**CO₂ Emissions**

**Forecasting**

*Team Members:*

Bercario Bodie, Christopher McManus, Lenka Raslová,

Catherine Wang, Xiaomeng Zhou

Course:

DSBA/MBAD 6211

Instructor:

Dr. Xue Guo

# **1** **Executive Summary:**

## **1.1** **Business Problem Statement**

## **1.2** **Business Goal**

## **1.3** **Data Profile**

## **1.4** **Results**

# **2** **Project Report:**

## **2.1** **Introduction**

## **2.2** **Background**

## **2.3** **Data**

## **2.4** **Method**

## **2.5** **Results**

## **2.6** **Conclusion**

# **3 References**

# **4 Appendix**

# 

# **1** **Executive Summary**

**1.1** **Business Problem Statement**

One of the most destructive fights against humanity is global warming. This is a worldwide issue that each of us has negatively contributed to at some point in time knowingly or unknowingly. However, due to the rise in concerns and the harmful effects that are currently occurring to the planet, many individuals, large corporations, and countries are making a conscious effort to minimize the effect of global warming. This initiative can be seen with their products (using more recyclable materials) in the packaging, like Amazon for example.

## **1.2** **Business Goal**

Our project will focus on CO₂ predictions and show data that illustrates the high level of emissions in the world. In addition, we will discuss the different initiatives some corporations are implementing, as this is a worldwide phenomenon.

## **1.3** **Data Profile**

Data profile is split into two sections - The First Dataset & The Second Dataset. The First Dataset is about data that was used for the proposal presentation. The Second Dataset discusses data that was used for time series analysis.

**1.3.1 The First Dataset**

The data chosen for the proposal presentation was from the World Bank. The World Bank is an international financial institution that provides financial services (such as loans, grants, advice, etc.) to governments. Governments that are using the services of the World Bank belong mainly between low- and middle-income countries. Governments are using these trades for pursuing capital projects that should help with economic growth in their countries. The mission of the World Bank is to decrease poverty and increase the incomes of the poorest 40 percent of people in every country (The World Bank, 2022).

The dataset that was used for the proposal presentation consisted of data about CO₂ emissions of individual countries. The dataset noted CO₂ in 266 countries and it was collected annually for 59 years (1960 - 2018). In total, data consisted of 59 records of CO₂ for each country. Unfortunately, 59 records were not enough for our time series analysis. Therefore, a new dataset had to be found and used.

**1.3.2 The Second Dataset**

The final dataset that was used for time series analysis is from the Climate Change Indicators Dashboard. The Climate Change Indicators Dashboard is managed by the IMF's Statistics Department (STA) with the cooperation of international organizations. The aim of the dashboard is to provide an aggregator for statistical indicators. Statistical indicators record climate changes, greenhouse gas emissions, green finance, government policies, and others. The indicators were created in collaboration with organizations such as Organization for Economic Co-operation and Development (OECD), the World Bank Group (WBG), the United Nations (UN), the European Commission, and others. (IMF Climate Changed Dashboard, 2022)

The data that was used for time series analysis, comes from the Climate Change Data category and it is about Atmospheric CO₂ Concentrations on a global level. The Atmospheric Carbon Dioxide concentrations were collected monthly from March 1958 to August 2021. In total, data consists of 762 observations. Dr. Pieter Tans from NOAA Global Monitoring Laboratory is the scientist who sourced the data. Atmospheric CO₂ Concentrations are measured at the Mauna Loa Observatory at an altitude of 3400 m. Two measurement techniques were used for data collection. An infrared absorption method was used for data until April 2019. An analyzer Cavity Ring-Down Spectroscopy (CRDS) was installed at the Mauna Loa Observatory and used from April 2019 for CO₂ collection (Pieter Tans and Kirk Thoning, 2008).

The original dataset has 1512 rows and 8 columns (*ObjectId, Country, ISO2, ISO3, Indicator, Unit, Date, Value*). The *Value* attribute consists of two measurements - 762 values of CO₂ in individual months and 750 values of Year on Year Percentage Change.

## **1.4** **Results**

The MAPE value for the exponential trend and ARIMA hybrid model was 0.0009970782. This was the lowest MAPE value of all models tested. This meant that the hybrid model had the highest accuracy, and was best at forecasting CO₂ emissions.

# 

# **2** **Project Report**

## **2.1** **Introduction**

What exactly is global warming? According to John Houghton, this is the effect of human activities through the burning of fossil fuels and deforestation that results in the emissions of carbon dioxide also known as (CO₂) which creates a “blanket” over the earth’s surface. This develops warmer temperatures, forcing unnatural climate changes. According to Kunda and Phiri, a large percentage of emission occurs from modes of transportation and factories.

This particular topic is extremely important because it affects the human race as we are rapidly depleting the earth's ozone layer. This layer is used to protect the earth’s stratosphere from harmful rays. Consequently, with the increase of CO₂ emissions, this layer is being eroded which causes the rise in sea levels, unusual weather activity, and other progressively negative reactions to the earth and its inhabitants.

Based on our research, the business issue is identifying and predicting CO₂ emissions. Our goal is to share information about the past and current levels of emission, and how these can be used in predicting future CO₂ emission levels by using the forecasting method. In addition, we will be suggesting and identifying ways companies are taking action to reduce these toxic levels.

## **2.2** **Background**

## This project is unique within itself. During the course of research we did not find an appropriate global Time Series CO₂ emission that can be compared to our findings, which will be shown. There are forecasting and time series information and scholarly articles that are available for specific countries; however, it would not be feasible to present such correlation because of the inaccuracy and the illegitimacy of a single country's statistics and compare it with the world. In addition, with the rise of concerns with CO₂ emissions there is a strong inclination that more information on data mining and time series will eventually become available, especially due to the fact that there was a disruption due to the pandemic.

## 

## **2.3** **Data**

The source of the dataset is the Climate Change Indicators Dashboard and the name of the dataset is Atmospheric CO₂ Concentrations on a global level. The original dataset has 1512 rows and 8 attributes.

