MA585 Final Project

Title: Time-series Forecasting for the US Merchant Wholesalers Alcohol Sales

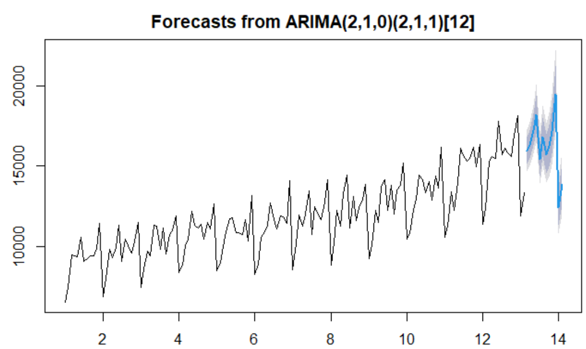
Name: Zehao Zhou

Dataset Name: Merchant Wholesalers, Except Manufacturers' Sales Branches and Offices: Nondurable Goods: Beer, Wine, and Distilled Alcoholic Beverages Sales

Data Source: https://fred.stlouisfed.org/series/S4248SM144NCEN

Final Model:

	Point Forecast <dbl>	Lo 80 <dbl>	Hi 80 <dbl>	Lo 95 <dbl>	Hi 95 <dbl>
Mar 13	15928.80	15111.68	16790.10	14696.23	17264.75
Apr 13	16315.38	15456.08	17222.46	15019.67	17722.88
May 13	17218.51	16265.92	18226.88	15783.17	18784.38
Jun 13	18175.65	17017.31	19412.83	16434.31	20101.49
Jul 13	15431.30	14409.50	16525.56	13896.27	17135.90
Aug 13	16783.82	15617.27	18037.49	15032.93	18738.62
Sep 13	15711.27	14550.32	16964.85	13970.88	17668.45
Oct 13	16213.63	14970.41	17560.10	14351.35	18317.57
Nov 13	17135.71	15769.02	18620.86	15090.22	19458.47
Dec 13	19486.04	17869.76	21248.52	17069.15	22245.16
Jan 14	12413.87	11350.49	13576.88	10824.95	14236.02
Feb 14	13836.45	12613.16	15178.39	12009.99	15940.68



Conclusion:

The goal of doing this report is to make a forecast of the US Alcohol sales in the next 12 months. In this report, I've used the SARIMA model selection and seasonal Holt-Winters as the two approaches to forecasting the time series. Based on the diagnostic and the accuracy check, I found the SARIMA could be a more appropriate way to forecast the Alcohol Sales data in the future. Therefore, I made a forecast for the next 12 months of Alcoholic beverages in the United States with this method. **[4] [5]** In the forecast, the sales of alcoholic beverages in the US are going to remain growing if it is without any intervention. And it's going to reach a higher peak in December(7.5% higher compared to the last peak in Dec 2021). Furthermore, industries related to sales of alcoholic beverages can also use the information and data provided in this report as a reference for their future planning.

Time-series Forecasting for the US Merchant Wholesalers Alcohol Sales

1. Introduction

Alcohol has long been a huge market branch of nondurable goods in the United States. As time goes by, the average demand for alcoholic beverages keeps growing. However, the sales of alcoholic beverages are not always good across the whole year. For merchant wholesalers, it becomes especially important to know about the change in the demand since they have limited space for their inventory. FRED(Federal Reserve Bank Economic Data), a very trusted source for economic data provides a detailed time series of changing alcoholic beverages' sales in the United States from Jan 1992 to Feb 2022. Within this time interval, the overall sales have increased by approximately 300%. In this report, I will apply seasonal ARIMA and seasonal Holt-Winters to forecast the sales of 2022.

2. Dataset

FRED has the data from the first month of 1992 to the beginning of 2022. However, not all data are useful for future prediction. From the graph given by FRED, there's a possible intervention from around 2007 to 2009. Thus, only the data starting from 2010 will be taken **(see Fig.1)**. The intervention could be the 2007 economic recession.

