EUROPA-UNIVERSITÄT VIADRINA FRANKFURT (ODER)

Time series of carbon dioxide emissions in the United States using the ARIMA(p,d,q) model

Data Analysis for Business and Economics

Submitted to: Prof. Dr. Simone Maxand

Iully Cegovia Rodrigues

122024

March 30th, 2023

Contents

Introduction

Global warming has been a widely debated and researched topic since the end of the last century. Jones et al. (1988) first reported a considerable increase in the earth's average temperature and pointed out that one of the factors for this phenomenon is the emission of carbon dioxide (CO^2) in high quantities into the atmosphere, besides other pollutants gases. The potential risks of global warming can already be seen by observing natural phenomena such as tsunamis, melting glaciers and rising sea levels (Jena et al., 2021). In order to avoid the possible increase of the earth's temperature by 2 degrees Celsius, it is necessary to take measures to reduce the emission of pollutants into the atmosphere (Dobrescu, 2009).

According to the Intergovernmental Panel for Climate Change (2014), studies tracking the increase in the emission of gases into the atmosphere show that human impact substantially helps heat waves to occur in certain areas. These heat waves, in turn, are predicted to have long-term consequences for nature. Since 1992, when it was realised that emerging countries would be the most affected by global warming, most industrialised countries have decided to commit themselves to reduce their pollution emissions and creating a fund for climate adaptation costs (Dobrescu, 2009).

The United States (US) is the second largest CO^2 emitter with 4.59 Gt of global emissions in 2019, second only to China with 10.06 Gt (Climate Watch, 2022). Despite being one of the biggest contributors of harmful gases, the position of the US in relation to global warming has varied a lot since the Kyoto Treaty, the first agreement to reduce pollution in 1992, where the country initially signed the agreement and later withdrew (Dobrescu, 2009). More recently in 2016, the Paris Agreement emerged as a treaty that treats decarbonisation policy not only as a climate issue but also as an economic one, with the adoption of technological measures to help achieve the expected goal (Freeman, 2019). Similar to the first Kyoto agreement, the US later withdrew its participation, only returning in 2021 under the current administration.

Furthermore, the current administration is proposing to reduce greenhouse gas emissions by 50-52% below 2005 levels by 2030 and, most importantly, to achieve the goal of 100% carbon-free electricity as stated by the White House in 2021. This topic is returned to later in the paper.

Given all the data and stated objectives, this research study seeks to forecast future CO^2 emissions in the United States using the ARIMA(p,d,q) model.

Visualising CO2 emissions throughout the years

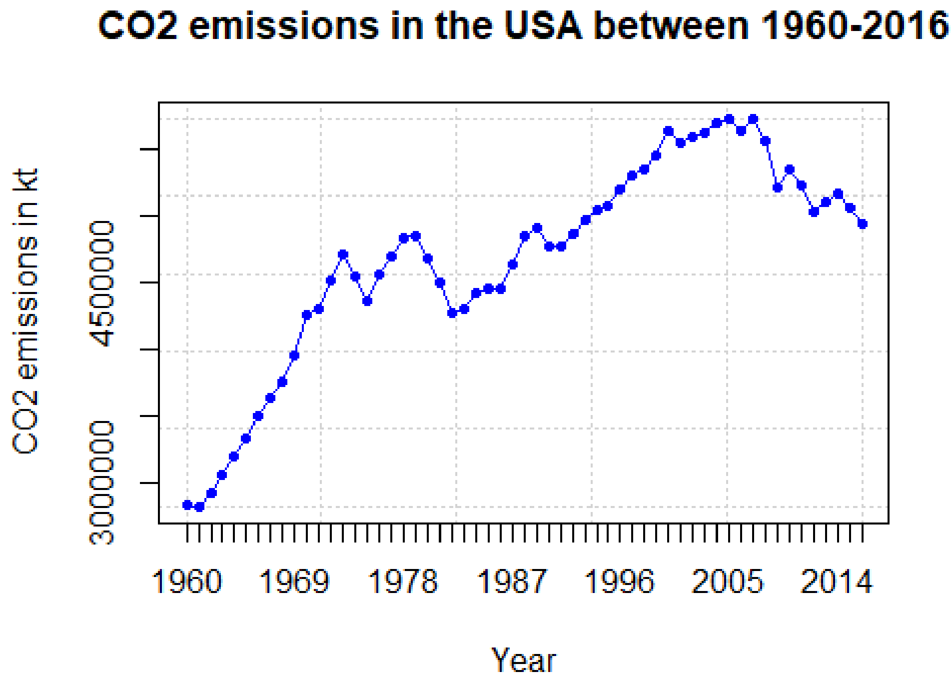

Figure 1: CO2 Emissions in the United States

Data

The data were taken from World Bank, in the World Development Indicator (WDI) section. The total CO^2 emissions were taken from total emissions in solid, gas and liquid form during the period of 1960-2016.

