Time Series Forecasting Models using ARIMA and ETS on US Retail Ecommerce Sales

<u>Submission by:</u>

Group 3

Ahmad Abdulaziz

Laurence Embiid

Prateek Gulati

Trang Khuat

Vidya Vinobha

Introduction

E-commerce sales are sales of goods/services where the buyer places an order and negotiates the price and terms of the sale over the Internet, mobile device (M-commerce), extranet, Electronic Data Interchange (EDI) network, electronic mail, or other comparable online system. The payment may or may not be made online. A dynamic retail economy is frequently in the interest of consumers as it evolves to meet their changing needs and to adapt to emerging technologies. The fast-paced evolution of e-commerce is particularly a noteworthy topic for industry experts, analysts, policymakers, and of course the consumers. As a result, we were interested in the Census Bureau's data set that shares the quarterly US retail ecommerce sales.

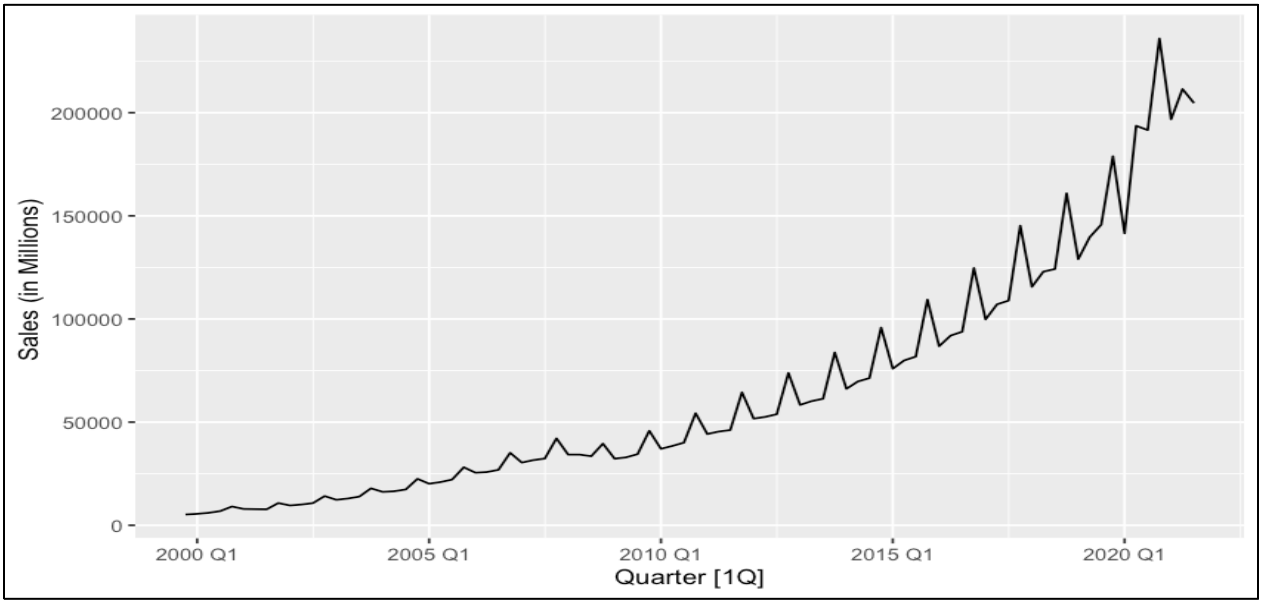


Figure 1

The dataset contains quarterly sales values from Q4 of 1994 up to Q3 of 2021. We split the dataset into a training set, which we use to fit our models, and testing set, which we use to test the forecasting power of our model. The training model has data points from 1994 Q4 to 2019 Q3, and the testing model has data points from 2019 Q4 to 2021 Q3.

As shown in Figure 1 the data tends to have a strong rising trend that can be correlated to the boom that has been happening in the ecommerce industry in the last decade. The seasonality component seem to come in more evident after 2009 and henceforth rising consistently with the level of the series. This can be confirmed by the STL decomposition as shown in Appendix 4.

Exponential Smoothing Model

The second model applied is the exponential smoothing method by using ETS model which refers to error, trend, and seasonality. The flexibility of the ETS model lies in its ability to extricate trend and seasonal components of different traits. This model usually performs better on a univariate time series forecast.

Based on the initial STL decomposition (Appendix 4), our data follows an upward trend with seasonal fluctuations. The seasonal variation seems to increase over time (multiplicative). Since the optimal lambda is close to 0 (-0.0768), we decided to go with log transformation in an effort to stabilize the variance. Prior to model development, a training set was created to build the model and a testing set to evaluate such model.