	DATE	Sales
1	1/1/2010	6558
2	2/1/2010	7481
3	3/1/2010	9475
4	4/1/2010	9424
5	5/1/2010	9351
6	6/1/2010	10552

Fig. 1 (a) Raw data captured by FRED

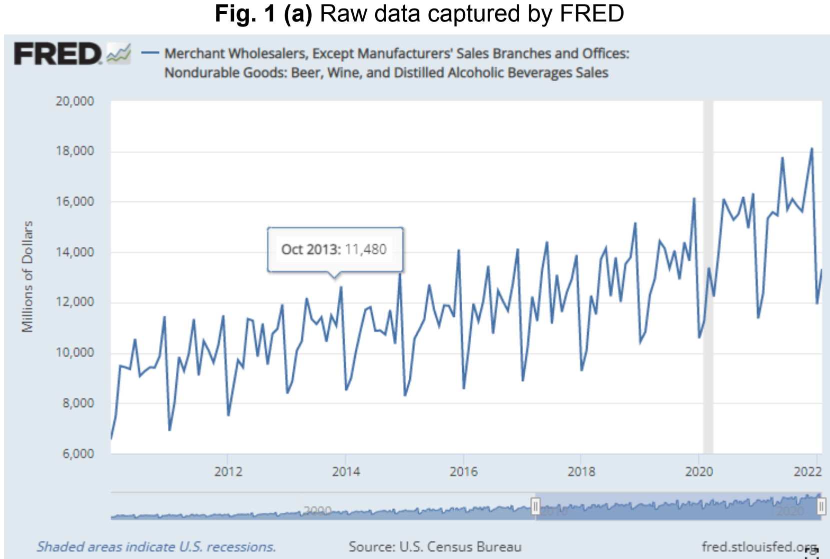


Fig. 2 (b) Time Series Plot Provided by FRED

3. **Results**

   a. **Visualize and evaluate the pattern of the data**

From the processed data of the time series captured by FRED**(see Fig. 2)**, the data does not appear stationary. There's an increasing trend and an obvious seasonality in the data. For confirmation, I've used a decomposition chart**(see Fig. 3)** that clearly shows the trend and seasonality part of the data. There's a period of 12 data points (12 months) in the data. Also, for the data I used, there could be a drift in the mean. Thus, The Dickey-Fuller test is used. Since the p-value of the test is 0.01 which rejects the null hypothesis that the time series is non-stationary, there's no need to correct for unit root.

