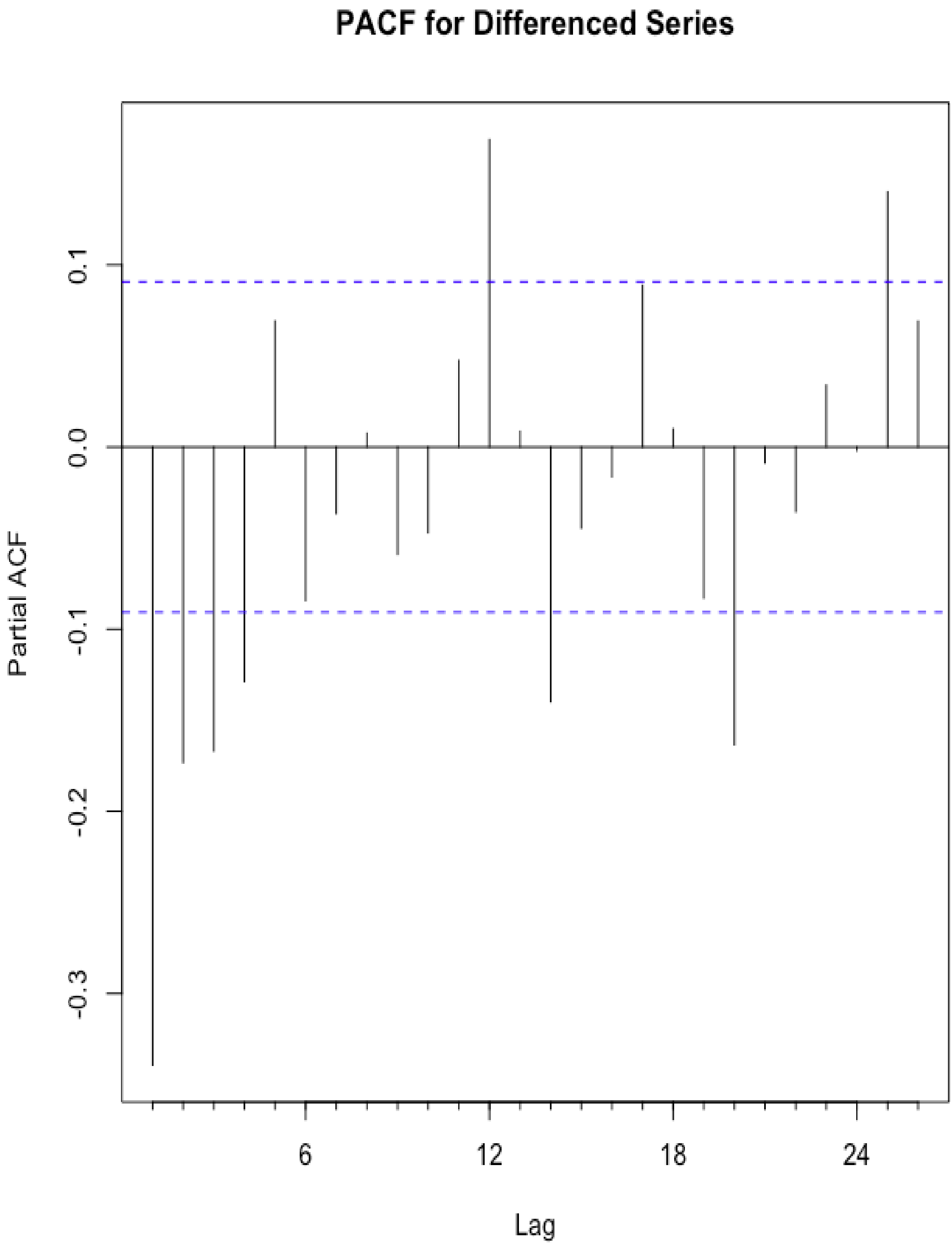
```
> count_d1 = diff(deseasonal_gas, differences = 1)
> plot(count_d1)
> adf.test(count_d1, alternative = "stationary")

        Augmented Dickey-Fuller Test

data:  count_d1
Dickey-Fuller = -9.1709, Lag order = 7, p-value = 0.01
alternative hypothesis: stationary
```