The ETS () function automatically optimizes the choice of model (by minimizing the Akaike information criterion or AICc) and necessary parameters. The AICc is useful for selecting between models in the same class. Note that on non-transformed data, the model selected was ETS (M, A, M). By taking the 'log', trend is linearized and thus an "additive" model can be used. With Log transformation, the model selected is ETS(A,Ad,A) which is the method with additive error, damped trend, and additive seasonality. The AIC for the selected model (ETS(A,Ad,A)) was -162.73 and the AICc was -159.5466. Looking at the ETS component plot below, we see still see our increasing trend. However, the slope shows a little variation. We noticed a decline in sales around 2008-2009 which might coincide with the financial crisis and after trend increases again.

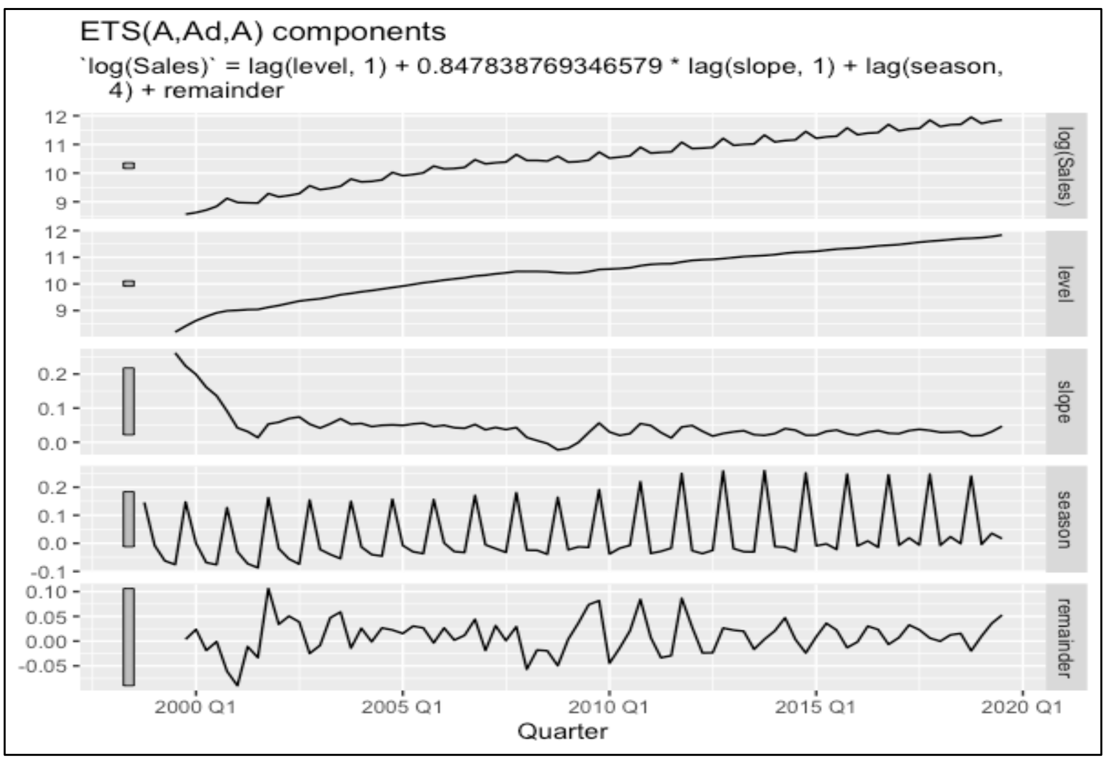


Figure 2

Using the residual plot below and the Ljung Box Test, we found our residuals to be uncorrelated (pvalue = 0.087 >0.05) meaning they are white noise.

