Rising CO_2 Levels in the Atmospheric Concentration Above Mauna Loa

Coleton Reitan

Overview

- Climate Change has become a primary topic of concern over the past decade.
- While some may argue against it, there is plenty of evidence pointing to climate change being a direct result of human interaction with the Earth.
- This can be shown through the increasing levels of greenhouse gases in the Earth's atmosphere.
- Greenhouse gases can easily trap heat from the sun within the Earth's atmosphere, resulting in the global warming and climate change being witnessed today.
- In 2019, the United States alone emitted 6.6 billion metric tons of greenhouse gases.
- Of this 6.6 billion tons, carbon dioxide (CO2) was the leading emission, totaling for about 80% of total US emissions. (c2es.org)

Objective

- Although emissions in the United States are high, the problem is not necessarily how much is emitted, but how much of that emitted gas stays within the atmosphere.
- So, in order to understand the true impact of these gas emissions in the United States, I will be performing a time series analysis of the atmospheric levels of carbon dioxide in the United States over the past several decades, above Mauna Loa, Hawaii.
- By the end of this analysis, I expect to have an understanding of not only the rate of increase in atmospheric CO_2 levels in the United States, but also where future levels are heading.

Previous Research

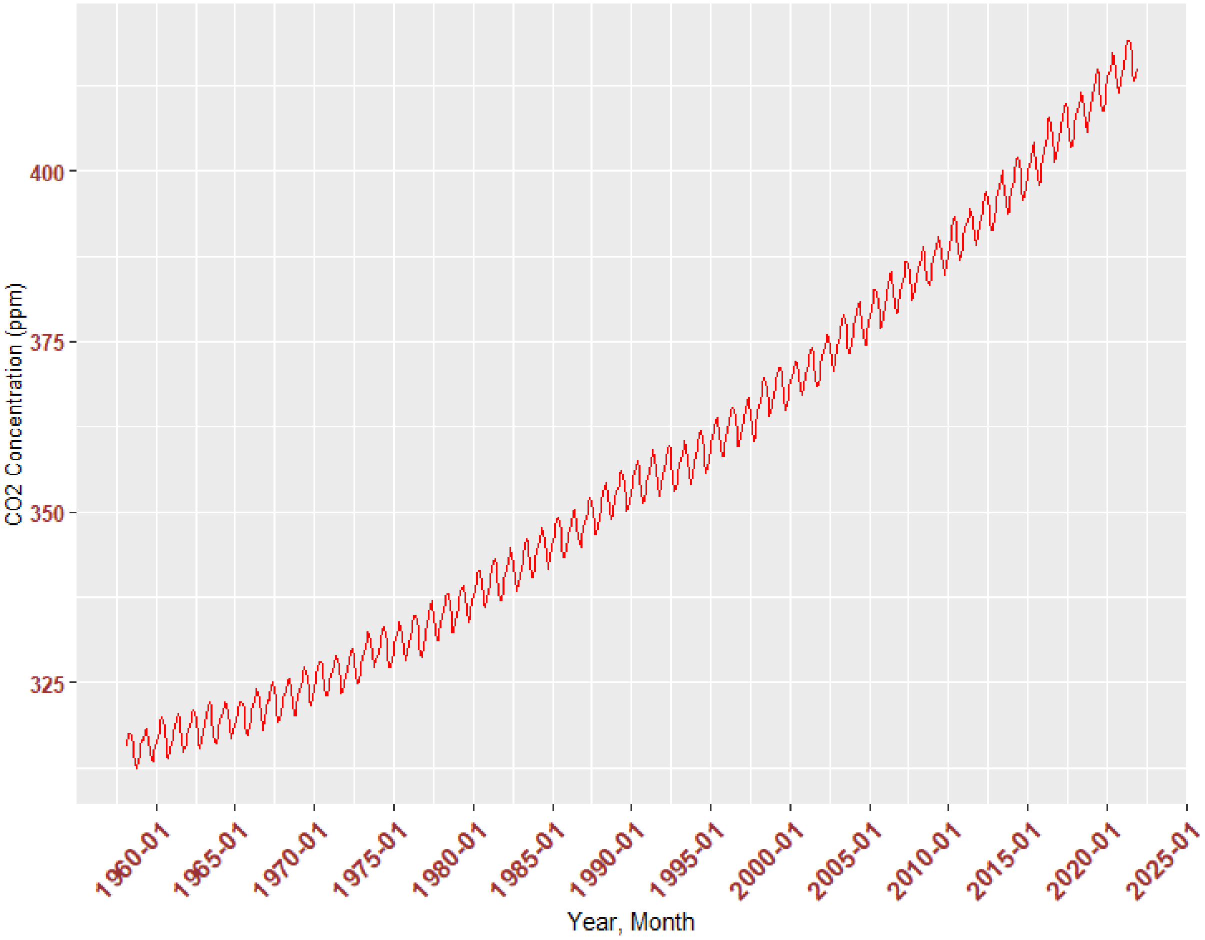
- There have been many studies on the increase of CO_2 levels, especially in recent times due to the increased social awareness of climate change.
- In the main study I based my work on, the researchers used Fourier Transformations to create a model for atmospheric CO_2 levels based off lab results across several sites in China.
- Another study that also uses lab results from Mauna Loa uses similar methods, but does not go into detail further than simple forecasting methods.
- Finally, in a more recent study, the researchers used more advanced methods such as a multilayer artificial neural network forecasting method to predict global atmospheric CO_2 levels.

Dataset

- The dataset being worked with is from the Global Monitoring Laboratory, given by Dr. Pieter Tans and Dr. Ralph Keeling (gml.noaa.gov).
- The data consists of monthly average levels of carbon dioxide, measured in Parts Per Million (PPM), dating back to 1960.
- The data is collected from an observatory in Mauna Loa, Hawaii. The lab is stationed in this location because it is said to have an accurate description for the true global atmospheric concentration. Reasons for this include:
    - The height above sea level is 3400 meters, so the measurements are representative of very large areas.
    - Consistent calibration made to all the measurements.
    - On-site measurement comparisons allow for more accurate data.

Dataset

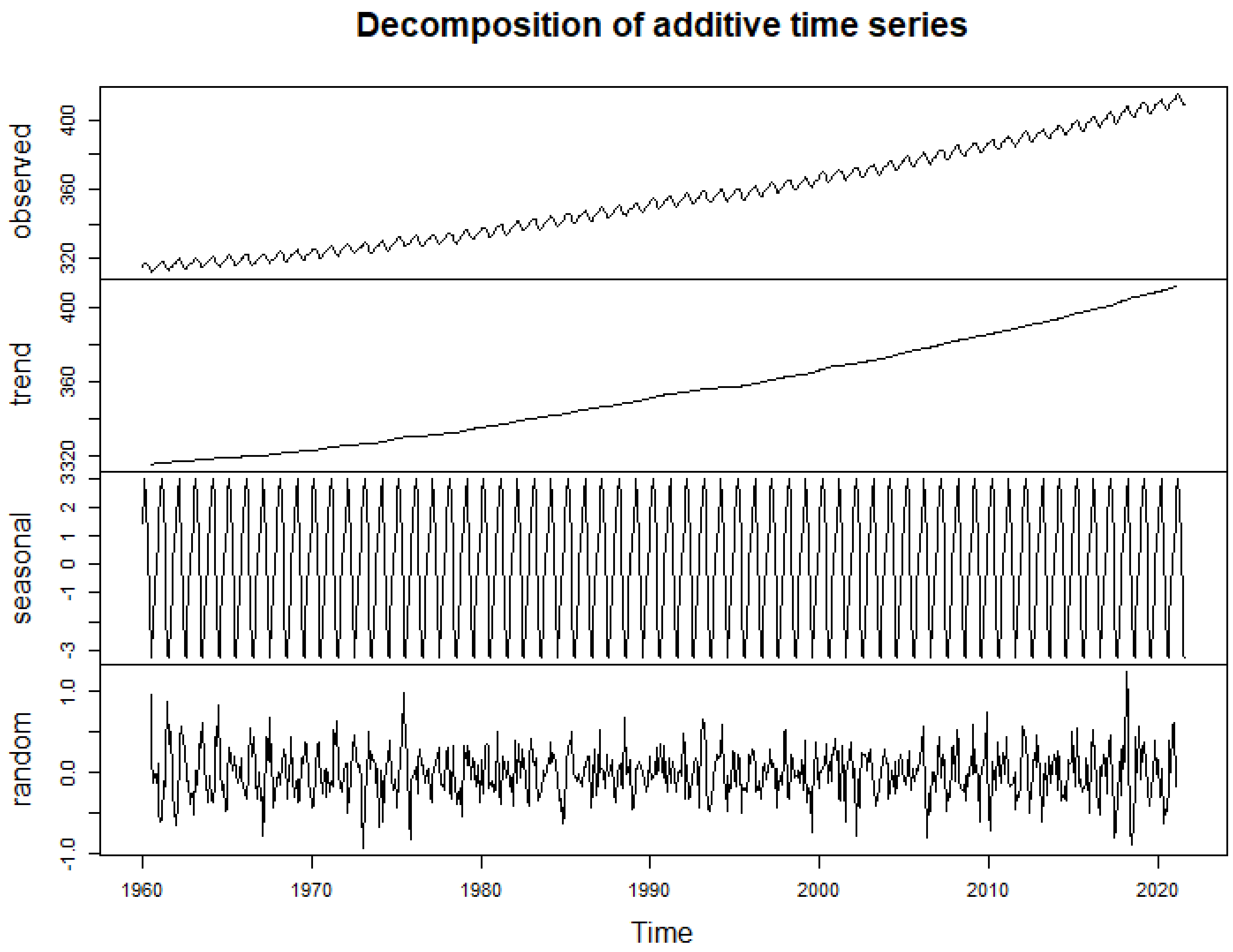
- The plot is showing the atmospheric concentration of CO_2 above Mauna Loa, HI, dating back to 1960.
- It can be seen that the minimum concentration was 312.4 around 1960, where the max is 414.9, dating around 2020.
- Just by looking at the plot, there is an evident problem in the increase of CO_2 levels.




```
> summary(mlts)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  312.4   328.9   352.0   354.7   377.7   414.9
```