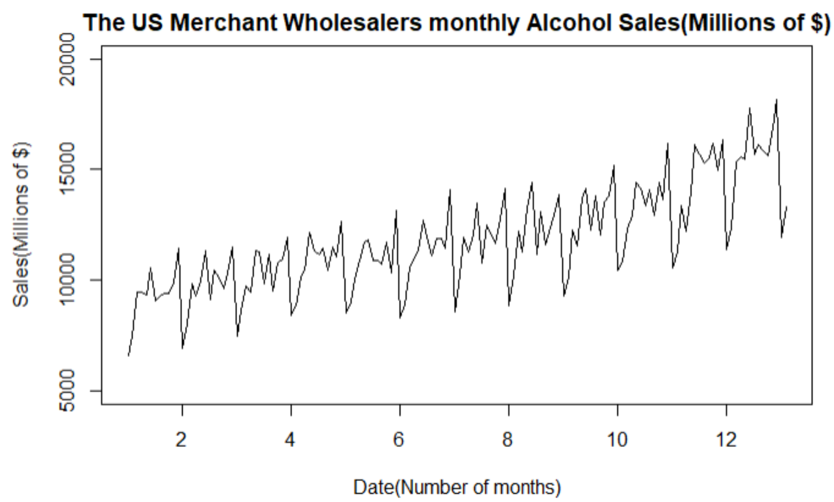


Fig. 2 The Time Series of The US Merchant Wholesalers' monthly Alcohol Sales

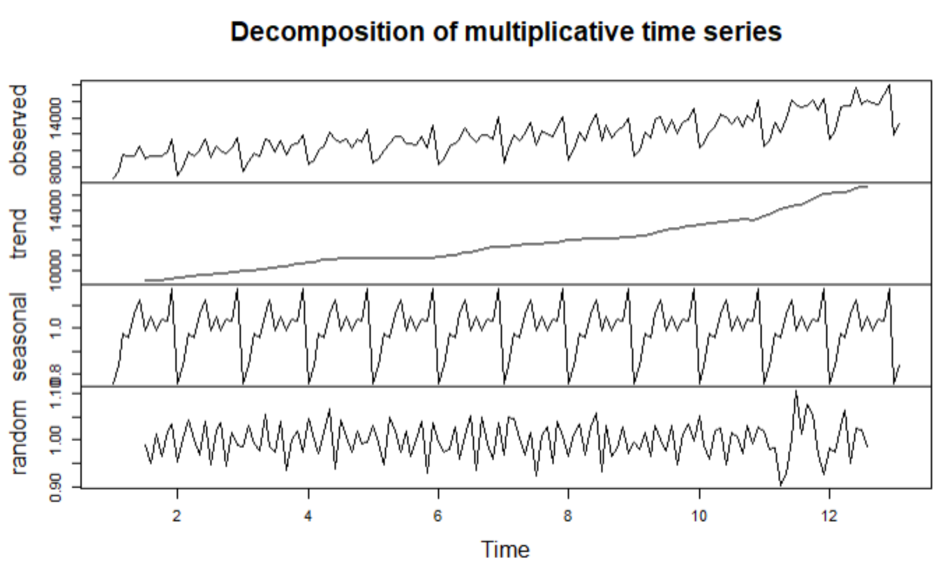


Fig. 3 The classical decomposition of the given time series

   b. **Transformation**

Based on the Dickey-Fuller test I've conducted, the data doesn't have to be corrected for the unit root. Therefore, the next step is to correct the effect of trend and seasonality; and there's no need to apply a variance stabilizing transformation. I applied order 1 and order 12 differencing **(see Fig. 4)** to correct both trend and seasonality.

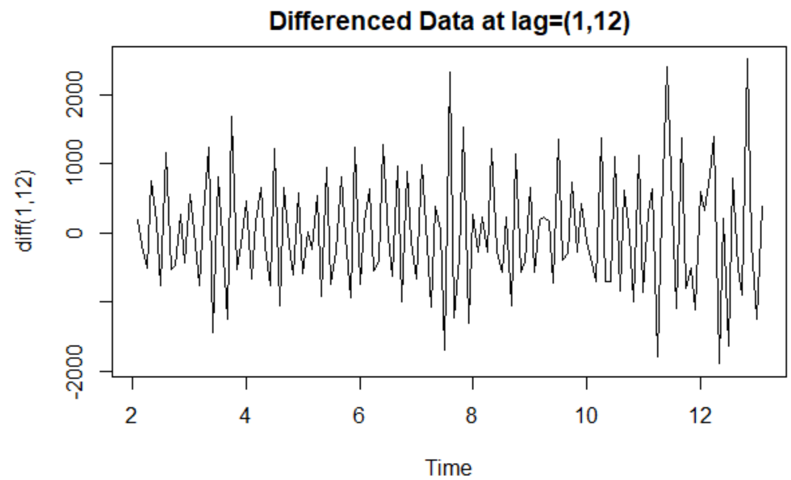


Fig. 4 differenced data at order 1 and order 12 of given time series data

c. Model-Based Forecast: ARIMA

The first forecasting approach that I used is by identifying the potential Seasonal ARIMA model of the given time series. From the ACF and PACF plot **(see Fig. 5)**, the time series seems to fit a low-order ARMA model. For the first few lags, ACF cuts off after lag 1 and PACF cuts off after lag 2. Thus, the non-seasonal part of the model could be AR(2), MA(1), and ARMA(2,1). For the seasonal part, ACF has a cut after lag 1 while PACF has a clear cut after lag 2. We can also interpret that ACF has a decreasing trend. Hence, seasonal AR(2) and ARMA(2,1) are appropriate. Thus, the potential models for this time series are

$(2, 1, 0) \times (2, 1, 0)_{12}$, $(0, 1, 1) \times (2, 1, 0)_{12}$, $(2, 1, 1) \times (2, 1, 0)_{12}$, $(2, 1, 0) \times (2, 1, 1)_{12}$, $(0, 1, 1) \times (2, 1, 1)_{12}$, $(2, 1, 1) \times (2, 1, 1)_{12}$.

