**STAT 5350**

**MOHAMMED RAZA SYED**

**ASSIGNMENT #3**

**Q1)**

The time series data for retail sales of beer, wine, and liquor shows both an increasing trend and an increasing magnitude of seasonal fluctuations over time. This suggests that the seasonal component is proportional to the overall level of the series, which is a key characteristic of a multiplicative relationship. In contrast, an additive model assumes constant fluctuations regardless of the series' level, which would not suit this data.

A graph showing the growth of sales

Description automatically generated

To confirm this, I performed a **median polish** on the sales data. The **Tukey Additivity Plot** (shown below) is a diagnostic tool that helps check whether the data follows an additive model. In the plot, the presence of structure in the residuals—particularly the curved pattern—indicates that an additive model does not fully capture the variations in the data. If the residuals were randomly scattered, it would suggest that the additive model was appropriate. However, the pattern observed supports the choice of a **multiplicative decomposition**.

Thus, the increasing seasonal variation and the diagnostic plot confirm that a multiplicative decomposition is more appropriate for this sales time series.

A graph with dots and numbers

Description automatically generated

**Q2)**

The seasonally adjusted sales were obtained by removing the seasonal component from the log-transformed data and converting it back to the original scale.

The plot below shows both the original sales (in blue) and the seasonally adjusted sales (in red). The original series exhibits clear seasonal spikes, while the seasonally adjusted series is smoother, revealing the underlying trend without the seasonal fluctuations.

A graph showing sales and sales

Description automatically generated

**Q3)**

The anticipated trend from the STL decomposition indicates a consistent increase in sales before March 2020. After Covid began, there was a significant increase in sales growth, followed by stabilization at a greater level than before. This suggests that Covid had a beneficial impact on long-term sales growth, most likely because of increased demand during the epidemic. The overall trend indicates that sales growth soared during the early months of Covid and stayed high afterward.