Decomposition

- After decomposing the data, it is evident that we are dealing with an additive time series dataset.
- First, we can see a clear positive trend.
- Next, we can see there is definitely a seasonal aspect to the data that should be accounted for.



Methods

Methods

- In order to have the best possible analysis of the data, I will be examining multiple forecasting models.
- These modeling methods include 3 "simple" methods:
    - Naïve Method
    - Drift Method
    - Holt-Winters Method
- For more accurate models, I will be using autoregressive moving average models:
    - SARIMA method
    - SARIMAX method

Naïve Method

- The first simple method that will be used in the analysis is the Naïve Method.
- The idea behind the naïve method is to set the forecasted values to be equal to the value of the last observation.
- Due to the exceptional seasonality seen within the dataset, I will be incorporating the seasonal aspect into my naïve forecasting.
    - Equation: $\mathrm{Y}_{T+h|T} = Y_{T+h-m(k+1)}$
- Seasonality is accounted for by setting the forecast equal to the last observed value from the same season of the year (same month of previous year…)

Drift Method

- The next simple method I will be using is a variation of the naïve method; the Drift Method.
- The Drift Method allows the forecast to increase and decrease (change) over time. This is change is calculated simply by using the average change from the historical data.
    - Equation: $Y_{T+h\,|\,T} = Y_T + (h/T-1)\ {}^{T}\sum_{t=2} (Y_t - Y_{t-1}) = Y_T + h(Y_T - Y1\ /\ T-1)$

Holt-Winters Method

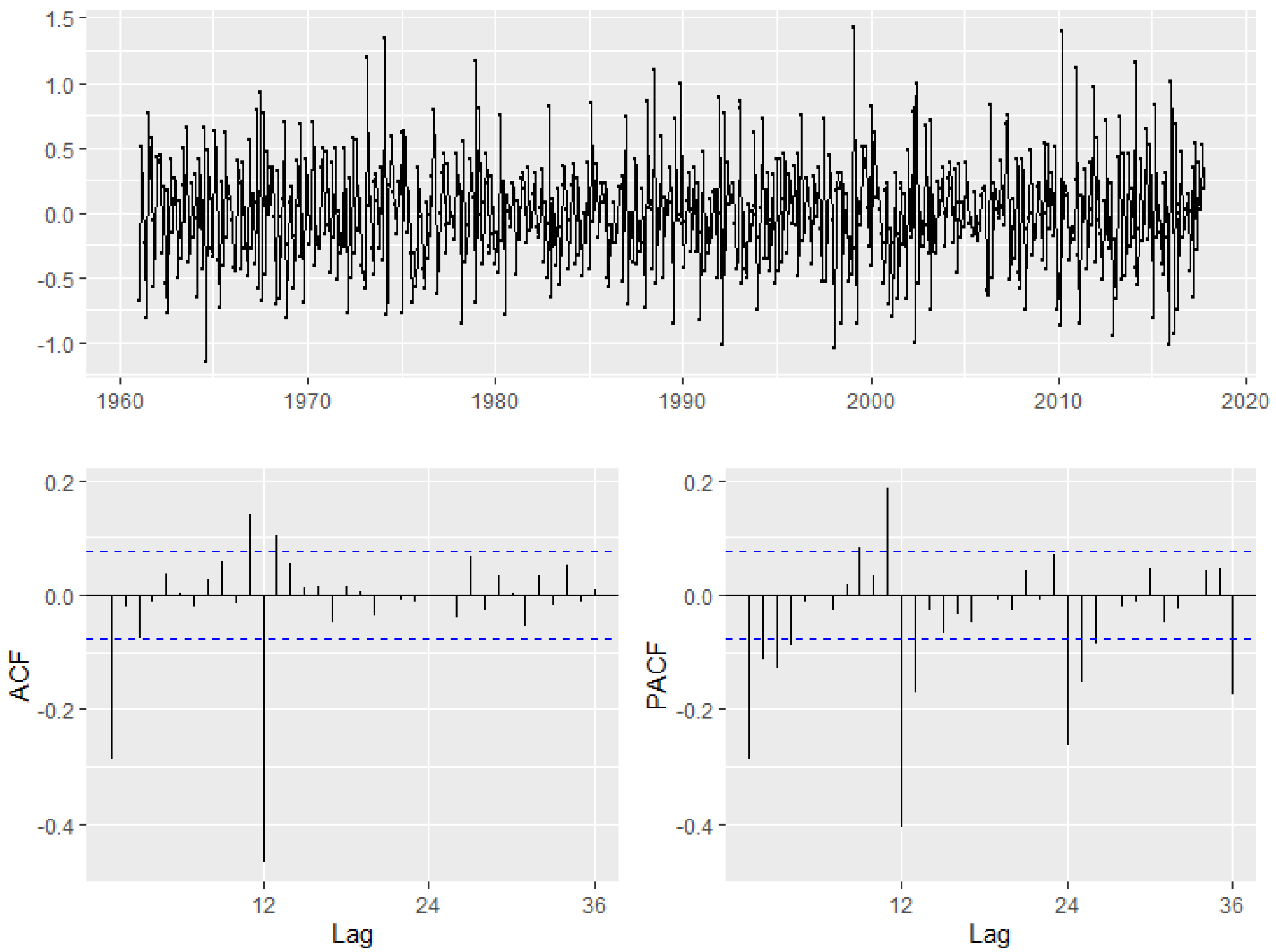
- The final simple forecasting method that will be used in the analysis is the Holt-Winters forecasting method.
- This method is expected to be the most accurate of the simple forecasting methods, as it incorporates three smoothing equations; one for level, one for trend, and one for seasonality.
- There is also an additive and multiplicative version of this method. As was shown before, this dataset is additive, so that is the version of the Holt-Winters method that will be used.
    - Equation: $Y_{t+h|t} = \ell_t + hb_t + s_{t+h-m(k+1)}$

SARIMA Model

- Moving on from simple forecasting methods and into more descriptive modeling, an ARIMA model will be used. However, due to the evident seasonal aspect in the data, the model will really be a SARIMA model.
- The first step to take preparing parameters for my SARIMA model is to check that the data is stationary. After taking an ADF test, the p-value was well above .05, telling me that the data was in fact non-stationary, and must be corrected and accounted for in my model.
- This can be corrected for by using differencing, which is computing the difference between consecutive observations.
- After taking a first order difference and reapplying the ADF test, the data was shown to be stationary.

