DSCI5340_HW1_Group2

2023-09-23

{r setup, include=FALSE} knitr::opts_chunk$set(echo = TRUE)

```
# Set the CRAN mirror URL
options(repos = "https://cran.r-project.org/")

# Install the "seasonal" package
install.packages("seasonal")

# Load the "seasonal" package
library(seasonal)
```

Install and Load Packages fpp3, dplys, GGally, gridExtra, tidyverse

```
pacman::p_load(dplyr, fpp3, GGally, gridExtra, tidyverse)
theme_set(theme_classic())
```

Understanding the Dataset:

```
canadian_gas
dim(canadian_gas)
head(canadian_gas)
tail(canadian_gas)
```

Description of the Data Set:

Monthly Canadian gas production, billions of cubic metres. Data is available from January 1960 - February 2005

** The data set has 542 records and 2 columns. The Index variable is “Month” and Measured Variable is “Volume”, Monthly Time series of class ‘tsibble’.

1. What type of Box-Cox transformation would be helpful for the data? Explain your answer.

```
cg <- canadian_gas |>
  summarise(Total_Production = sum(Volume))
cg

p1 <- cg |> autoplot(Total_Production) +
  labs(y = "Total Gas Production in billion of cubic metres")+
  ggtitle("Total Gas Production without Transformation")
p1

# Box-Cox Transformed Total Gas Production with Lambda Value = 1
```

```
pl1 <- cg |> autoplot(box_cox(Total_Production, 1)) +
  labs(y = "Box=Cox transformed Total Production")+
  ggtitle("Box-Cox Transformed Total Gas Production with Lambda Value
= 1")
pl1

# Box-Cox Transformed Total Gas Production with Lambda Value = 0.5

pl0.5 <- cg |> autoplot(box_cox(Total_Production, 0.5)) +
  labs(y = "Box=Cox transformed Total Production")+
  ggtitle("Box-Cox Transformed Total Gas Production with Lambda Value
= 0.5")
pl0.5


# Box-Cox Transformed Total Gas Production with Lambda Value = 0
pl0 <- cg |> autoplot(box_cox(Total_Production, 0)) +
  labs(y = "Box=Cox transformed Total Production")+
  ggtitle("Box-Cox Transformed Total Gas Production with Lambda Value
= 0")
pl0


# Box-Cox Transformed Total Gas Production with Lambda Value = -1
plminus1 <- cg |> autoplot(box_cox(Total_Production, -1)) +
  labs(y = "Box=Cox transformed Total Production")+
  ggtitle("Box-Cox Transformed Total Gas Production with Lambda Value
= -1")
plminus1


## Box-Cox Transformation. Using "guerrero" feature to get optimal
lambda

cg |> features(Total_Production, features = guerrero) |>
pull(lambda_guerrero)

p2 <- cg |> autoplot(box_cox(Total_Production, 0.5767648)) +
  labs(y = "Box=Cox transformed Total Production")+
  ggtitle("Box-Cox Transformed Total Gas Production with Optimal
Lambda Value")
p2

# Comparing the Time Series plot for Lambda = 0.5 & Optimal Lambda
value obtained
grid.arrange(p2, pl0.5, nrow = 2)
```

In the graph title “Total Gas Production without Transformation” ((without Box Cox Transformation and Mathematical Transformation) the data shows different variation at different levels of the series. Applied Box Cox Transformation to stabilize the variation.

For Lambda = 1, there is no transformation change observed with respect to the original time series plot without any transformation.

For Lambda = 0.5: Almost a constant variation is observed throughout the time series plot.

For Lambda = 0, the variations found increased in the beginning of the time series and the variation got decreased drastically at the end of the time series.

For lambda = -1, there is a inverse variation observed across the time series.

For Optimal Lambda = 0.5767648, the variations found almost similar to the variations observed in the time series plot plotted with Lambda value = 0.5

Hence, for the above observations, the Box Cox Transformation with Lambda value = 0.5 could be helpful for the data.