*Table 1: Dataset Variables Description*

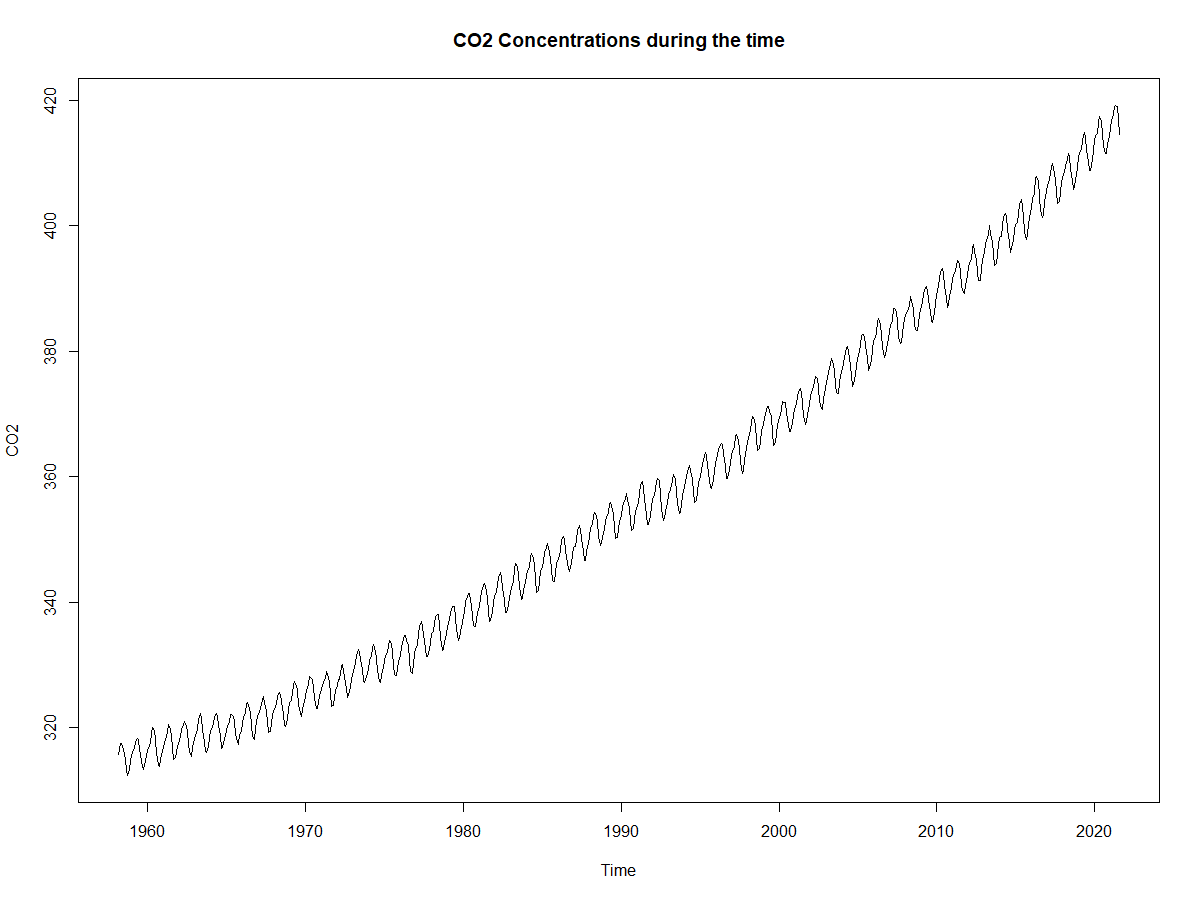
|  |  |
| --- | --- |
| **Attribute** | **Description** |
| Object Id | The identification # of records, a row of numbers from 1 to 1,512 |
| Country | The type of country for which was measured CO₂, it has just one category–World |
| ISO2 | Two-letter country codes, it does not have any values in it– it is an empty column |
| ISO3 | Three-letter country codes, it has just one value WLD (World) |
| Indicator | Description of Value, it has two categories, whether the value is Monthly Atmospheric Carbon Dioxide Concentrations or Year on Year Percentage Change |
| Unit | Unit description of metrics of Value, whether a value is Parts Per Million or Percent |
| Date | Date of Value measurement (format: YearMonth) |
| Value | Value of CO₂ Parts Per Million or CO₂ Year on Year Percentage Change |

For the project purposes, variables that did not bring any insights were skipped. Object Id was deleted because it is just the identification number of objects. The country variable was skipped because it has only one value–World. IS02 consists of 1,512 missing values, therefore, it was not used for time series analysis (the column is empty). ISO3 variable has the same value for all rows, so it was deleted. Based on the Indicator variable, the variable Value is split into two columns (one column for Monthly Atmospheric Carbon Dioxide Concentrations and another column for Percentage Change), the first column consists of 762 records, and the second consists of 750 records. The new column of Percentage Change has the first 12 values empty. This is because there were no recorded values for Atmospheric Carbon Dioxide Concentrations for the previous 12 months (it is not possible to calculate the Year on Year Percentage Change). Variable Date was kept without change.

The final dataset consists of three variables–Date, Monthly Atmospheric Carbon Dioxide Concentrations, and Year on Year Percentage Change (Table 2).

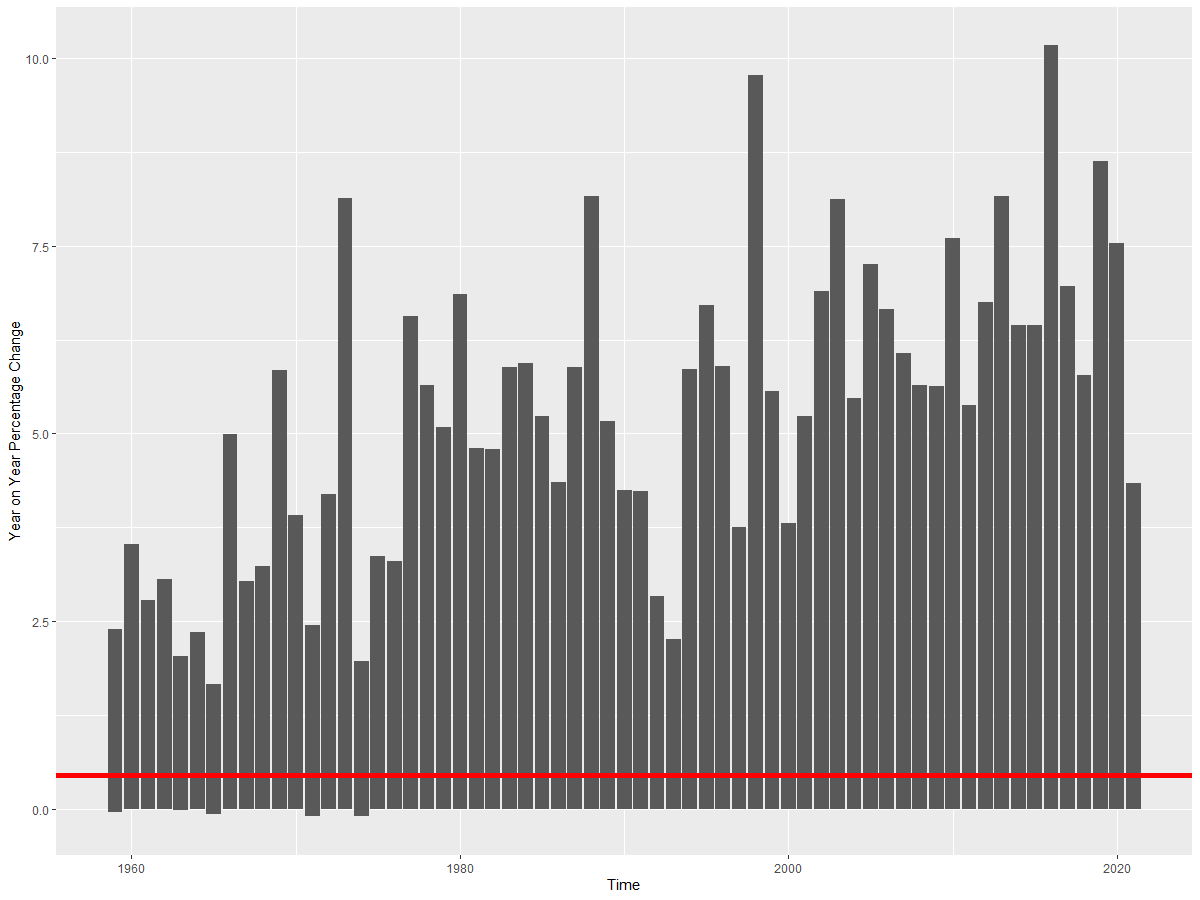
*Table 2: Summary Statistics of Kept Variables*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variable** | **Observations** | **Mean** | **Std. Dev.** | **Min** | **Max** |
| Date | 762 | character | character | 3/1/1958 | 8/1/2021 |
| Monthly Atmospheric Carbon Dioxide Concentrations | 762 | 356.5 | 29.6613 | 312.4 | 419.1 |
| Year on Year Percentage Change | 750 | 0.4433 | 0.2004 | -0.1036 | 1.0537 |