SARIMA Model

- After differencing, I then looked towards the ACF and PACF plots to help determine the AR and MA terms. (With differencing and seasonality, it is determined that d=D=I(1)
- The ACF plot shows a recommendation of using a q=MA(3) model, as there are 3 spikes before the first lag, with one significant spike in the seasonal lag, so Q=MA(1). The PACF plot shows signs of a seasonal P=AR(3) model, so that will be tried first.



SARIMA Model

- With those parameters in place, a preliminary equation can now be put together.
    - Equation: $(1-\phi_1 B)\ (1-\Phi_1 B^{12})(1-B)(1-B^{12})Y_t = (1+\theta_1 B)\ (1+\Theta_1 B^{12})\varepsilon t$
- It can be seen on the left side of the equation are the AR terms, along with the appropriate differencing not accounting for as well as accounting for seasonality.
- On the right side of the equation are the MA terms, seasonality being shown be the right most.
- The SARIMA model I will be testing will be an ARIMA(0,1,3)(3,1,1)[12] model.
- I will not be limited to just testing this one model, but as I said before, this is my preliminary model.

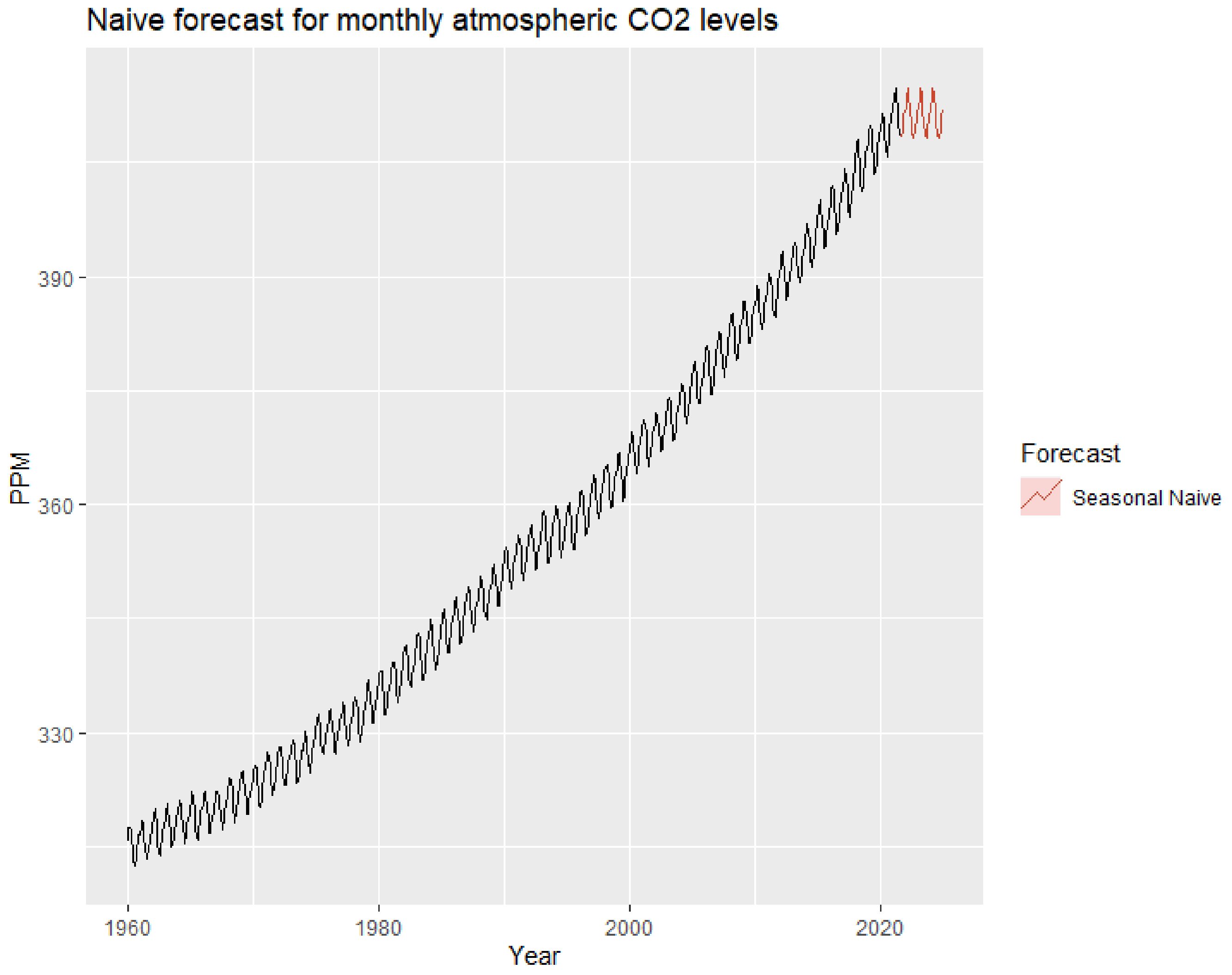
SARIMAX Model

- Following the implementation of the SARIMA model, I will be using a SARIMAX model in order to enhance the forecasting of the SARIMA model.
- SARIMAX models incorporate exogenous variables in attempt to have a better model prediction.
- For my model, I will be testing with 2 exogenous; US average income as well as US unemployment rates.
- Both variables will need to be differenced in an effort to make them stationary.
- The equation for the SARIMAX model is very similar to that of the SARIMA model, only difference is the addition of an exogenous covariate.
    - Equation: $(1-\phi_1 B)(1-\Phi_1 B^{12})(1-B)(1-B^{12})Y_t = (1+\theta_1 B)(1+\Theta_1 B^{12})\varepsilon t+\beta xt$

Results

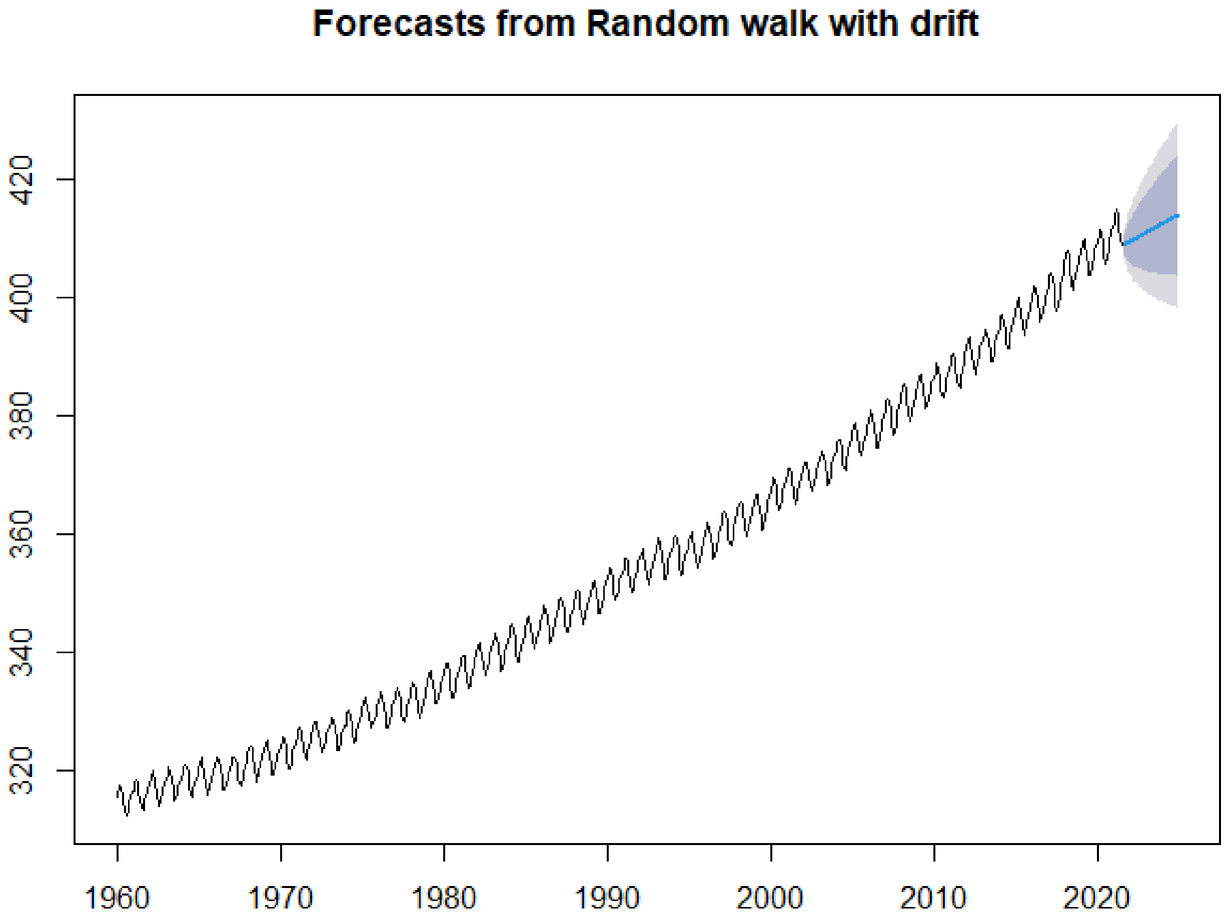
Naïve Results

- I'll start by examining the results given by the simple methods, with the first being the naïve method results.
- It's seen that the plot does a good job of keeping the seasonality but does not do a very good job of keeping the positive trend.



