Modeling and forecasting carbon dioxide emissions in China using Autoregressive Integrated Moving Average (ARIMA) models

NYONI, THABANI and MUTONGI, CHIPO

University of Zimbabwe, Midlands State University

6 May 2019

Modeling and Forecasting Carbon Dioxide Emissions in China Using Autoregressive Integrated Moving Average (ARIMA) Models

Nyoni, Thabani
Department of Economics
University of Zimbabwe
Harare, Zimbabwe
Email: nyonithabani35@gmail.com

Dr. Chipo, Mutongi (Ph.D)
Department of Business Management
Midlands State University
Gweru, Harare
mutongic@gmail.com

ABSTRACT

This research uses annual time series data on CO_2 emissions in China from 1960 to 2017, to model and forecast CO_2 using the Box – Jenkins ARIMA approach. Diagnostic tests indicate that China CO_2 emission data is I (2). The study presents the ARIMA (1, 2, 1) model. The diagnostic tests further imply that the presented best model is stable and hence acceptable for predicting carbon dioxide emissions in China. The results of the study reveal that CO_2 emissions in China are likely to increase and thereby exposing China to a plethora of climate change related challenges. 4 main policy prescriptions have been put forward for consideration by the Chinese government.

Key Words: ARIMA model, China, CO_2 emissions

JEL Codes: C53, P28, Q47, Q52, Q53, Q54

INTRODUCTION

Climate change has been one of the top issues on international political agendas in recent years for global warming. Global warming is one of the most gripping and complicated problems facing the world. It is generally caused by greenhouse gas – mainly CO_2 emission in the atmosphere (Hossain *et al*, 2017). The forecasts of CO_2 emissions constitute a vital part of a clean energy economy (Pao *et al*, 2012). It is therefore invaluable to have a deeper understanding of China's past CO_2 emission path in order to make a reliable prediction of its future emission. This paper seeks to model and forecast CO_2 emission in China.

LITERATURE REVIEW

In China, Sun (2009) studied CO_2 emission patterns for all 30 provinces using ARIMA models and concluded that by 2010 CO_2 emission in China would be approximately 1990 mmt. In Iran, Lotfalipour *et al* (2013) modeled and predicted CO_2 emissions using Grey and ARIMA models over the period 1965 to 2010 and discovered that the amount of carbon dioxide emissions will reach up to 925.68 million tons in 2020 in Iran. In Bangladesh, Rahman & Hasan (2017), using time series data of 44 years from 1972 - 2015 based on ARIMA models; uncovered that the ARIMA (0, 2, 1) model is the optimal model for modeling and forecasting carbon dioxide in Bangladesh. In another Bangladesh study, Hossain *et al* (2017) analyzed carbon dioxide emissions in Bangladesh using the Box-Jenkins ARIMA technique over the period 1972 - 2013 and concluded that the ARIMA (12, 2, 12), ARIMA (8, 1, 3) and the ARIMA (5, 1, 5) are the best fit models for forecasting CO_2 emission from GFC, LFC and SFC rather the other methods of forecasting – HWNS and ANN models. In Thailand, Pruethsan (2017) examined CO_2 emissions using the VARIMAX technique over the period 2000 - 2015 and discovered that the VARIMAX (2, 1, 2) and VARIMAX (2, 1, 3) models are optimal models for modeling CO_2 emissions in Thailand. This study will make use of the ARIMA technique in modeling and forecasting CO_2 emissions in China.

MATERIALS & METHODS

ARIMA Models

ARIMA models are often considered as delivering more accurate forecasts then econometric techniques (Song *et al*, 2003b). ARIMA models outperform multivariate models in forecasting performance (du Preez & Witt, 2003). Overall performance of ARIMA models is superior to that of the naïve models and smoothing techniques (Goh & Law, 2002). ARIMA models were developed by Box and Jenkins in the 1970s and their approach of identification, estimation and diagnostics is based on the principle of parsimony (Asteriou & Hall, 2007). The general form of the ARIMA (p, d, q) can be represented by a backward shift operator as:

$$\emptyset(B)(1-B)^d CC_t = \theta(B)\mu_t \quad \ldots [1]$$

Where the autoregressive (AR) and moving average (MA) characteristic operators are:

$$\emptyset(B) = \left(1 - \emptyset_1 B - \emptyset_2 B^2 - \cdots - \emptyset_p B^p\right) \quad \ldots [2]$$

$$\theta(B) = \left(1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q\right) \quad \ldots [3]$$

and

$$(1-B)^d CC_t = \Delta^d CC_t \quad \ldots [4]$$

Where $\emptyset$ is the parameter estimate of the autoregressive component, θ is the parameter estimate of the moving average component, Δ is the difference operator, d is the difference, B is the backshift operator and μ_t is the disturbance term.