Acf (count_d1, main='ACF for Differenced Series')



Pacf(count_d1, main='PACF for Differenced Series')



Splitting Data Set :
gasTStrain = window(deseasonal_gas, start=1956, end=c(1980,12))
> gasTStest= window(deseasonal_gas, start=1981, end=c(1994,12))

- Develop an ARIMA Model to forecast for next 12 periods. Use both manual and auto.arima (Show & explain all the steps) **(20 marks)**
- Report the accuracy of the model **(5 marks)**

Arima Model:

gasARIMA = arima(gasTStrain, order=c(0,1,0))

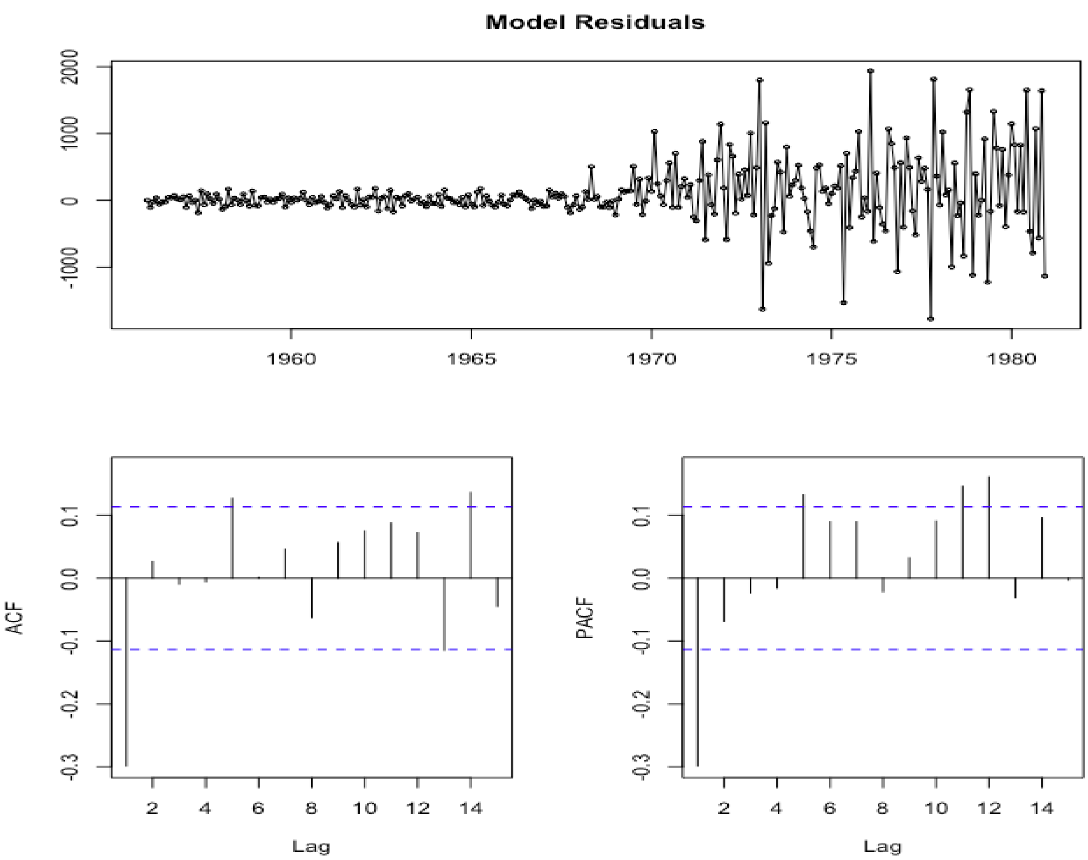
gasARIMA

Call:

arima(x = gasTStrain, order = c(0, 1, 0))

sigma^2 estimated as 235166: log likelihood = -2273.29, aic = 4548.57

Model Residuals:



Auto-Arima model:

> autoarima1<-auto.arima(gasTStrain, seasonal=FALSE)
> autoarima1
Series: gasTStrain
ARIMA(1,2,2)