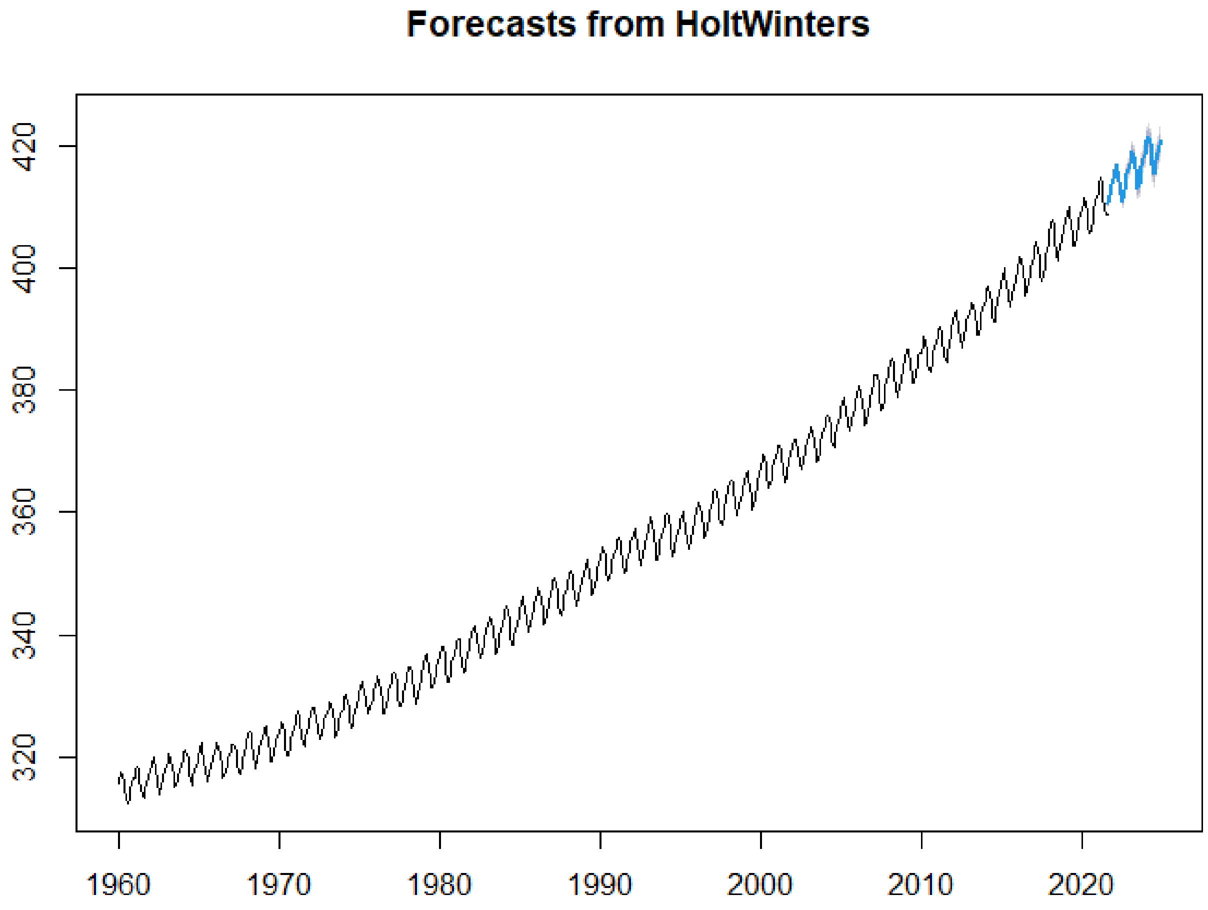
Drift Results

- Now, we'll take a look at the drift method and its respective plot.
- As you can see, the drift method doesn't quite capture seasonality as well as the naïve method does but does a better job of showing the possibility of the intervals for where the values could go.
- We can see that the trend is kept in the drift method.



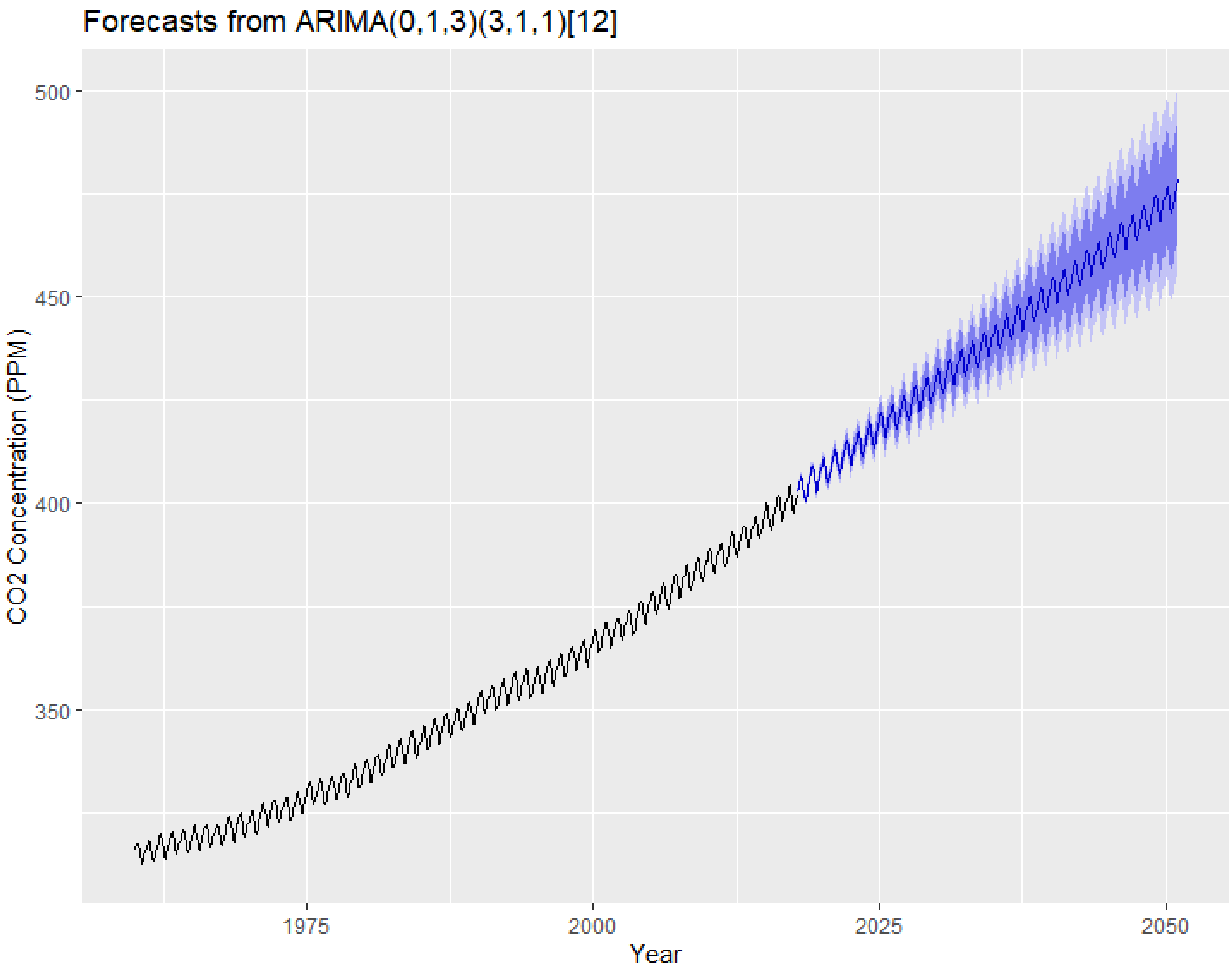
Holt-Winters Results

- Finally, we'll take a look at the Holt-Winters method and its corresponding plot.
- Just by looking at the plot, it is evident that Holt-Winters forecasting is the most accurate of the 3 methods used.
- Not only did this method keep the seasonality, but also kept trend as well and shows a realistic prediction of where atmospheric CO2 levels could go.