A graph showing the growth of sales

Description automatically generated

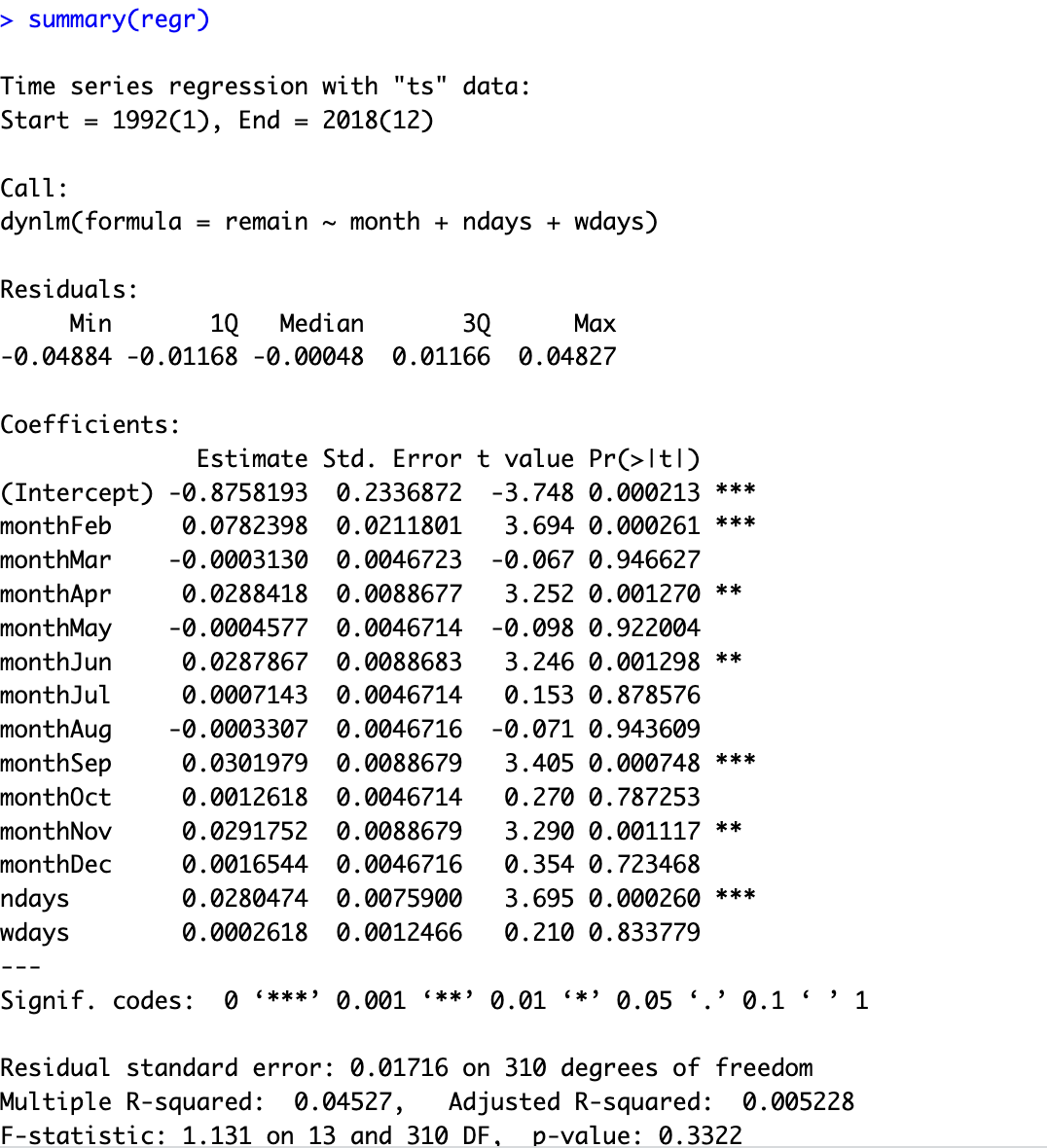
**Q4)**

There are no variations in the seasonal pattern found by STL before and after COVID-19. The seasonal components had less year-to-year fluctuations and were more consistent prior to the Covid-19 pandemic. Post-Covid (2022–2023), there are minor upward movements in the early months (January to March) and minor changes in the mid-year months (e.g., May, August), even though the overall form of the seasonal pattern stays the same. The seasonal high in December, however, doesn't change between the two times. As a result, despite some fluctuation brought about by the pandemic, the basic seasonal pattern has mainly continued same with little to no modification or changes

**A graph of a number of covid-19

Description automatically generated**

**Q5)**



The regression analysis reveals that the remainder from the STL decomposition still contains **seasonal information**. Several **month variables** (February, April, June, September, and November) were statistically significant with p-values below 0.05, indicating that the remainder still exhibits some seasonal effects. Additionally, the **number of days in a month (ndays)** was also highly significant, suggesting that the length of the month influences the remainder.

Despite the STL decomposition, which should have removed seasonal patterns, these results indicate that **some seasonality remains** in the remainder component. This suggests that the decomposition did not fully capture all seasonal variations, particularly those tied to specific months and the number of days in a month. However, the **number of weekdays (wdays)** did not show any significant effect, meaning that the number of weekdays in a month does not impact the remainder.

**Q6)**

1. From the visual inspection, it can be inferred that From the visual inspection, it can be inferred that the **ACF plot** shows a **gradual decay**, indicating an autoregressive process. Meanwhile, the **PACF plot** exhibits a **sharp cutoff after lag 2**, suggesting the presence of **2 autoregressive terms** with **p = 2**. Since there is no significant moving average component indicated by the ACF, the time series can be identified as following an **ARMA(2, 0)** process.

A graph of a series of lag

Description automatically generated with medium confidence

1. From the calculation of AIC and BIC, it can be observed that **AIC suggests an ARMA (4, 1)** process, while **BIC indicates an ARMA (2, 1)** process.