*Figure 1: CO*₂ *Concentrations during the time*

Based on Figure 1, CO₂ Concentrations have an increasing trend and seasonalities in the observed time period.



*Figure 2:* Year on Year Percentage Change

Figure 2 shows Year on Year % Changes of CO₂ during the time period. The red line represents the mean of % changes for the whole time period.

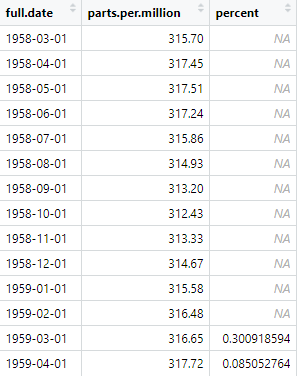
## 

## **2.4** **Method**

**Data Preparation**

To prepare the data for analysis, three rounds of pre-processing were conducted. First, RStudio was used to split the *Value* column into *parts.per.million* and *percent* (first photo under Appendix A)*.* Next, Excel was used to split the text in the *full.date* column into a date format of MM/DD/YYYY, using the assumption that the data was gathered on the first day of the month (second photo under Appendix A). Finally, the extraneous data columns (*id, Country, Indicator*) were removed using R Studio. The resulting data set used for modeling can be seen below in Table 3.

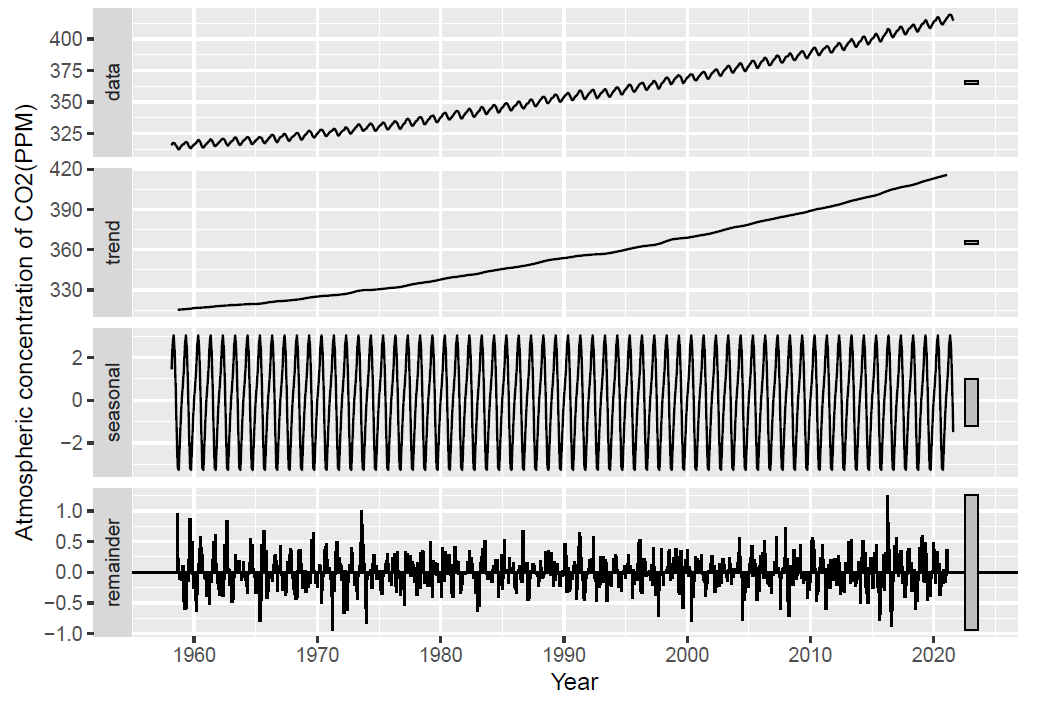
*Table 3: Dataset after all pre-processing steps*



The only missing data in the monthly CO₂ concentrations dataset occurred in the *percent* column for March 1958 through February 1959. Since the values in the percent column is the percentage difference between the previous year’s CO₂ concentration and the current year’s CO₂ concentration, not having data from March 1957 through February 1958 yielded NA data values for March 1958 through February 1959 (ex: (March 1959 - March 1958)/(March 1958) \* 100). Since the percent column essentially duplicates the information from the CO₂ concentration column, we decided to conduct our time series modeling only on the CO₂ concentration data.

**Preparation for Modeling**

Before data modeling, Exploratory Data Analysis (EDA) was performed to better understand the behavior of our dataset. For a time series dataset, the best EDA to perform is a Decomposition, which breaks down the dataset into different components–data, trend, seasonal, and remainder (Figure 3). This allows more efficient analysis as the Decomposition provides a general direction of where to start.



*Figure 3: EDA–Decomposition of dataset*

After Decomposition, the following questions can be asked to help in identifying which model would suffice in fitting the data:

1. Does the time series data have a trend? If yes, does the trend change linearly or exponentially?
2. Does the time series data have seasonality? If yes, does the seasonal component change in magnitude over time?

It can be seen through conducting the Decomposition that our data is linear but contains a constant yearly seasonal trend.

Another important step before modeling is splitting up the original data set into two subsets: a “training” set (used to create the model) and a “validation” set (used to test the accuracy of the model). The training set includes data from 1958-03 to 2017-12, while the validation set consists of the remaining, more recent data, from 2018-01 to 2021-08. Using this subset method allows us to more effectively validate the performance of our models.

**Modeling the Data**

Knowing this, we plan to use four different class of methods to forecast:

1. Deterministic Trend Model (Regression-based Model)
2. ETS Model (Error, Trend, Seasonality)
3. ARIMA Model
4. Hybrid Model (Combining ETS and ARIMA)

In effect, we are transforming “non-stationary” data into “stationary” data. Non-stationary data is difficult to analyze because it doesn’t have a consistent probability distribution, making statistical inference very difficult. Transforming non-stationary data into stationary data allows for a time series whose properties do not depend on time–the data has a constant mean, constant variance (homoscedasticity), and constant covariance–providing a simpler platform for data modeling.