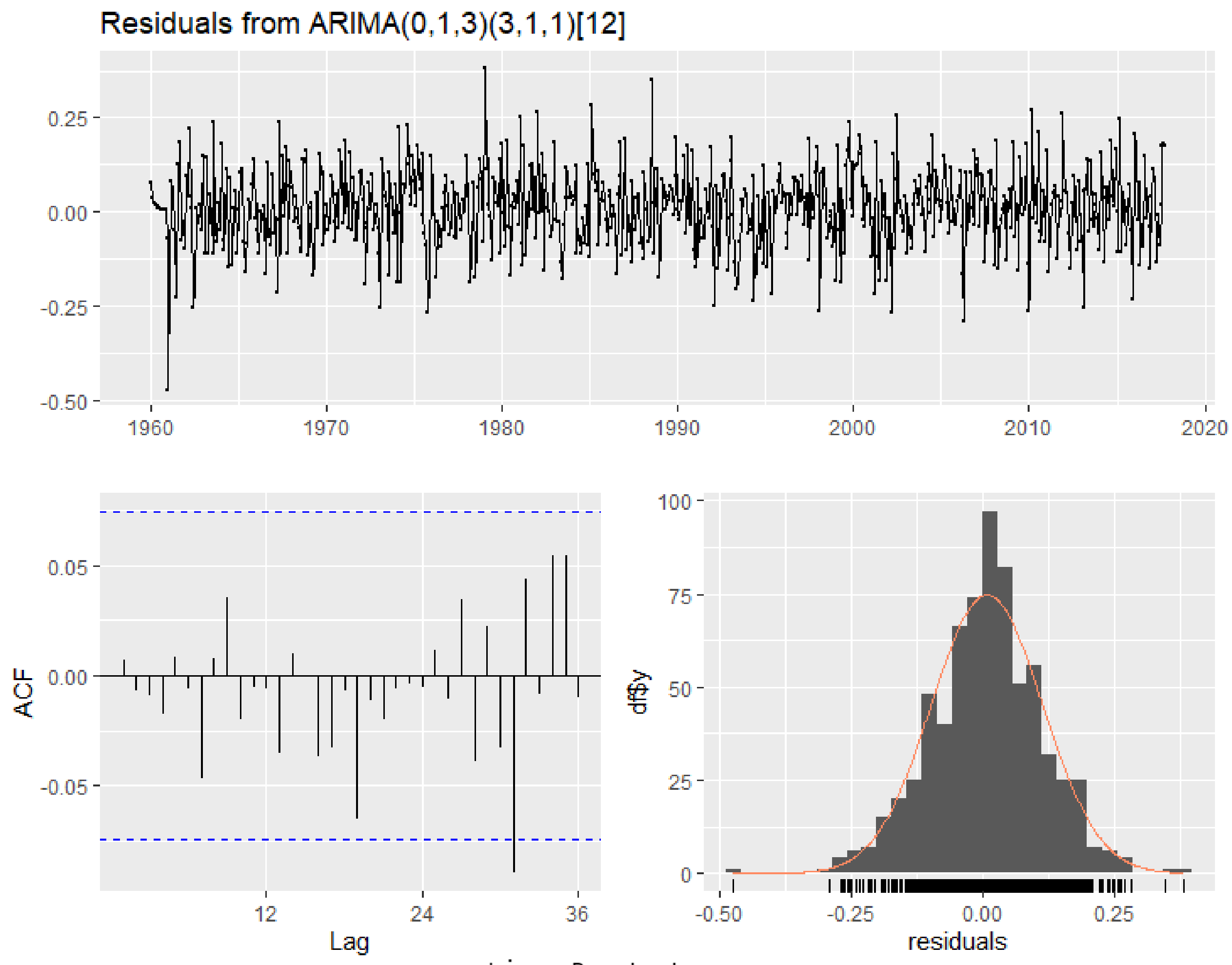
SARIMA Results

- Now we'll look at the SARIMA modeling results, forecasting to 2050.
- Recall: The model parameters that I had come up with was an ARIMA(0,1,3)(3,1,1)[12] model.
- The AIC for this model was -1095.1323, which is relatively good compared to other models that will be shown.
- The plot proves that the model is strong, as the trend and seasonality is kept in the forecast.



SARIMA Results

- It’s important to check the residuals for the model, to ensure that the model adequately captured information in the data.
- Looking at the residuals plot, it can be seen that the residuals are fairly balanced.
- Another important value to look at would be in the Ljung-Box Test.
- Seeing a value above .05 assures me that this model is good.



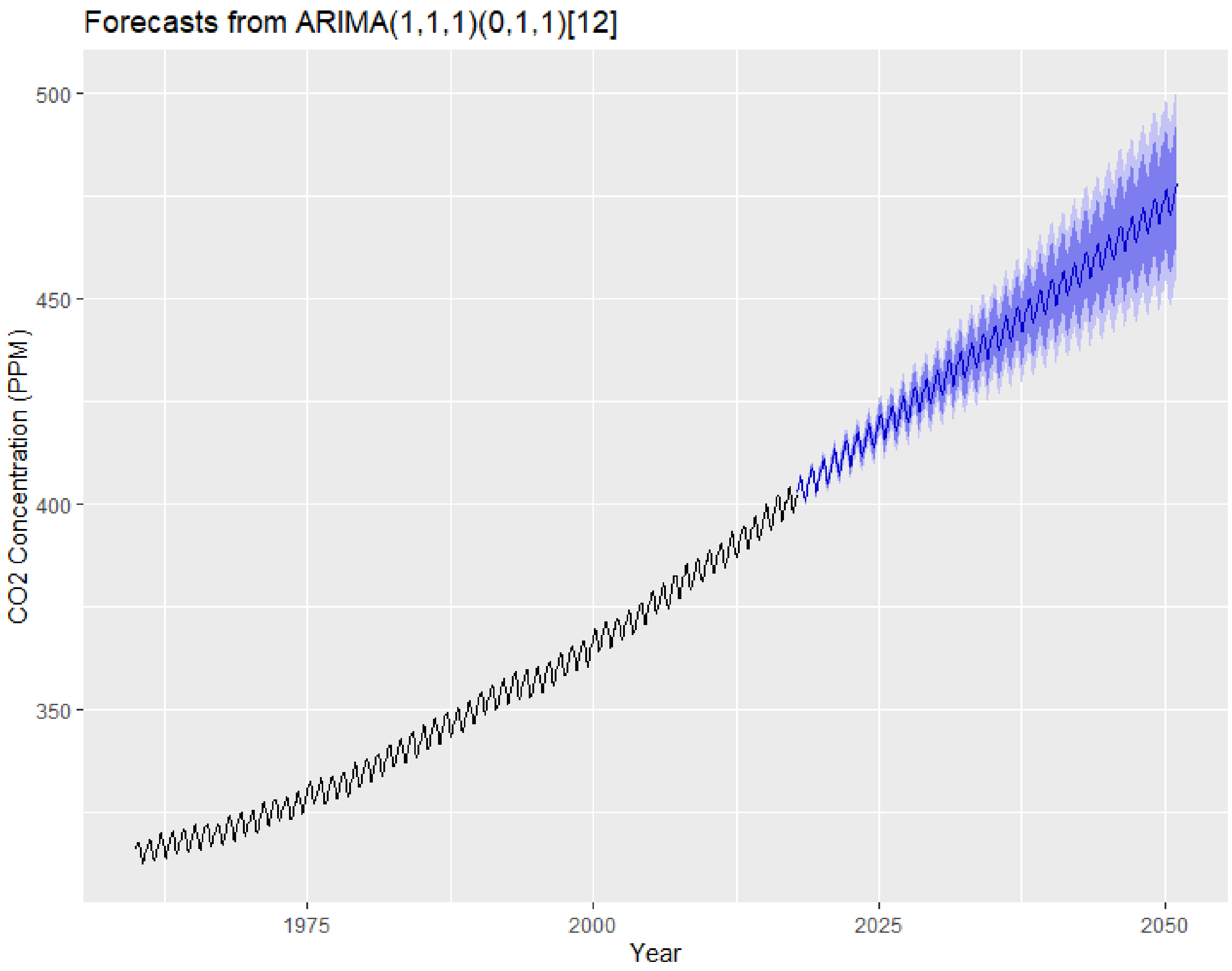

```
	Ljung-Box test

data:  Residuals from ARIMA(0,1,3)(3,1,1)[12]
Q* = 24.46, df = 29, p-value = 0.7059

Model df: 7.   Total lags used: 36
```