ARIMA Model

The autoregressive integrated average model, or ARIMA model, is a commonly used forecasting model using a combination of the autoregressive (AR) model and the moving average (MA) model to predict nonstationary sequences, i.e., the mean and variance do not change over time (Zhang et al., 2023). The ARIMA model has three parameters ARIMA(p, d, q) that are defined by the following factors:

1. To be able to predict a sequence n, it is necessary to reach its stationarity, i.e., it is necessary to take the differential of the time series until a stable sequence is obtained. This value corresponds to the parameter **d** and is usually between 0, 1 and 2.

2. The parameter **p** corresponds to the AR model which regresses the time series using its past values. These values will be determined using the partial autocorrelation coefficient (PACF).

3. The parameter **q** corresponds to the MA model where the time series with the residuals (errors) of past observations. The visualization of this model can also be achieved through the autocorrelation coefficient (ACF). To test the necessity of transforming the time series into stationary, there are some tests that can be performed to check unit roots and observe if there is any dependency in the sequence.

Stationarity and unit roots tests

The most usual test is the Augmented Dickey-Fuller Test or more commonly called the ADF test. This test is based on the presence of unit roots (Zhang et al., 2023) and if present in the sequence, it is necessary to keep differentiating until it is not possible to identify unit roots present. In this paper, the Phillips-Perron Unit Root Test will also be utilised which also has the function of identifying unit roots, as well as the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test which is used to identify stationarity in time series (Musbah et al., 2023).

For PP and ADF tests, it is assumed that the null hypothesis in the series is non-stationary. That is, if p-value >0.05 the null hypothesis is accepted and if p-value is <0.05 the null hypothesis is rejected. The difference between these two tests and KPSS is that in this model, the null hypothesis assumes that the series is stationary. In addition, the ARIMA(p,d,q) model is estimated according to parameters such as AIC (Akaike's Information Criteria) which is an estimator that predicts error thus defining the quality of the chosen model. As well as AIC, the parameter BIC (Bayesian Information Criterion) is an estimator that is similarly based on the likelihood function.

The Box-Pierce test is a simplified version of the Ljung-Box test and it is commonly applied in econometrics to determine whether there exists an autocorrelation in the time series (Dare et al., 2022). The application of this statistical test will be conducted after selecting the ARIMA(p,d,q) model to verify if there is still an autocorrelation in the residuals. If the p-value is < 0.05, it could be a signal that the time series could not be following the white noise process, i.e. it does not have a zero mean or is autocorrelated (Nyoni et al., 2019).

Model Evaluation

```
##
##  Augmented Dickey-Fuller Test
##
## Dickey-Fuller = -1.9192, Lag order = 3, p-value = 0.6072
## alternative hypothesis: stationary
```

```
##  Phillips-Perron Unit Root Test
##
## Dickey-Fuller Z(alpha) = -3.142, Truncation lag parameter = 3, p-value
## = 0.9257
## alternative hypothesis: stationary

##
##  KPSS Test for Level Stationarity
##
## data:  co2ts
## KPSS Level = 1.2099, Truncation lag parameter = 3, p-value = 0.01
```