Once a model is created, we use the mean absolute percentage error (MAPE) as the main measure to compare performances between models since our data does not contain any extremes nor any zeros. It measures forecasting accuracy as a percentage, and can be calculated as the average absolute percent error for each time period minus actual values divided by actual values (Stephanie, 2017). The closer a model’s MAPE value is to zero, the more accurate the model is to predicting the data set.

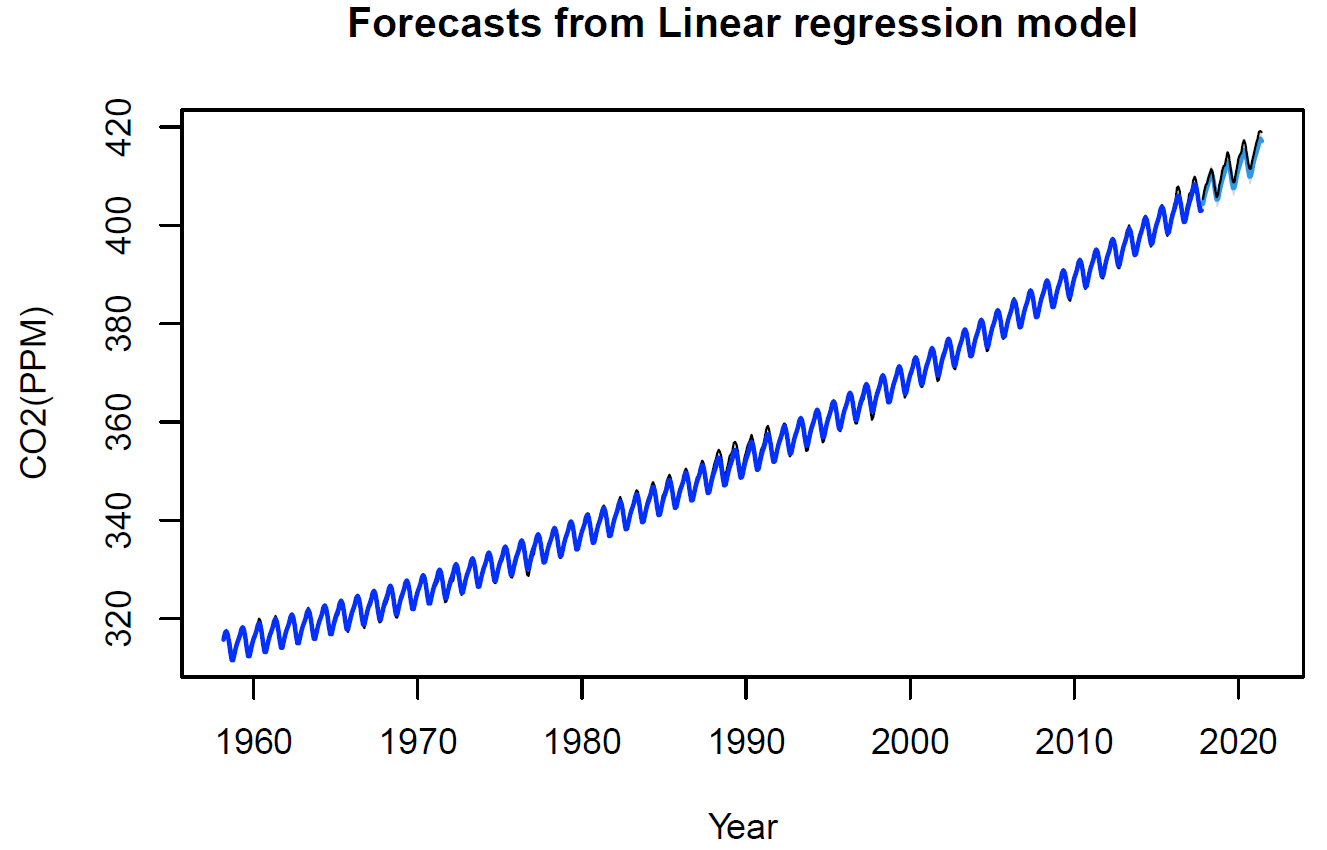
When developing our models, it is helpful to look at the graphs of the models’ residuals. A high-performing model’s residual graph should closely resemble white noise–containing a mean of zero, constant variance, and residuals that are independent from each other. If the model’s residuals show a trend, then it is obvious that something was not accurately captured from the data.

In addition to residual graphs, the Autocorrelation Function (ACF) can also be used to measure the performance of a model. The ACF shows the lagged variables that have a statistically significant effect on the present value of the dependent variable . Significant and positive values in the ACF suggest that we should include lagged dependent variables into the model, while significant and negative values in the ACF suggest we need to include lags of the error term in the model. If there are many significant lags, there is a high likelihood that an autoregressive component has been left out of the model and needs to be included to better represent the dataset.

In the following paragraphs we will delineate the four methods that we performed to find the most optimal model for the CO₂ Emissions data.

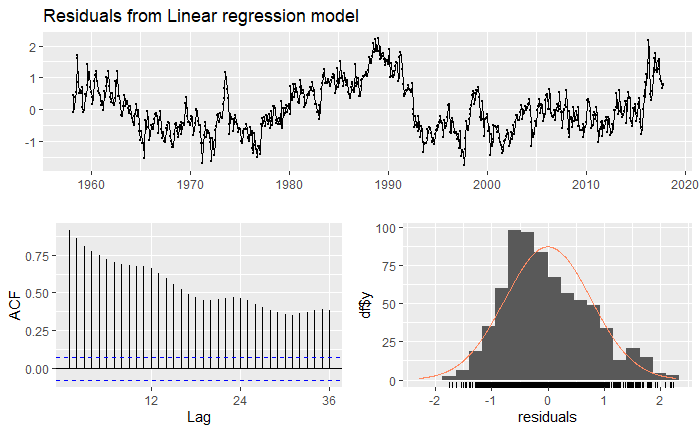
**Model 1: Deterministic Trend, Regression-Based**

The first method we chose is the regression-based model, which uses time-step features. To best capture our data, we chose to include a linear, a quadratic, and a seasonality term in our tslm() function in RStudio (CO₂.lm.trend.season <- tslm(train.ts ~ trend + I(trend^2) + season)). The summary and accuracy results of this model can be found in Appendix B, and the visual representation of the forecast results can be seen in Figure 4.



*Figure 4: Graph of forecast results from regression-based modeling method*

The existing trend that can be seen in the residuals graph coupled with the large ACF values in Figure 5 are evidence that Model 1 does not fully capture the trends in the dataset.