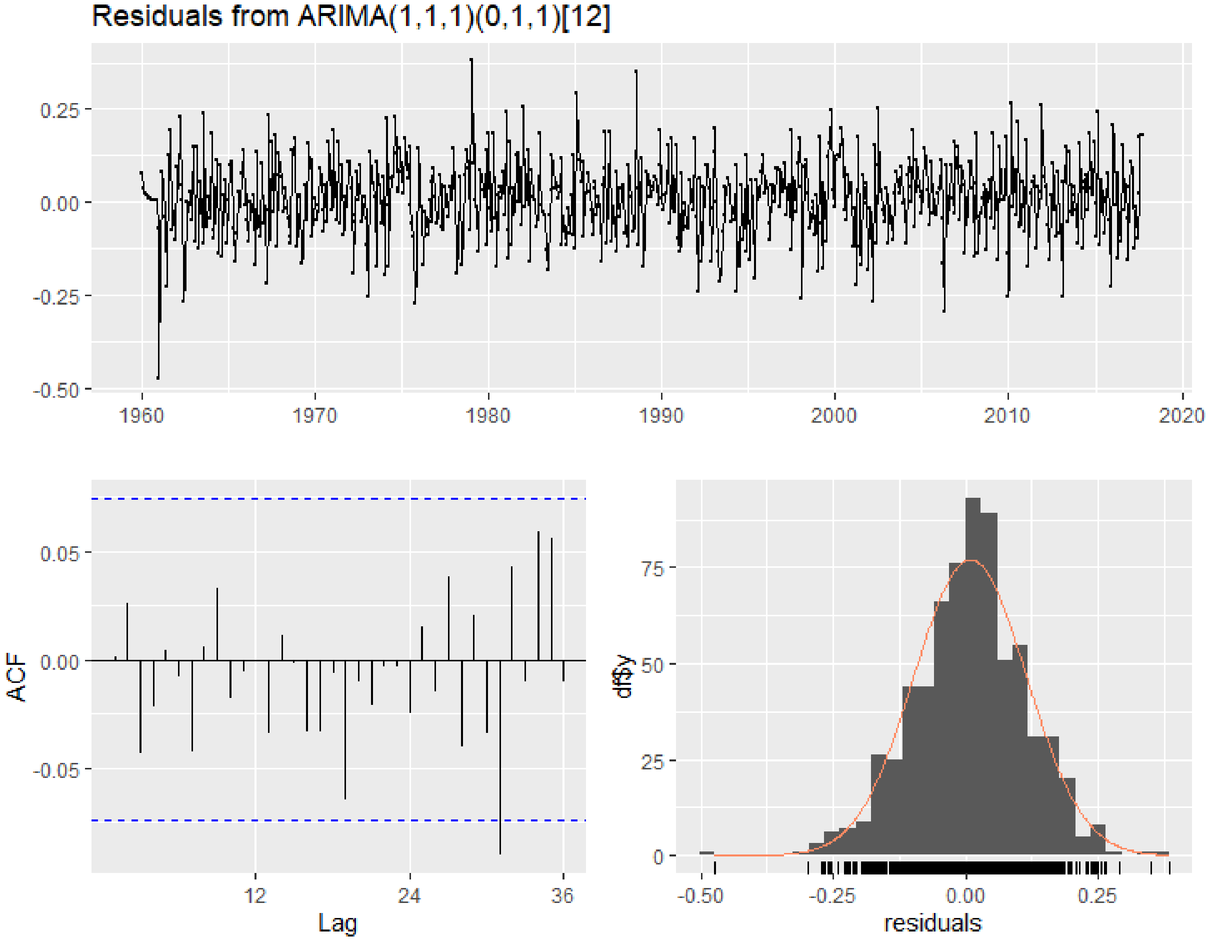
SARIMA Results

- I think it would be beneficial to see how the auto-arima model compares to my own.
- The model parameters that the auto-arima had come up with was an ARIMA(1,1,1)(0,1,1)[12] model.
- The AIC for this model was -1101.2607, which is better than my model’s AIC of -1095.1323, but not by much.



SARIMA Results

- Looking at the residuals for the auto-arima model, it's evident that the residuals for this model are also well balanced.
- The Jjung-Box Test has a p-value well above .05, telling me that statistically, this model is good.




```
        Ljung-Box test

data:  Residuals from ARIMA(1,1,1)(0,1,1)[12]
Q* = 26.865, df = 33, p-value = 0.7656

Model df: 3.   Total lags used: 36
```