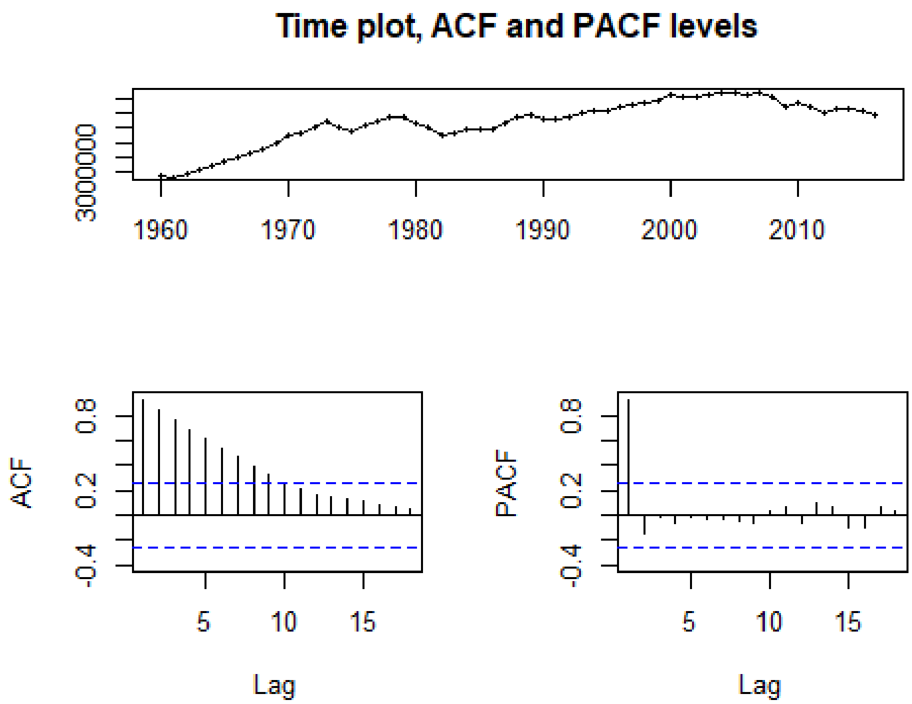

Figure 2: Stationarity Evaluation

The above results indicate that the Augmented Dickey-Fuller Test and Phillips-Perron Unit Root Test have a higher p-value showing that the data is not stationary as the high p-value fails to reject the null stationary hypothesis. The KPSS test shows that there is a presence of unit root.

The ACF decays slowly as it has significant spikes in higher lags and the PACF has a sharp cut after the first lag. It is necessary to take the first and second differences in order to achieve stationarity.

1^{st} Difference Correlogram and Stationarity Test

```
##
##  Augmented Dickey-Fuller Test
##
## data:  diff(co2ts)
## Dickey-Fuller = -3.5515, Lag order = 3, p-value = 0.04526
## alternative hypothesis: stationary

##
##  Phillips-Perron Unit Root Test
##
## data:  diff(co2ts)
## Dickey-Fuller Z(alpha) = -40.89, Truncation lag parameter = 3, p-value
## = 0.01
## alternative hypothesis: stationary

##
##  KPSS Test for Level Stationarity
##
## data:  diff(co2ts)
## KPSS Level = 0.4897, Truncation lag parameter = 3, p-value = 0.04399
```

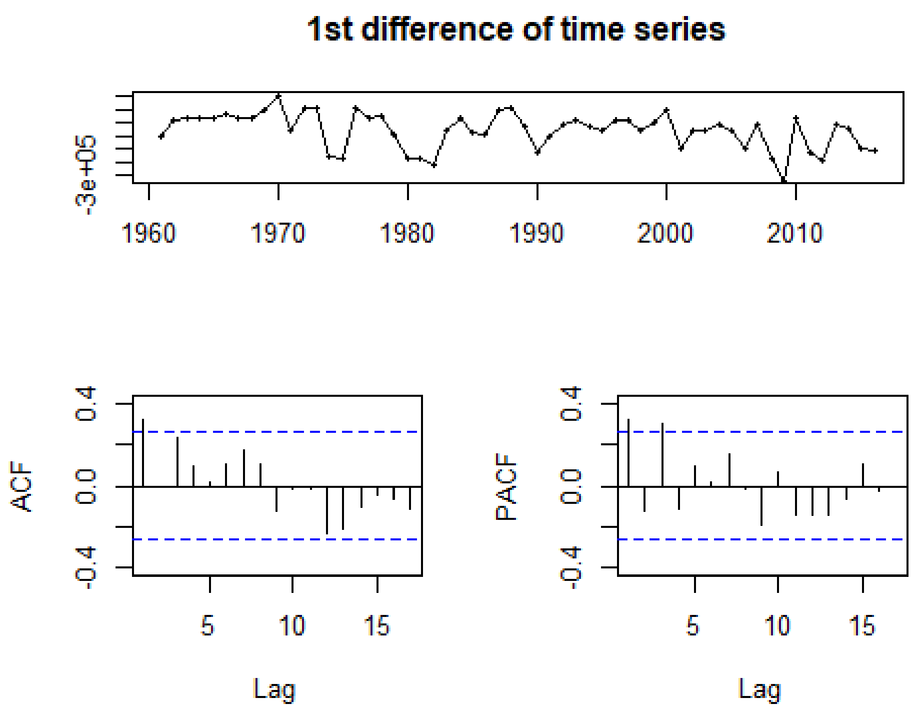

Figure 3: 1st Difference of the Series

2nd Difference Correlogram and Stationarity Test

```
##
##  Augmented Dickey-Fuller Test
##
## data:  diff(diff(co2ts))
## Dickey-Fuller = -5.5606, Lag order = 3, p-value = 0.01
## alternative hypothesis: stationary
```

```
##
##  Phillips-Perron Unit Root Test
##
## data:  diff(diff(co2ts))
## Dickey-Fuller Z(alpha) = -53.754, Truncation lag parameter = 3, p-value
## = 0.01
## alternative hypothesis: stationary
```

```
##
##  KPSS Test for Level Stationarity
##
## data:  diff(diff(co2ts))
## KPSS Level = 0.056818, Truncation lag parameter = 3, p-value = 0.1
```