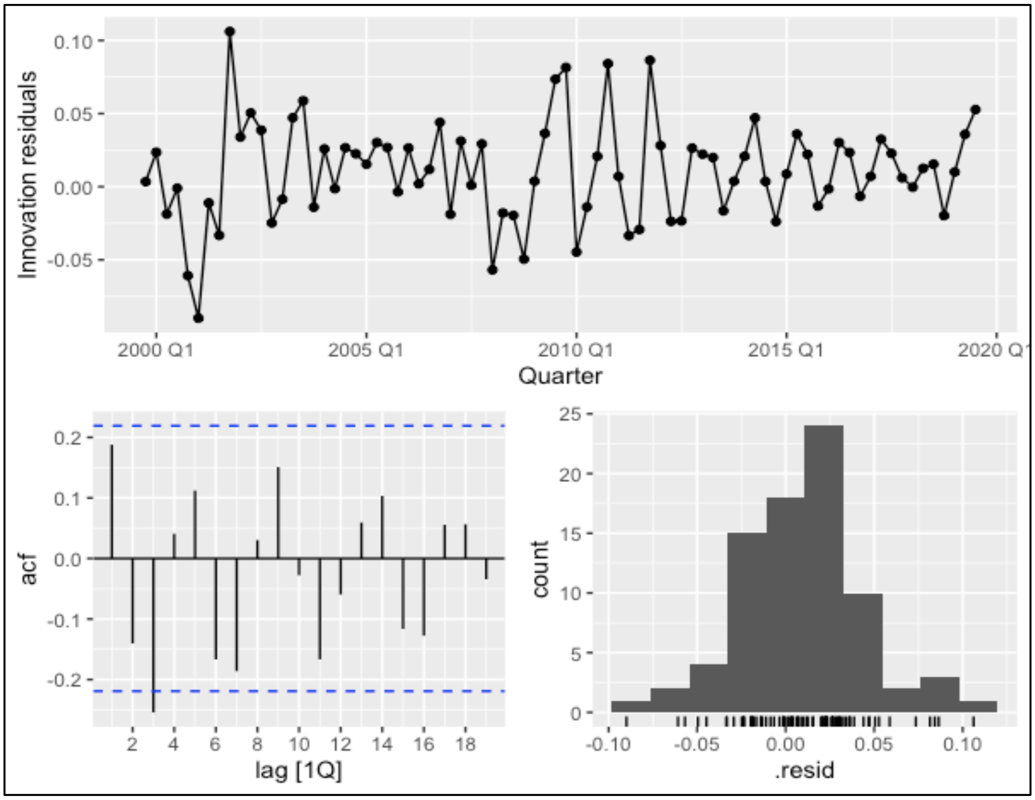


Figure 3

We were able to obtain forecasts for 2 years (h=8 quarters) from our ETS model, we used the forecast () function to do so. The table below shows the performance metrics (RMSE, MAPE, MASE, etc.) used to evaluate our model on training and test set.

Table 1

Model	Type	RMSE	MAE	MAPE	MASE	RMSSE
ETS(log(Sales))	Training	**1803**	1239	2.74	0.173	0.213
ETS(log(Sales))	Test	28989	26286	13.1	3.68	3.42

The training RMSE is much smaller than test RMSE, showing that our model performs far better on training set than test set.

Next, we will continue our analysis with one of the most performant forecasting models, ARIMA.

ARIMA Model

ARIMA model provides a complementary approach to exponential smoothing model applied previously. While ARIMA models attempt to explain the autocorrelation in the data, exponential smoothing models are developed based on description of trend and seasonality in the data.

As the initial time series plots (Figure 1) suggests, the e-commerce sales data is clearly trended shows evidence of seasonality, though not obvious, suggesting that the data is non-stationary and requires differencing to remove the trend and seasonality.