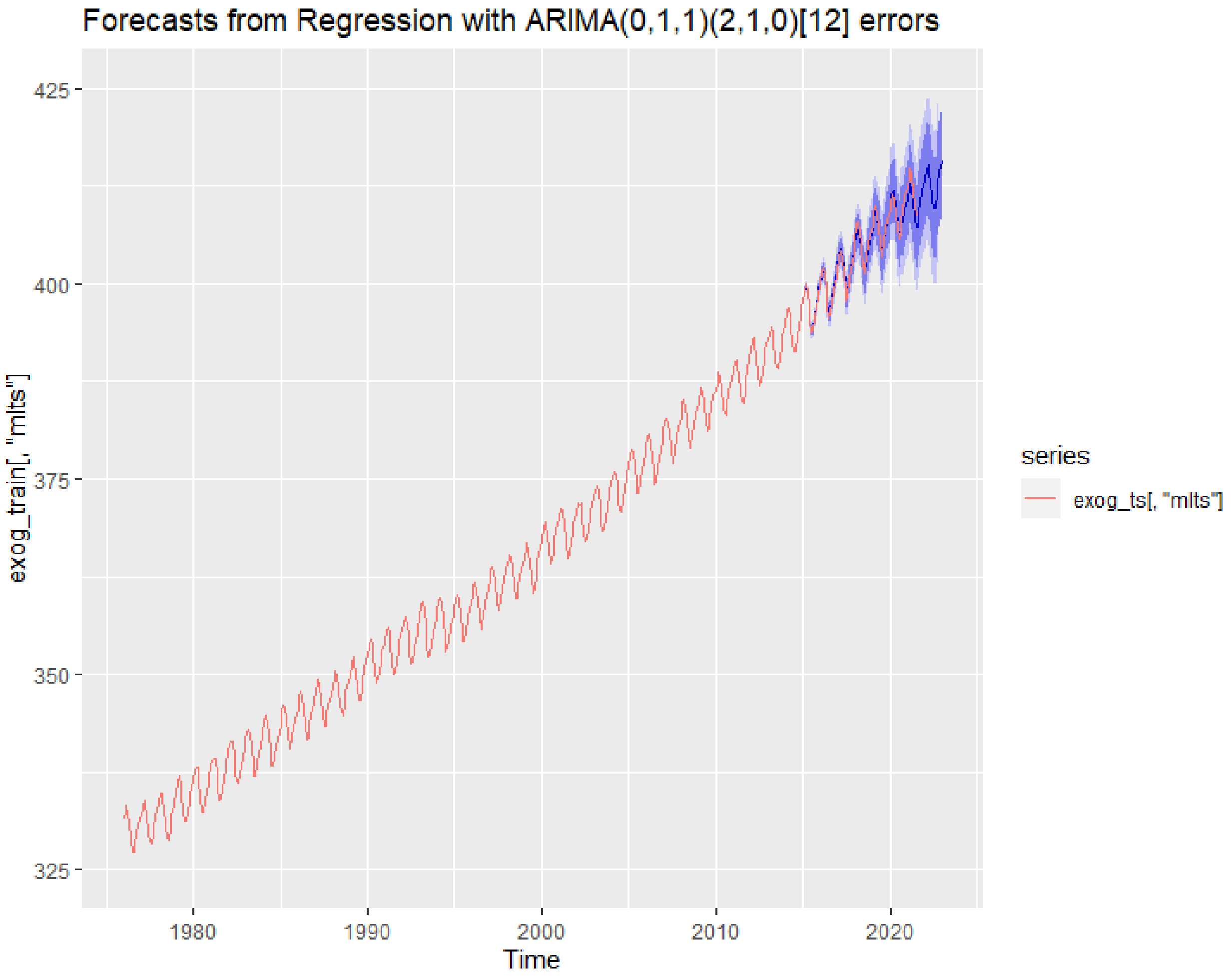
SARIMA Results

- There were other models tested as well but only one other model had a better AIC than my predicted model (which was ARIMA(1,1,3)(0,1,1) with an AIC of -1098.8099) and the autoARIMA model, as seen in the AICs listed here.

	Model_name	AIC
1	ARIMA(0,1,3)(3,1,1)[12]	-1095.1323
2	ARIMA(0,1,1)(3,1,1)[12]	-1093.8877
3	ARIMA(1,1,0)(1,1,0)[12]	-898.9803
4	ARIMA(1,1,2)(1,1,0)[12]	-915.4286
5	ARIMA(1,1,3)(0,1,1)[12]	-1098.8099
6	ARIMA(1,1,1)(1,1,0)[12]	-917.1378
7	ARIMA(1,1,1)(1,1,0)[12]	-917.1378
8	ARIMA(1,1,0)(1,1,1)[12]	-1083.5009
9	ARIMA(1,1,1)(0,1,1)[12]	-1101.2607

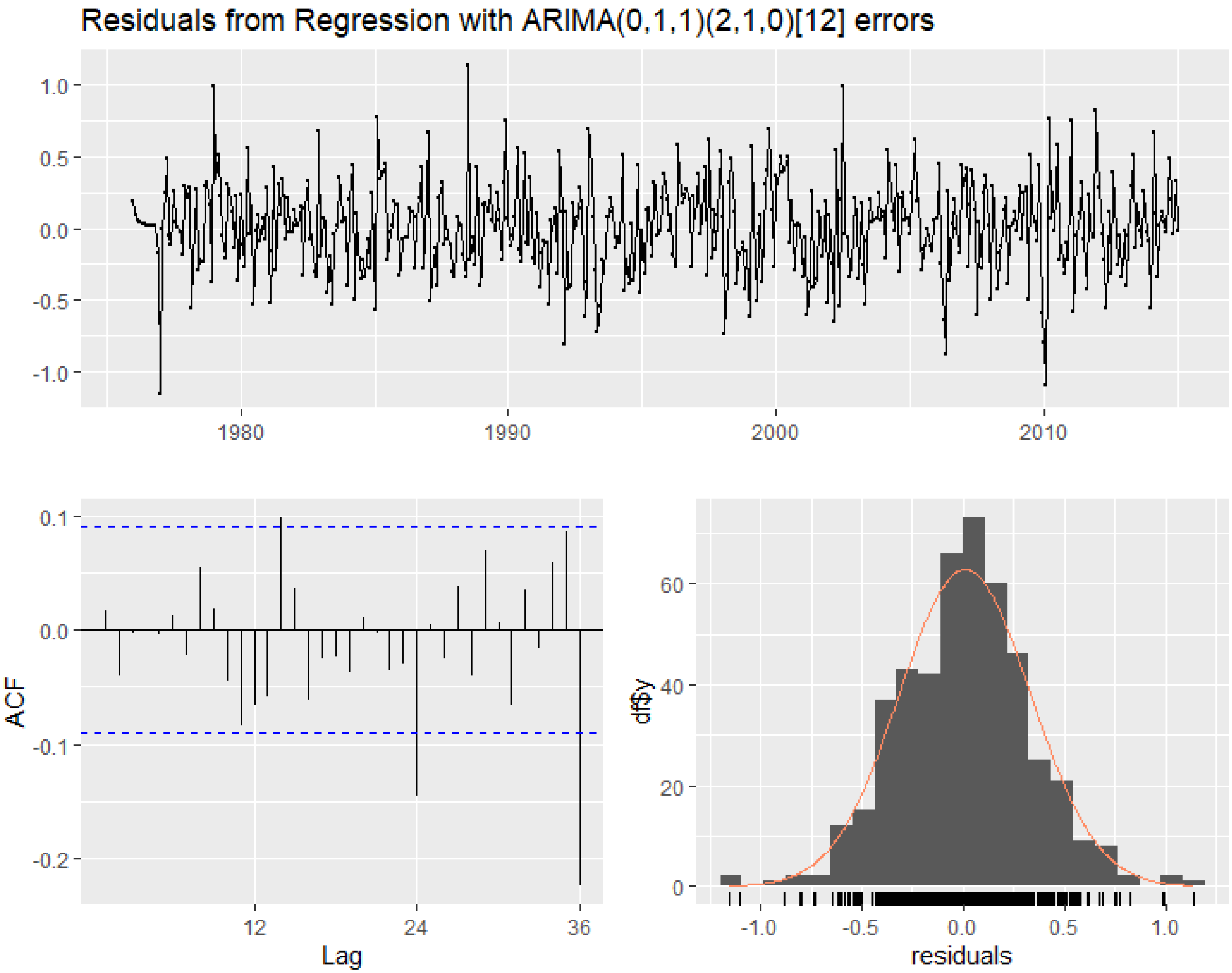
SARIMAX Results

- Now we'll take a look at the SARIMAX modeling results.
- For the first model, the exogenous variable of average income in the US was used. The data was originally non-stationary, so that had to be corrected.
- Looking at the plot, the model is shown to be very good, as the values resemble those very similar to the original data.



SARIMAX Results

- Looking at the residuals, the model looks good.
- The Ljung-box Test is also above .05 (just slightly), which tells me the model just meets the criteria to be a good fit.
- The AIC of the model is 283.94, so the model has a very good fit.