The Box – Jenkins Methodology

The first step towards model selection is to difference the series in order to achieve stationarity. Once this process is over, the researcher will then examine the correlogram in order to decide on the appropriate orders of the AR and MA components. It is important to highlight the fact that this procedure (of choosing the AR and MA components) is biased towards the use of personal judgement because there are no clear – cut rules on how to decide on the appropriate AR and MA components. Therefore, experience plays a pivotal role in this regard. The next step is the estimation of the tentative model, after which diagnostic testing shall follow. Diagnostic checking is usually done by generating the set of residuals and testing whether they satisfy the characteristics of a white noise process. If not, there would be need for model re – specification and repetition of the same process; this time from the second stage. The process may go on and on until an appropriate model is identified (Nyoni, 2018i).

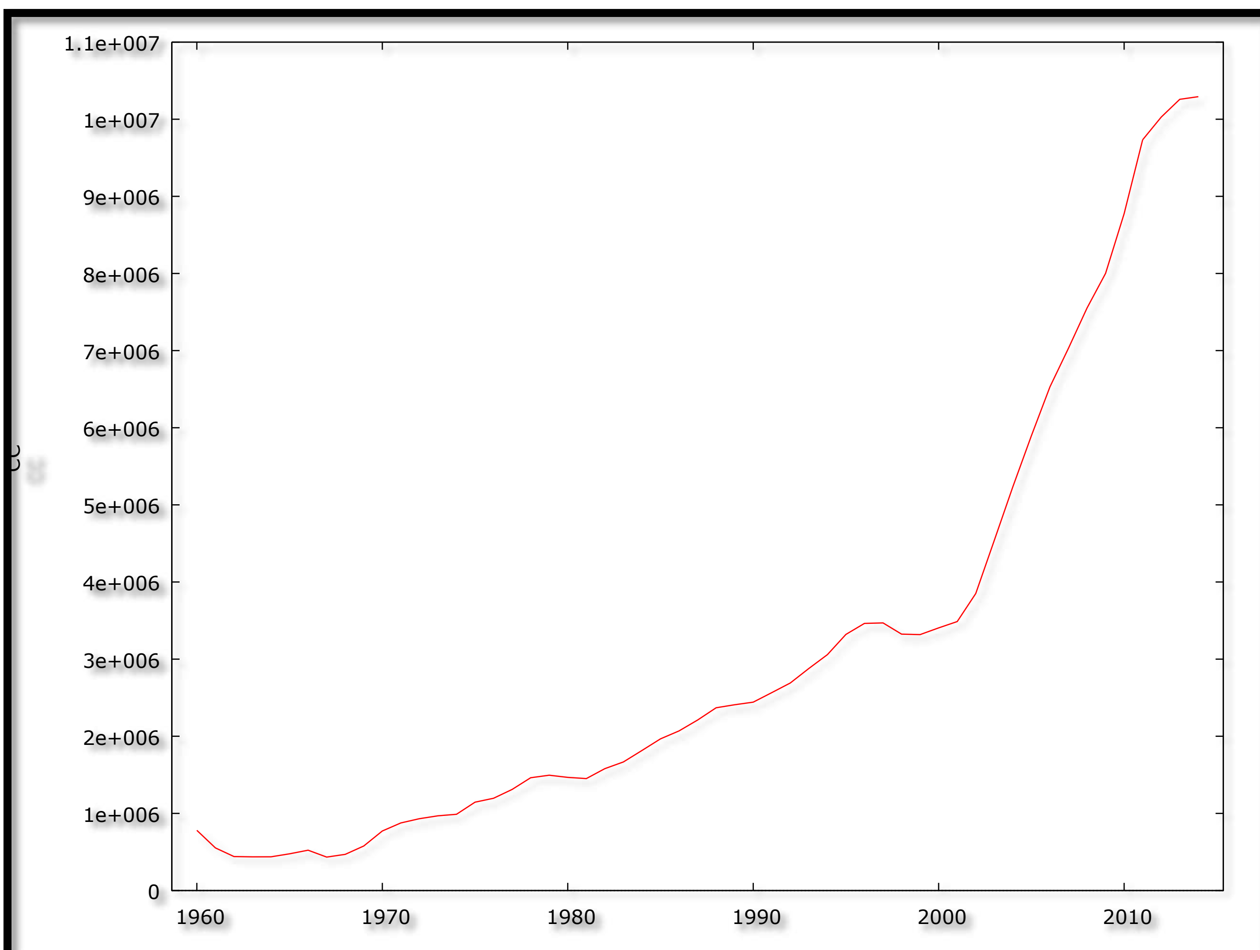
Data Collection

This study is based on 55 observations of annual total carbon dioxide emissions in China, i.e. 1960 – 2014.