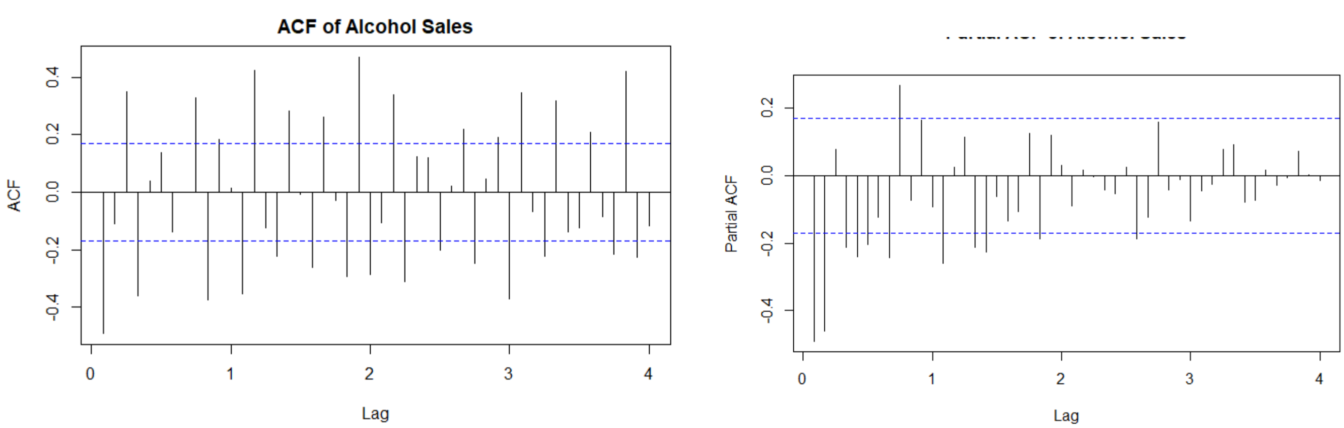


Fig. 5 ACF and PACF plot of the given time series

Fitting the data with each model we can interpret, we can get a table of AICc values for each SARIMA we try to fit **(see Table. 1)**.

Model	AICc
$(2,1,0) \times (2,1,0)_{12}$	-384.7
$(0,1,1) \times (2,1,0)_{12}$	-357.8
$(2,1,1) \times (2,1,0)_{12}$	-383.1
$(2,1,0) \times (2,1,1)_{12}$	-405.64
$(0,1,1) \times (2,1,1)_{12}$	-392.39
$(2,1,1) \times (2,1,1)_{12}$	-404.05

Table.1 AICc value for each seasonal ARIMA model

Since the fourth model**[1]** has the smallest AICc value, it is going to be used for time series forecasting.

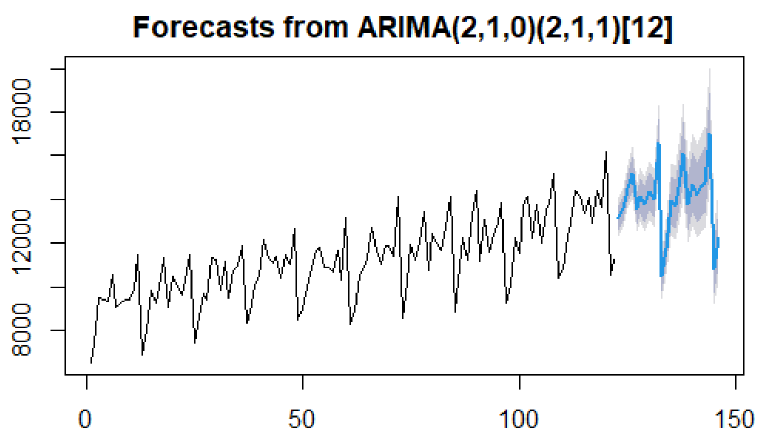


Fig.6 Forecasts from ARIMA(2,1,0)(2,1,1)[12]

d. Model Diagnostics

For the SARIMA $(2,1,0) \times (2,1,1)_{12}$ model, the diagnostic plots show that neither ACF residual nor p-values are significant.**[2]** And the QQ-plot**[3]** of this model shows that the population of the captured data is normal. Thus, the model is adequate. Moreover, the parameters of the model we choose for this time series are all significant.