2. Subset the Canadian_gas data and keep observations from March 1990 to February 2005. Plot the data using autoplot(), gg_subseries(), and gg_season(). Describe the presence (or absence) of seasonality in the data.

```
subset_data <- canadian_gas %>%  filter(year(Month) >= 1990 &
(year(Month) < 2005 | (year(Month) == 2005 & month(Month) <= 2)))
cg2 <- subset_data %>%  filter(Month %in% seq(as.Date("1990-03-01"),
as.Date("2005-02-28"), by="months"))
cg2

head(cg2)
tail(cg2)

##Autoplot

cg2 |> autoplot(Volume)+
  xlab("Month") + ylab("Total Gas Production in billion of cubic
metres")+
  ggtitle("Autoplot of the Gas Production from Mar 1990 to Feb 2005")

## Sub series Plot

cg2 |> gg_subseries(Volume)+
  ylab("Total Gas Production in billion of cubic metres")+
  xlab("Years")+
  ggtitle("Sub Series Plot of the Gas Production between Mar 1990 to
Feb 2005")

cg2 |> gg_season(Volume, labels = "both")+
```

```
   ylab("Total Gas Production in billion of cubic metres")+
   xlab("Month")+
  ggtitle("Seasonal Plot of the Gas Production between Mar 1990 to Feb 2005")
```

Describe the presence (or absence) of seasonality in the data

The subset data has 180 records and 2 variables.Applied head() and tail() to ensure the data is as per the subset condition.

Overall there is an increase in the gas production from Mar-1990 to Feb-2005. There is a positive trend in gas production over the period. Further there are variations i.e., peaks and troughs indicating there is a seasonality.

From the sub series plot, it is observed that the Average gas production in Canada is highest in Jan & Dec across the time period from Mar-1990 to Feb-2005 and lowest in June across time period from Mar-1990 to Feb-2005.

From Seasonal Plot, seasonality is observed as there is increase in gas production in January & December of every year across the given time period. Further, lowest gas production is observed in June month of every year across all the years of the given time period. Overall there is seasonality.

3. Perform an STL decomposition of the data. Plot the seasonal component from STL decomposition using gg_season(). Does the seasonal shape change over time? Explain.

STL Decomposition of the data:

```
dcmp <- canadian_gas |>
  model(stl = STL(Volume))
components(dcmp)
```

The out of the STL decomposition shows the components of STL decomposition (trend, season_year, remainder, season_adjust)

STL Decomposition

```
components(dcmp) |> autoplot() + xlab("Year")
```

All the components of the STL decomposition is plotted above using autoplot().

seasonal component from STL decomposition using gg_season()

```
components(dcmp) |> gg_season(season_year)
```

By observing the season component using gg_season(), change in season shape observed, as there are peaks and troughs throughout the time series. Peak magnitude is observed in January & December Months of every year and highest magnitude of troughs is observed in most of the June Months of ever year in the given time series. Hence,because of the above reasons, the seasonal shape changes over time.

4. Plot a seasonally-adjusted series using the results from the model above. Using this plot, explain why the seasonally-adjusted series is different from the original series.

```
canadian_gas |> autoplot(Volume, color = 'gray') +
  autolayer(components(dcmp), season_adjust, color = 'blue')+
    labs(title = "Total Gas production",
       subtitle = "Seasonally - Adjusted",
       y = "Total Gas Production in billion of cubic metres")
```

Actual model is Volume = trend + season_year + remainder. Seasonally adjusted Model= Volume - season_year.

From the above, as the seasonally adjusted series contain the remainder component as well as the trend-cycle component, the series is not smooth. Further, the presence of downturns and upturns are misleading to interpret the change in direction. The trend component of the original series is smooth where as the trend component of the seasonally adjusted series is not smooth. Hence, the seasonally adjusted series is different from the original series.

5. Next, perform decomposition using SEATS and X-11 decomposition methods. How are they different from the results obtained using the STL decomposition?

```
install.packages("seasonal")
library(seasonal)
```

X-11 Decomposition Method:

```
x11_dcmp <- canadian_gas |>
  model(x11 = X_13ARIMA_SEATS(Volume ~ x11())) |>
  components()
autoplot(x11_dcmp) +
  labs(title = "Decomposition of Total Gas Production using X-11
method")
```

SEATS Decomposition Method:

```
seats_dcmp <- canadian_gas |>
  model(seats = X_13ARIMA_SEATS(Volume ~ seats())) |>
  components()
autoplot(seats_dcmp) +
  labs(title = "Decomposition of Total Gas Production using SEATS
method")
```

Difference we noticed:

1) Seasonal and irregular component is centered around 1 for X-11 and SEATS decomposition where as seasonal and irregular component is centered around 0 for STL decomposition.

2) STL decomposition is additive decomposition where as X-11 & SEATS decomposition is multiplicative decomposition.
3) Variation in the seasonality in the beginning of the series is high is X-11 & SEATS decomposition where as the variation in seasonality is small in the beginning of series in STL decomposition.
4) Variation in the seasonality in the end of the series is small is X-11 & SEATS decomposition where as the variation in seasonality is high in the end of series in STL decomposition.