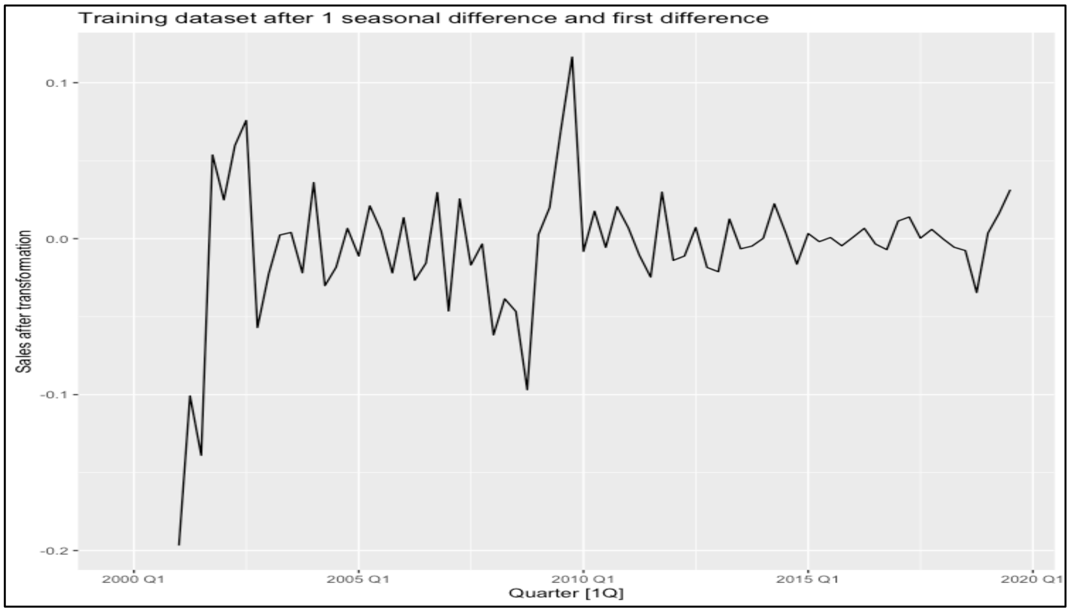


Figure 4

Continuing applying log transformation and estimating the seasonal strength, F_S, we found that $F_S = 0.981$, which is very close to 1, suggesting that strength of seasonality is strong, and that seasonal difference can be applied first. Resulted plot (Figure 4) after transformation still contains evidence of trends, thus a first difference is applied to remove the remaining non-stationarity. Performing KPSS test confirms that the after-transformed data is stationary at 0.05 confidence level (p-value = 0.1).

Evaluating the spikes and decaying pattern in the ACF and PACF plots after each differencing step (Appendix 8&9), two candidate models were selected. In addition, we built another model with default setting in R, which searches through the model space and presumably returns the best ARIMA model with the lowest AICc.

Model details can be found in Appendix 10. The three chosen models for comparison are:

- Model 1 with only seasonal differencing, non-seasonal AR(1) process: ARIMA(1,0,0)(0,1,0)[4]
- Model 2 includes first differencing, non-seasonal and seasonal MA(1) process: ARIMA(0,1,1)(0,1,1)[4]
- Model 3 or "best" model found by ARIMA algorithm in R: ARIMA(4,1,0)(0,1,0)[4]

Comparing the goodness of fit of three models on the training set confirms that Model 3 with non-seasonal AR(4) process is the best fitted model with the lowest AICc (Table 2). Interestingly, Model 2's AICc does not vary much from Model's 3.

Table 2

Model	Variance of residual	AIC	AICc	BIC
Model 1: ARIMA(1,0,0)(0,1,0)[4]	0.00186	-256	-255	-249
Model 2: ARIMA(0,1,1)(0,1,1)[4]	0.00147	-271	-271	-264
Model 3: ARIMA(4,1,0)(0,1,0)[4]	0.00128	-279	**-278**	-267

As seen in App 8&9, the autocorrelation of residual produced by Model 3 does not have any spikes and is within confidence interval. Residual follows a normal distribution and shows a horizontal trend. Hence, they suggest that the residual resembles white noise. Ljung Box test

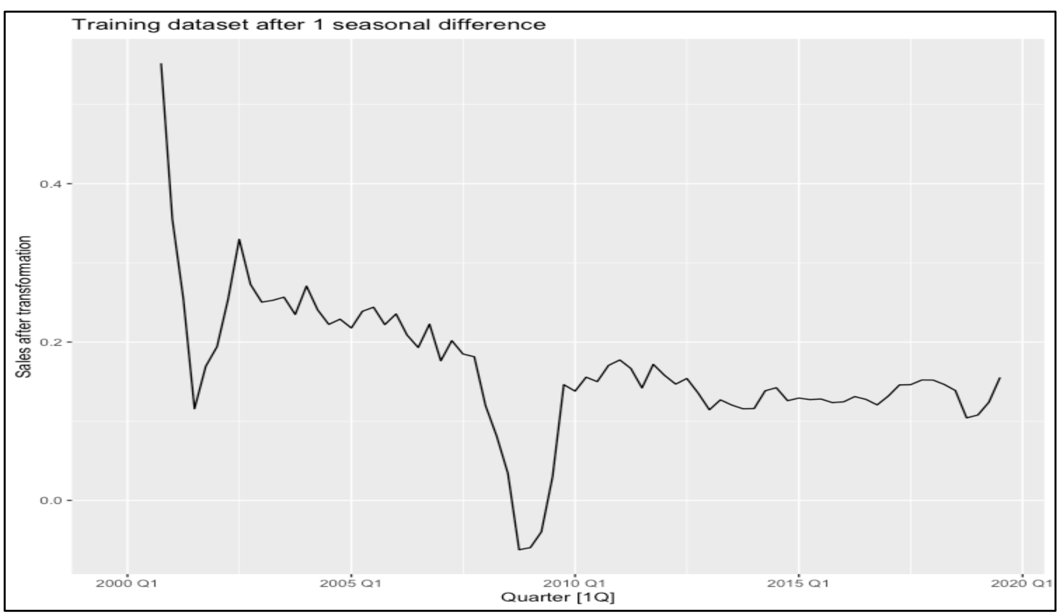


Figure 5

confirms that the residual is white noise (p-value = 0.516; Ho: residual is white noise). Model 1 (p-value = 0.154) and Model 2 (p=value = 0.890) also produce residuals that are similar to white noise. Thus, these three models are able to capture the trend and seasonality of the data and are considerably good models for forecasting purposes.