**

*Figure 5: Graph of the residuals and ACF from regression-based modeling method*

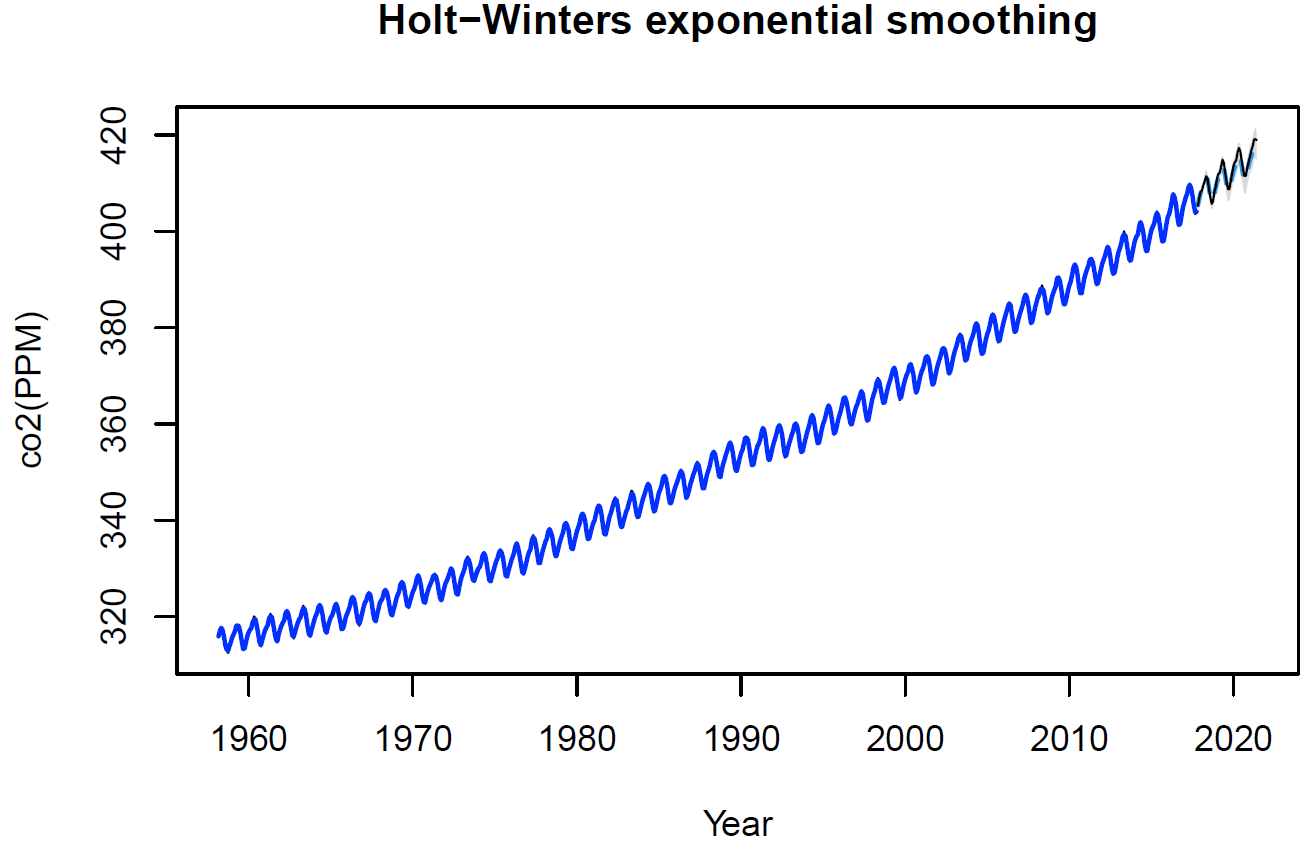
**Model 2: Error, Trend, Seasonality (ETS)**

The ETS models refer to the exponential smoothing methods. There are several ETS models:

1. *Simple Exponential Smoothing*: finds the level of the time series which is suitable for time series. It is best used when the data does not include trend nor seasonality.
2. *Holt-Winters Exponential Smoothing*: finds the level of the time series and adds the term for linear trend. This is used when the time series has a linear trend but no seasonality.
3. *Exponential Trend*: finds the level of the time series and adds the multiplicative term for exponential trend. This method is best used when the time series has an exponential trend but no seasonality.
4. *Holt-Winters Seasonal Method*: finds the level of the time series and adds the additive and/or multiplicative terms for trend and seasonality.

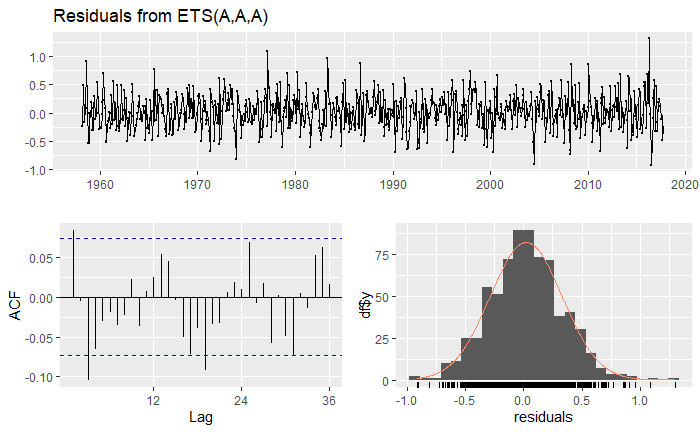
Based on the decomposition analysis completed for the CO₂ emissions data, it can be determined that the best ETS model to use is the Holt-Winters Seasonal Model. Holt-Winters Seasonal Model contains three components: one for level (error), one for trend, and one for seasonality.

When building a Holt-Winters Seasonal Model, one must determine if each component will be additive or multiplicative. An additive term is used when the trend and/or seasonal variation are relatively constant over time, while a multiplicative term is used when the trend and/or seasonal variation increases or decreases at a different magnitude over time. Since the CO₂ emissions data has a relatively constant trend and constant seasonality, and the error term is close to white noise, we will use the additive term for each component. This yields the final equation to be: hwin <- ets(train, model = “AAA”) in RStudio, with the accuracy results in Appendix C and the graph of the forecast in Figure 6.



*Figure 6: Graph of forecast results for ETS seasonal modeling method*

In Figure 7 it can be seen that the residuals of Model 2 more closely resemble white noise compared with the residuals in Model 1. The ACF values of Model 2 are also significantly smaller, with only a couple of lags missing from the model that need to be included.

****

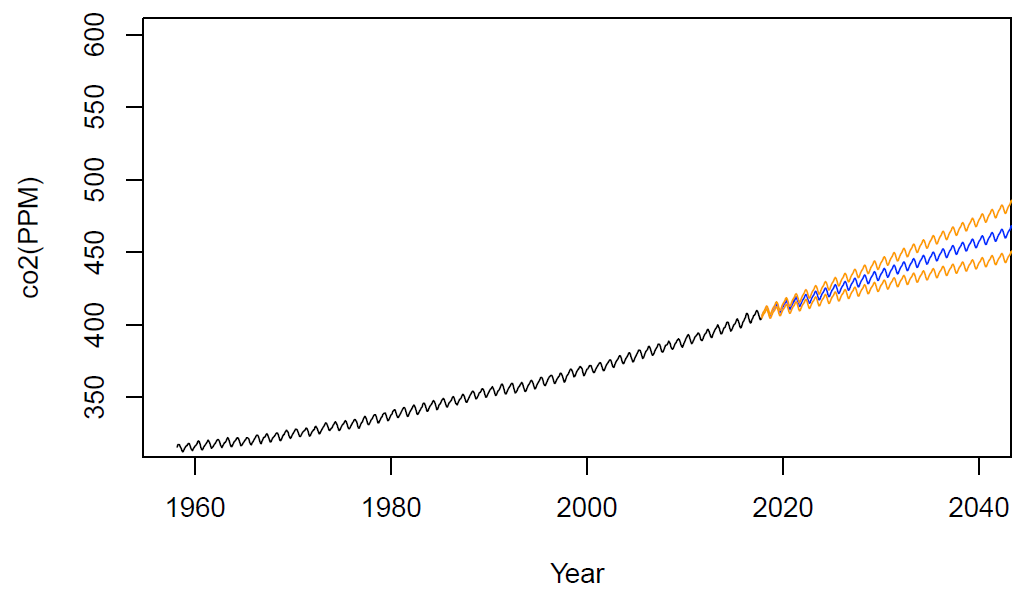
*Figure 7: Graph of residuals and ACF for ETS seasonal modeling method*