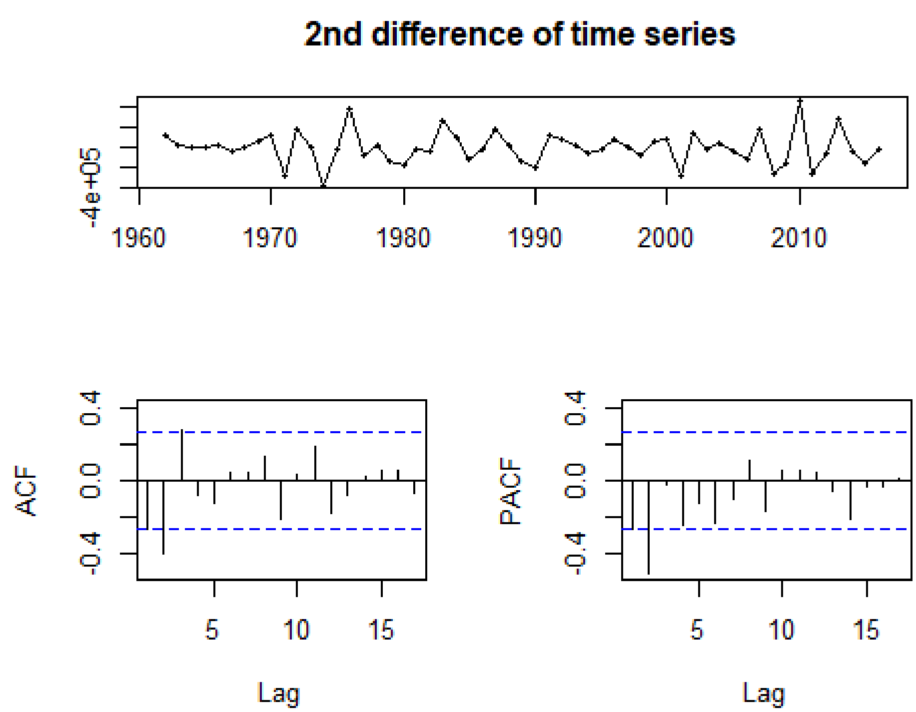

Figure 4: 2nd Difference of the Series

The ADF and PP tests have a p-value <0.05 i.e. it does not confirm the existence of remaining unit roots. The KPSS test does not reject the stationarity of the series, confirming that d = 2.

The PACF might suggest AR(0) and ACF suggests either MA(2) or MA(3). It is necessary to further investigate the most appropriate values to fit the model.

Evaluation of ARIMA Model

Model	**AIC**
ARIMA(0,2,0)	1474.665
ARIMA(0,2,1)	1459.462
ARIMA(0,2,2)	1456.083
ARIMA(0,2,3)	1455.762
ARIMA(0,2,4)	1457.993
ARIMA(0,2,5)	1460.085
ARIMA(1,2,0)	1473.045
ARIMA(1,2,1)	1459.256
ARIMA(1,2,2)	1456.011
ARIMA(1,2,3)	1458.11
ARIMA(1,2,4)	1460.195
ARIMA (2,2,0)	1459.181
ARIMA(2,2,1)	1458.969
ARIMA(2,2,2)	1457.772
ARIMA(3,2,0)	1461.449
ARIMA(3,2,1)	1458.452
ARIMA(3,2,2)	1459.539
ARIMA(4,2,0)	1460.733
ARIMA(4,2,1)	1459.69
ARIMA(5,2,0)	1462.279

A recommended model is a model that has the lowest AIC (Nyoni et al., 2019). Therefore, the recommended ARIMA model is ARIMA (0,2,3) which is consistent with the second difference correlogram.

In order to check the validity of the test, a Ljung-Box Q-test will be performed and check if the residuals are independent.