Looking at the forecasting result across 3 models (Figure 6), Model 3 produces the closest point forecast to the hold-out sample in the testing dataset. Likewise, Model 3 produces the lowest RMSE but slightly higher MAE and MAPE than Model 1 (Table 3).

Table 3

Model	ME	RMSE	MAE	MPE	MAPE
Model 1: ARIMA(1,0,0)(0,1,0)[4]	14136	19394	16176	6.83	8.15
Model 2: ARIMA(0,1,1)(0,1,1)[4]	15838	21331	19160	7.58	9.69
Model 3: ARIMA(4,1,0)(0,1,0)[4]	11537	**17647**	16200	5.36	8.36

Despite producing slightly higher MAE and MAPE, Model 3 provides a much better fit to the training data than Model 1 and achieves the lowest RMSE, Model 3 is a better model than Model 1. Therefore, Model 3 is chosen to compare with the other models developed above to find the optimal forecasting model for this data.

Discussion & Conclusion

Since ARIMA and ETS models take on different approaches to the forecasting task, the most effective method of comparison is to compare their forecasting accuracy on the hold-out sample as opposed to their fit on the training sample. The forecasting accuracy can be measured by several metrics that summarize the forecast errors in different ways. Among them, root mean squared error (RMSE) and mean absolute error (MAE) are frequently used to measure forecast errors that are on the same scale as the data, allowing for easy interpretation. Scaled or percentage errors methods such as mean absolute percentage error

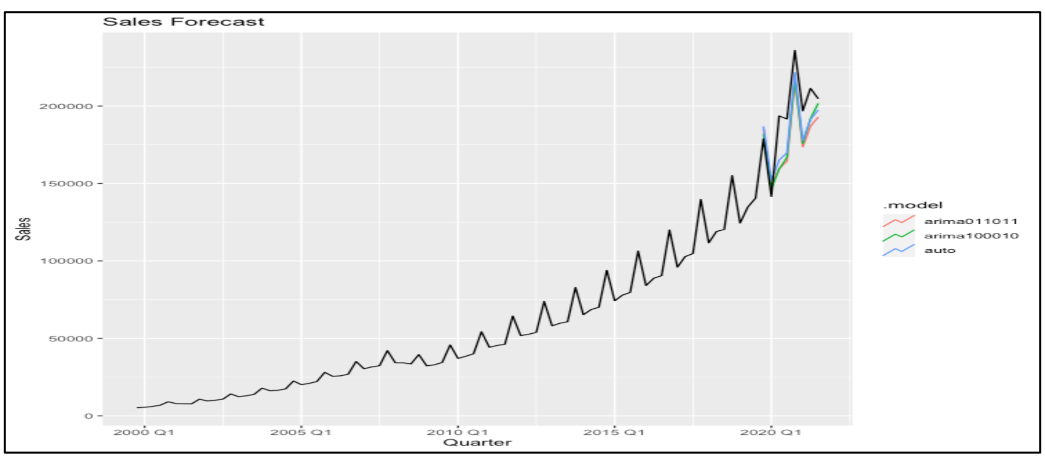


Figure 6

(MAPE) are also beneficial in comparison across different units. For our purpose, we will employ all metrics but will give more weight to scale-dependent metrics, RMSE and MAE, as our forecasted data is on the same scale. The comparison is shown in Table 4.

Table 4

	ARIMA(4,1,0)(0,1,0)[4]	ETS(A, Ad, A)
MAE	16176	28989
RMSE	**19394**	26286
MAPE	8.15	13.1

As the result suggests, ARIMA model outperforms ETS model in all metrics on the testing set as it yields the smallest errors. Interestingly, even though not listed in Table 4, the forecasting from all three tested ARIMA models yield smaller errors than the optimal ETS model. Therefore, we can safely assume that ARIMA is generally a better forecasting approach to predict the U.S. e-commerce sales than ETS approach.