**Model 3: ARIMA**

ARIMA models, or autoregressive integrated moving average models, are regression models that use lagged values of the dependent variable and/or random disturbance term as explanatory variables. ARIMA is composed of three parts, and can be interpreted as ARIMA(p,d,q) for nonseasonal models and ARIMA(p,d,q)(P,D,Q) for seasonal models:

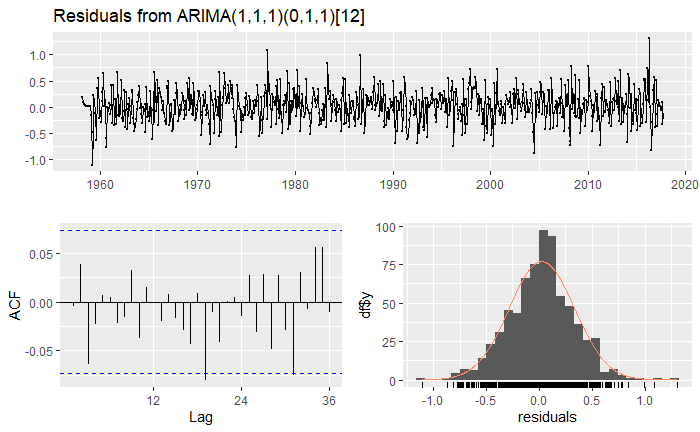
1. AR (“autoregressive”, “p” indicating the number of autoregressive terms) terms: lags of the series (considers the relationship between and )
2. I (“integrated”, “d” indicating the number of nonseasonal differences) terms: uses a method called “differencing” to ensure an outcome/predictor variable is stationary
3. MA (“moving average”, “q” indicating the number of moving-average terms) terms: lags of forecast errors (considers the prediction errors from prior periods–, etc.)

ARIMA is made very simple to calculate in RStudio, where we are able to use the function auto.arima() from the *forecast* package to search for the best fit model (ARIMAfit <- auto.arima(train, approximation = FALSE, trace = TRUE), which yielded the result of ARIMA(1,1,1)(0,1,1). The summary and accuracy of Model 3 can be found in Appendix D and the graph of the forecast is shown in Figure 8.



*Figure 8: Graph of forecast results for ARIMA modeling method*

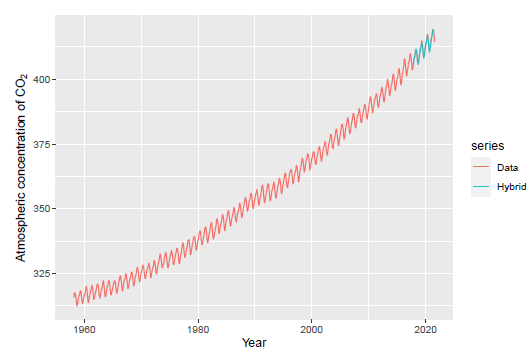
By using the function *checkresiduals(ARIMAfit)* we are able to generate an ACF graph of the residuals for ARIMA(1,1,1)(0,1,1) that can be viewed in Figure 9. We can use the ACF graph to further prove that ARIMA has yielded an effective model since all of ARIMA’s ACF lines are within the upper and lower control limits (blue dotted lines); this confirms that the residuals are close to, if not are exactly, white noise and that there is no autocorrelation with any lags.



*Figure 9: Graph of residuals and ACF for ARIMA model*

**Model 4: Hybrid Method Combining ETS and ARIMA**

For the fourth model, we decide to use the Hybrid Method by using the formula *mixture()* from the *opera* package in R. The Hybrid method combines the forecasts from Model 2 (ETS) and Model 3 (ARIMA), allowing us to capitalize on the strengths of both Model 2 and Model 3. The Hybrid method does this by dynamically computing and adjusting the weights when combining both forecasts by looking at how well each method has performed up to that data point. Results are in Appendix E and graphs of forecast results are in Figure 10.



*Figure 10: Graph of forecast results for hybrid modeling method*

## **2.5** **Results**

Of the four models, the hybrid model had the best ability to forecast CO₂ emissions. This was verified by comparing the accuracy measures of the other models to the accuracy measure of the hybrid model. Although the other models also had relatively low test MAPE values, the hybrid model had the lowest test MAPE value. The deterministic trend model had a MAPE value of 0.3346419077, the Holt-Winters exponential smoothing model had a MAPE value of 0.137174444, and the ARIMA model had a MAPE value of 0.00143381, but the hybrid model had the lowest MAPE value of all the models at 0.0009970782.

After verifying each model’s prediction capabilities, it was also important to compare the plots of each model to see which one captured the trends of the dataset greatest. The forecasting results of the deterministic model appeared to be minimal and constant but missed some of the trends. Likewise, the Holt-Winters also did not fully capture the trends of the dataset because around 2020, the predicted values became short of the observed values. Lastly, the predicted values of the ARIMA model support the notion that the hybrid model performed best. Initially, the predicted values appear to be close to the actual data, the wide confidence intervals suggest that the ARIMA model has a greater margin of error than the hybrid model and may not be the best at capturing the trends of the data.

One carbon emissions study conducted by Marcelo Bohrer used two ARIMA models to predict global CO₂ emissions between different power sectors such as coal and natural gas (Bohrer, n.d.). The first model yielded the result of ARIMA(2,1,1)(0,1,1). The second model yielded the result of ARIMA(1,1,1)(2,1,1). Using cross-validation to test accuracy, Marcelo’s second model was found to perform better than the first model. However, not only did our ARIMA model outperform both of Marcelo’s models, but our hybrid model outperformed his models as well. This was supported by our models’ high accuracies, and closer predicted values to the observed ones.

## **2.6** **Conclusion**

Global warming threatens the existence of humanity. People as owners of the planet have to do maximum of protecting the environment. One global warming indicator is the CO₂ Concentration level in the world. The aim of this paper is to find a model that predicts future CO₂ emissions. This model should help open the eyes of people and motivate them to protect our planet. Varies time series models were used for finding the best model for predicting CO₂ Concentration. The dataset used for the project comes from the Climate Change Indicators Dashboard. The data that was used for time series analysis were collected monthly on a global level from March 1958 to August 2021 (in total 762 records). The four different time series models were used for forecasting CO₂ emissions. Among the Deterministic Trend Model, ETS Model, ARIMA Model, and Hybrid Model, the best performance had the Hybrid Model. The Hybrid Model had the lowest test MAPE value which made it the best forecast predictor for future CO₂ emissions.