Diagnostic Tests & Model Evaluation

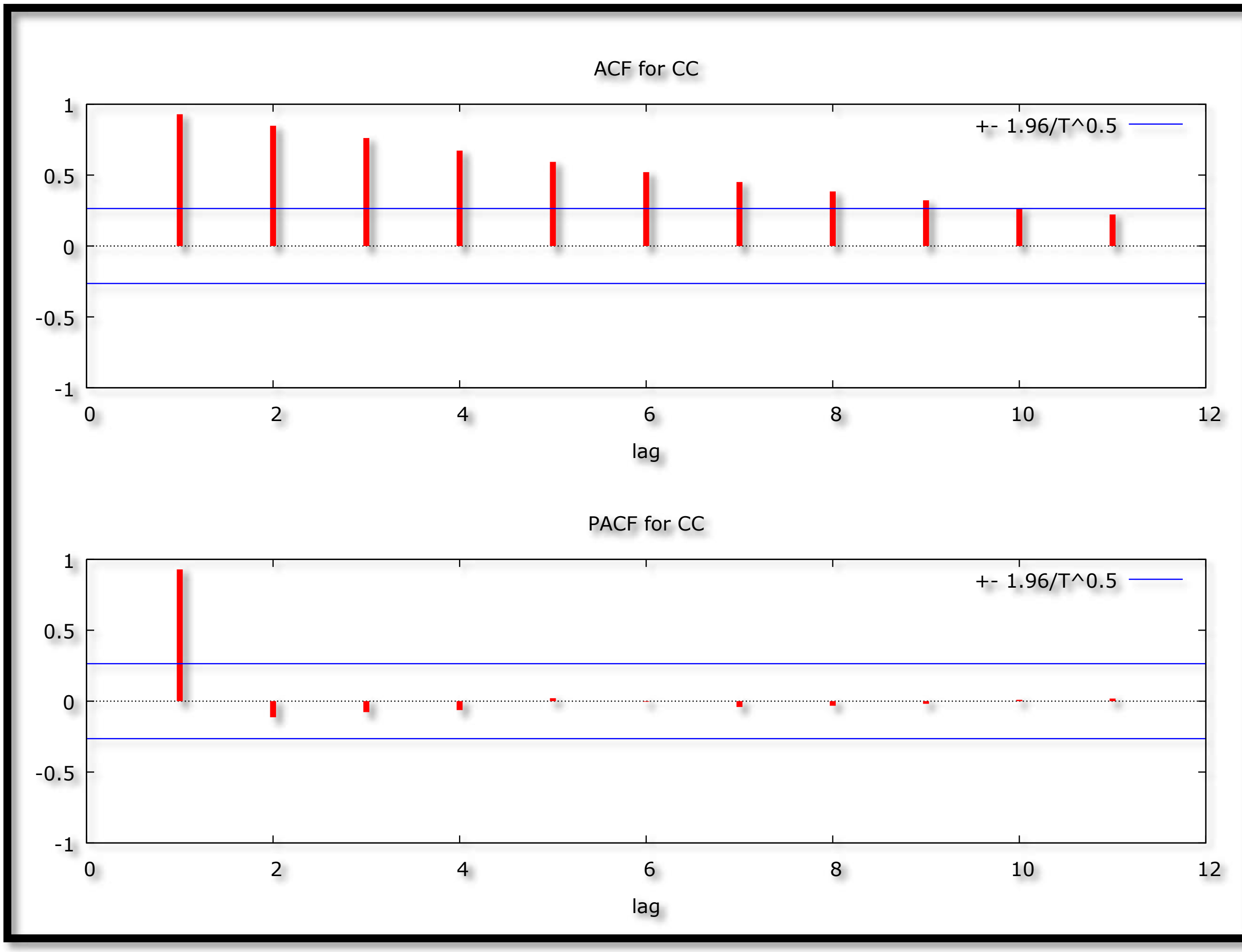
Stationarity Tests: Graphical Analysis

Figure 1



The Correlogram in Levels

Figure 2



The ADF Test

Table 1: Levels-intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
CC	0.054551	0.9591	-3.560019	@1%	Not stationary
			-2.917650	@5%	Not stationary
			-2.596689	@10%	Not stationary

Table 2: Levels-trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
CC	-1.374600	0.8571	-4.140858	@1%	Not stationary
			-3.496960	@5%	Not stationary
			-3.177579	@10%	Not stationary

Table 3: without intercept and trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
CC	1.011003	0.9158	-2.609324	@1%	Not stationary
			-1.947119	@5%	Not stationary
			-1.612867	@10%	Not stationary