Coefficients:

	ar1	ma1	ma2
	0.1520	-1.5403	0.5632
s.e.	0.1387	0.1187	0.1177

sigma^2 estimated as 194414: log likelihood=-2237.73
AIC=4483.46   AICc=4483.6   BIC=4498.25

> autoarima2<-auto.arima(gasTStrain, stationary=TRUE)
> autoarima2
Series: gasTStrain
ARIMA(0,0,2)(0,0,2)[12] with non-zero mean

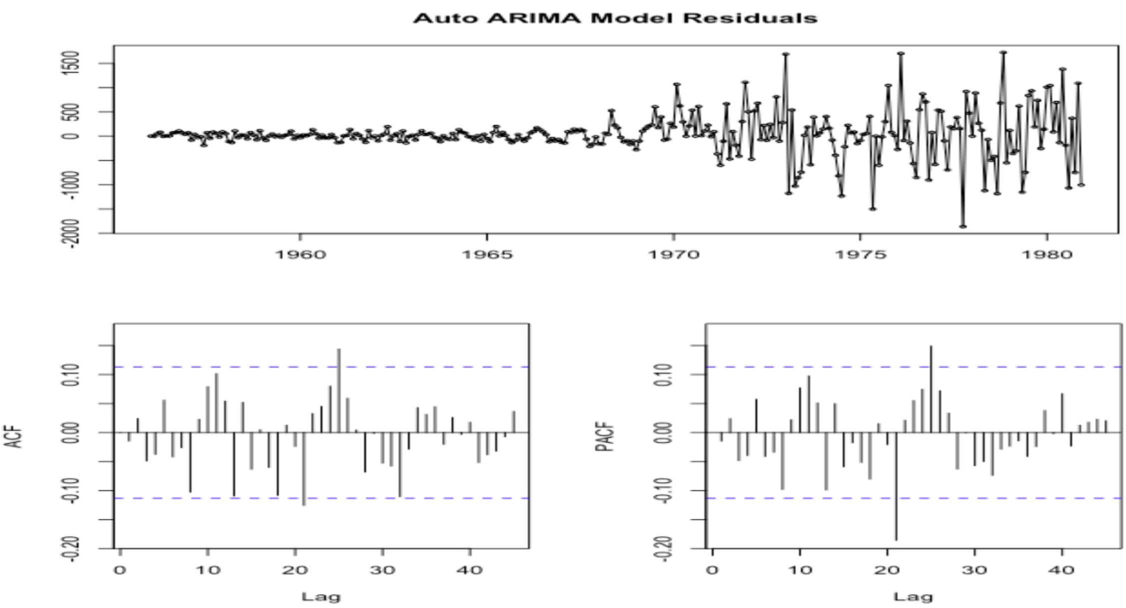
Coefficients:

```
        ma1     ma2    sma1    sma2       mean
     1.0029  0.9646  1.2704  0.6789  9360.0421
s.e.  0.0181  0.0149  0.0531  0.0458   703.9277
```

sigma^2 estimated as 2126358: log likelihood=-2622.19
AIC=5256.37   AICc=5256.66   BIC=5278.6