e. Holt-Winters Forecast

Seasonal Holt-Winters is another approach we can take to forecast the time series. Different from the ARIMA model selection method, it extrapolates the existing data based on the pattern it has. After applying this method to the training data, we can see it is a good forecast for the data. **(see Fig. 7)**

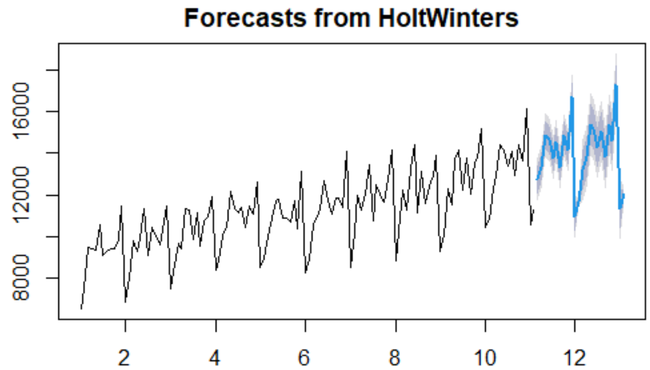


Fig.7 Forecasts from seasonal Holt-Winters

f. Evaluation of Forecast Accuracy

From SARIMA selection and Seasonal Holt-Winters forecasting plots generated, we can see they are both reasonable forecasts for the data. Thus, to pick a better method of forecasting, we have to compare the forecasting data with the true data to know which one has a larger deviation. To do this, I split the original dataset into two parts, training data(1st to 122nd) and testing data(123rd to 146th). By calculating the Root Mean Square(RMSE), Mean Absolute Error(MAE), and Mean Absolute Percentage Error(MAPE), we can see which one is a better approach to forecasting this time series. (see Table. 2)

	SARIMA	Holt-Winters
RMSE	1333.734	1389.162
MAE	1217.696	1201.845
MAPE	8.0382	7.850262

Table. 2 RMSE, MAE, MAPE comparison between SARIMA and Holt-Winters

The two models we have are showing similar performance. The RMSE of the SARIMA method is less than the one of Holt-Winters. However, the MAE and MAPE of Holt-Winters are slightly lower than the ones of the SARIMA method. Since RMSE has more importance to the largest errors, I will choose for SARIMA model for time series forecasting.

4. Conclusion

The goal of doing this report is to make a forecast of the US Alcohol sales in the next 12 months. In this report, I've used the SARIMA model selection and seasonal Holt-Winters as the two approaches to forecasting the time series. Based on the diagnostic and the accuracy check, I found the SARIMA could be a more appropriate way to forecast the Alcohol Sales data in the future. Therefore, I made a forecast for the next 12 months of Alcoholic beverages in the United States with this method. **[4] [5]** In the forecast, the sales of alcoholic beverages in the US are going to remain growing if it is without any intervention. And it's going to reach a higher peak in December(7.5%

higher compared to the last peak in Dec 2021). Furthermore, industries related to sales of alcoholic beverages can also use the information and data provided in this report as a reference for their future planning.