A screenshot of a computer

Description automatically generated

1. The models chosen in parts (a) and (b) do not fully agree. From visual inspection in part (a), an **ARMA (2, 0)** model was suggested, while in part (b), **AIC** selected an **ARMA (4, 1)** model, and **BIC** suggested an **ARMA (2, 1)** model. After fitting both the **ARMA (2, 0)** **ARMA (4, 1)** and **ARMA (2, 1)** models, it was found that the **AIC** for **ARMA (2, 1)** (3.0114) is lower than for **ARMA(2, 0)** (3.1445), indicating that **ARMA(2, 1)** provides a better fit according to AIC. Similarly, the **BIC** for **ARMA (2, 1)** (3.1118) is lower than for **ARMA (2, 0)** (3.2247), meaning **BIC** also favors the **ARMA (2, 1)** model. Furthermore, in the **ARMA (2, 1)** model, both the AR and MA terms are statistically significant, whereas in the **ARMA (2, 0)** model, only the AR terms are significant, with no MA component. Also **ARMA (4,1)**  does indicate a decrease in BIC value as compared to our visual inspected model but **ARMA (2,1)** shows a much better model with lowest BIC.

Given these calculations we favor the **ARMA(2, 1)** model, and it provides a better overall fit with significant coefficients, I recommend using the **ARMA(2, 1)** model. This model includes both autoregressive and moving average components, offering a more balanced and effective fit for the time series.

A screenshot of a computer program

Description automatically generated

**Q7)**

1. From the visual inspection, it can be inferred that the **ACF plot** shows a sharp cutoff at lag 2, indicating a moving average component with **q = 2**. On the other hand, the **PACF plot** exhibits a slow decay, suggesting the presence of an autoregressive process with a small **p value of 1**. Therefore, the time series can be identified as following an **ARMA (1, 2) process**.

**A graph of a series of lag

Description automatically generated with medium confidence**

1. Now From the calculation of AIC and BIC, it can be seen that **AIC selects an ARMA(1, 3) model**, while **BIC selects an ARMA(0, 2) model**.

**A screenshot of a computer

Description automatically generated**

1. The models selected in parts (a) and (b) do not fully agree. Based on the visual inspection in part (a), an **ARMA (1, 2)** model was suggested. However, in part (b), **AIC** selected an **ARMA (1, 3)** model, while **BIC** preferred a simpler **ARMA (0, 2)** model.

When the results disagree, it is generally advisable to consider **BIC**, as it tends to favor simpler models with fewer parameters, thereby avoiding overfitting. In this case, we would recommend the **ARMA (0, 2)** model as chosen by BIC, as it strikes a balance between model simplicity and fit. The ARMA(0, 2) model adequately captures the moving average characteristics suggested by the ACF plot, without adding unnecessary complexity.

Also after running and fitting these two models using sarima we get:

In terms of BIC, the **ARMA (0, 2)** model has a lower BIC (2.9999) compared to the **ARMA(1, 2)** model (3.0329), and **ARMA (1,3)** model (3.0189) suggesting that the simpler model is preferred by BIC. Additionally, the **AR (1)** term in the ARMA(1, 2) model is not statistically significant (p-value = 0.7937), indicating that the autoregressive term adds little value. Both models have a similar error variance (sigma²), supporting the case for the simpler **ARMA(0, 2)** or **MA(2)** model.

A screenshot of a computer program

Description automatically generated

**Q8)**

1. From visual inspection, it can be inferred that the model is **ARMA(2, 0)**. This conclusion is based on the fact that the **ACF plot** shows a gradual decay, rather than a sharp cutoff, while the **PACF plot** exhibits a sharp cutoff after lag 2. These characteristics are typical of an AR process, where the **PACF** cuts off after lag **p**. Therefore, based on these observations, the process is identified as **ARMA(2, 0)**.

**A graph of a series of lag

Description automatically generated with medium confidence**

1. It can be seen it From the calculation of AIC and BIC, we can see that the results align with our visual inspection, both indicating an ARMA(2, 0) process. Both AIC and BIC have their minimum values at p = 2 and q = 0, confirming that AR(2) is the best model choice.

**A screenshot of a computer screen

Description automatically generated**

1. since both the visual inspection in part (a) and the AIC/BIC values in part (b) suggest the same model, **ARMA (2, 0)**, the two methods agree. Therefore, using the **AR (2)** model would be the best approach. Both methods indicate that this model best captures the structure of the time series.