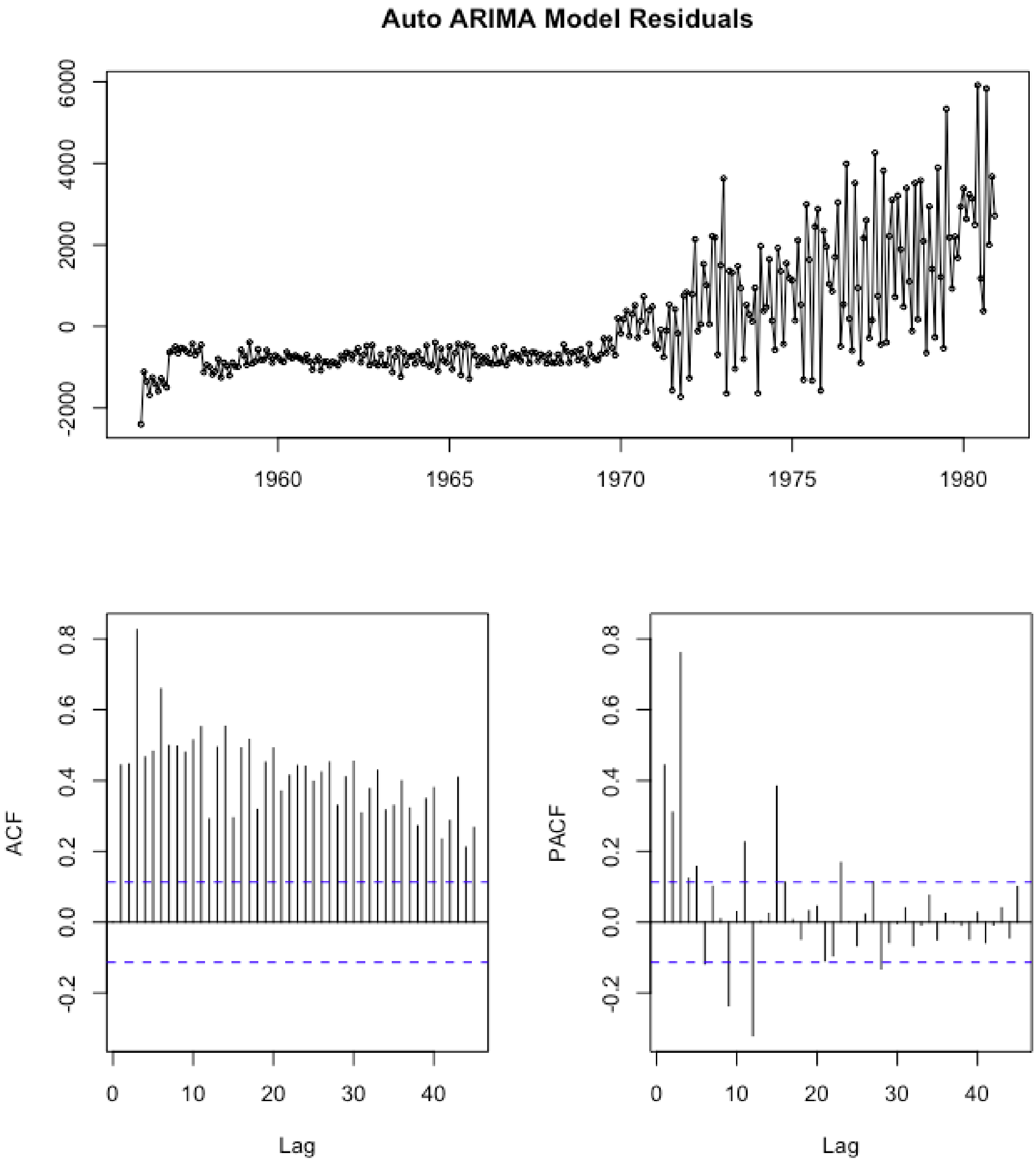
tsdisplay(residuals(autoarima1), lag.max=45, main='Auto ARIMA Model Residuals')

Model Residuals:



```
tsdisplay(residuals(autoarima2), lag.max=45, main='Auto ARIMA Model Residuals')
```

Model Residual Auto-Arima Model 2:



#Ljung box test
####H0: Residuals are independent####
####Ha: Residuals are not independent######

1.Box.test(gasARIMA$residuals)