ARIMA(0,2,3) Model Results and Validity

```
## Series: co2ts
## ARIMA(0,2,3)
##
## Coefficients:
##           ma1      ma2     ma3
##       -0.5096  -0.5214  0.2071
## s.e.   0.1299   0.1309  0.1303
##
## sigma^2 = 1.61e+10:  log likelihood = -723.48
## AIC=1454.96   AICc=1455.76   BIC=1462.99
##
## Training set error measures:
##                      ME   RMSE      MAE        MPE    MAPE      MASE
ACF1
## Training set -11633.25 121193 95678.32 -0.1761931 2.02405 0.7941673 -
0.02087018
```

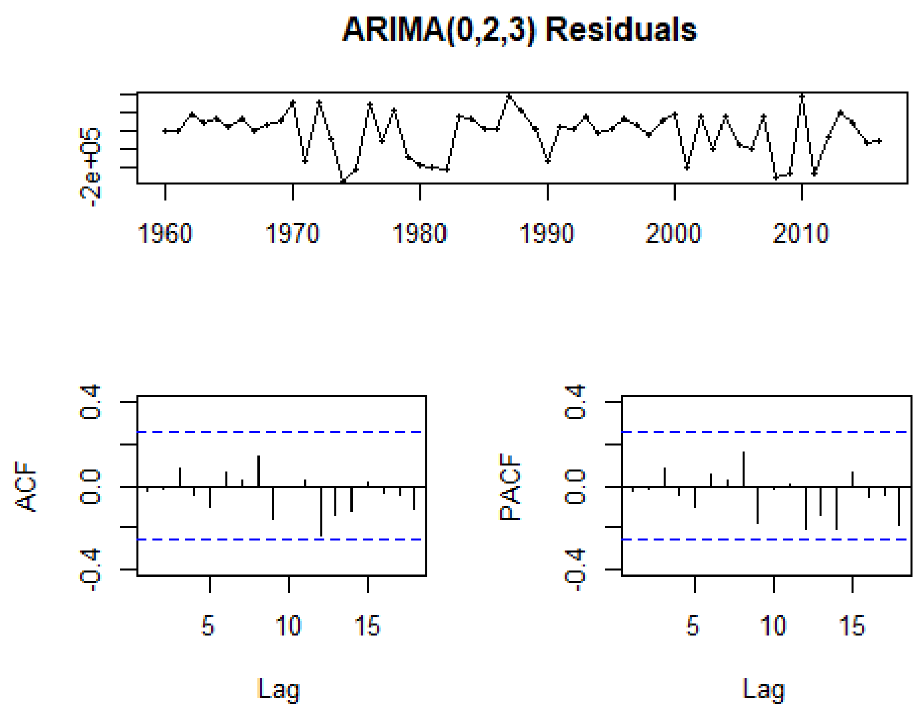

Figure 5: ARIMA(0,2,3) Residuals and Stability Test

```
##
##  Box-Ljung test
##
## data:  arima1$residuals
## X-squared = 0.026157, df = 1, p-value = 0.8715
```

It is possible to observe in this test that the p-value is >0.05 indicating that the residuals are only white noise.

The residuals are within the box and close to zero, so it could be stated that the series are not autocorrelated.

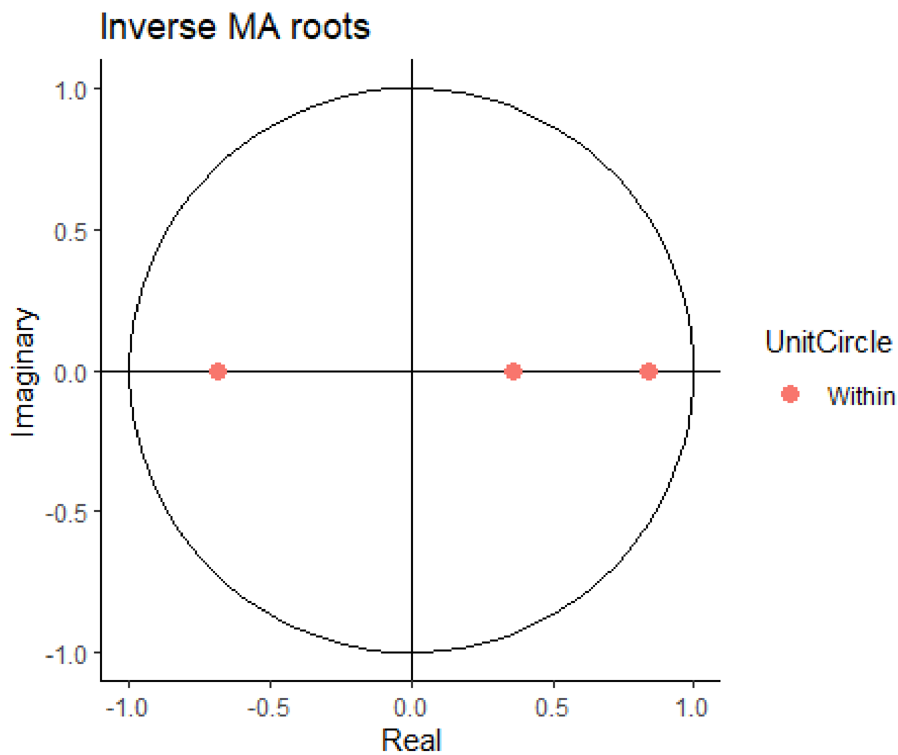

Figure 6: Inverse of moving average (MA) roots

Descriptive Statistics

Mean	*Median*	*Minimum*	*Maximum*	*Sd*	*Skewness*	*Kurtosis*
4684616	4844056	2823484	5732999	776154.9	-0.8520843	3.074317

In the table above it is possible to access the time series summary.

