ECON 427: HOMEWORK #2

1. Using the series from sportinggoods.csv

Chart

Description automatically generated

Figure 1.1

* 1. In the Figure 1.1, we could see an upward sloping trend over the years as well as the magnitude of retail sales looking increasing. Along with that, if we zoom in the time-series graph, we could also observe seasonal variation within a year and the business cycles observed do not follow a certain pattern of any kind – it is quite irregular and random.
  2. The time-series in Figure 1.1 does not appear mean stationary because the trend that is shown is not horizontal which depicts that mean would be different at different points of time in the graph.
  3. The series in Figure 1.1 does not appear variance stationary as the magnitude is shown to be increasing over the period of time, i.e., the difference between ups and downs in retail sales are seen to be increasing over the period of time.

Chart

Description automatically generated

Figure 1.5

In a correlogram, dotted lines correspond to normally distributed autocorrelation, which means correlation of a variable with itself at different points in time. In Figure 1.5, we could obviously see that more than 5% of our data resides outside the dashed lines and also at lag 1.0 and 2.0, there seems to be a spike in sales of Sporting Goods which depicts seasonal variation. Due to these reasons, we would reject the null hypotheses.

Chart, histogram

Description automatically generated

Figure 1.6

My interpretation is same in f. as it was in e. as it is shown in Figure 1.6, almost all data is outside the dotted lines and spikes are also visible at lag of 2, 4, 6, etc.

* 1. Random component of the series

Chart

Description automatically generated

Figure 1.7

* 1. In Figure 1.7, the random component appears to be mean stationary as it is quite horizontal over the period of time even though the magnitude of Retail sales is fluctuating a lot between 1992 and 2020. Over the time period, it looks like the mean of all observations is going to be same.
  2. In Figure 1.6, the random component isn’t variance stationary as the magnitude of retail sales shows trend as well as seasonal variation during some years. The difference between various business cycles throughout the business seems to be quite important for this data.
  3. Correlogram of the random component

Chart

Description automatically generated

Figure 1.8

The correlogram of the random component shows that some sort of seasonal variation over time – spikes at 1.0 and 2.0. also, more than 5% of the lines (data) is outside the dashed lined which concludes that we should reject the null hypotheses and say that there is no autocorrelation visible in the random component of Sporting Goods Store.

1. Using the series from UNRATENSA.csv

Chart, line chart

Description automatically generated

Figure 2.1

I have seasonally adjusted the series and plotted it along with the raw data in Figure 2.1. looking at both the additive and multiplicative decomposition, it looks like additive composition seasonally adjusts the national unemployment rate data from January 2000 to August 2020.

Chart, line chart

Description automatically generated

Figure 2.2a

Chart, line chart

Description automatically generated

Figure 2.2b

Figure 2.2a compares seasonally adjusted series that government calculated with the multiplicative decomposition and Figure 2.2b compares it with additive decomposition. All the figures are almost similar with minor variations over the period of time. Additive decomposition is more similar to the series that government calculated as it seems to be more smoothed out compared to multiplicative decomposition.

1. Use the series from inventory.csv and prices.csv
   1. Prices.csv file presents median listing price of housing inventory in Summit County from July 2016 to August 2020 whereas Inventory.csv presents median days of housing inventory on market in Summit County from July 2016 to August 2020. Leading variable means a related variable that leads the variable we want to forecast by one or more-time intervals. As per the given data of Prices and Inventory in Summit County, I believe inventory is the leading variable for prices. Prices depend on sale of inventory (if there are more sales, prices would be eventually got higher and vice versa) other economic factors.

Chart, line chart, histogram

Description automatically generated

Figure 3.2a

Chart, line chart, histogram

Description automatically generated

Figure 3.2b

Both the figures seem to be moving in opposite direction, except for following the magnitude of changes in both the graphs. Figure 3.2a shows the housing inventory and it indicates that the trend of inventory staying in market for more days was higher in the beginning of 2000 and eventually towards 2020, it seems to have been decreased with some seasonal variation. The lowest inventory in market was observed around 2020.

Contrary, in Figure 3.2b, prices of houses seem to be increasing over the years with some seasonal variation and it was quite low around 2017.

* 1. Cross-correlation between Prices and Inventory

Chart, box and whisker chart

Description automatically generated

Figure 3.3

In Figure 3.3, Prices.ts is the dependent variable and Inventory.ts is the independent variable. We can see how values of Inventory.ts at different periods of time correlate with the current time (value) of Prices.ts. the log 0 at x-axis shows the correlation between inventory and prices at the same time period. The first bar to the left of lag 0 shows correlation between inventory at time t-1 and prices as time t, so on and so forth. The dashed lines would help us figure out the significance of the correlation. As observed in Figure 3.3, all bars on top are significant as none of the data is present outside the dashed lines.it shows that past values of inventory are correlated with the current value of prices which concludes that inventory is a leading variable for prices. Also, the lines below ACF 0.0 displays the negative cross-correlation between inventory and prices. As more lines are crossing the dashed blue lines the variables are statistically significant and we can use inventory to forecast prices.

This confirms the hypotheses I did in part(a).

1. Use the series from sporttinggood.csv
   1. Smoothed series on top of raw data

Chart

Description automatically generated

Figure 4.1

* 1. (i) alpha = 0.2

Chart

Description automatically generated

Figure 4.2a

(ii) alpha = 0.05

A picture containing text, antenna, screenshot

Description automatically generated

Figure 4.2b

(iii) alpha = R chose automatically

Chart

Description automatically generated

Figure 4.2c

The SSE of all three different alphas with values such as 0.2, 0.005, and whatever R chose are 125273140, 120960948, and 117266312 respectively. According to that, the alpha chosen by R automatically minimizes the sum of squared predictions.

* 1. (i) Model generated series on the same plot as raw data:

Chart

Description automatically generated

Figure 4.3a

(ii) Forecasted series 5 years into future

Chart

Description automatically generated

Figure 4.3b

(iii)

Looking at Figures 4.3a and 4.3b, it looks like a good model to forecast retail sales of sporting goods, because when we plotted the generated model on the raw data it looks like it followed the seasonal variation and trend that was present in the original data. Black lines show the actual data and red line shows the forecasted data, so if we extend the model to 5 years later, it will pretty much give a good estimate of sales to be expected.

The values for smoothing parameters are as follows:

alpha – 0.688854

beta – 0

gamma – 1

According to this observation, it seems that Gamma has the most weight (100%) to observe the most recent data compared to past data (0%).

68.9% weight for calculating a new value of alpha is coming from the most recent data observed and the old value of alpha gets 31.1% weight, whereas for gamma 0% weight is used for calculating the most recent value and the old value of beta gets 100% weight.