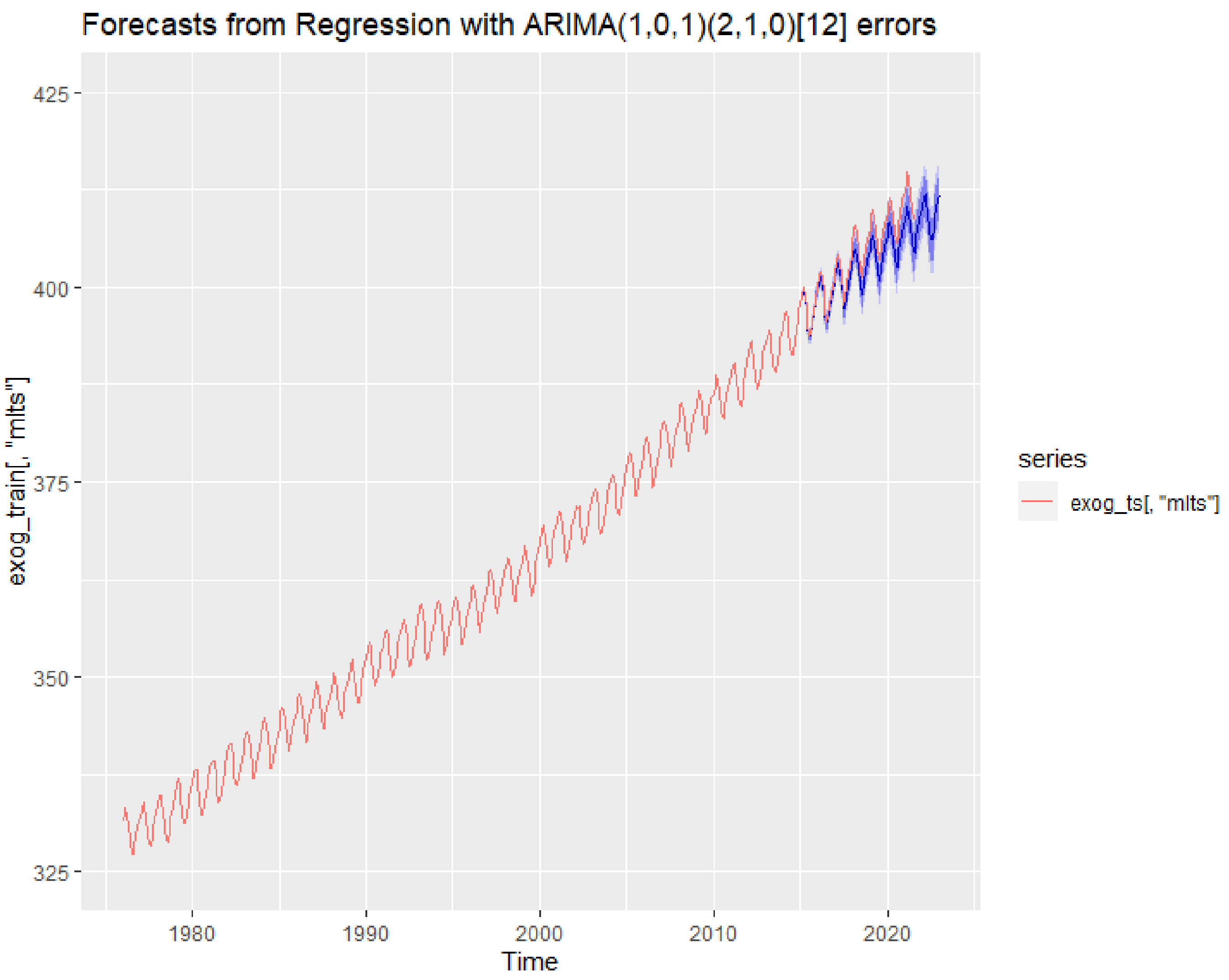

```
        Ljung-Box test

data:  Residuals from Regression with ARIMA(0,1,1)(2,1,0)[12] errors
Q* = 30.934, df = 20, p-value = 0.05607

Model df: 4.   Total lags used: 24
```

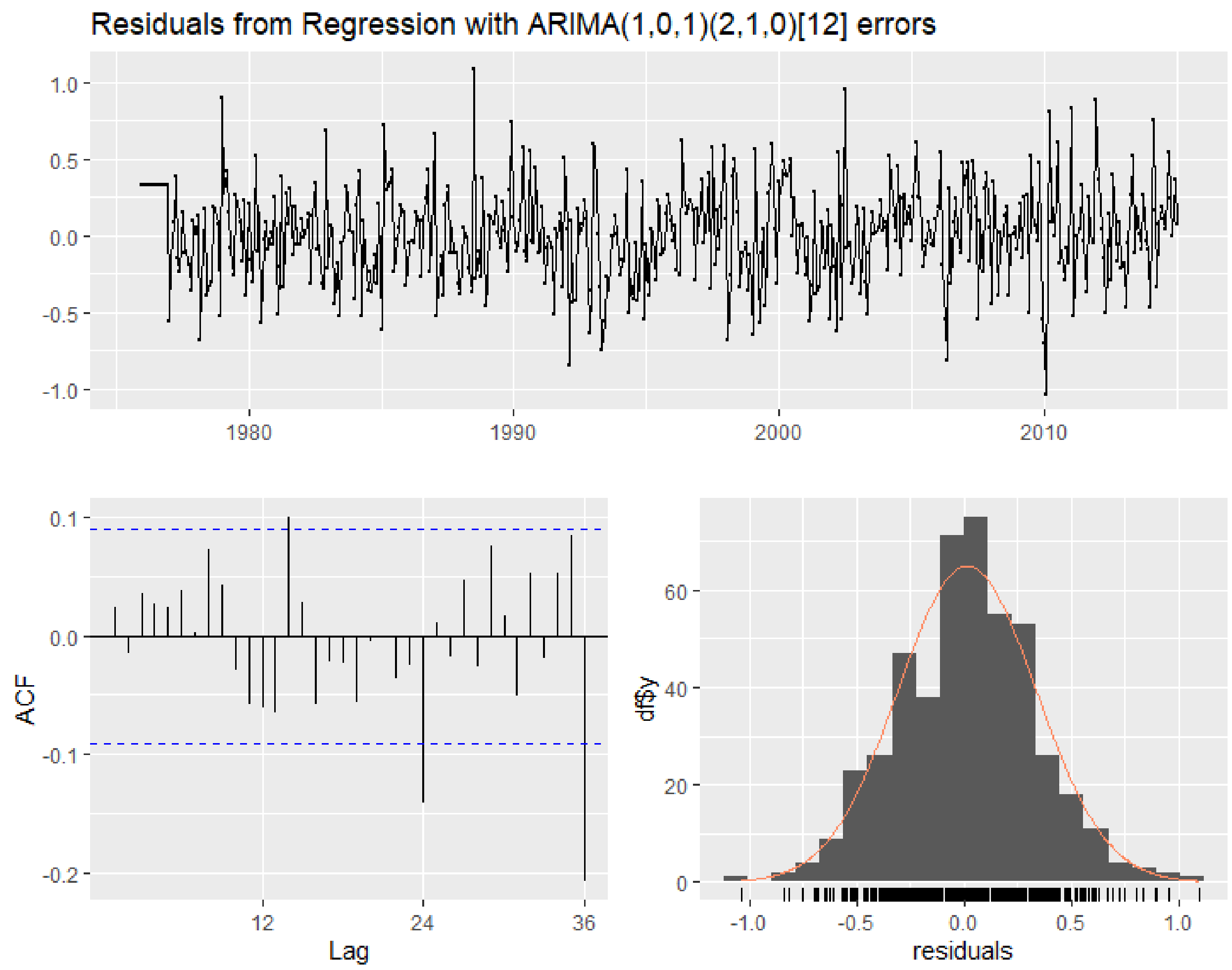
SARIMAX Results

- The second SARIMAX uses the exogenous variable US unemployment.
- Just by looking at the plot, it is evident that the model is not as good as the previous SARIMAX model, as you can see the values are not as precisely on the trend, as the predicted blue lines predict a little lower than the actual.



SARIMAX Results

- Looking at the residuals, we can also see the model is not as good, as it is not nearly as smooth as other models we've seen.
- Unsurprisingly, the Ljung-Box Test proved that this model should not be used, as the p-value is below .05 telling me that there is still correlation that is unaccounted for in the model.




```
	Ljung-Box test

data:  Residuals from Regression with ARIMA(1,0,1)(2,1,0)[12] errors
Q* = 31.528, df = 18, p-value = 0.02499

Model df: 6.   Total lags used: 24
```

Limitations

- The modeling performed for this analysis was good, but not perfect, as there were some limitations when it came to creating forecasting models.
- In the ARIMAX modeling, the preferred dataset I wanted to use was monthly average GDP, but I could not find any that were suitable. This may have been a better and more constructive option.
- I also recommend further research uses more advanced exploratory techniques that may show a more accurate representation of where future global atmospheric levels could be heading towards.

Discussion

- After a very detailed exploratory time series analysis of the atmospheric concentration of CO_2 above Mauan Loa, Hawaii, it is evident that the trend of ever-increasing CO_2 levels will continue.
- The models used within this study were good, but for future studies, it may be beneficial for them to use more advanced forecasting methods, such as spectral analysis or hidden Markov modeling.
- Aside from the study itself, the continual increase in the atmospheric concentration of CO_2 is not a good sign for the planet. Climate change will continue to be a problem and will more than likely get much worse before it starts to get better.
- As a society, it is imperative that we slow this growth of CO_2 over the next few years and try to prove the "accurate" forecasts wrong.

References

- U.S. Emissions | Center for Climate and Energy Solutions (c2es.org)
- Global Monitoring Laboratory - Carbon Cycle Greenhouse Gases (noaa.gov)
- energies-14-06336-v2.pdf
- Temporal and spatial variations of the atmospheric CO2 concentration in China (wiley.com)
- Time Series Modeling for Atmospheric CO2 Concentration(ppm), 1958–2019 | by Peyman Kor | Towards Data Science