With the results of our model supporting our claim that the world’s CO2 emissions are predicted to increase in the future, it is now more important than ever for companies to reduce their carbon footprint. Companies can do this by becoming ISO 14001 certified as well as using suppliers that also hold this certification, resulting in a sustainable supply chain (“11 ways”, 2021). ISO 14001 certification means that an organization is using environmental management techniques to improve reduce waste and improve resource efficiency (“11 ways”, 2021). Additionally, companies can promote virtual meetings to reduce unnecessary travel-related emissions such as cars on the road, or planes in the sky (“11 ways”, 2021). Lastly, increasing the use of recycled paper products can help reduce deforestation by limiting the need to cut down additional trees (“11 ways”, 2021).

As consumers become more environmentally conscious, these companies are taking initiatives to better align themselves with their customer base. In 2020, Apple made breaking news by removing the charging brick and earphones from their iPhone packaging (“Apple”, 2020). This was an effort to reduce shipping-related emissions by increasing the number of iPhones that could fit on a shipping pallet (“Apple”, 2020). More recently, FedEx began taking deliveries of its first electric delivery trucks in an effort to promote “green logistics” (Holland, 2021). With FedEx signaling their transition away from gas- and diesel-powered delivery trucks, competitors like UPS, and USPS will hopefully follow suit. Finally, Coca-Cola has pledged to reduce its carbon footprint by utilizing sustainable agricultural techniques for key ingredients in its products (Morgan, 2021). By using sustainable agricultural techniques Coca-Cola can improve its supply chain and even become ISO 14001 certified as previously mentioned. These efforts show that companies are interested in making a change in their practices for the greater good.

**3 References**

*11 ways businesses can reduce their carbon emissions*. Shred Station®. (2021, December 7). Retrieved May 8, 2022, from https://www.shredstation.co.uk/blog/11-ways-businesses-can-reduce-carbon-emissions/

*About*. (n.d.). IMF Climate Changed Dashboard. Retrieved May 6, 2022, from <https://climatedata.imf.org/pages/about>

*Bohrer, M*. (n.d.). *Carbon Emissions - Basic Time Series Analysis*. Retrieved May 5, 2022, from <https://rstudio-pubs-static.s3.amazonaws.com/284152_c7d74c4d9f694725a40c5fe192509a1f.html>

*CO₂ emissions (kt)* | Data. (n.d.). The World Bank. Retrieved May 6, 2022, from [https://data.worldbank.org/indicator/EN.ATM.CO₂E.KT?end=2018&start=2000](https://data.worldbank.org/indicator/EN.ATM.CO2E.KT?end=2018&start=2000)

*Holland, F*. (2021, December 17). Fedex gets first of 500 electric trucks from GM's EV unit in a major advance for Green Logistics. CNBC. Retrieved May 8, 2022, from https://www.cnbc.com/2021/12/17/fedex-gets-first-of-500-electric-trucks-from-gms-ev-unit-in-move-to-green-logistics.html

*Houghton, J*. (2005). Global warming. Reports on progress in physics, 68(6), 1343.

*Kunda, D., & Phiri, H*. (2017). An approach for predicting CO₂ emissions using data mining techniques. International Journal of Computer Applications, 172(13), 7-10.

*Morgan, B*. (2021, December 10). 101 companies committed to reducing their carbon footprint. Forbes. Retrieved May 8, 2022, from https://www.forbes.com/sites/blakemorgan/2019/08/26/101-companies-committed-to-reducing-their-carbon-footprint/?sh=3999ef1b260b

*Pieter Tans and Kirk Thoning*, NOAA Global Monitoring Laboratory, Boulder, Colorado, September 2008. Updated December 2016; March 2018, September 2020, from [https://gml.noaa.gov/ccgg/about/CO₂\_measurements.html](https://gml.noaa.gov/ccgg/about/co2_measurements.html)

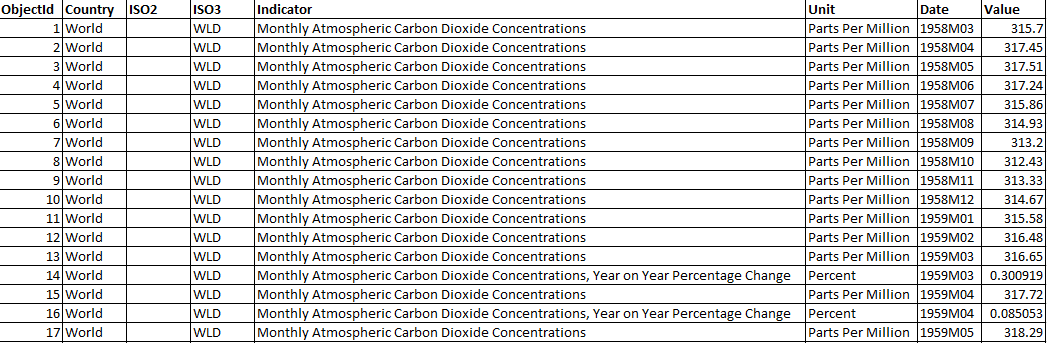
*Stephanie*. (2017, September 8). Mean absolute percentage error (MAPE). Statistics How To. <https://www.statisticshowto.com/mean-absolute-percentage-error-mape/>

*Who we are*. World Bank. (n.d.). Retrieved May 6, 2022, from <https://www.worldbank.org/en/who-we-are>

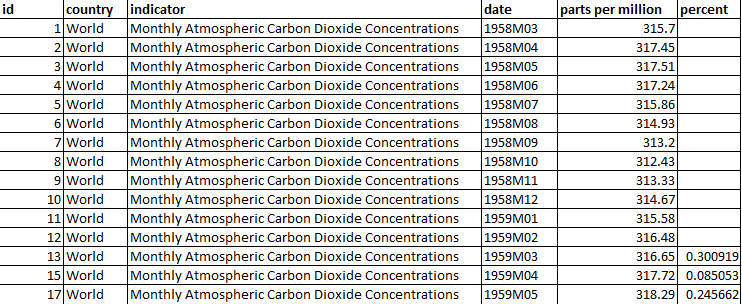
# **4 Appendix**

**Appendix A**:

Screenshot of initial data before any processing:

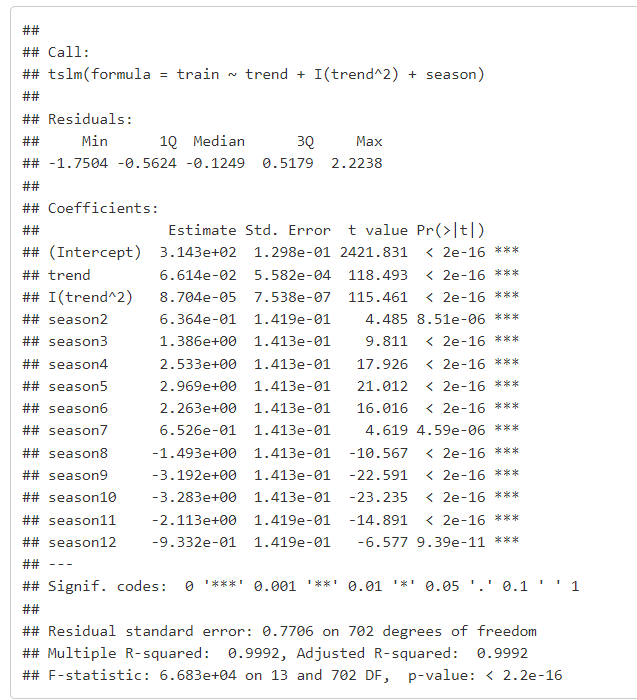


Screenshot of data after processing with Python:

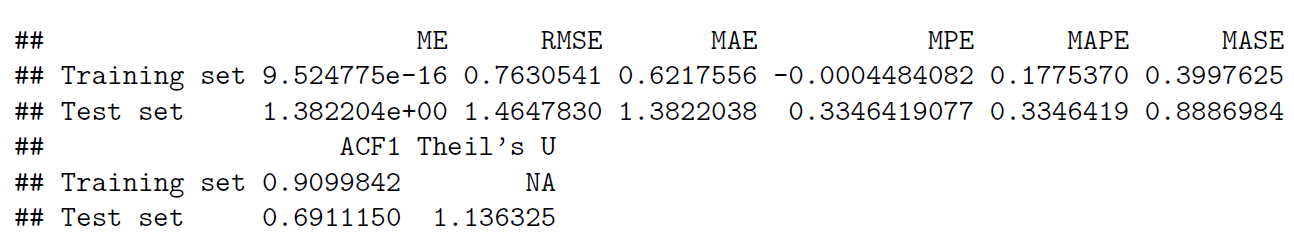


**Appendix B**: Results of Model 1: Deterministic Trend/Regression-Based Method

Model 1 Summary:

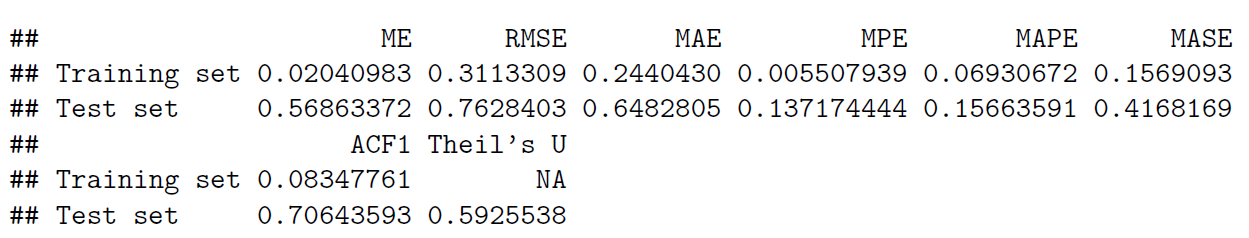


Model 1 Accuracy:



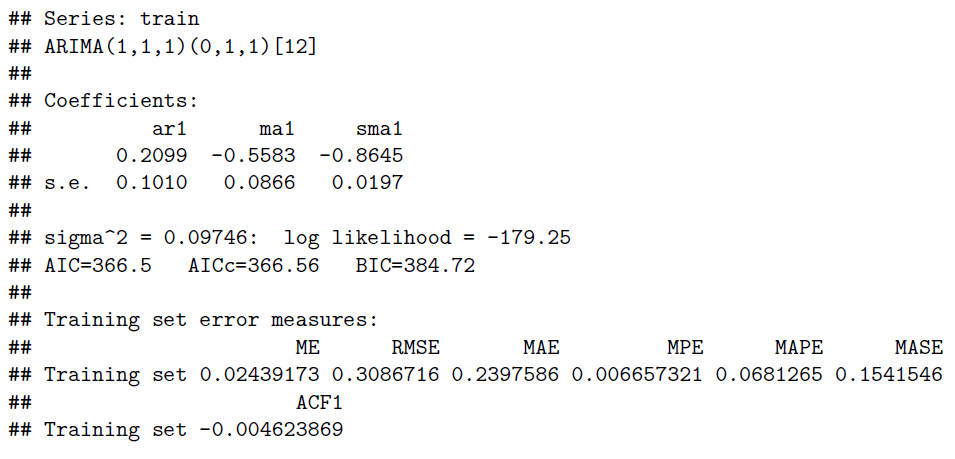
**Appendix C**: Model 2: ETS Seasonal Method

Model 2 Accuracy:

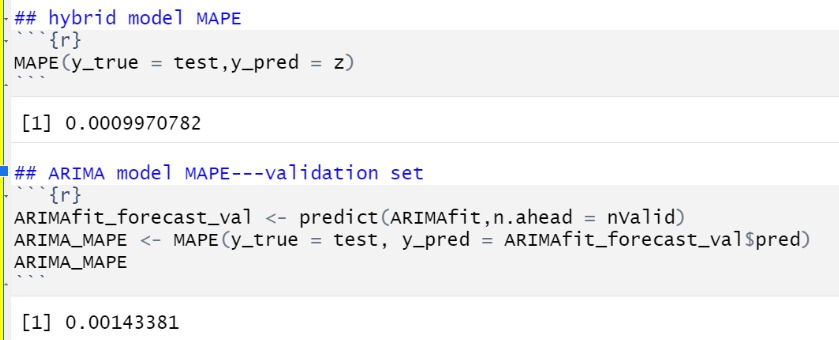


**Appendix D**: Model 3: ARIMA Method

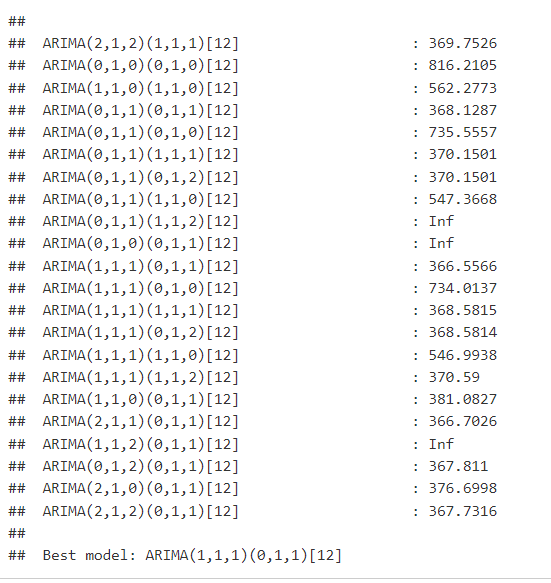
Model 3 Summary and Accuracy:



Model 3 Validation Set MAPE:

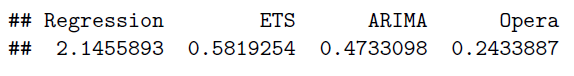


ARIMAfit, auto.arima() function output:



**Appendix E:** Model 4: ETS and ARIMA Hybrid Model

Model 4 accuracy results (MSE):



Model 4 Validation Set MAPE:

