Term Paper for
Time Series Analysis

Forecasting Carbon Dioxide Emissions of China, the United States, and the European Union: Comparing ARIMA and Naïve Predictions

Maximilian Hildebrandt

Berlin, August 31st, 2021

Instructor: Prof. Dr. Simone Maxand, Chair of Econometrics, Humboldt University of Berlin

Module: Time Series Analysis

Summer term 2021

Contact:

Maximilian Hildebrandt

Einbecker Str. 78

10315 Berlin

Email: maximilian.hildebrandt@hu-berlin.de

Degree program: M.Sc. Statistics

Matriculation number: 616088

Contents

1. Introduction

Changes in the earth climate have become apparent in many regions across the world (IPCC, 2021). For example, scientists claim that the permafrost melting will increase, with a thawing of glaciers and ice sheets (IPCC, 2021). Also, oceans are changing, with acid levels in the seas rising and oxygen levels decreasing, which negatively affect the species inhabiting the sea. Also, sea levels are projected to rise, which are expected to severely disrupt low-lying areas (IPCC, 2021). Furthermore, water cycles are expected to change, with more intense rainfalls and flooding catastrophes, as recently observed in Rhineland-Palatinate and North Rhine-Westphalia, Germany (FAZ, 2021). Scientists argue that some of these changes might exceed critical climate thresholds, which might lead to irreversible changes (e.g., Artic Sea melt; Banarjee, 2015). The drastic temperature change that was observed is likely to exceed 1.5°C of warming over the next 20 years (IPCC, 2021), intensifying many of the problems observed in the present.

One of the primary levers to deaccelerate climate change is shrinking global CO_2 emissions. Indeed, scientists attribute the rise in temperatures to well-mixed greenhouse gas emissions. In radiative forcing studies, carbon dioxide emissions, and to a lesser extend non-CO_2 greenhouse gases (e.g., Methane, halogenated gases etc.) appear to be the primary driving forces of the drastic increase in temperatures (IPCC, 2021). The top three emitters by annual CO_2 emissions in 2017 are China (27% of global emissions), the United States (US; 15%), and the 27 member states of the European Union (EU-27; 10%), accounting for more than 50% of global CO_2 emissions in 2017 (Ritchie & Roser, 2020).

Recognizing their influence in driving global climate change, countries across the globe have committed to reducing CO_2 emissions in a series of agreements. The most recent agreement has been the Paris Agreement (2015), which was adopted by 196 parties in Paris, on December 12th, 2021. Its stated aim is to limit global warming to well below 2°C, preferably 1.5°C above pre-industrial levels. It represents a binding agreement to combat climate change, and as such was a significant step forward in multilateral relationships.

As indicated by the outlined consequences and the political debate on interventions to tackle climate change, there is a strong need of policymakers to monitor the current status of the CO_2 emissions as well as to where it is going. Forecasting methods can serve as guidance for policymakers in providing such a transparency and to support the development of effective policy development (Stock & Watson, 2017). Methods for time series data have been developed intensively in the past decades (Stock & Watson, 2017), and now leverage improved computational power, data availability, and more rigorous theory (Stock & Watson, 2017). Opponents of climate interventions have cited the high cost of carbon reductions to refrain from joining the global forces in addressing the climate issue (Auffhammer & Steinhäuser, 2012).

Having adequate baseline forecasts for U.S. CO_2 emissions is of crucial importance, since they are used to calculate the expected cost of emission reductions in future periods (Auffhammer & Steinhauser, 2012) or can serve as an important input for climate risk management strategies (Keller & Nicholas, 2015).

In line with this need, this paper aims to model and forecast CO_2 emissions for the top three CO_2 emitters, both for one year ahead and for 2030 and by deriving concrete policy implications. The next section will elaborate on the theoretical background and formalizes the technical problem that is to be addressed in this term paper.

2. Theoretical Background

This section is divided into three parts to describe the theoretical background of the time series analysis. First, relevant climate policies on the multilateral level as well as for China, the US, and the EU will be reviewed. Second, the empiric literature on CO_2 forecasting will be briefly reviewed. Third, the technical background on time series analysis, *autoregressive integrated moving average* (ARIMA) models, the baseline model and relevant residual as well as diagnostic checks are outlined, before concluding with specifying the technical problem.

2.1 Climate Policy Background

The emergence of climate change as a policy problem has been a rather recent phenomenon, with the first global alignment in 1979, on the *World Climate Conference* (Gupta, 2010). In 1988, the *International Panel on Climate Change* (*IPCC*) was established to synthesize scientific findings on climate change and to provide a foundation for any climate change-centred political conversations. Subsequently, the IPCC published a series of assessment reports, with the first assessment report being published in 1990, which suggested greenhouse gas emissions to be maintained at the 1990 level. An additional major milestone was achieved in 1992, when United Nations *Framework Convention on Climate Change* (FCCC) was signed by 154 parties in Rio de Janeiro, thereby establishing an overarching objective to stabilize the climate and formulating a series of shared principles and commitments (United Nations, 1992). The signing parties committed themselves to monitor and report their greenhouse gas emissions and to take precautionary measures to prevent causes of climate change (United Nations, 1992).

One of the key weaknesses of this convention were the weakly formulated emission targets for 2000 in comparison with 1990. This was addressed partially by the *Kyoto Protocol*, which was adopted in December 1997. Complementing the FCCC, it laid out a series of policies and measures, from which countries can choose the most appropriate ones (e.g., energy efficiency policies; Gupta, 2010). Developed countries committed to jointly reduce their total emissions of six greenhouse gas emissions by 5.2% during the period of 2008-2012. In addition, the protocol established a series of five mechanisms, such as the *emissions trading scheme* (*ETS*) to enable international trade of emission certificates (Gupta, 2010). The effectiveness of the protocol was impaired when the US decided to step back from the protocol. This was negatively perceived by developing countries, and impaired implementation (Gupta, 2010).

With the implementation of the Kyoto commitment period in the close future, there was need to re-ignite climate change efforts by passing a novel legally binding agreement. However, the key meeting in Copenhagen in 2009 led to ambivalent results, generally falling short of expectations (Gupta, 2010). In the informal *Copenhagen accord*, which was formally passed in the *Cancun Agreement* one year later, countries for the first time agreed upon a 2°C objective. Furthermore, developing countries now also committed to monitor and publish their emissions in regular intervals and both developed as well as undeveloped countries formulated quantified emission reduction targets for 2020 (Gupta, 2010; Lefevere, 2015).

The most recent breakthrough was achieved in Paris in 2015, when parties strengthened the previously stated objective of limiting temperature increase, now stating to keep the increase well below 2°C, preferably to 1.5°C (United Nations, 2015). At the core of the *Paris Agreement* are *nationally determined contributions* (*NDCs*), which countries formulate independently, and which represent national plans and targets for greenhouse gas emission reductions (Schleussner et al., 2016). Parties committed to update these NDCs every five years. In addition, parties agreed to periodically assess the collective progress. The agreement faced a severe impairment, when the Trump administration decided to leave the agreement on June 1st, 2017.

Next to these international global agreements, countries or economic unions also issue independent policy initiatives. Recently, the European Commission published its set of policy initiatives to achieve climate neutrality in 2050, the so-called *European Green Deal* (European Commission, 2020), with implementation measures being proposed, but not yet passed, such as a stricter and expanded ETS (European Union, n.d.). Based on the Paris agreement, the EU aims to cut emissions by 55% compared to 1990 levels by 2030. The Chinese government has also stated clear climate objectives to stay consistent with the 1.5°C objective. In September 2020, the Chinese president Xi Jinping announced the objective to achieve climate neutrality in 2060 (Ministry of Foreign Affairs of the People's Republic of China, 2020). In its current 5-year plan, the Chinese government states it wants to achieve peak carbon emissions by 2030 (Xinhua News Agency, 2021). The United States has also issued new targets for CO_2 emission reductions. In January 2021, the US government announced to re-join the Paris Agreement. In April 2021, it announced a new target of 50-52% CO_2 emission reductions by 2030 as compared with the 2005 levels (The White House, 2021). In addition, the administration announced its aim of net-zero emissions in 2050.

2.2 CO_2 Emission Forecasting

Many studies have applied statistical forecasting methods to CO_2 emissions, although ARIMA models are only rarely the main technique to generate forecasts. For example, Dritsaki and Dritsaki (2020) used an ARIMA(1,1,1)-autoregressive conditional heteroscedasticity (ARCH) (1) model to predict CO_2 emissions per capita in the EU. Their forecast predicts 5.678 tons of per capita emissions in 2020. In another study, Pao et al. (2012) built models to forecast carbon emissions, energy consumption, and real outputs using a *nonlinear grey Bernoulli model* (NGBM), comparing its performance to grey models and ARIMA models. The NGBM models achieved a strong er performance than the ARIMA baseline model, as measured by MAPE. The main model forecasted 12.46 billion tons of CO_2 emission for 2020. Furthermore, Nyoni and Mutongi (2019) used an (1, 2, 1) model, predicting 10.011.298 kt emissions for 2024. Li et al. (2020) compared the ARIMA model to three-grey system-based models for forecasting of CO_2 emission intensity (i.e., the amount of emissions per unit GDP). The ARIMA (1, 2, 4) model showed the smallest MAPE across all models. They predict a CO_2 emission intensity of 0.3711 tons per 10 thousand RMB. In addition, Auffhammer and Steinhauser (2012) used reduced form models to predict CO_2 emissions for the US using state-level information. In contrast with ARIMA models, they used predictors such as income per capita, population density, along with other variables to create a large model universe of over 27000 models, with the best model predicting 1.689 million tons of carbon for 2011. Moreover, Bennedson et al. (2021) used a structural augmented dynamic model for US CO_2 emissions, researching relevant predictors from a large set of 226 macroeconomic indicators using automatic variable selection. They found that industrial production indices best explain emissions, predicting a 2019 emission change of minus 1%. Their model had a lower *root mean square error* (RMSE) and *mean average error* (MAE), when compared in a one-step ahead forecast with an (1,1) ARMA baseline model.

2.3 Methodological Background

In the following sections, the methodological background of this paper is outlined.

2.3.1 ARIMA models

The ARIMA model is one of the methods put forward by Box and Jenkins (1970) to model time series data and to generate forecasts. It is widely employed in econometrics and aims to describe the autocorrelations in the data (Hyndman & Athanasopoulos, 2021). The ARIMA model (*p, d, q*) is defined as follows (adapted from Hamilton, 1994):

$$\left(1-\sum_{i=1}^{p}\phi_i L^i\right)(1-L)^d y_t = c + \left(1+\sum_{i=1}^{q}\theta_i L^q\right)\varepsilon_t \tag{1}$$

This general process has three parameters. The first parameter *p* defines the number of autoregressive lags, the second parameter *d* specifies the order of integration, and the last parameter *q* equals the number of moving average lags (Hamilton, 1994). y_t is the variable of interest, measured at time *t*, ε_t is the random error at point in time t, c is the transformed mean of the model $(1-L)^p y_t$ (Hyndman & Athanasopoulos, 2021). L denotes the lag operator such that the first difference operator is given by $\Delta y_t = (1-L)y_t = y_t - y_{t-1}$.

2.3.2 Maximum-likelihood estimation

The primary estimation method for ARIMA models is *maximum likelihood estimation* (Hamilton, 1994). The population parameters are denoted by the vector $\boldsymbol{\theta} = (c, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \sigma^2)'$.

For this estimation purpose, let *T* be the sample size, so that $\boldsymbol{y_t} = y_1, \dots, y_T$ represents the sample. For the estimation, the joint density function is represented as

$$f_{Y_T, Y_{T-1}, \dots, Y_1}(Y_T, Y_{T-1}, \dots, Y_1; \boldsymbol{\theta}) \tag{2}$$

The maximum likelihood estimate (MLE) $\widehat{\boldsymbol{\theta}}$ equals the values of the parameters which maximise the probability of obtaining the data that we have observed (Hamilton, 1994). The approach relies heavily on an appropriate construction of the random error process ε_t. Commonly, one would assume a Gaussian white noise:

$$\varepsilon_t \sim i.i.d.\, N(0, \sigma^2) \tag{3}$$

Mathematically, conducting the maximum likelihood estimation involves two steps (Hamilton, 1994). First, the *likelihood function* as a product of the density functions needs to be computed. Second, the MLE is obtained by globally maximizing the likelihood function with respect to the parameter vector $\boldsymbol{\theta}$. Maximizing the likelihood function is equivalent to maximizing the *log-likelihood function* (because log-transformation is a monotone transformation).

To generate the relevant log likelihood function requires conditioning on both y's and ε's. As recommended by Hamilton (1994), values for both are initiated, such that $\mathbf{y_0} = (y_0, y_{-1}, \dots, y_{-p+1})$ and $\boldsymbol{\varepsilon_0} = (\varepsilon_0, \varepsilon_0, \dots, \varepsilon_0)$. Consequently, under the Gaussian error process, the *conditional log likelihood function* is given by (Hamilton, 1994):

$$\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{y_t}) = -\frac{T}{2} log(2\pi) - \frac{T}{2} log(\sigma^2) - \sum_{t=1}^{T} \frac{\varepsilon_t^2}{\sigma^2} \tag{4}$$

2.3.3 Baseline model

The ARIMA model performance will be compared to a simple baseline model to test if the model provides a significant improvement in forecasting performance on out-of-sample data. A *naïve forecast* is used as a baseline model, since it is considered quite effective despite its simplicity and computational efficiency (Hyndman & Athanasopoulos, 2021). A naïve forecast sets all forecasted values to the value of the last observation (Hyndman & Athanasopoulos, 2021):

$$\hat{y}_{T+h|T} = y_T \tag{5}$$

2.3.4 Stationarity assessment, evaluation metrics & diagnostic tests

Thorough application of the ARIMA model requires assessing the stationarity of the time series, selecting the appropriate model, and conducting residual checks to validate the model. Time series stationarity was assessed through application of the *augmented Dicker-Fuller-test* (ADF; Dickey & Fuller, 1979; Dickey & Fuller, 1981; Said & Dickey, 1984). One of the key technical decisions in ARIMA time series analysis is the choice of the parameters *p, d*, and *q* (i.e., *lag length selection*). The appropriate parameters are identified using *autocorrelation functions* (ACFs), *partial autocorrelation functions* (PACFs) and two in-sample metrics displayed in *Table 1* (Akaike; 1973; Schwarz, 1978):

Table 1.
In-sample criteria used for ARIMA model selection

Criterion	Formula
1. Akaike information criterion (*AIC*)	$\log(\hat{\sigma}_{\varepsilon}^2) + 2\frac{p+q+1}{T}$
2. Bayesian information criterion (*BIC*)	$\log(\hat{\sigma}_{\varepsilon}^2) + \frac{p+q+1}{T}\log(T)$

The BIC differs from the AIC in that it selects more parsimonious models (Hyndman & Athanasopoulos, 2021). After having identified the best performing models according to in-sample characteristics, it is necessary to test if the model is correctly specified. For that reason, residuals are examined using diagnostic tests. Residual tests are used to confirm that the model assumptions are met (i.e., the residuals represent white noise). Specifically, serial correlation is tested using *Ljung-Box Portmanteau test* (1978), conditional heteroscedasticity is examined using the *McLeod-Li* test statistic (1983) for the squared residuals, and the *Jarcque-Bera test* (1980) is used to assess the normality of the residuals. Additionally, residuals are plotted visually and the *auto-correlation function* (*ACF*) for the residuals and squared residuals are examined to further test the model assumptions.

Furthermore, it was tested if the main model outperforms the baseline model on test data using the *Diebold-Mariano* test (1995). In addition, the metrics in *table 2* were computed on out of sample data (Hyndman & Köhler, 2006; Hyndman & Athanasopoulos, 2021). These metrics are used to compare the main model with the baseline model and to quantify the forecasting performance on the test set.

Table 2.
Forecast evaluation metrics

Criterion	Formula
Root mean squared error (*RMSE*)	$\sqrt{\frac{1}{K}\sum_{k=1}^{K} e_k^2}$
Mean absolute deviation (*MAD*)	$\frac{1}{K}\sum_{k=1}^{K} \lvert e_k \rvert$
Mean absolute percentage error (*MAPE*)	$\frac{1}{K}\sum_{k=1}^{K} \left\lvert 100\frac{e_k}{y_k} \right\rvert$
Mean absolute scaled error (*MASE*)	$\frac{1}{K}\sum_{k=1}^{K} \frac{\lvert e_k \rvert}{\frac{1}{T-1}\sum_{t=2}^{T} \lvert y_t - y_{t-1} \rvert}$

Note. e_k denotes the forecast errors in the time periods 1, … *k*, for which the forecasts have been computed.

Next, the technical problem will be detailed.

2.3.5 Technical problem specification

The objective of this paper is to generate and evaluate appropriate forecasts for the CO_2 emissions for the EU, China, the US, and to derive policy implications based on these forecasts. Technically speaking, the CO_2 emissions of the three countries or unions can be considered realizations of an unknown stochastic process. We have a sample size of $T = 57$ observations of three random variables X_t, Y_t, Z_t for the emissions of the EU, China, and the US, respectively:

$$\{x_1, x_2, \dots, x_{57}\} \tag{6}$$

$$\{y_1, y_2, \dots, y_{57}\} \tag{7}$$

$$\{z_1, z_2, \dots, z_{57}\} \tag{8}$$

The technical objective of this paper is to create three separate ARIMA models that sufficiently capture the structure of the respective stochastic processes. Specifically, this involves determining the appropriate parameters *p, d, q* of the ARIMA models and estimating the parameter $\boldsymbol{\theta} = (c, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \sigma^2)'$ for each of the three models. In addition, the one-step ahead forecasted values as well as the forecast for 2030 will be computed:

$$X_{2017|T} := E_T(X_{2017}) \tag{9}$$

$$Y_{2017|T} := E_T(Y_{2017}) \tag{10}$$

$$Z_{2017|T} := E_T(Z_{2017}) \tag{11}$$

Where *T* will be the last time value of the training dataset.

$$X_{2030|T} := E_T(X_{2030}) \tag{12}$$

$$Y_{2030|T} := E_T(Y_{2030}) \tag{13}$$

$$Z_{2030|T} := E_T(Z_{2030}) \tag{14}$$

The problem can be considered solved appropriately, if the transformations to the data as well as the model parameters are chosen in such a way, that the model residuals resemble Gaussian white noise and that the forecast can appropriately forecast out-of-sample observations of the CO_2 emissions. Generally, the aim of the forecast modeling procedure is to minimize the defined loss functions in Table 1. More specifically, the forecast is considered appropriate, if it performs better than the baseline model on the test data, as quantified by the in-sample metrics displayed in Table 2 (e.g., RMSE) and supported by the Diebold-Mariano test (1995).

3. Methods

In this section, all methodological aspects of the term paper will be presented. First, the data sample will be described before the measures will be addressed. Eventually, the analytical strategy will be outlined.

3.1 Sample

The emission dataset was downloaded as a .csv file on August 20th, 2021 on the World Bank data server (World Bank, 2021). The raw dataset contains annual data from 1960 to 2018 for total greenhouse gas emissions by individual countries as well as country sets. The data was filtered for the time period 1970 to 2016 (T = 57 years) and for China, the United States, and European Union data. No data was excluded from the analysis and there were no missing values.

3.2 Measures

In this section, the key dimensions of the term paper dataset are described. The dataset contains data for *annual* CO_2 *emissions*, broken down into separate time series for *China*, the *United States*, and *European Union.* The unit of measurement is *kt* (kiloton). Carbon dioxide emissions are those stemming from the burning of fossil fuels and the manufacture of cement. They include carbon dioxide produced during consumption of solid, liquid, and gas fuels and gas flaring (World Bank, 2021). The emission data excludes emissions from land use such as deforestation (World Bank, 2021). No control variables are included in the dataset.

3.3 Analytical Strategy

The data was analyzed with the *R version 4.1.4.* software environment for statistical computing (R Core Team, 2021). The R software package *forecast (version 8.15)* was used to compute the time series analyses (Hyndman & Khandakar, 2008).

Overall, eight steps were used in deriving the analysis results. First, the time series data was explored to assess the statistical properties of the time series data using descriptive statistics and statistical charts. During this step, it was assessed if the series contains any irregular changes that require further inspection or obvious trends. Due to its annual periodicity, the time series contains no seasonality. Second, the data was pre-processed to enable a sound time series analysis. Due to heteroscedasticity, the logarithm was applied to the EU and the China CO_2 emission data. For evaluation purposes, the data was split into a *training* (*in-sample*) and *test* (*out-of-sample*) dataset (Hyndman & Athanasopoulos, 2021). The test set size was determined based on forecast length. It is recommended to have a test set size that is at least as large as the forecast horizon (Hyndman & Athanasopoulos, 2021). Given the forecasting target for 2030, the test set size is $T_{test} = 13$. Consequently, the training dataset size is $T_{train} = 44$. Models will be fitted to the training dataset and forecasts will be compared on the test set. The final predictions for the one-step ahead forecast and the 2030 forecast are done using the model trained on the training data, and then forecasting the period of 2017-2030. Third, it was assessed if the series is non-stationary by using visual plots and unit root tests. Specifically, it was assessed if the non-stationarity in the data is due to a deterministic time trends or a unit root using the *Augmented Dickey-Fuller test* (Dickey & Fuller, 1979). Afterwards, the data was differenced once for the China and US data. Second order differences were computed for the EU time series to achieve stationarity. Fourth, a modeling approach was chosen. Selecting an appropriate model is necessary in order to produce appropriate forecasting results (Hyndman & Athanasopoulos, 2021). For the univariate time series data, the main model was chosen to be an ARIMA model, since these models can be used to model non-stationary univariate time series data (Hamilton, 1994). In addition, a *baseline model* was included to test if the ARIMA is superior in predicting the CO_2 emissions. The *naïve forecast* was selected as a baseline model, since it is a simple procedure that can act as a comparison for more nuanced methods (Hyndman & Athanasopoulos, 2021). Fifth, the model parameters p and q were based on selecting the most parsimonious model with the lowest *AIC* and *BIC*. In comparison with the AIC, the BIC selects more parsimonious models. Both metrics implied the same models, requiring no decision to choose between the metrics. Sixth, the best performing models per time series were fitted using the maximum likelihood algorithm. Specifically, the parameters $c, \phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q$ were computed to obtain the fitted ARIMA models. Seventh, model residuals were estimated to assess if the residuals resemble a white noise process. Residuals are tested for the following properties: (1) Residuals are uncorrelated. (2) residuals have a mean of zero to ensure unbiasedness, (3) residuals have equal variance (i.e., homoscedasticity), (4) residuals are normally distributed. (3) and (4) ensure valid prediction intervals. ACFs of the residuals are used to test for uncorrelatedness. Furthermore, the Ljung-Box Portmanteau test (1978) is used to also assess uncorrelatedness. The mean level is assessed visually in residual plots. Remaining conditional heteroscedasticity is tested by ACFs using the squared residuals as well as the McLeod-Li test (1983). Normality of residuals is assessed by applying the Jarque-Bera test (1980) for normality using standardized residuals. Eighth, forecasts for the different countries or country sets are computed. Out-of-sample metrics are calculated to quantify and compare the forecast performance. Models are compared using the *Mariano-Diebold* test (1995), assessing the alternative hypothesis that the ARIMA models perform better than the baseline model. Finally, the final forecasts are generated through the model trained on the training data forecasting the period from 2017-2030. *Fan plots* using 30%, 60%, and 90% prediction intervals are computed to quantify the uncertainty of the forecasts.

4. Results

Table 4

Descriptive statistics for annual CO_2 emissions per country or economic and political union

Country	*M*	*SD*	*CV*	*Min* (year)	*Max* (year)
European Union	3263529	564107	17.29%	1687911 (1960)	4114434 (1979)
China	3234343	2969449	91.81%	433234 (1967)	9936680 (2013)
United States	4727344	776677	16.43%	2880506 (1961)	5776410 (2000)

Note. *M* = sample mean, *SD* = standard deviation, *CV* = coefficient of variation

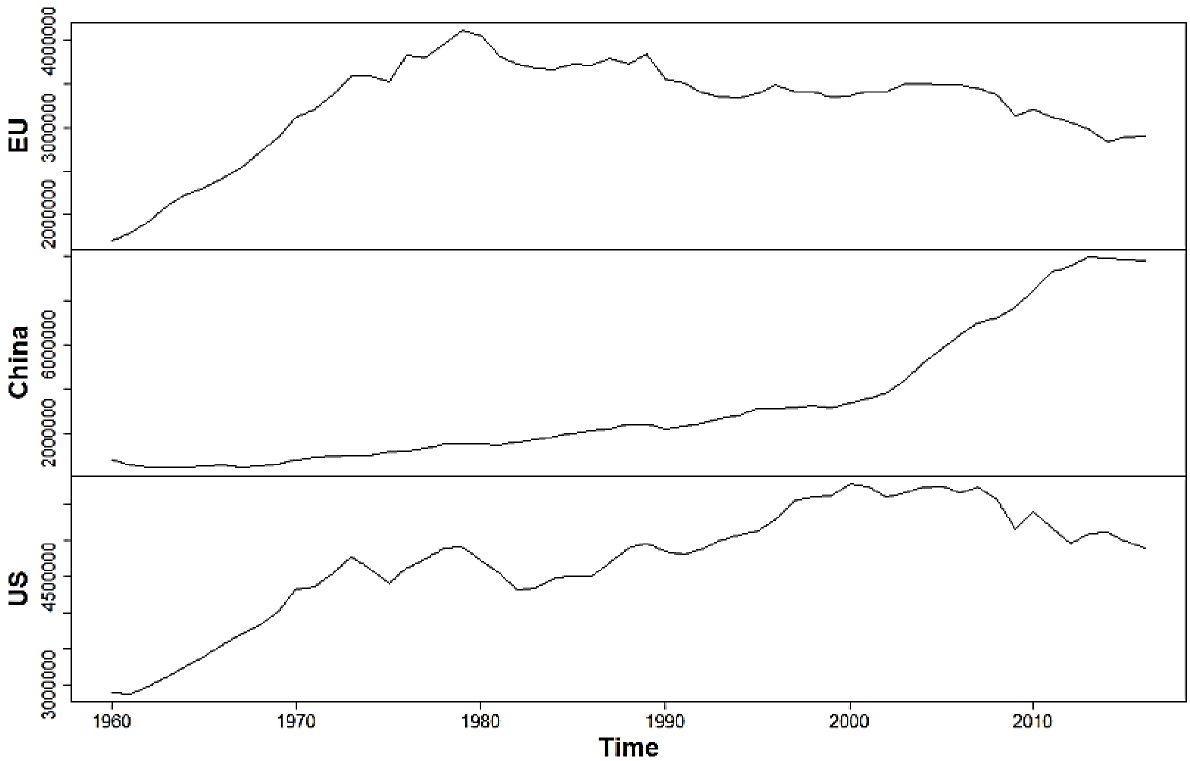

Figure 1. Amount of carbon dioxide emissions (kilo tons) over 1960-2016 in EU-27, China, and US. Note that the scale is unstandardized to provide a better view on the variation per time series.

The descriptive statistics for the time series data are shown in *Table 4*. *Figure 1* displays the annual CO_2 emissions over the whole time period of 1960 to 2016. The emissions in the EU are characterized by a period of substantial growth until around 1980, with a slow decline afterwards. In contrast, the Chinese emissions started at a substantially lower level in 1960, followed by a slowly increasing growth trend until around 2002, after which emissions grew strongly, even surpassing US emissions in 2005. After hitting their peak in 2013, emissions have remained at an approximately stable level until 2016. Starting with a higher base of emissions in 1960 than the other two countries or unions, the CO_2 emissions of the US also show a period of approximately steady growth until 1980, followed by a decline in emissions in the following three years. Up until 2000, emissions have risen steadily, after which they stayed at a similar level until 2008, and then started declining.

In the sections that follow, results for time series stationarity tests are described, before selecting the most appropriate among competing models. Afterwards, the best models are fitted to the time series and the persistence of the models is tested. Finally, forecasting results are presented.

4.1 Stationarity-Tests

Table 5

Augmented Dickey-Fuller (ADF) unit root test results

	Constant without trend			Constant with trend		
CO_2 emitter	Levels	1st diff	2nd diff	Levels	1st diff	2nd diff
EU (log)	-2.12 (1)	-2.95** (1)	-5.97***	-3.04 (0)	-3.03 (0)	-5.97***
China (log)	-1.47 (1)	-4.49*** (1)	-	-2.45 (0)	-4.53*** (0)	-
US	-.56 (1)	-4.14*** (1)	-	-2.08 (0)	-4.11** (1)	-

Note. *** p ≤.01 ** p ≤ .05; * p ≤ .10; the optimal lag length was determined using the Bayesian information criterion (BIC).

Test results of the ADF are displayed in *Table 5*. Based on the results, stationarity is implied for the 1st order differences for the log-transformed China data and the US data, already being achieved with a constant, indicating no need to include a trend term to achieve stationarity. On an α-level of 5%, the null hypothesis of unit root is rejected for the EU data. Consequently, results of the ADF indicate that the first order differences of the EU emission data are stationary without a deterministic trend. However, when assessing the time series plot, there is still some negative trend visible (see appendix), consequently the second order differences used, which show strong signs of stationarity. ACF/PACF plots for the three time series are included in the appendix and interpreted based on Nau (2014). The ACF for EU CO_2 emission data shows many significant autocorrelations, supporting the decision to apply differencing at least once. The ACF for the second order differences shows a sharp cutoff and the lag-1 autocorrelation is negative, consequently an MA term of lag 1 is indicated. The PACF plot for the 2nd order differences shows a few significant autocorrelations. Similar to the EU data, level data for Chinese emissions shows a high number of significant lags, indicating the need for further differencing. The PACF shows a sharp cutoff and a negative lag-1, indicating an MA signature. Given the significant first lag in the ACF, an additional MA term could potentially further enhance the model. US level data shows a high number of significant autocorrelations, thereby also supporting differencing of at least order one. The ACF plot for the first order differences of US CO_2 emissions shows a sharp cutoff after lag one, indicating the need for an MA term of lag 1. The PACF shows no sharp cutoff, but a significant first lag, indicating that an additional AR term might be viable if the AR and MA term don't cancel each other out.

4.2 Model Selection

Table 6

Comparison of competing ARIMA models using AIC and BIC for the different time series

EU (without constant)			China (without constant)			US (with constant)		
Model	AIC	BIC	Model	AIC	BIC	Model	AIC	BIC
(2, 2, 2)	-157.48	-148.91	(2, 1, 2)	-71.25	-62.56	(2, 1, 2)	1104.33	1114.75
(0, 2, 0)	-146.99	-145.28	(0, 1, 0)	-63.91	-62.17	(0, 1, 0)	1113.70	1117.18
(1, 2, 0)	-158.18	-154.75	**(1, 1, 0)**	**-76.96**	**-73.48**	(1, 1, 0)	1108.72	1113.93
(0, 2, 1)	**-163.06**	**-159.63**	(0, 1, 1)	-73.96	-70.48	**(0, 1, 1)**	**1102.97**	**1108.18**
(1, 2, 1)	-161.11	-155.96	(2, 1, 0)	-75.12	-69.90	(1, 1, 1)	1104.95	1111.91
(0, 2, 2)	-161.10	-155.96	(1, 1, 1)	-75.10	-69.80	(0, 1, 2)	1104.95	1111.90

Note. Best-performing model highlighted in bold based on lowest AIC/BIC values.

Table 6 displays the results of different ARIMA models that were fitted to the training data, and the corresponding AIC and BIC values for different parameters p and q. Results indicate that the most suitable model based on AIC and BIC for the EU emission data is the (0, 2, 1) model. According to the same criteria, the best performing model for the China emission time series is the (1, 1, 0) model. Finally, the most suitable model for the US emission time series is a (0, 1, 1) model.

4.3 Residual Diagnostics & Model Results

Model diagnostic test results are displayed in *Table 7*. Based on the test results, the (0, 2, 1)-model shows no significant autocorrelations, no heteroscedasticity, and no violation of the normal distribution assumption. Residual diagnostic plots can be found in the appendix, which support these conclusions, showing a white noise structure, no significant autocorrelations of residuals in the ACF plot and an approximately normally shaped distribution of the residuals. The Box-Ljung and McLeod-Li test indicate that the (1, 1, 0)-model for Chinese CO_2 emissions is a good fit to the data, displaying no significant autocorrelations in the residuals as well as no significant heteroscedasticity. However, the Jarque-Bera test indicates a strong violation of the normal distribution assumption, implying that no standard prediction intervals can be used. Consequently, *bootstrapping prediction intervals* will be used in the following forecasts. The histogram of model residuals (in the appendix) indicates a shape of the distribution with a negative skew. For US data, the tests and plots show no violation of the assumptions of no autocorrelations as well as normally distributed residuals, but a rather strongly heteroskedastic structure of the residuals. The plot of squared residuals indicates no linearly increasing variance, thereby rendering transformations via the logarithm or Box-Cox-transformations ineffective. Another model class would be required to take into account the heteroscedastic variance structure (i.e. *GARCH models*).

Table 7

Model residual diagnostic test statistics

Test	df	EU	China	US
Box-Ljung	5	5.46	7.08	2.23
McLeod-Li	6	3.08	9.28	18.03**
Jarque-Bera	2	4.36	15.79**	.78

Note. ** $p \leq .01$, * $p \leq .05$.

The final models that have passed the residual diagnostics are presented below:

$$\text{EU ARIMA (0, 2, 1):} \quad \Delta^2 X_t = -.72^{***}\varepsilon_{t-1} + \varepsilon_t \tag{15}$$

$$\text{China ARIMA (1, 1, 0):} \quad \Delta Y_t = .62^{***}\Delta^2 Y_{t-1} + \varepsilon_t \tag{16}$$

where *** indicates $p \leq 0.001$ and Δ and Δ^2 denote first and second order differences, respectively.

4.4 Persistance

The *impulse response functions* are displayed in Figure 2. Following a one unit shock, the fitted model for the EU CO_2 emissions shows a slight change in the opposite direction, but then a return to the mean after around 7 lags, demonstrating it is persistent. In contrast, the fitted model

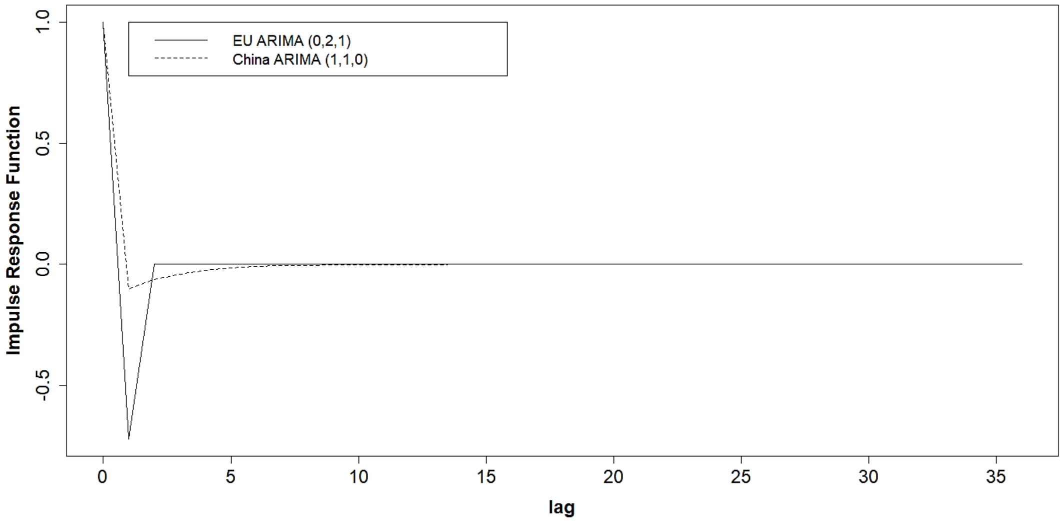

Figure 2. Impulse response function for the fitted ARIMA models.

for the Chinese CO_2 emissions shows a strong change in the opposite direction following a one-unit shock, but then quickly returning to its mean in the second lag, indicating that it is strongly persistent.

4.5 Forecasts

Forecasts of the main and baseline model for log-transformed out of sample data for the period 2002-2016 are displayed in *Figure 3*. The naïve baseline model used the last observed

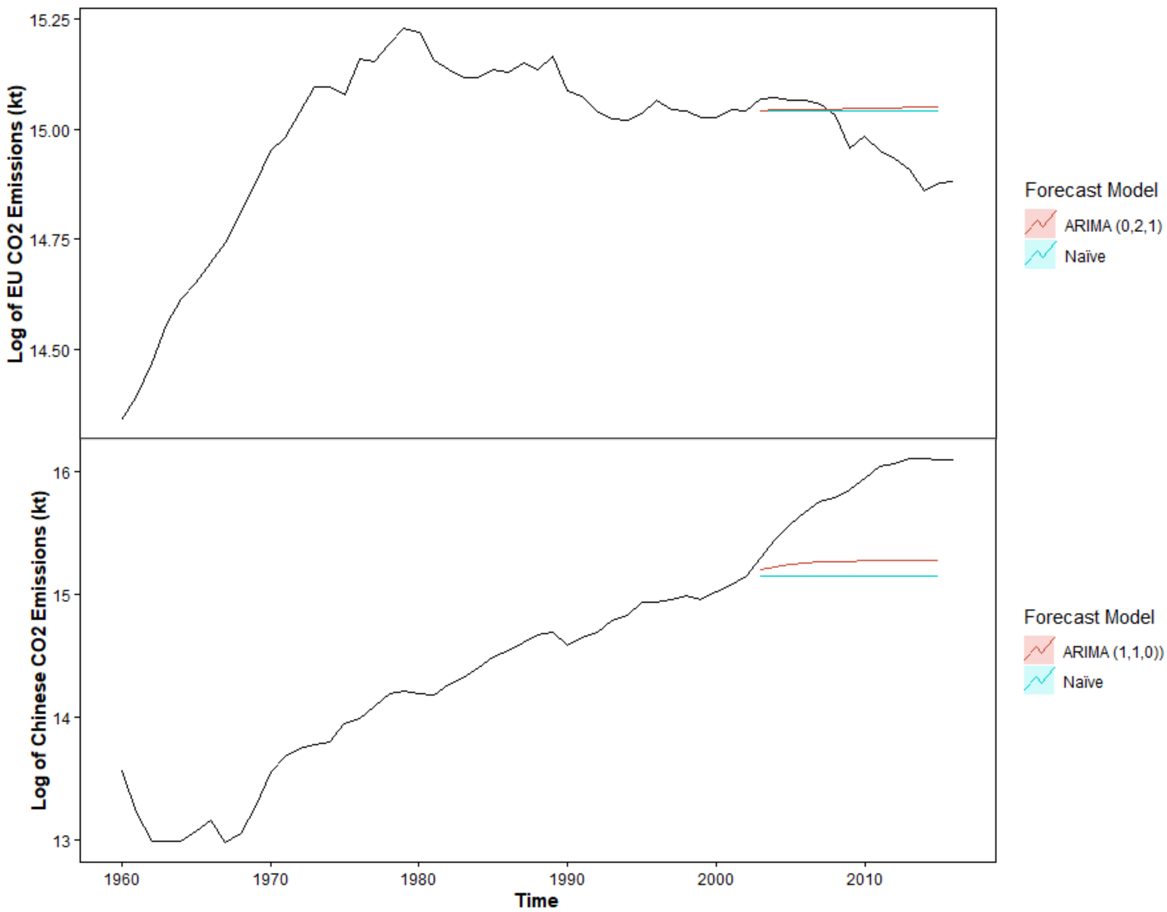

Figure 3. Out-of-sample forecasts for the EU and China CO_2 Emission data, comparing the best ARIMA model and a naïve baseline model.

value for the whole prediction period. For the EU data, the ARIMA (0,2,1) model predicted a slight increase of the CO emissions in the forecasted time period, which is not in line with the true values. For the Chinese emission data, the ARIMA (1,1,0) model predicted a slight increase of the CO_2 emissions, which does not account for the strong increase in the CO_2 emissions in the test set time period but showing a slightly better forecast than the naïve forecast.

The evaluation criteria for the prediction of the log-transformed out-of-sample data are displayed in *Table 8*. In line with the visual results, the four error criteria are lower for the naïve method for the EU time series, whereas the ARIMA model generates lower forecast error criteria values than the baseline model for the Chinese emission data.

Table 8

Forecast evaluation results

Forecasting method	RMSE	MAE	MAPE	MASE
Europe				
Naïve method	**.093**	**.073**	**.490**	**2.203**
ARIMA (0, 2, 1)	.097	.076	.510	2.292
China				
Naïve method	.725	.677	4.255	8.246
ARIMA (1, 1, 0)	**.616**	**.568**	**3.565**	**6.913**

Note. Best performing model highlighted in bold.

The Diebold-Mariano tests indicates that the ARIMA (0, 2, 1) model performs equally or worse than the naïve baseline method for the EU CO_2 emission data (DM = 0.73, p = .24). Similarly, the null hypothesis cannot be rejected for the test of the Chinese emission data models (DM = -1.45, p = .96), indicating that the ARIMA (1, 1, 0) model performs equally or worse than the naïve baseline method for the Chinese CO_2 emission data as well.

Next, the actual forecasts will be presented. The forecast point forecasts and *prediction intervals* (PI) for the ARIMA models are displayed in fan charts in *Figure 4*. For the EU in 2017, the ARIMA (0, 2, 1) model predicts 1.677.122 kt of CO_2 emissions (90% PI 2.729.202-3.030.338 kt). For 2030, the model forecasts a total emission amount of 2.854.991 kt (90% PI 1.398.582-4.585.498 kt). For China in 2017, the ARIMA (1, 1, 0) model predicts 770205 kt emissions (90% PI 8.414.963-11.423.131 kt). For 2030, the model forecasts 9.790.720 kt of CO_2 emissions for China (90% PI 2.531.665-37.844.416 kt)

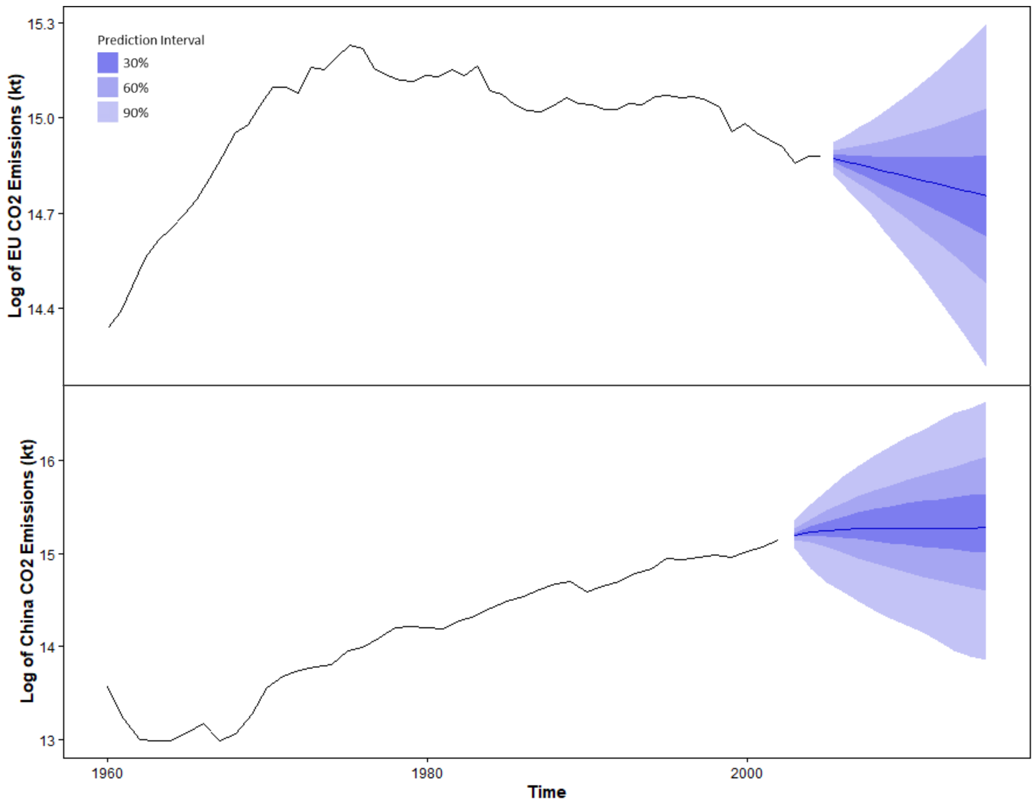

Figure 4. Fan plot of forecasts for CO_2 emissions in EU and China up to 2030 with 30%, 60%, 90% prediction intervals.

5. Discussion

This term paper contributes to the literature on CO_2 forecasting by applying the univariate ARIMA model to the total CO_2 emission data from the three largest global emitters. The technical problem consisted of transforming the data to achieve stationarity, selecting the appropriate model parameters, achieving white noise properties for the model residuals, and finally evaluating model forecasting performance on out of sample data. Overall, it has to be noted that the ARIMA model did not appear to be an appropriate model class for all the three time series. Specifically, the US CO_2 emission data, when sliced into training and test data as done in this study, exhibited a strong heteroscedastic error structure, requiring specific model classes to model this error process appropriately. The remaining time series for EU and China CO_2 emissions could be modelled in such a way that white noise residual characteristics could be achieved, and a long-term forecast was generated, although it has to be noted that the residuals for the Chinese time series model exhibited non-normality, requiring the application of bootstrapping prediction intervals to ensure a valid uncertainty quantification. Surprisingly, the ARIMA models did not outperform a naïve baseline model when tested on out-of-sample data for both the EU and the Chinese emission data. This could be due to a number of reasons. First, the Chinese data showed a strong increase in growth in the post-2002-period, which is equivalent with the test set period. This regime shift could not be modelled by the ARIMA model, since it was trained on data before that period. Second, researchers have noted that using goodness of fit indices are not necessarily related to post-sample accuracy (Makridakis et al., 2020), questioning the Box-Jenkins method that was used in this term paper. Finally, the weak performance could be due to the long forecasting horizon in the test set, covering over a decade of years. ARIMA models predict the future by making a series of one-period-ahead forecasts, that are combined. An accurate long-term forecast might rely on using extra information being incorporated into the model.

Furthermore, the empirical results reported herein should be considered in the light of some limitations. First, the Box-Jenkins-methodology involves a high degree of subjectivity in choosing the best model, thereby negatively impacting the validity of the results. For example, there are different tests to assess the non-stationarity of the time series data and the interpretation of visual plots is relying on the expertise of the statistician to identify the best model. Second, ARIMA models suffer from optimizing model selection based on statistical in-sample criteria, which might not be related to out-of-sample performance. Forecasting competitions cast doubt on the effectivity of this goodness of fit approach (Makridakis et al., 2020). Finally, the time series itself might be difficult to forecast, due to the possibility of structural breaks in the forecast period (Bårdsen & Nymoen, 2013). These structural breaks can have their source in technology and political decisions (Bårdsen & Nymoen, 2013). Given the high public attention that climate change has gathered in recent years, political actions are highly likely to influence the path of CO_2 emissions. The present paper does not include any conditioning assumptions to reflect potential policy paths, as outlined in Wieland and Wolters (2013), for example.

Future work can improve the CO_2 emission forecasts in several ways. First, an appropriate model class should be used to capture the heteroscedastic error structure in the US CO_2 emission time series. As a next step, a *GARCH model* could be applied (Bollerslev, 1986). Second, *forecast combinations* could be used, which have proven successful in further reducing forecasting errors (Bates & Granger, 1969; Clemen, 1989). This technique involves using several methods on the same time series and then combining the forecasting values through averaging, for example (Hyndman & Athanasopoulos, 2021). Indeed, state-of-the-art models for forecasting also strongly rely on combinations, as evident in the *M4 competition*, where largely combinations of mostly statistical models were performing best in terms of point forecasts and prediction intervals (Makridakis et al., 2020). Future work could identify an appropriate pool of statistical and machine learning models for CO_2 forecasting and optimize the weighting scheme. The former can range from simple equal-weighted forecast, over weighting based on models' past forecasting performance (Timmermann, 2018) to more sophisticated machine-learning based optimization of weightings as successfully applied in the M4 competition (Montero-Manso et al., 2020). Third, future work could leverage information from the different CO_2 emission time series to predict the individual time series (so-called *cross-learning*), as has proven a fruitful approach in practice (Makridakis et al., 2020). Specifically, the information from CO_2 emission time series from different countries could be used to guide model selection or combination as well as parameter estimation (Makridakis et al., 2020). Fourth, forecasts could be further tweaked by including relevant covariates in the model. This could be implemented via a *regression with ARMA errors* or *transfer function models* (Hyndman, 2010). Finally, a more sophisticated evaluation could be applied. In *time series cross-validation*, a series of test sets with one observation is used. The corresponding training sets consists of observations up to this test set observation, with too small training sets being discarded as they cannot be used to estimate a forecast reliably (Hyndman & Athanasopoulos, 2021). The final forecast accuracy is then based on averaging over the performance for each test set. Bergmeir and Benítez (2012) demonstrate that cross-validation leads to more robust model selection results, when compared to choosing the series' last part.

Due to the low forecast performance of the ARIMA models compared to the baseline model and the uncertainty related to long-term forecasts, this term paper can highlight policy implications only carefully. Due to the model assumptions not being met for the US data, no implications are derived for this country. If the point forecast results are taken as a true approximation of the 2030 CO_2 emission levels, the EU would miss its stated objective of 55% reduction compared to 1990 by a large margin (approximately 78% larger emissions than stated), although it has to be noted that the stated objective falls within the outer ends of the prediction intervals. Nevertheless, it becomes apparent that the EU needs to take measures to increase the certainty of achieving its reduction aim. To facilitate the attainment of these objectives, measures should differentiate between the ETS and non-ETS sectors. For the ETS sector, the main levers include the coverage of sources as well as the total emission cap and its time path (Harrison et al., 2012). The EU could expand the coverage of the ETS and define a stricter time path for the CO_2 emission reductions to facilitate goal attainment. Indeed, the EU plans to revise the ETS as part of its Green Deal initiative, although the proposed measures have not yet been passed into supra- or national laws (E.U., n.d.). The non-ETS sectors could implement a carbon tax or implement certain technology or performance standards to cut emission (Duval, 2008). Long-term measures such as R&D support are not considered feasible in supporting the 2030 objectives, given their long investment horizons. For China, the forecast predicts a slight increase for 2030, in line with China's stated objective to reach peak emissions by 2030. Given its objective to increase net zero in 2060, a massive amount of CO2 will need to be reduced within the following three decades. China can focus its policies to support this massive transition towards a low-carbon economy and energy system. First, providing a long-term incentive for investment in the development of clean energy technologies could help to guide research efforts. Second, known knowledge-market failures preventing talented individuals to engage in research could be addressed through patent protection measures, R&D tax cuts, and funding for basic research could support the development of innovative technologies in the long-run (Popp, 2010). Thirdly, setting a clear cap on total CO_2 emissions could support the reductions of CO_2 emissions effectively (Duval, 2008; Harrison et al., 2011).

This term paper applied ARIMA models to CO_2 emission data from the EU, China, and the US. ARIMA models are sophisticated statistical techniques that leverage past and current observations to predict future values. However, it was shown that simple baseline models already provide an effective starting point, requiring more elaborate forecasting solutions to achieve a substantial outperformance.

References

Akaike, H. (1973). Maximum likelihood identification of gaussian autoregressive moving average models. *Biometrika*, *60*(2), 255–265. https://doi.org/10.1093/biomet/60.2.255

Auffhammer, M., & Steinhauser, R. (2012). Forecasting the path of U.S. CO_2 emissions using state-level information. *The Review of Economics and Statistics*, *94*(1), 172–185. https://doi.org/10.1162/REST_a_00152

Banerjee, L. (2015). Climate thresholds, weather extremes, and catastrophic losses. In L. Bernard & W. Semmler (Eds.), *The Oxford handbook of the macroeconomics of global warming.* Oxford University press.

Bårdsen, G., & Nymoen, R. (2009). Macroeconometric modeling for policy. In Mills, T., & Patterson, K. (Ed.), *Palgrave handbook of econometrics: Applied econometrics* (pp. 851–916). Springer.

Bates, J. M., & Granger, C. W. J. (1969). The combination of forecasts. *Journal of the Operational Research Society*, *20*(4), 451–468. https://doi.org/10.1057/jors.1969.103

Bennedsen, M., Hillebrand, E., & Koopman, S. J. (2021). Modeling, forecasting, and nowcasting U.S. CO2 emissions using many macroeconomic predictors. *Energy Economics*, *96*, 105118. https://doi.org/10.1016/j.eneco.2021.105118

Bergmeir, C., & Benítez, J. M. (2012). On the use of cross-validation for time series predictor evaluation. *Information Sciences*, *191*, 192–213. https://doi.org/10.1016/j.ins.2011.12.028

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327. https://doi.org/10.1016/0304-4076(86)90063-1

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: Forecasting and control* (5th ed.). John Wiley & Sons.

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, *5*(4), 559–583. https://doi.org/10.1016/0169-2070(89)90012-5

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, *74*(366a), 427–431. https://doi.org/10.1080/01621459.1979.10482531

Dickey, D. A., & Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*, *49*(4), 1057–1072. https://doi.org/10.2307/1912517

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, *13*(3), 253–263. https://doi.org/10.1198/073500102753410444

Dritsaki, M., & Dritsaki, C. (2020). Forecasting European Union CO_2 emissions using autoregressive integrated moving average-autoregressive conditional heteroscedasticity models. *International Journal of Energy Economics and Policy*, *10*(4), 411–423.

Duval, R. (2008). A taxonomy of instruments to reduce greenhouse gas emissions and their interactions. *OECD Economics Department Working Papers.* Advance online publication. https://doi.org/10.1787/236846121450

European Comission. (2020, December 9). *European Green Deal*. https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1588580774040&uri=CELEX:52019DC0640

European Union. (n.d.). *Delivering the European Green Deal*. https://ec.europa.eu/clima/policies/eu-climate-action/delivering_en

FAZ (2021, July 16). Flutkatastrophe im Westen: Talsperren laufen über, Häuser sind zerstört. *Frankfurter Allgemeine Zeitung*. https://www.faz.net/aktuell/gesellschaft/ungluecke/hochwasser-in-deutschland-die-katastrophe-in-bildern-17438475.html

Gupta, J. (2010). A history of international climate change policy. *Wiley Interdisciplinary Reviews: Climate Change*, *1*(5), 636–653. https://doi.org/10.1002/wcc.67

Hamilton, J. D. (1994). *Time series analysis*. Princeton University Press.

Harrison, D., Foss, A., Klevnas, P., & Radov, D. (2011). Economic policy instruments for reducing greenhouse gas emissions. In J. S. Dryzek & Norgaard, R. B., Schlosberg, D. (Eds.), *The oxford handbook of climate change and society* (pp. 536–549). Oxford University press. https://doi.org/10.1093/oxfordhb/9780199566600.003.0035

Hyndman, R. J [R. J.]. (2010). *The ARIMAX model muddle*. https://robjhyndman.com/hyndsight/arimax/

Hyndman, R. J [R. J.], & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for r. *Journal of Statistical Software*, *27*(1), 1–22.

Hyndman, R. J [Rob J.], & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, *22*(4), 679–688. https://doi.org/10.1016/j.ijforecast.2006.03.001

Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and practice* (3rd ed.). OTexts.

IPCC. (2021). *Climate change 2021: The physical science basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

Jarque, C. M., & Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, *6*(3), 255–259. https://doi.org/10.1016/0165-1765(80)90024-5

Keller, K., & Nicholas, R. (2015). Improving climate projections to better inform climate risk management. In L. Bernard & W. Semmler (Eds.), *The Oxford handbook of the macroeconomics of global warming.* Oxford University press.

Lefevere, J., Runge-Metzger, A., & Werksman, J. (2015). The EU and international climate change policy. In J. Delbeke & P. Vis (Eds.), *EU climate policy explained* (pp. 94–107). Routledge.

Li, Y., Wei, Y., & Dong, Z. (2020). Will China achieve its ambitious goal? - Forecasting the CO_2 emission intensity of China towards 2030. *Energies, 13*(11), 2924. https://doi.org/10.3390/en13112924

Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, *65*(2), 297–303. https://doi.org/10.1093/biomet/65.2.297

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). The M4 competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting*, *36*(1), 54–74. https://doi.org/10.1016/j.ijforecast.2019.04.014

McLeod, A. I., & Li, W. K. (1983). Diagnostic checking ARMA time series models using squared-residual autocorrelations. *Journal of Time Series Analysis*, *4*(4), 269–273. https://doi.org/10.1111/j.1467-9892.1983.tb00373.x

Ministry of Foreign Affairs of the People's Republic of China. (2020, September 22). *Statement by H.E. Xi Jinping President of the People's Republic of China At the general debate of the 75th Session of The United Nations General Assembly* [Press release]. https://www.fmprc.gov.cn/mfa_eng/zxxx_662805/t1817098.shtml

Montero-Manso, P., Athanasopoulos, G., Hyndman, R. J [Rob J.], & Talagala, T. S. (2020). FFORMA: Feature-based forecast model averaging. *International Journal of Forecasting*, *36*(1), 86–92. https://doi.org/10.1016/j.ijforecast.2019.02.011

Nau, R. (2014). *Identifying the numbers of AR or MA terms in an ARIMA model*. https://people.duke.edu/~rnau/411arim3.htm

Nyoni, T., & Mutongi, C. (2019). Modeling and forecasting carbon dioxide emissions in China using autoregressive integrated moving average (ARIMA) models. *Munich Personal RePEc Archive*(Paper No. 93984). https://mpra.ub.uni-muenchen.de/93984/

Pao, H.-T., Fu, H.-C., & Tseng, C.-L. (2012). Forecasting of CO_2 emissions, energy consumption and economic growth in China using an improved grey model. *Energy*, *40*(1), 400–409. https://doi.org/10.1016/j.energy.2012.01.037

Popp, D. (2010). Innovation and climate policy. *Annual Review of Resource Economics*, *2*(1), 275–298. https://doi.org/10.1146/annurev.resource.012809.103929

R Core Team. (2021). *R* [Computer software]. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/.

Ritchie, H., & Roser, M. (2020). *CO_2 and greenhouse gas emissions*. Our World in Data. https://ourworldindata.org/co2-emissions#citation

Said, S. E., & Dickey, D. A. (1984). Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika*, *71*(3), 599–607. https://doi.org/10.1093/biomet/71.3.599

Schleussner, C.-F., Rogelj, J., Schaeffer, M., Lissner, T., Licker, R., Fischer, E. M., Knutti, R., Levermann, A., Frieler, K., & Hare, W. (2016). Science and policy characteristics of the Paris Agreement temperature goal. *Nature Climate Change*, *6*(9), 827–835. https://doi.org/10.1038/nclimate3096

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464. http://www.jstor.org/stable/2958889

Stock, J. H., & Watson, M. W. (2017). Twenty years of time series econometrics in ten pictures. *Journal of Economic Perspectives*, *31*(2), 59–86. https://doi.org/10.1257/jep.31.2.59

The White House. (2021, April 22). *FACT SHEET: President Biden sets 2030 greenhouse gas pollution reduction target aimed at creating good-paying union jobs and securing U.S. leadership on clean energy technologies* [Press release]. https://www.whitehouse.gov/briefing-room/statements-releases/2021/04/22/fact-sheet-president-biden-sets-2030-greenhouse-gas-pollution-reduction-target-aimed-at-creating-good-paying-union-jobs-and-securing-u-s-leadership-on-clean-energy-technologies/

Timmermann, A. (2018). Forecasting methods in finance. *Annual Review of Financial Economics*, *10*(1), 449–479. https://doi.org/10.1146/annurev-financial-110217-022713

United Nations. (1992). *United Nations Framework Convention on Climate Change*. *United Nations Treaty Collection. Chapter XXVII 7.* https://unfccc.int/resource/docs/convkp/conveng.pdf

United Nations. (2015). *Paris Agreement*. *United Nations Treaty Collection, Chapter XXVII 7 d*. https://treaties.un.org/Pages/ViewDetails.aspx?src=IND&mtdsg_no=XXVII-7-d&chapter=27&clang=_en

Wieland, V., & Wolters, M. (2013). Chapter 5 - forecasting and policy making. In G. Elliott & A. Timmermann (Eds.), *Handbook of economic forecasting* (Vol. 2, pp. 239–325). Elsevier. https://doi.org/10.1016/B978-0-444-53683-9.00005-0

World Bank. (2021). *Total greenhouse gas emissions (kt of CO_2 equivalent)*. https://data.worldbank.org/indicator/EN.ATM.GHGT.KT.CE

Xinhua News Agency. (2021, March 12). *Outline of the People's Republic of China 14th five-year plan for national economic and social development and long-range objectives for 2035* [Press release]. http://www.gov.cn/xinwen/2021-03/13/content_5592681.htm

Appendix

1. Plots

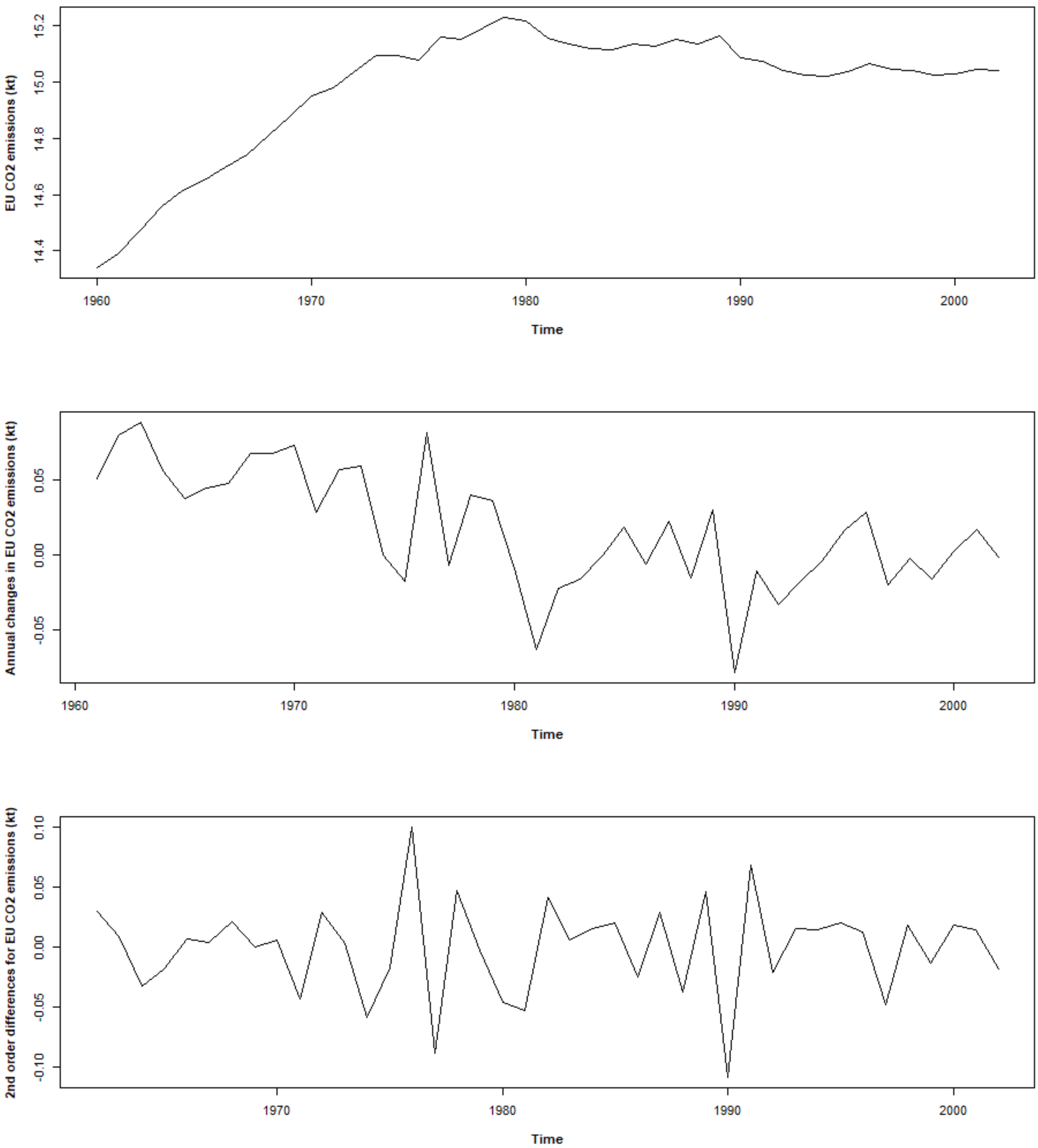

Figure A1. Time series plots of log-transformed EU annual CO_2 emission data for I(0), I(1), I(2).

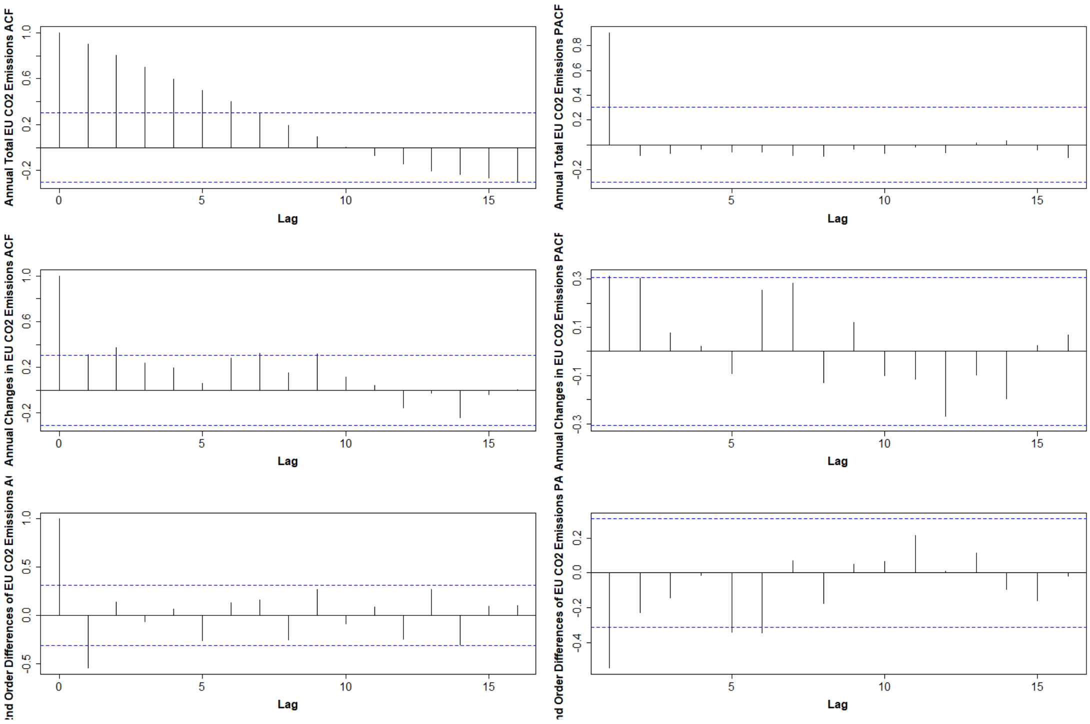

Figure A2. ACF and PACF plots for (1) level EU CO2 Emission data, (2) first order differences, (3) second order differences.

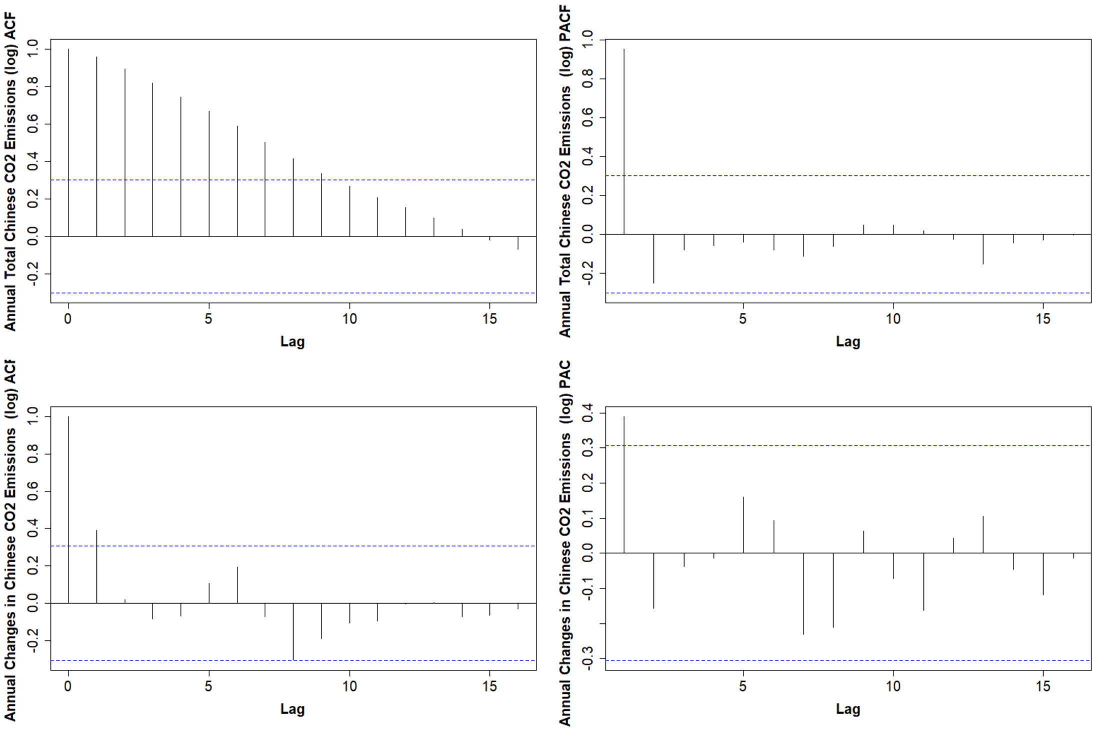

Figure A3. ACF and PACF plots for (1) level China CO_2 emission data, (2) first order differences, (3) second order differences.

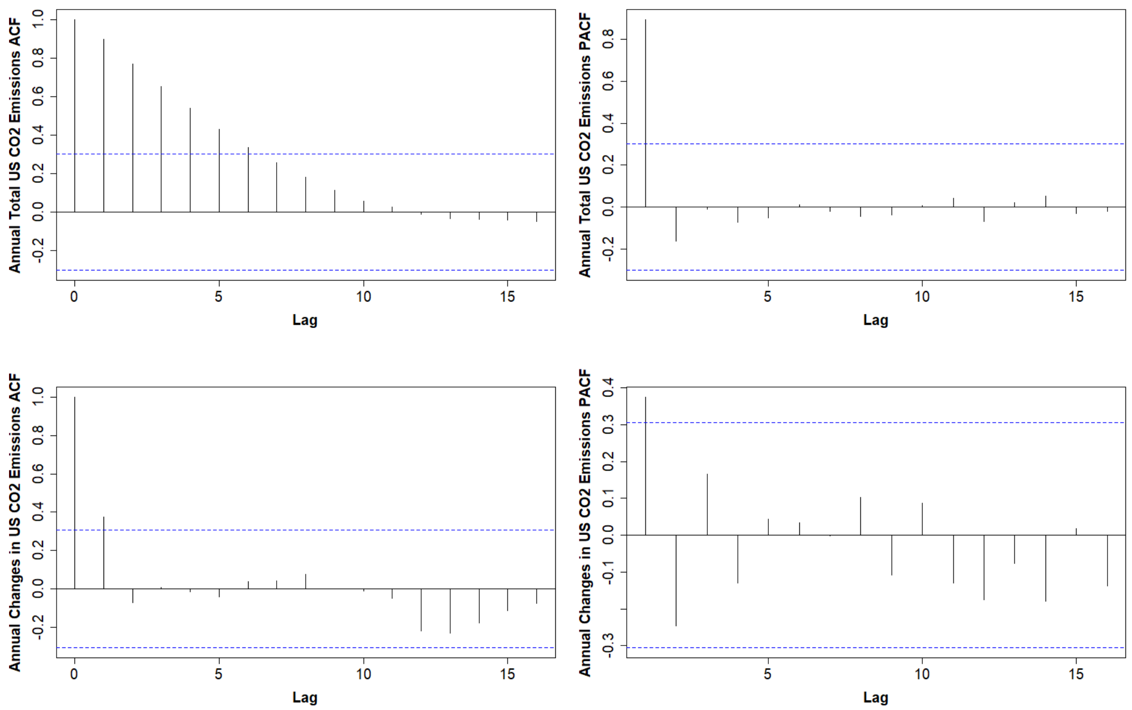

Figure A4. ACF and PACF plots for (1) level US CO_2 emission data, (2) first order differences, (3) second order differences.

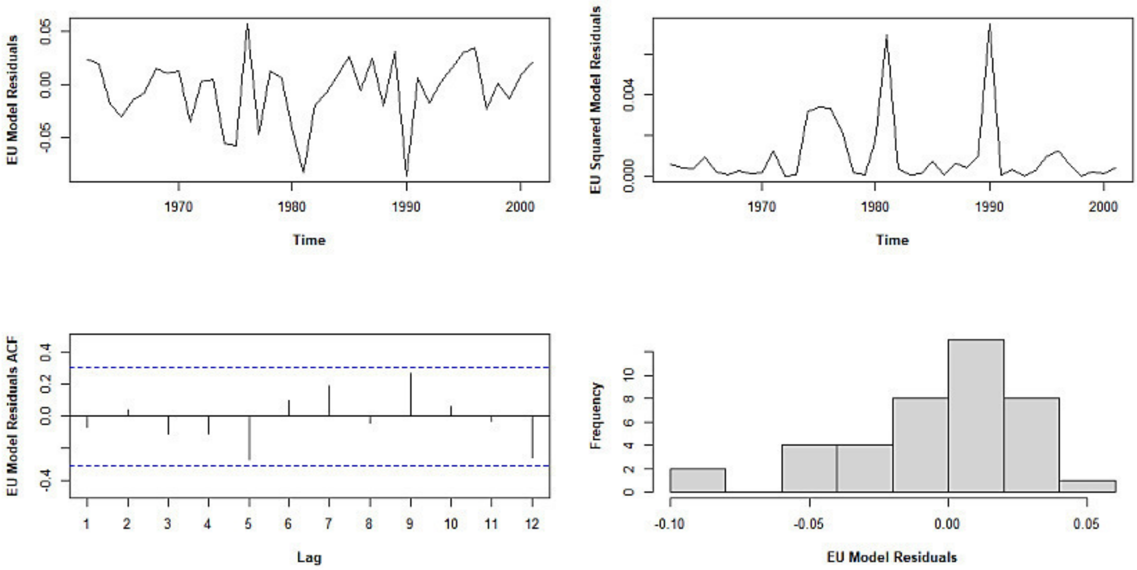

Figure A5. Residual diagnostic plots for the EU model with the (1) residual plot, (2) the squared residual plot, (3) the ACF plot for the residuals, and the (4) histogram of the residuals.

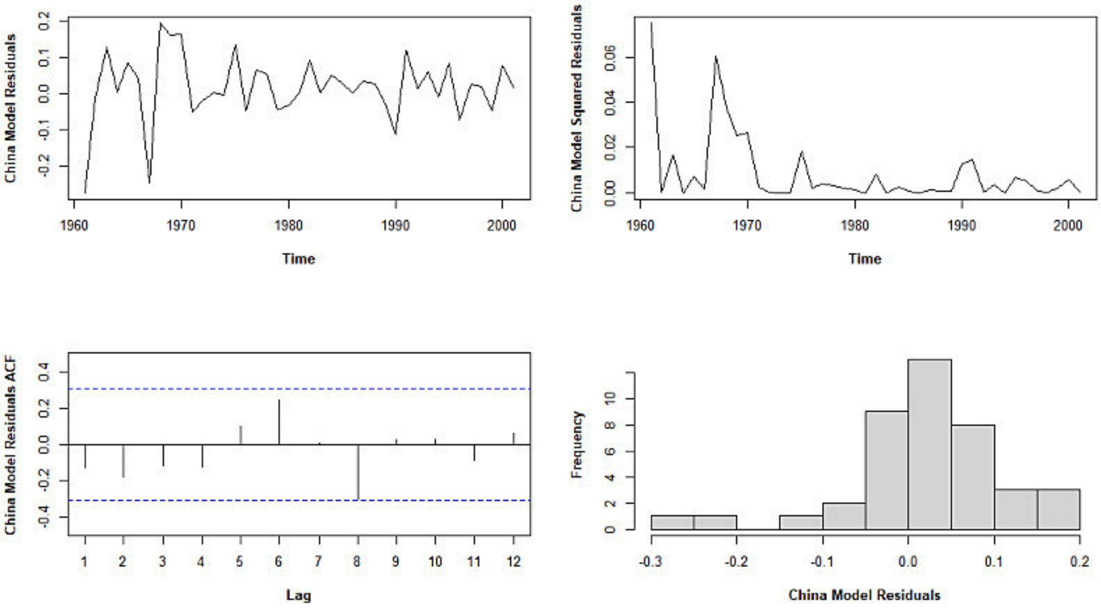

Figure A6. Residual diagnostic plots for the China model with the (1) residual plot, (2) the squared residual plot, (3) the ACF plot for the residuals, and the (4) histogram of the residuals.

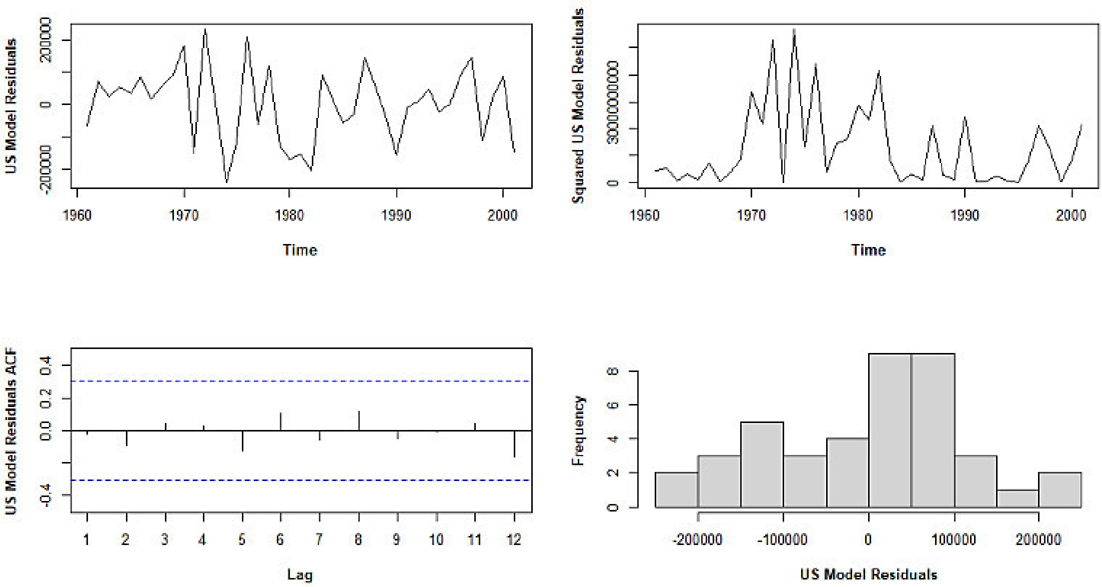

Figure A7. Residual diagnostic plots for the US model with the (1) residual plot, (2) the squared residual plot, (3) the ACF plot for the residuals, and the (4) histogram of the residuals.

2. Code

```
######################################################
# Time Series Analysis, Summer term 2021
# CO2 Emission Forecasting
######################################################

# 1. Loading packages & data
# Install & load relevant packages
#install.packages("forecast")
#install.packages("stats")
#install.packages("urca")
#install.packages("tseries")
#install.packages("ggplot2")

library(forecast)

## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo

library(stats)
library(urca)
library(tseries)
library(ggplot2)
library(lmtest)

## Lade nötiges Paket: zoo

##
## Attache Paket: 'zoo'

## Die folgenden Objekte sind maskiert von 'package:base':
##
##     as.Date, as.Date.numeric

# Import data
dataset <- read.table('C:/Users/M/Dropbox/05 Statistik-Master/2. Semester/Tim
e Series Analysis/04 Hausarbeit Time Series Analysis/Data Analysis/Worldbank
CO2 Emissions Processed vf.csv', header=TRUE, sep=';')  # load the data
ts <- dataset[,2:4]
head(dataset)

# 2. Define parameters
# Define parameter settings: Set maximum number of lags of AR and MA polynomi
als to
pmax <- 6
qmax <- 6
pmax_ac <- 6 # and max number of lags in Ljung-Box test

# Define parameters for train-test-split, test set size equal forecasting tim
e period (i.e., 2030-2017 = 13 years)
train_start_year <- 1960
train_end_year <- 2002
test_start_year <- 2003
test_end_year <- 2016
```

```r
# 3. Preprocess data
# Full range (for plotting)
date <- ts(dataset$ï..Jahr)
EU <- ts(dataset$European.Union, frequency=1, start=c(1960,1), end=c(2016,1))
# Set dependent variable
CN <- ts(dataset$China, frequency=1, start=c(1960,1), end=c(2016,1))
US <- ts(dataset$United.States, frequency=1, start=c(1960,1), end=c(2016,1))

# Define training set (in-sample) to estimate model parameters
# Apply log transformation to stabilize variance for EU and China based on pl
ots and residual diagnostics
EU_train_set <- window(log(EU),start=c(train_start_year,1),end=c(train_end_ye
ar,1))
CN_train_set <- window(log(CN),start=c(train_start_year,1),end=c(train_end_ye
ar,1))
US_train_set <- window(US,start=c(train_start_year,1),end=c(train_end_year,1)
)

# Define test set (out-of-sample) to assess forecast accuracy
EU_test_set <- window(log(EU), start=c(test_start_year,1),end=c(test_end_year
,1))
CN_test_set <- window(log(CN),start=c(test_start_year,1),end=c(test_end_year,
1))
US_test_set <- window(US,start=c(test_start_year,1),end=c(test_end_year,1))

# 4. Conduct descriptive analyses (to full dataset)
# EU
summary(EU)
M <- mean(EU)
SD <- sd(EU)
CV <- SD/M*100 # Coefficient of variation
print(SD)
print(CV)

# CN
summary(CN)
M <- mean(CN)
SD <- sd(CN)
CV <- SD/M*100 # Coefficient of variation
print(SD)
print(CV)

# US
summary(US)
M <- mean(US)
SD <- sd(US)
CV <- SD/M*100 # Coefficient of variation
print(SD)
print(CV)

# Plot integrated time series plot
China <- CN
dat_int <- ts.intersect(EU, China, US)
plot.ts(dat_int, cex.lab=1.5, cex.axis=1.5, main="")
```

```
# 5. Assess stationarity of time series as a foundation for ARIMA modeling
# Plot specific plots to assess stationarity of training set
options(scipen=10) # reformat axis numbers in order to ensure that full numbe
rs are shown
par(font.lab=2, font.axis=1) # Create bold font titles
plot(EU_train_set, ylab="EU CO2 emissions (kt)")

plot(CN_train_set, ylab="Logarithmized China CO2 emissions (kt)")

plot(US_train_set, ylab="US CO2 emissions (kt)")

plot(diff(EU_train_set), ylab="Annual changes in EU CO2 emissions (kt)")

plot(diff(CN_train_set), ylab="Annual changes in China CO2 emissions (%)")

plot(diff(US_train_set), ylab="Annual changes in US CO2 emissions (kt)")

plot(diff(diff(EU_train_set)), ylab="2nd order differences for EU CO2 emissio
ns (kt)")

# Compare different orders for EU data to decide on stationarity
layout(matrix(c(1,2, 3), nrow=3, byrow=TRUE))
plot(EU_train_set, ylab="EU CO2 emissions (kt)")
plot(diff(EU_train_set), ylab="Annual changes in EU CO2 emissions (kt)")
plot(diff(diff(EU_train_set)), ylab="2nd order differences for EU CO2 emissio
ns (kt)")

# Test for stationarity using ADF
# Determine number of differences required
print("Number of differences required based on trend (+intercept for ADF) com
ponent:")
ndiffs(EU_train_set, alpha = 0.05, test = "adf", type ="level")
ndiffs(CN_train_set, alpha = 0.05, test = "adf", type ="level")
ndiffs(US_train_set, alpha = 0.05, test = "adf", type ="level")
ndiffs(EU_train_set, alpha = 0.05, test = "adf", type ="trend")
ndiffs(CN_train_set, alpha = 0.05, test = "adf", type ="trend")
ndiffs(US_train_set, alpha = 0.05, test = "adf", type ="trend")

# Compute test statistics for ADF test
# For EU emissions:
pmax_adf = round(12*(length(EU_train_set)/100)^0.25, 0)  # truncation lag acc
ording to Schwert (1989); length equal across countries
adf_drift_EU = ur.df (EU_train_set, type="drift" , selectlags="BIC", lags=pma
x_adf)
print("ADF test results (with drift) for the EU emission data:")
summary (adf_drift_EU)
punitroot(q=adf_drift_EU@teststat[1], N=length(EU_train_set), trend="c")
length(adf_drift_EU@testreg$coefficients[,1])-2 #lag length
adf_drift_EU_diff = ur.df (diff(EU_train_set), type="drift" , selectlags="BIC
", lags=pmax_adf)
print("ADF test results (with drift) for the EU emission data (first order di
fferences):")
summary (adf_drift_EU_diff)
punitroot(q=adf_drift_EU_diff@teststat[1], N=length(diff(EU_train_set)), tren
d="c")
length(adf_drift_EU_diff@testreg$coefficients[,1])-2 #lag length
adf_drift_EU_diff_diff = ur.df (diff(diff(EU_train_set)), type="drift" , sele
```

```
ctlags="BIC", lags=pmax_adf)
print("ADF test results (with drift) for the EU emission data (second order d
ifferences):")
summary (adf_drift_EU_diff_diff)
punitroot(q=adf_drift_EU_diff_diff@teststat[1], N=length(diff(diff(EU_train_s
et))), trend="c")
length(adf_drift_EU_diff_diff@testreg$coefficients[,1])-2 #lag length
adf_trend_EU = ur.df (EU_train_set, type="trend" , selectlags="BIC", lags=pma
x_adf)
print("ADF test results (with trend) for the EU emission data:")
summary (adf_trend_EU)
punitroot(q=adf_trend_EU@teststat[1], N=length(EU_train_set), trend="ct")
length(adf_drift_EU@testreg$coefficients[,1])-3 #lag length, due to trend tes
t there is one additional parameter that needs to be subtracted
adf_trend_EU_diff = ur.df (diff(EU_train_set), type="trend" , selectlags="BIC
", lags=pmax_adf)
print("ADF test results (with trend) for the EU emission data (first order di
fferences):")
summary (adf_trend_EU_diff)
punitroot(q=adf_trend_EU_diff@teststat[1], N=length(diff(EU_train_set)), tren
d="ct")
length(adf_drift_EU@testreg$coefficients[,1])-3 #lag length, due to trend tes
t there is one additional parameter that needs to be subtracted
adf_drift_EU_diff_diff = ur.df (diff(diff(EU_train_set)), type="trend" , sele
ctlags="BIC", lags=pmax_adf)
print("ADF test results (with trend) for the EU emission data (second order d
ifferences):")
summary (adf_drift_EU_diff_diff)
punitroot(q=adf_drift_EU_diff_diff@teststat[1], N=length(diff(diff(EU_train_s
et))), trend="ct")
length(adf_drift_EU_diff_diff@testreg$coefficients[,1])-3 #lag length

#China
adf_drift_CN = ur.df (CN_train_set, type="drift" , selectlags="BIC", lags=pma
x_adf)
print("ADF test results (with drift) for the China emission data:")
summary (adf_drift_CN)
punitroot(q=adf_drift_CN@teststat[1], N=length(CN_train_set), trend="c")
length(adf_drift_CN@testreg$coefficients[,1])-2 #lag length
adf_drift_CN_diff = ur.df (diff(CN_train_set), type="drift" , selectlags="BIC
", lags=pmax_adf)
print("ADF test results (with drift) for the China emission data (first order
differences):")
summary (adf_drift_CN_diff)
punitroot(q=adf_drift_CN_diff@teststat[1], N=length(diff(CN_train_set)), tren
d="c")
length(adf_drift_CN@testreg$coefficients[,1])-2 #lag length
adf_trend_CN = ur.df (CN_train_set, type="trend" , selectlags="BIC", lags=pma
x_adf)
print("ADF test results (with trend) for the China emission data:")
summary (adf_trend_CN)
punitroot(q=adf_trend_CN@teststat[1], N=length(CN_train_set), trend="ct")
length(adf_drift_CN@testreg$coefficients[,1])-3 #lag length, due to trend tes
t there is one additional parameter that needs to be subtracted
adf_trend_CN_diff = ur.df (diff(CN_train_set), type="trend" , selectlags="BIC
", lags=pmax_adf)
```

```
print("ADF test results (with trend) for the China emission data (first order
differences):")
summary (adf_trend_CN_diff)
punitroot(q=adf_trend_CN_diff@teststat[1], N=length(diff(CN_train_set)), tren
d="ct")
length(adf_drift_CN@testreg$coefficients[,1])-3 #lag length, due to trend tes
t there is one additional parameter that needs to be subtracted

#US
adf_drift_US = ur.df (US_train_set, type="drift" , selectlags="BIC", lags=pma
x_adf)
print("ADF test results (with drift) for the US emission data:")
summary (adf_drift_US)
punitroot(q=adf_drift_US@teststat[1], N=length(US_train_set), trend="c")
length(adf_drift_US@testreg$coefficients[,1])-2 #lag length
adf_drift_US_diff = ur.df (diff(US_train_set), type="drift" , selectlags="BIC
", lags=pmax_adf)
print("ADF test results (with drift) for the US emission data (first order di
fferences):")
summary (adf_drift_US_diff)
punitroot(q=adf_drift_US_diff@teststat[1], N=length(diff(US_train_set)), tren
d="c")
length(adf_drift_US_diff@testreg$coefficients[,1])-2 #lag length
adf_trend_US = ur.df (US_train_set, type="trend" , selectlags="BIC", lags=pma
x_adf)
print("ADF test results (with trend) for the US emission data:")
summary (adf_trend_US)
punitroot(q=adf_trend_US@teststat[1], N=length(US_train_set), trend="ct")
length(adf_trend_US@testreg$coefficients[,1])-3 #lag length, due to trend tes
t there is one additional parameter that needs to be subtracted
adf_trend_US_diff = ur.df (diff(US_train_set), type="trend" , selectlags="BIC
", lags=pmax_adf)
print("ADF test results (with trend) for the US emission data (first order di
fferences):")
summary (adf_trend_US_diff)
punitroot(q=adf_trend_US_diff@teststat[1], N=length(diff(US_train_set)), tren
d="ct")
length(adf_trend_US_diff@testreg$coefficients[,1])-3 #lag length, due to tren
d test there is one additional parameter that needs to be subtracted

# Conduct diagnostic (partial) autocorrelation tests to assess non-stationari
ty and support decision for best model
par(mfrow=c(3,2))
acf_EU <- acf(EU_train_set, plot = FALSE)
plot(acf_EU, ylab="Annual Total EU CO2 Emissions ACF", main="", cex.lab=1.5,
cex.axis=1.5)
pacf_EU <- pacf(EU_train_set, plot = FALSE)
plot(pacf_EU, ylab="Annual Total EU CO2 Emissions PACF", main="", cex.lab=1.5
, cex.axis=1.5)
acf_EU_diff <- acf(diff(EU_train_set), plot = FALSE)
plot(acf_EU_diff, ylab="Annual Changes in EU CO2 Emissions ACF", main="", cex
.lab=1.5, cex.axis=1.5)
pacf_EU_diff <- pacf(diff(EU_train_set), plot = FALSE)
plot(pacf_EU_diff, ylab="Annual Changes in EU CO2 Emissions PACF", main="", c
ex.lab=1.5, cex.axis=1.5)
acf_EU_diff_diff <- acf(diff(diff(EU_train_set)), plot = FALSE)
```

```
plot(acf_EU_diff_diff, ylab="2nd Order Differences of EU CO2 Emissions ACF",
main="", cex.lab=1.5, cex.axis=1.5)
pacf_EU_diff_diff <- pacf(diff(diff(EU_train_set)), plot = FALSE)
plot(pacf_EU_diff_diff, ylab="2nd Order Differences of EU CO2 Emissions PACF"
, main="", cex.lab=1.5, cex.axis=1.5)

par(mfrow=c(2,2))
acf_CN <- acf(CN_train_set, plot = FALSE)
plot(acf_CN, ylab="Annual Total Chinese CO2 Emissions (log) ACF", main="", ce
x.lab=1.5, cex.axis=1.5)
pacf_CN <- pacf(CN_train_set, plot = FALSE)
plot(pacf_CN, ylab="Annual Total Chinese CO2 Emissions  (log) PACF", main="",
cex.lab=1.5, cex.axis=1.5)
acf_CN_diff <- acf(diff(CN_train_set), plot = FALSE)
plot(acf_CN_diff, ylab="Annual Changes in Chinese CO2 Emissions  (log) ACF",
main="", cex.lab=1.5, cex.axis=1.5)
pacf_CN_diff <- pacf(diff(CN_train_set), plot = FALSE)
plot(pacf_CN_diff, ylab="Annual Changes in Chinese CO2 Emissions  (log) PACF"
, main="", cex.lab=1.5, cex.axis=1.5)

par(mfrow=c(2,2))
acf_US <- acf(US_train_set, plot = FALSE)
plot(acf_US, ylab="Annual Total US CO2 Emissions ACF", main="", cex.lab=1.5,
cex.axis=1.5)
pacf_US <- pacf(US_train_set, plot = FALSE)
plot(pacf_US, ylab="Annual Total US CO2 Emissions PACF", main="", cex.lab=1.5
, cex.axis=1.5)
acf_US_diff <- acf(diff(US_train_set), plot = FALSE)
plot(acf_US_diff, ylab="Annual Changes in US CO2 Emissions ACF", main="", cex
.lab=1.5, cex.axis=1.5)
pacf_US_diff <- pacf(diff(US_train_set), plot = FALSE)
plot(pacf_US_diff, ylab="Annual Changes in US CO2 Emissions PACF", main="", c
ex.lab=1.5, cex.axis=1.5)

# 6. Model selection based on statistical criteria
# Identify the best ARMA(p,q) model through computation
modelEU.aic <- auto.arima(diff(diff(EU_train_set)), ic="aic", d=0, max.p=pmax
, max.q=qmax, seasonal=FALSE, trace=TRUE, allowmean=FALSE)  # AIC
modelEU.bic <- auto.arima(diff(diff(EU_train_set)), ic="bic", d=0, max.p=pmax
, max.q=qmax, seasonal=FALSE, trace=TRUE, allowmean=FALSE)  # BIC
modelCN.aic <- auto.arima(diff(CN_train_set), ic="aic", d=0, max.p=pmax, max.
q=qmax, seasonal=FALSE, trace=TRUE, allowmean=FALSE)  # AIC
modelCN.bic <- auto.arima(diff(CN_train_set), ic="bic", d=0, max.p=pmax, max.
q=qmax, seasonal=FALSE, trace=TRUE, allowmean=FALSE)  # BIC
modelUS.aic <- auto.arima(diff(US_train_set), ic="aic", d=0, max.p=pmax, max.
q=qmax, seasonal=FALSE, trace=TRUE, allowmean=TRUE)  # AIC
modelUS.bic <- auto.arima(diff(US_train_set), ic="bic", d=0, max.p=pmax, max.
q=qmax, seasonal=FALSE, trace=TRUE, allowmean=TRUE)  # BIC

# Select and store the most relevant models
EU_p <- modelEU.aic$arma[1]
EU_q <- modelEU.aic$arma[2]
CN_p <- modelEU.aic$arma[1]
CN_q <- modelEU.aic$arma[2]
US_p <- modelEU.aic$arma[1]
US_q <- modelEU.aic$arma[2]
```

```
# 7. Conduct residual diagnostics & store model results:
# For EU
layout(matrix(c(1,2,3,4), nrow=2, byrow=TRUE)) # Create a compact layout for
the resulting charts
EU_e <- modelEU.aic$residuals # Store residuals
plot (EU_e, type="l", ylab="EU Model Residuals") # Residual plot
plot (EU_e^2, type="l", ylab="EU Squared Model Residuals") # Squared residual
plot
Acf(EU_e,12, main="", ylab="EU Model Residuals ACF") # ACF plot to test for r
emaining auto-correlation
hist(EU_e, main="", xlab="EU Model Residuals") # Histogram to assess normalit
y of residuals

LB.modelEU.aic <- Box.test(EU_e, lag=pmax_ac, type="Ljung-Box", fitdf=(EU_p+E
U_q)) # Ljung-box test for auto-correlation
print(LB.modelEU.aic)
MCL.modelEU.aic <- Box.test(EU_e^2, lag=pmax_ac, type="Ljung-Box")  # McLeod-
Li test for conditional heteroscedasticity
print(MCL.modelEU.aic)
JB.modelEU.aic <- jarque.bera.test(EU_e) # Jarque-Bera test for normality of
residuals
print(JB.modelEU.aic)
# Print model
coeftest(modelEU.aic) # Test significance of coefficients & print coefficient
s

# For China
layout(matrix(c(1,2,3,4), nrow=2, byrow=TRUE))
CN_e <- modelCN.aic$residuals    # Compute the residuals
plot (CN_e, type="l", ylab="China Model Residuals") # Residual plot
plot (CN_e^2, type="l", ylab="China Model Squared Residuals") # Squared resid
ual plot
Acf(CN_e,12, main="", ylab="China Model Residuals ACF") # ACF plot to test fo
r remaining auto-correlation
hist(CN_e, main="", xlab="China Model Residuals") # Histogram to assess norma
lity of residuals

LB.modelCN.aic <- Box.test(CN_e, lag=pmax_ac, type="Ljung-Box", fitdf=(CN_p+C
N_q)) # Ljung-box test for auto-correlation
print(LB.modelCN.aic)
MCL.modelCN.aic <- Box.test(CN_e^2, lag=pmax_ac, type="Ljung-Box")  # McLeod-
Li test for conditional heteroscedasticity
print(MCL.modelCN.aic)
JB.modelCN.aic <- jarque.bera.test(CN_e) # Jarque-Bera test for normality of
residuals
print(JB.modelCN.aic)
# Print model
coeftest(modelCN.aic) # Test significance of coefficients & print coefficient
s

# For US
layout(matrix(c(1,2, 3, 4), nrow=2, byrow=TRUE))
US_e <- modelUS.aic$residuals    # Compute the residuals
plot (US_e, type="l", ylab="US Model Residuals") # Residual plot
plot (US_e^2, type="l", ylab="Squared US Model Residuals") # Squared residual
```

```
plot
Acf(US_e,12, main="", ylab="US Model Residuals ACF") # ACF plot to test for r
emaining auto-correlation
hist(US_e, main="", xlab="US Model Residuals") # Histogram to assess normalit
y of residuals

LB.modelUS.aic <- Box.test(US_e, lag=pmax_ac, type="Ljung-Box", fitdf=(US_p+U
S_q)) # Ljung-box test for auto-correlation
print(LB.modelUS.aic)
MCL.modelUS.aic <- Box.test(US_e^2, lag=pmax_ac, type="Ljung-Box")  # McLeod-
Li test for conditional heteroscedasticity
print(MCL.modelUS.aic)
JB.modelUS.aic <- jarque.bera.test(US_e) # Jarque-Bera test for normality of
residuals
print(JB.modelUS.aic)

# 8. Compute impulse response functions to test persistance of model
# Set parameters
max.lag <- 36
ar.coef <- 0
ma.coef <- 0

# EU
model <- modelEU.aic
p <- model$arma[1]
q <- model$arma[2]
if (p>0) {ar.coef <- model$coef[1:p]}
if (q>0) {ma.coef <- model$coef[(p+1):(p+q)]}
IRF_EU <- c(1, ARMAtoMA(ar=ar.coef, ma=ma.coef, lag.max=max.lag))

#China
model <- modelCN.aic
p <- model$arma[1]
q <- model$arma[2]
if (p>0) {ar.coef <- model$coef[1:p]}
if (q>0) {ma.coef <- model$coef[(p+1):(p+q)]}
IRF_CN <- c(1, ARMAtoMA(ar=ar.coef, ma=ma.coef, lag.max=max.lag))

#US model is not tested due to heteroscedastic error structure

# Generate plot of impulse response functions
par(mfrow=c(1,1), font.lab = 2)
plot(seq(0,max.lag,1),IRF_EU, type="l",lty=1, xlab="lag", cex.lab=1.5, cex.ax
is=1.5, ylab="Impulse Response Function")
lines(seq(0,max.lag,1),IRF_CN, type="l",lty=2, xlab="lag")
legend(1,legend=c("EU ARIMA (0,2,1)", "China ARIMA (1,1,0)"),lty=1:2, cex=1.2
)

# 9. Conduct forecasting comparison on test set
# Set parameters
level = c(30, 60, 90)
h <- 13 # equals length of forecasting time period (Hyndman & Athanasopoulos,
2021)

# EU - Generate forecasts for test set
EU_forecast_baseline <- naive(EU_train_set, h=h, lambda=0)
```

```
EU_forecast_ARIMA <- forecast(EU_train_set, model=Arima(EU_train_set, order =
c(0,2,1), lambda=0), h=h, level=level)
EU_forecast_ARIMA_residuals <- EU_forecast_ARIMA$fitted - EU_train_set

# EU - Create plot including main model, baseline model, and actuals
autoplot(window(log(EU), start=train_start_year))+
  autolayer(EU_forecast_baseline, series = "Naïve", PI=FALSE) +
  autolayer(EU_forecast_ARIMA, series = "ARIMA (0,2,1)", PI=FALSE) +
  guides(colour=guide_legend(title="Forecast Model")) +
  theme_linedraw() + theme(panel.grid.major = element_blank(), panel.grid.min
or = element_blank(), axis.title.x = element_text(face="bold"), axis.title.y
= element_text(face="bold")) +
  xlab("Time") + ylab("Log of EU CO2 Emissions (kt)")

# EU - Compute evaluation metrics and test statistics for model comparison
accuracy(EU_forecast_baseline, EU_test_set)
accuracy(EU_forecast_ARIMA, EU_test_set)
dm.test(residuals(rwf(EU_train_set), h=h, lambda=0), EU_forecast_ARIMA_residu
als, h=h, alternative="greater")


# EU - Prediction: One-step ahead and 2030
EU_prediction_ARIMA <- forecast(log(EU), model=Arima(EU_train_set, order = c(
0,2,1)), h=14, level=level) # Use model fitted on training data to predict fu
ture values (after test set period)
autoplot(EU_prediction_ARIMA, series = "ARIMA (0,2,1))") + # Create plot with
prediction intervals
  theme_linedraw() + theme(panel.grid.major = element_blank(), panel.grid.min
or = element_blank(), axis.title.x = element_text(face="bold"), axis.title.y
= element_text(face="bold")) +
  xlab("Time") + ylab("Log of EU CO2 Emissions (kt)")

EU_prediction_ARIMA # log values
exp(EU_prediction_ARIMA$fitted) # back-transformed point forecasts
exp(EU_prediction_ARIMA$lower) # back-transformed lower prediction intervals
exp(EU_prediction_ARIMA$upper) # back-transformed upper prediction intervals

# China - Generate forecasts for test set
CN_forecast_baseline <- rwf(CN_train_set, h=h, lambda=0)
CN_forecast_ARIMA <- forecast(Arima(CN_train_set, order = c(1,1,0),lambda=0),
h=h, level=level, bootstrap =TRUE) #due to non-normal errors, bootstrapping i
s used for the prediction intervals
CN_forecast_ARIMA_residuals <- CN_forecast_ARIMA$fitted - CN_train_set

# China - Create plot including main model, baseline model, and actuals
autoplot(window(log(CN), start=train_start_year)) +
  autolayer(CN_forecast_baseline, series = "Naïve", PI=FALSE) +
  autolayer(CN_forecast_ARIMA, series = "ARIMA (1,1,0))", PI=FALSE) +
  guides(colour=guide_legend(title="Forecast Model")) +
  theme_linedraw() + theme(panel.grid.major = element_blank(), panel.grid.min
or = element_blank(), axis.title.x = element_text(face="bold"), axis.title.y
= element_text(face="bold")) +
  xlab("Time") + ylab("Log of Chinese CO2 Emissions (kt)")

# China - Compute evaluation metrics and test statistics for model comparison
accuracy(CN_forecast_baseline, CN_test_set)
```

```r
accuracy(CN_forecast_ARIMA, CN_test_set)
dm.test(residuals(rwf(CN_train_set, h=h, lambda=0)), CN_forecast_ARIMA_residu
als, h=14, alternative="greater")

# China - Prediction: One-step ahead and 2030
CN_prediction_ARIMA <- forecast(log(CN), model=Arima(CN_train_set, order = c(
1,1,0)), h=14, level=level)
autoplot(CN_forecast_ARIMA, series = "ARIMA (1,1,0)") + # Create plot with pr
ediction intervals
  theme_linedraw() + theme(panel.grid.major = element_blank(), panel.grid.min
or = element_blank(), axis.title.x = element_text(face="bold"), axis.title.y
= element_text(face="bold")) +
  xlab("Time") + ylab("Log of China CO2 Emissions (kt)")

CN_prediction_ARIMA #log values
exp(CN_prediction_ARIMA$fitted) # back-transformed point forecast
exp(CN_prediction_ARIMA$lower) # back-transformed lower prediction interval
exp(CN_prediction_ARIMA$upper) # back-transformed upper prediction interval
```

Declaration of Authorship

I hereby confirm that I have authored this document independently and without use of others than the indicated sources. All passages which are literally or in general matter taken out of publications or other sources are marked as such.

I have attached the code used to produce the analysis in the appendix. I confirm that I have written and executed the analysis, and that the code is complete and executable.

31.08.2021 M. Hildebrandt

Location, Date Signature