The mean is positive and non-zero. The difference between minimum (2823484 kt) and maximum (5732999 kt) corroborates the increase of emissions in the US from 1960 to the mid-2000s when environmental issues started to be more debated.

The negative skewness and kurtosis of 3.074317 indicate that the time series is not normally distributed.

Forecast for one period ahead (2020)

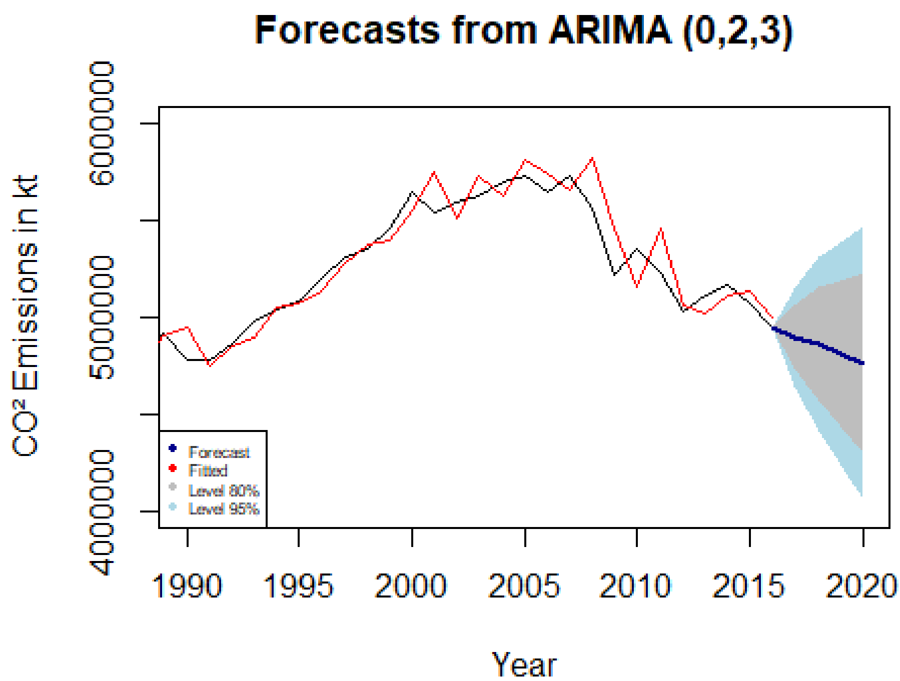

Figure 7: Forecast short-term using ARIMA(0,2,3)

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2017	4898258	4735647	5060869	4649566	5146950
2018	4860040	4568186	5151893	4413688	5306391
2019	4811480	4435318	5187642	4236190	5386770
2020	4762921	4302241	5223601	4058371	5467471

Forecast long term (2030)

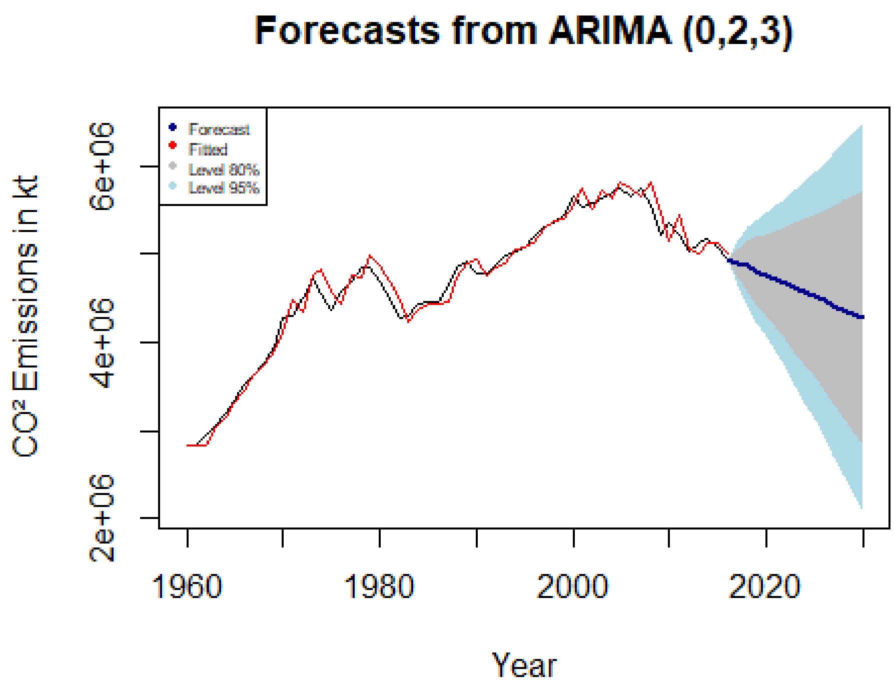

Figure 8: Forecast long term using ARIMA(0,2,3)

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2017	4898258	4735647	5060869	4649566	5146950
2018	4860040	4568186	5151893	4413688	5306391
2019	4811480	4435318	5187642	4236190	5386770
2020	4762921	4302241	5223601	4058371	5467471
2021	4714362	4167549	5261174	3878084	5550639
2022	4665802	4030609	5300995	3694359	5637246
2023	4617243	3891114	5343373	3506724	5727762
2024	4568684	3748912	5388455	3314951	5822416
2025	4520124	3603940	5436309	3118941	5921308
2026	4471565	3456179	5486951	2918666	6024465
2027	4423006	3305638	5540374	2714139	6131873
2028	4374446	3152342	5596550	2505400	6243493
2029	4325887	2996327	5655448	2292500	6359274
2030	4277328	2837630	5717025	2075501	6479155

Evaluating ARIMA and Holt-Winters Exponential Smoothing

Another forecasting model is Holt-Winters Exponential Smoothing, as it is also an effective model in adapting linearity in time series (Da Veiga et al., 2014).

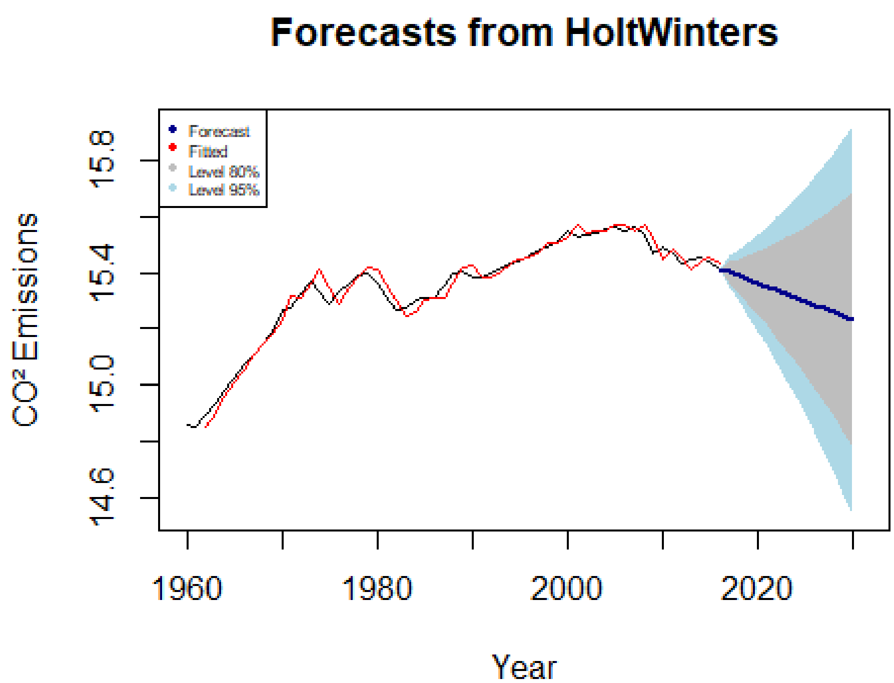

Figure 9: Forecast long term using Holt-Winters model

```
##
##  Box-Ljung test
##
## data:  forecast
## X-squared = 8.322, df = 1, p-value = 0.003917

##
##  Box-Ljung test
##
## data:  hwfpred
## X-squared = 8.2313, df = 1, p-value = 0.004117
```

Here it can be seen that the difference between ARIMA(0,2,3) and Holt-Winters is not very significant, however, the Box-Pierce test points out that the ARIMA(0,2,3) model has a higher validity as the p-value is closer to 0.

Holt-Winters model is a recommended method when the time series is seasonal and thus not recommended when the forecast exceeds the seasonality of the time series (Da Veiga et al., 2014). Hence, the ARIMA(0,2,3) model is more effective for the type of time series forecasting in this paper.