5. Reference

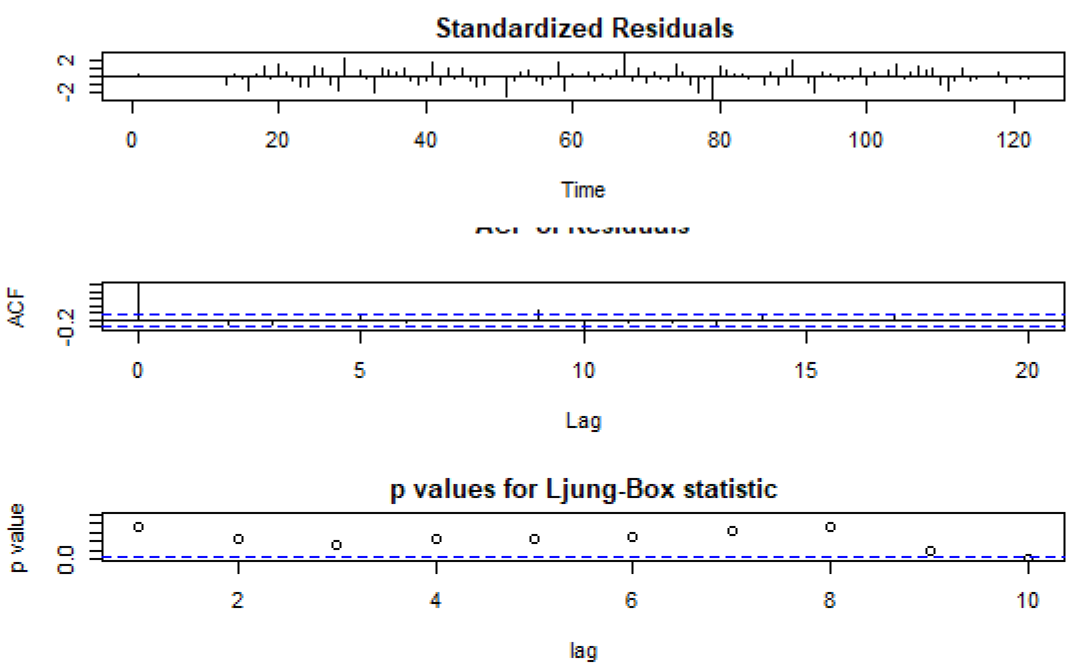
- U.S. Census Bureau, Merchant Wholesalers, Except Manufacturers' Sales Branches and Offices: Nondurable Goods: Beer, Wine, and Distilled Alcoholic Beverages Sales [S4248SM144NCEN], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/S4248SM144NCEN, May 1, 2022.

6. Appendices

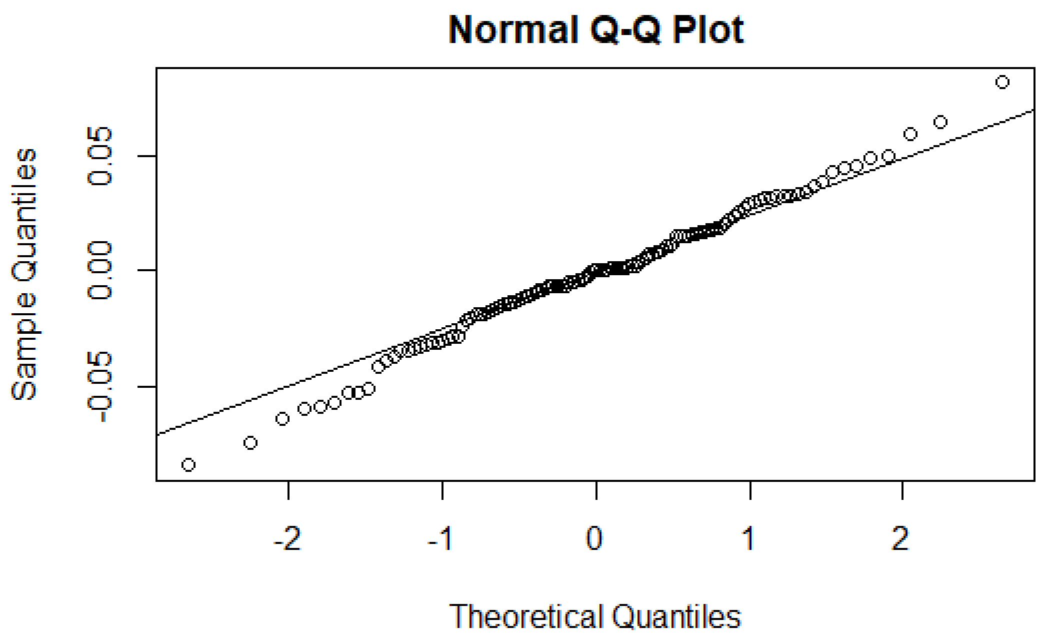
[1]

```
Series: a_train
ARIMA(2,1,0)(2,1,1)[12]
Box Cox transformation: lambda= 0

Coefficients:
          ar1      ar2    sar1     sar2     sma1
      -0.9470  -0.6332  0.3417  -0.3588  -1.0000
s.e.   0.0821   0.0854  0.1085   0.1100   0.1536

sigma^2 = 0.0009881:  log likelihood = 209.23
AIC=-406.46   AICc=-405.64   BIC=-390.32
```

[2]



[3]

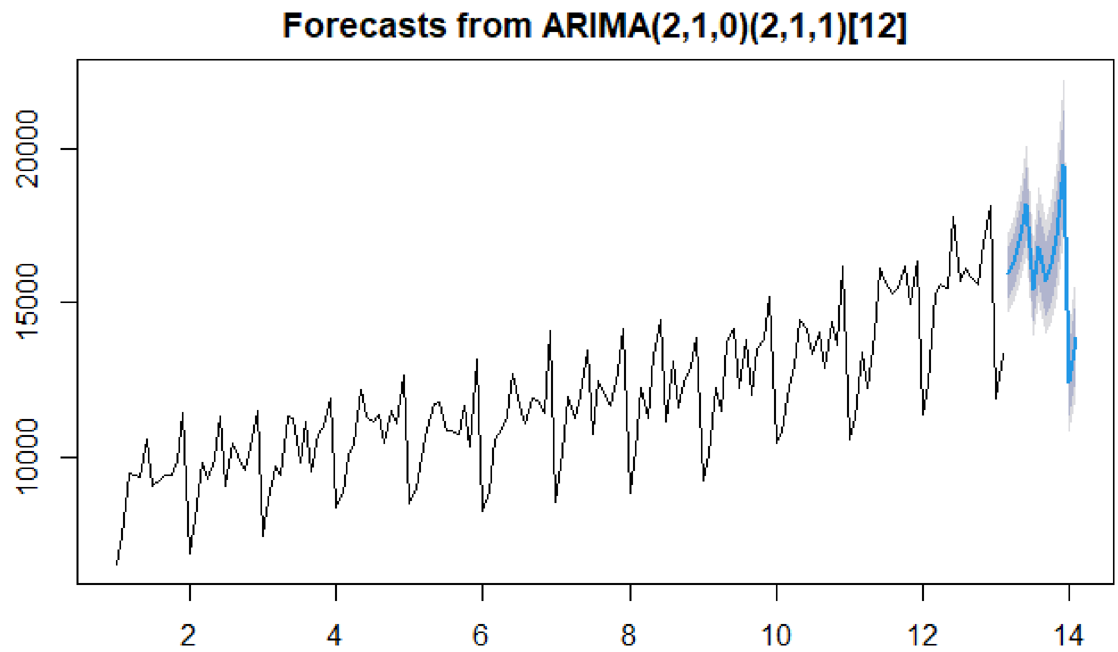


[4]

	Point Forecast <dbl>	Lo 80 <dbl>	Hi 80 <dbl>	Lo 95 <dbl>	Hi 95 <dbl>
Mar 13	15928.80	15111.68	16790.10	14696.23	17264.75
Apr 13	16315.38	15456.08	17222.46	15019.67	17722.88
May 13	17218.51	16265.92	18226.88	15783.17	18784.38
Jun 13	18175.65	17017.31	19412.83	16434.31	20101.49
Jul 13	15431.30	14409.50	16525.56	13896.27	17135.90
Aug 13	16783.82	15617.27	18037.49	15032.93	18738.62
Sep 13	15711.27	14550.32	16964.85	13970.88	17668.45
Oct 13	16213.63	14970.41	17560.10	14351.35	18317.57
Nov 13	17135.71	15769.02	18620.86	15090.22	19458.47
Dec 13	19486.04	17869.76	21248.52	17069.15	22245.16
Jan 14	12413.87	11350.49	13576.88	10824.95	14236.02
Feb 14	13836.45	12613.16	15178.39	12009.99	15940.68

(13 represents the 13th year of the time series(2022). 14 represents the 14th year(2023))

[5]



Forecasts from ARIMA(2,1,0)(2,1,1)[12]
10000
15000
20000
2
4
6
8
10
12
14