Box-Pierce test

data: gasARIMA$residuals
X-squared = 26.721, df = 1, p-value = 2.351e-07

2. Box.test(autoarima1$residuals)

Box-Pierce test

data: autoarima1$residuals
X-squared = 0.064832, df = 1, p-value = 0.799

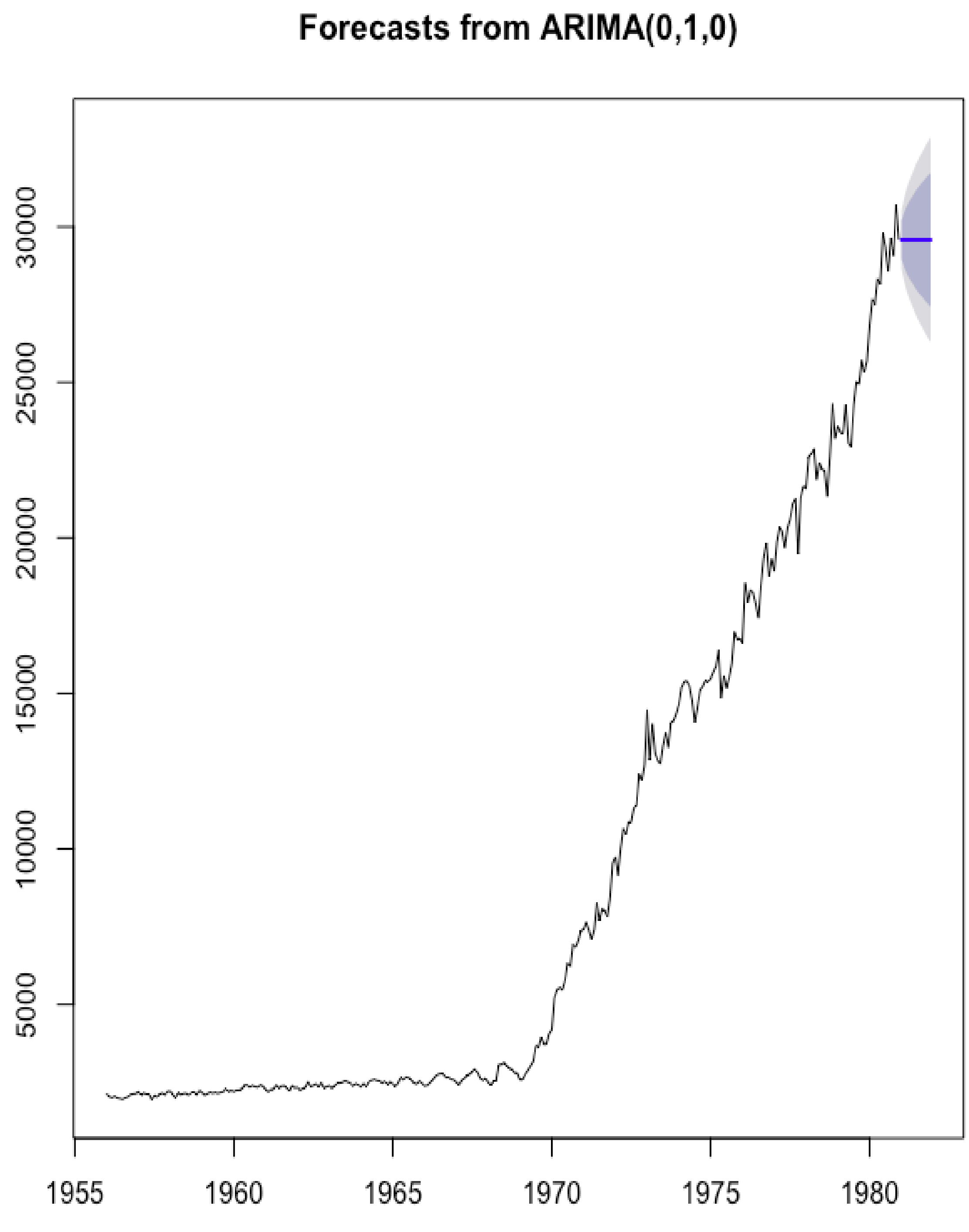
3. > Box.test(autoarima2$residuals)