Appendix

Data Source:

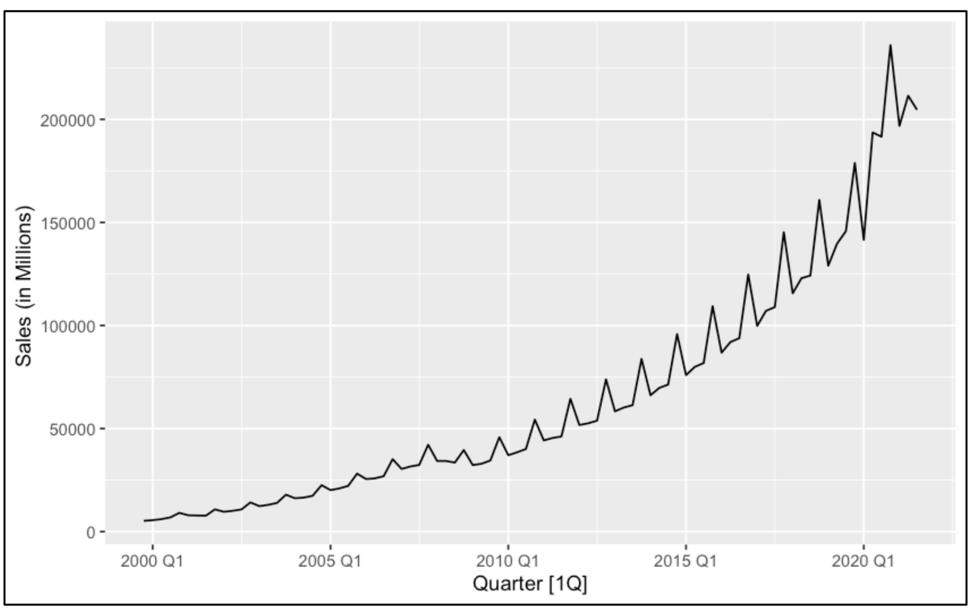
https://www.kaggle.com/census/e-commerce-retail-sales-series-data-collection

https://fred.stlouisfed.org/

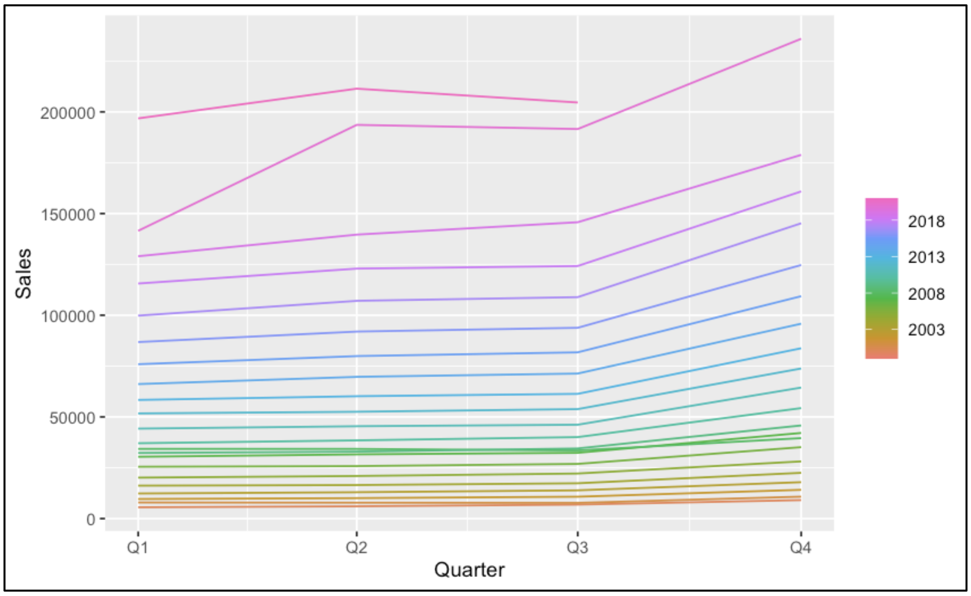
https://www.census.gov/retail/index.html

Exponential smoothing model

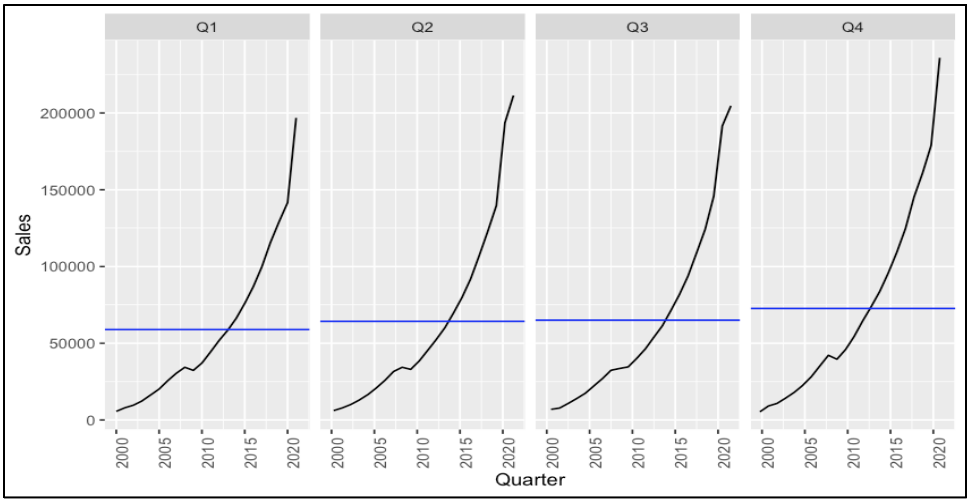
Appendix 1:



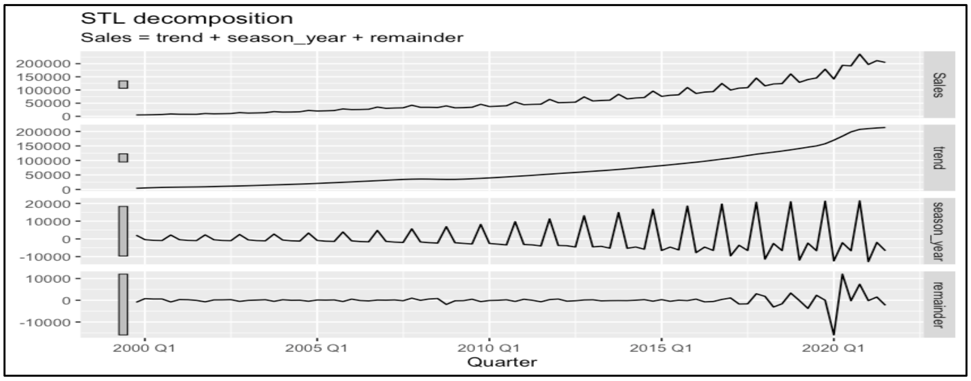
Appendix 2:



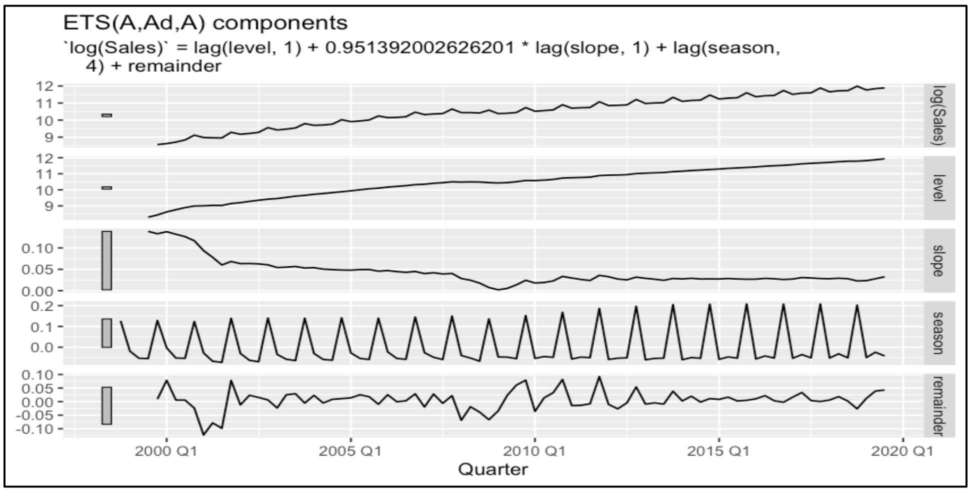
Appendix 3:



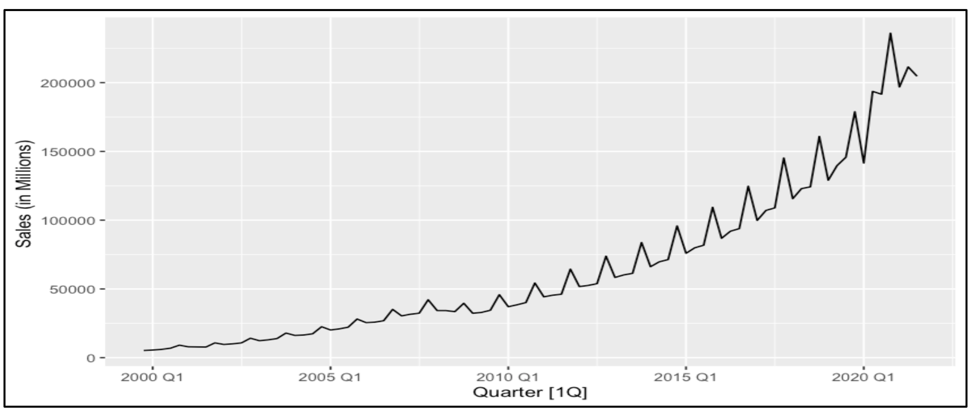
Appendix 4:



Appendix 5:



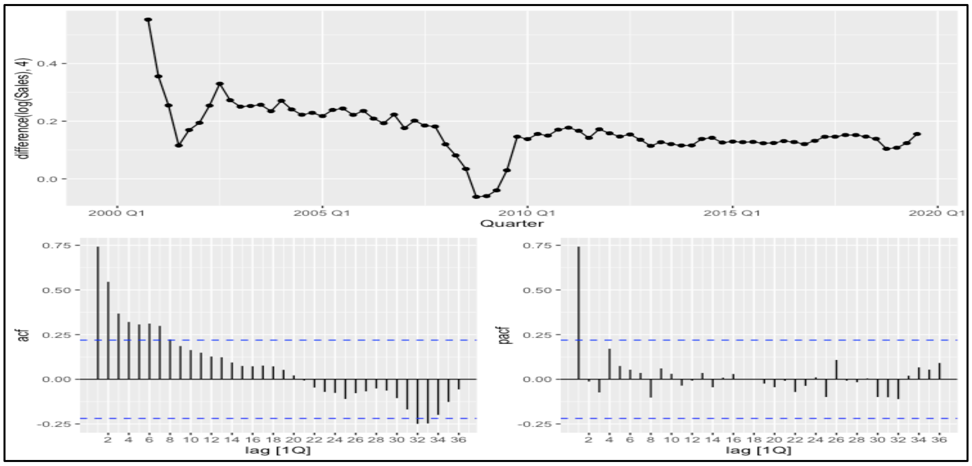
Appendix 6:



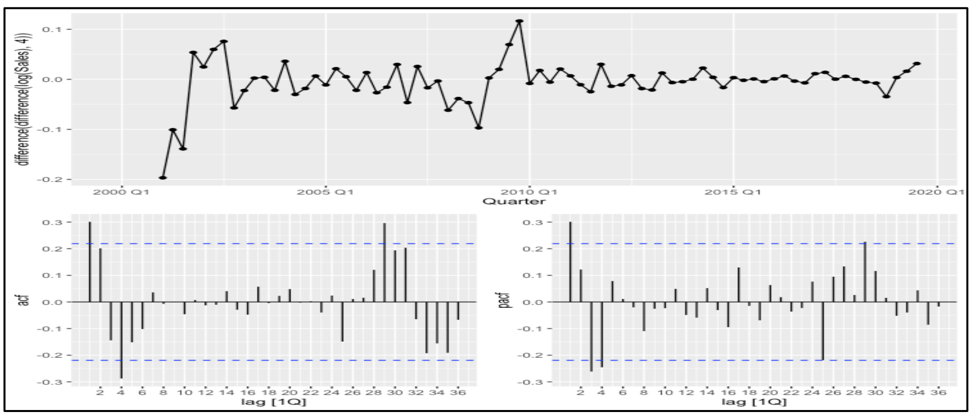
ARIMA model

Appendix 8 & 9:

ACF and PACF Analysis with 1 seasonal difference on log transformed data



ACF and PACF Analysis with both seasonal and first difference on log transformed data



Appendix 10:

Model 1: ARIMA(1,0,0)(0,1,0)[4]

```
Series: Sales
Model: ARIMA(1,0,0)(0,1,0)[4] w/ drift
Transformation: log(Sales)

Coefficients:
         ar1  constant
      0.9562    0.0103
s.e.  0.0420    0.0043

sigma^2 estimated as 0.001864:  log likelihood=130.85
AIC=-255.71   AICc=-255.37   BIC=-248.71
```

Model 2: ARIMA(0,1,1)(0,1,1)[4]

```
Series: Sales
Model: ARIMA(0,1,1)(0,1,1)[4]
Transformation: log(Sales)

Coefficients:
         ma1     sma1
      0.2229  -0.4721
s.e.  0.1000   0.1117

sigma^2 estimated as 0.001472:  log likelihood=138.71
AIC=-271.42   AICc=-271.08   BIC=-264.46
```

Model 3: ARIMA(4,1,0)(0,1,0)[4]

```
Series: Sales
Model: ARIMA(4,1,0)(0,1,0)[4]
Transformation: log(Sales)

Coefficients:
         ar1     ar2      ar3      ar4
      0.3063  0.2971  -0.1736  -0.4755
s.e.  0.1129  0.1212   0.1302   0.1263

sigma^2 estimated as 0.001284:  log likelihood=144.49
AIC=-278.98   AICc=-278.11   BIC=-267.39
```