**Q9**

1. set.seed(54)

xt <- arima.sim(n = 800, list(ar = c(1.6, -0.8)), sd = 1)

acf2(xt)

sarima(xt, p=2,d=0,q=0,details=FALSE)

A graph of a graph of a line

Description automatically generated with medium confidence

After generating the time series, we plotted the **ACF** and **PACF** to confirm that they exhibited the expected characteristics of an AR(2) process: a gradual decay in the ACF and a sharp cutoff after lag 2 in the PACF. This behavior aligns with the properties of an AR(2) process. To further confirm this, we estimated the model parameters using the **sarima()** function, which yielded AR(1) and AR(2) coefficients close to the true values of **1.6** and **-0.8**, respectively.

A screenshot of a computer code

Description automatically generated

**set.seed(62)**

**yt <- xt + rnorm(length(xt))**

**acf2(yt)**

**z <- fit\_arma\_models(yt)**

**round(z$aic, 3)**

**round(z$bic, 3)**

**sarima(yt, p=2,d=0,q=2,details = FALSE)**

Upon inspecting the **ACF** and **PACF** of the resulting time series and confirmation upon calculating the **AIC** and **BIC** values, both of which indicated that the **ARMA(2, 2)** model provided the best fit. Therefore, the time series, after the addition of the white noise, is best described as following an **ARMA(2, 2)** process.

**A graph of a series of data

Description automatically generated with medium confidence**

**A screenshot of a math document

Description automatically generated**

**Q10)**



For model 1:

Below are the polyroots of AR and MA components. The Magnitude of 2nd root of AR is close to the MA Root. Hence we can simply them to reduce redundancy. To further confirm our evidence we also fitted the model with and without redundancy as shown below

A screenshot of a computer

Description automatically generated

A screenshot of a computer program

Description automatically generated

For **Model 1**, an **ARMA (2, 1)** model was fitted to the simulated data with AR coefficients of 0.80 and -0.15, and an MA coefficient of -0.30. The results showed that none of the AR or MA terms were statistically significant, and the AIC (2.8517) was relatively high. In contrast, the reduced **AR(1)** model had a lower AIC (2.8479), and the AR(1) term was highly significant (p-value = 0.0000), indicating a better fit. Given the higher AIC and lack of significance in the ARMA(2, 1) model, it suggests that the **AR(1)** model is a more appropriate and simplified representation. Therefore, **Model 1** exhibits **parameter redundancy** and can be reduced to an **AR(1)** process.

Here is our final reduced model : xt​=-0.8xt−1​+wt​

For model 2:

Below are the polyroots of AR and MA components. The magnitude of the AR roots are not close to the MA roots hence it cannot be said the model has parameter redundancy. To confirm this we perform and fitted model with and without redundant parameters.

A computer code with text

Description automatically generated

A screenshot of a computer program

Description automatically generated

For **Model 2**, an **ARMA(2, 1)** model was fitted with AR coefficients of 1 and -0.50, and an MA coefficient of -1. All terms in the model were highly significant (p-value = 0.0000), and the AIC (2.8559) indicated a good fit. When the reduced **AR(1)** model was fitted, the AIC was much higher (3.2698), and the error variance increased significantly. This suggests that the **ARMA(2, 1)** model captures the full structure of the data, while the simpler AR(1) model fails to do so. Therefore, there is **no parameter redundancy** in **Model 2**, and the model should remain an **ARMA(2, 1)** process.

Model 1:

The reduced form of **Model 1** is an **AR(1)** process: 0.80xt−1​+wt​. To check for causality, the root of the AR polynomial was calculated to be **1.25**, which is greater than 1. Therefore, the model is **causal**. Since the model has no MA component, invertibility is not applicable to this model. Hence, the reduced **AR(1)** model is **causal**, and invertibility does not apply.

Model 2:

**Model 2 has** **ARMA(2, 1)** process. The AR ) has complex roots 1+ i, 1 – i, with magnitudes of approximately **1.41**, which are greater than 1, indicating that the model is **causal**. However, the MA polynomial has a root of **1**, which is on the boundary of invertibility. Therefore, the model is **causal**, but it is **not invertible**, as the MA root is on the edge of the invertibility condition.