Box-Pierce test

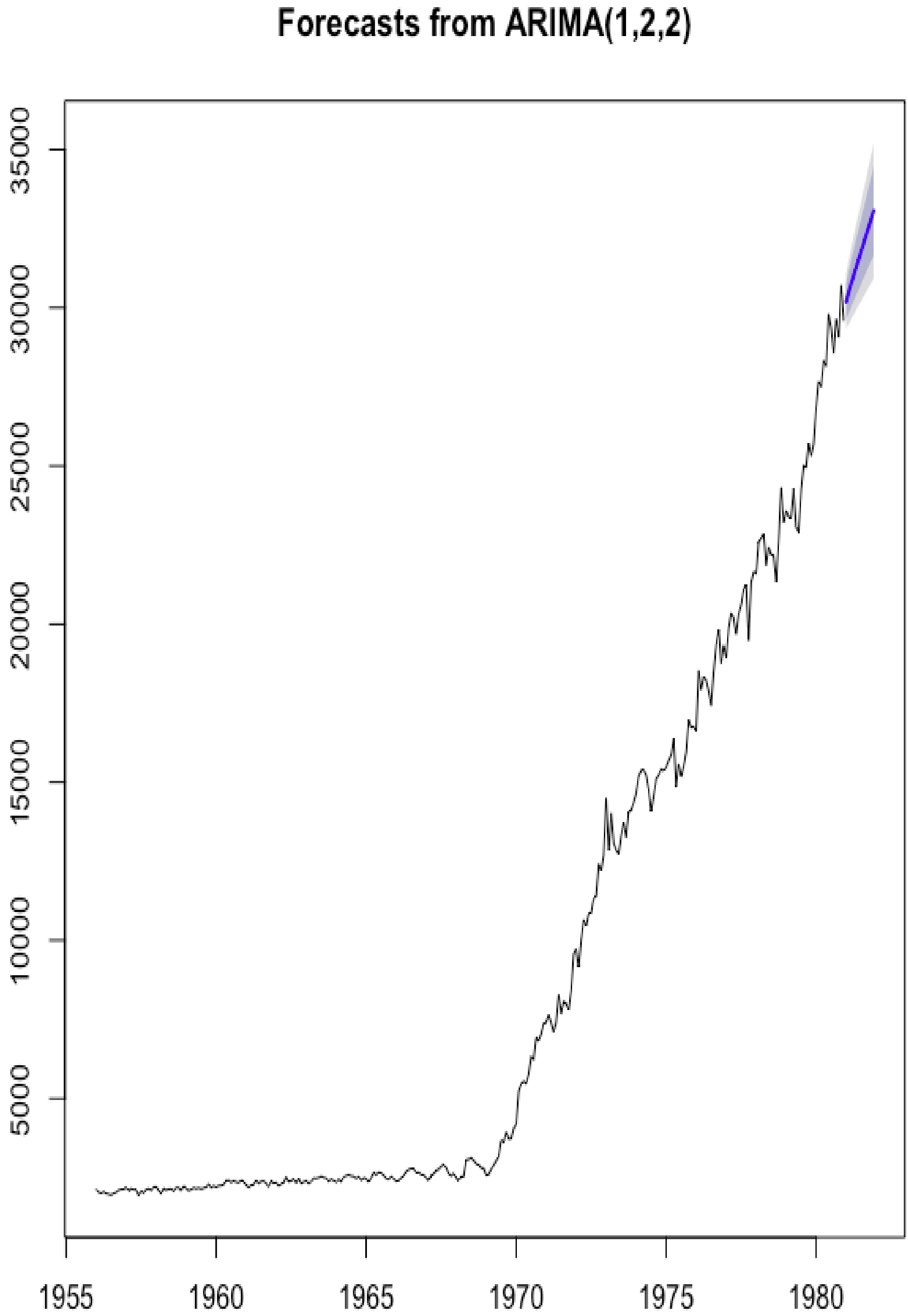
data: autoarima2$residuals
X-squared = 58.923, df = 1, p-value = 1.643e-14

Forecasting With Arima Model:

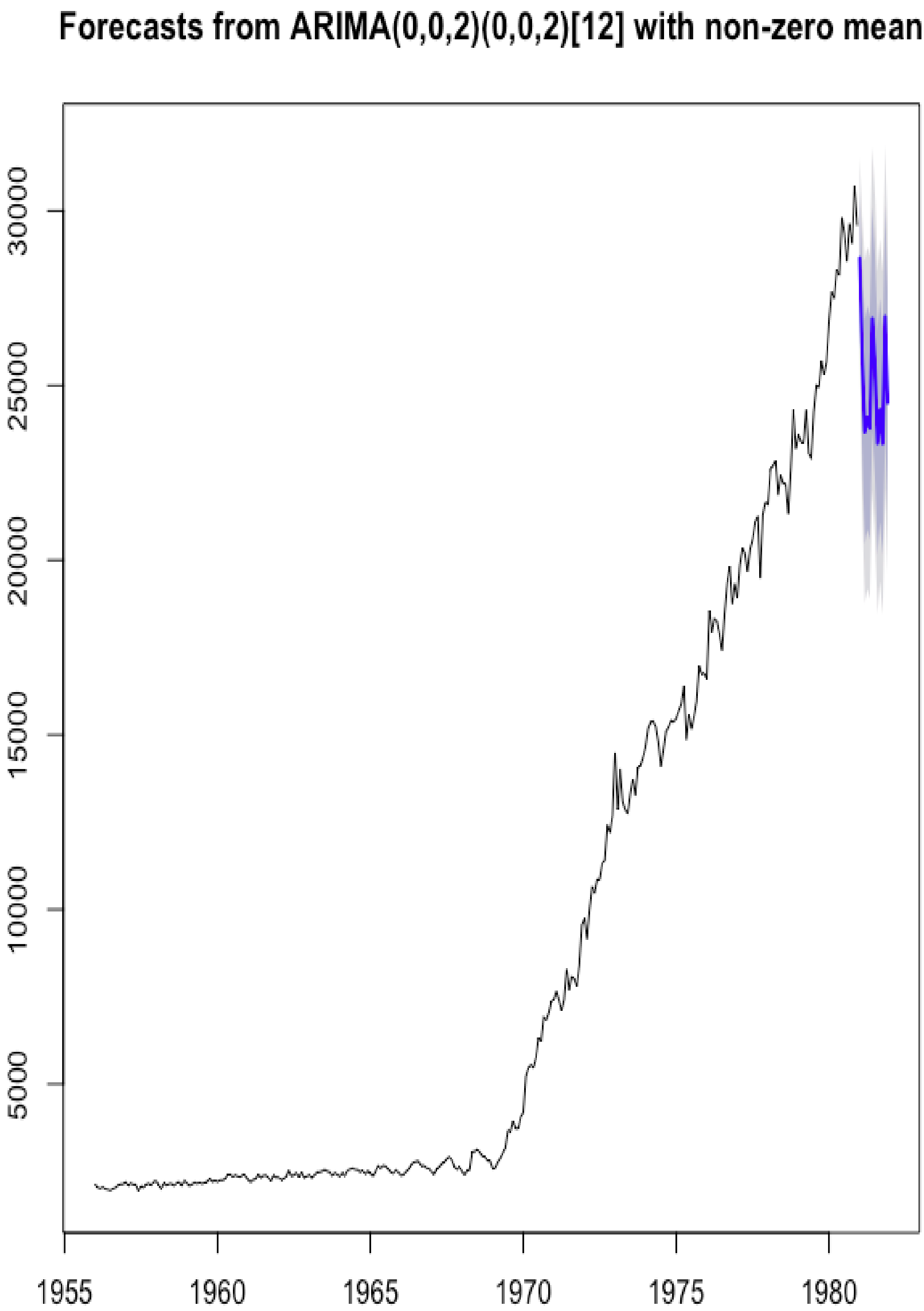
```
fcast <- forecast(gasARIMA, h=12)

plot(fcast)
```



```
fcast1 <- forecast(autoarima1, h=12)
plot(fcast1)
```



```
fcast2<- forecast(autoarima2, h=12)
plot(fcast2)
```



Accuracy of the model:

> f5=forecast(gasARIMA)
> accuracy(f5, gasTStest)

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Training set	91.47174	484.1298	287.7958	0.7786581	3.327277	0.2549334	-0.2984452	NA
Test set	5925.56223	6166.1954	5925.5622	16.4959721	16.495972	5.2489420	0.2474454	2.916241

> f6=forecast(autoarima1)
> accuracy(f6, gasTStest)

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Training set	32.67425	437.2339	260.1139	0.4336466	3.231609	0.2304123	-0.01470057	NA
Test set	2316.34348	3167.0653	2790.8694	6.3608766	7.774813	2.4721893	0.41476592	1.458487

f7=forecast(autoarima2)
accuracy(f7, gasTStest)

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Training set	47.40557	1446.001	1106.272	-16.62639	22.90379	0.9799501	0.4431819	NA
Test set	14792.25620	15709.469	14792.256	41.43893	41.43893	13.1031777	0.8179726	7.488328