Conclusion and Political Implications

Despite the projection showing a decrease in CO^2 emissions, measures remain insufficient to achieve carbon neutrality in the long term, although there could be an attainment of the 50–52% objective set by the US as of 2005. However, the President of the United States has now approved the “Willow Project” in March 2023, which allows for the construction of 5 drill pads for a total of 250 oil wells in Alaska (Land Management, 2023), thereby reducing the country’s dependence on other nations for sources of energy.

This project not only goes against efforts to reduce CO^2 emissions, but it also poses risks to native Americans, wildlife and waters in the region. According to (Nong et al., 2018), there are different risks associated with oil spills that could harm local species, reduce food supplies, and contribute to the melting of glaciers. The long-term goal of this project is to produce approximately 200,000 barrels of oil per day over a 30-year period. There would be approximately 600 million barrels of oil and 287 million tonnes of CO^2 emissions by the end of this forecasted period (Land Management, 2023).

Thereby, based solely on what is shown in Figure 8 it is recommended:

a) endorsement of the importance of education in renewable energy and the construction of renewable energy grids such as solar panels;

b) a “carbon tax” for industries that emit CO^2 through polluting agents on a daily basis, as a mandatory incentive to change energy sources. This measure has already been implemented in some US states, but it is not a national requirement;

c) more importantly, the recognition that current oil exploration plans, such as the Willow Project, will drastically change the forecast for CO^2 emissions.

While the ARIMA(0,2,3) model has been predicted to be efficient and stable and to yield positive results in the future, it cannot be applied in multivariate time series, i.e. it is not possible to use ARIMA(p,d,q) to predict CO^2 emissions and other gases simultaneously or the possible increase of fossil fuel emissions due to the new project.

Furthermore, CO^2 emissions control remains heavily dependent on policy and economic measures that unfortunately cannot be predicted, particularly when it comes to a country that is very volatile in its decisions concerning the environment. The United States’ pursuit of energy independence through the Willow Project may result in long-term consequences that could be irreversible.

References

Da Veiga, C. P., Da Veiga, C. R. P., Catapan, A., Tortato, U., & Da Silva, W. V. (2014). Demand forecasting in food retail: A comparison between the holt-winters and ARIMA models. *WSEAS Transactions on Business and Economics*, *11*(1), 608–614.

Dare, J., Patrick Olabanji, A., & Oyewola, D. (2022). Comparison of stationarity on ljung box test statistics for forecasting. *Earthline Journal of Mathematical Sciences*, 325–336. doi: 10.34198/ejms.8222.325336

Dobrescu, E. (2009). Reducing the polluting emissions. A source for diminishing the climate changes. *Romanian Journal of Economics*, *29*, 107–121.

Emissions, C. W. H. G. (2022). *Climate watch*. Retrieved from https://www.climatewatchdata.org/ghg-emissions

Freeman, D. (2019). *The EU-US-china triangles and the paris agreement* (pp. 174–191). doi: 10.4324/9780429026201-10

House, T. W. (2021). *Executive order on tackling the climate crisis at home and abroad*. Retrieved from https://www.whitehouse.gov/briefing-room/presidential-actions/2021/01/27/executive-order-on-tackling-the-climate-crisis-at-home-and-abroad/

Jena, P. R., Managi, S., & Majhi, B. (2021). Forecasting the CO2 emissions at the global level: A multilayer artificial neural network modelling. *Energies*, *14*(19). doi: 10.3390/en14196336

Jones, P., Wigley, T., Folland, C., Parker, D., Angell, J., Lebedeff, S., & Hansen, J. (1988). Evidence for global warming in the past decade. *Nature*, *332*, 790–790.

Land Management, B. of. (2023). *Willow master development plan*. doi: DOI-BLM-AK-0000-2018-0004-EIS BLM/AK/PL-22/032+1610+F010

Musbah, H., Aly, H., & Little, T. (2023). A proposed novel adaptive DC technique for non-stationary data removal. *Heliyon*, *9*, e13903. doi: 10.1016/j.heliyon.2023.e13903

Nong, D., Countryman, A. M., & Warziniack, T. (2018). Potential impacts of expanded arctic alaska energy resource extraction on US energy sectors. *Energy Policy*, *119*, 574–584. doi: https://doi.org/10.1016/j.enpol.2018.05.003

Nyoni, T., & Mutongi, C. (2019). *Modeling and forecasting carbon dioxide emissions in china using autoregressive integrated moving average (ARIMA) models*.

Stocker, T. (2014). *Climate change 2013: The physical science basis: Working group contribution to the fifth assessment report of the intergovernmental panel on climate change*. Cambridge university press.

Zhang, Y., & Meng, G. (2023). Simulation of an adaptive model based on AIC and BIC ARIMA predictions. *Journal of Physics: Conference Series*, *2449*, 012027. doi: 10.1088/1742-6596/2449/1/012027