The Correlogram (at 1st Differences)

Figure 3

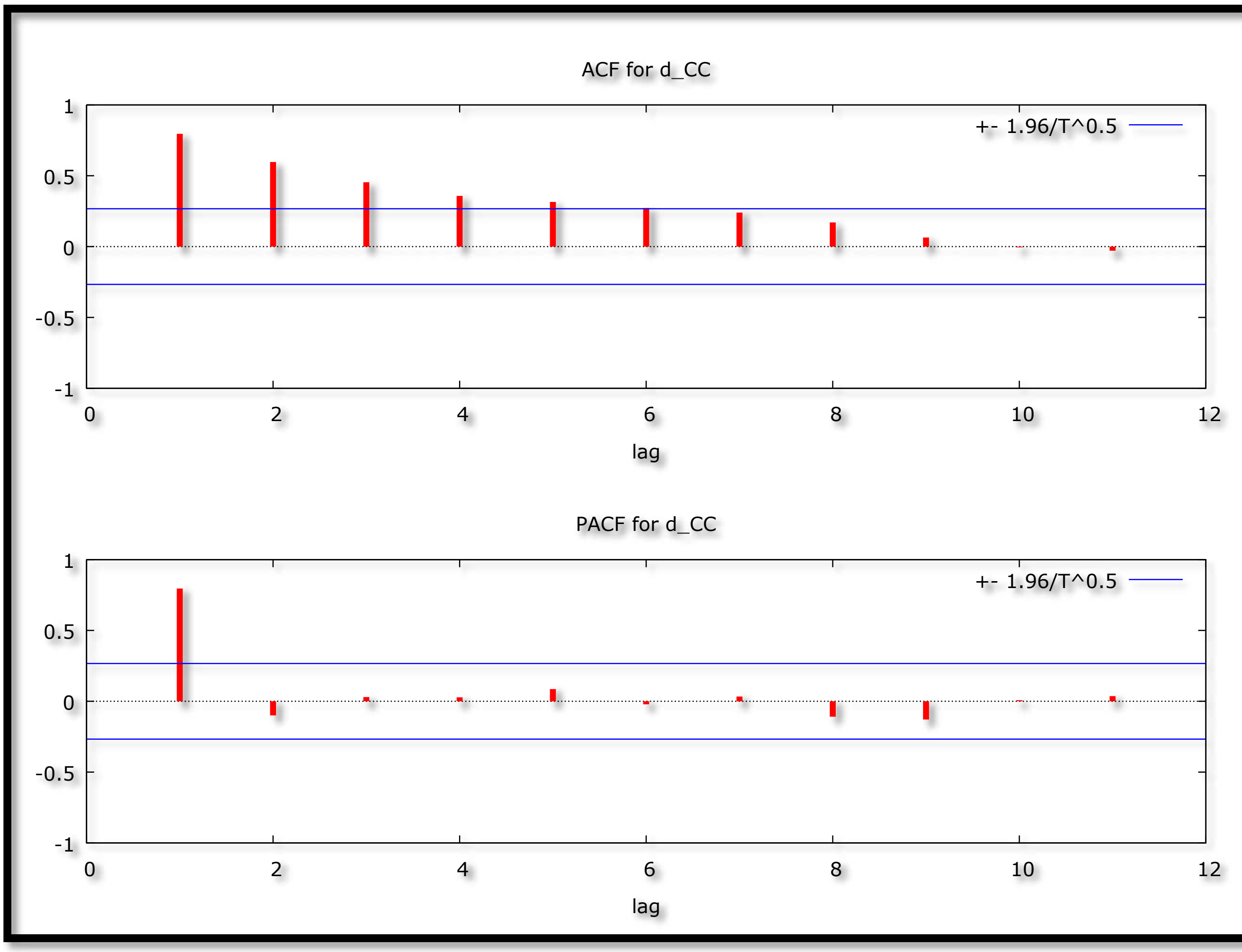


Table 4: 1st Difference-intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
CC	-2.544668	0.1110	-3.560019	@1%	Not stationary
			-2.917650	@5%	Not stationary
			-2.596689	@10%	Not stationary

Table 5: 1st Difference-trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
CC	-2.628233	0.2700	-4.140858	@1%	Not stationary
			-3.496960	@5%	Not stationary
			-3.177579	@10%	Not stationary

Table 6: 1st Difference-without intercept and trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
CC	-1.867940	0.0594	-2.609324	@1%	Not stationary
			-1.947119	@5%	Not stationary
			-1.612867	@10%	Not stationary

Figure 1 – 3 and tables 1 – 6 indicate the CC series is neither I (0) nor I(1).

The Correlogram in (2nd Differences)

Figure 4

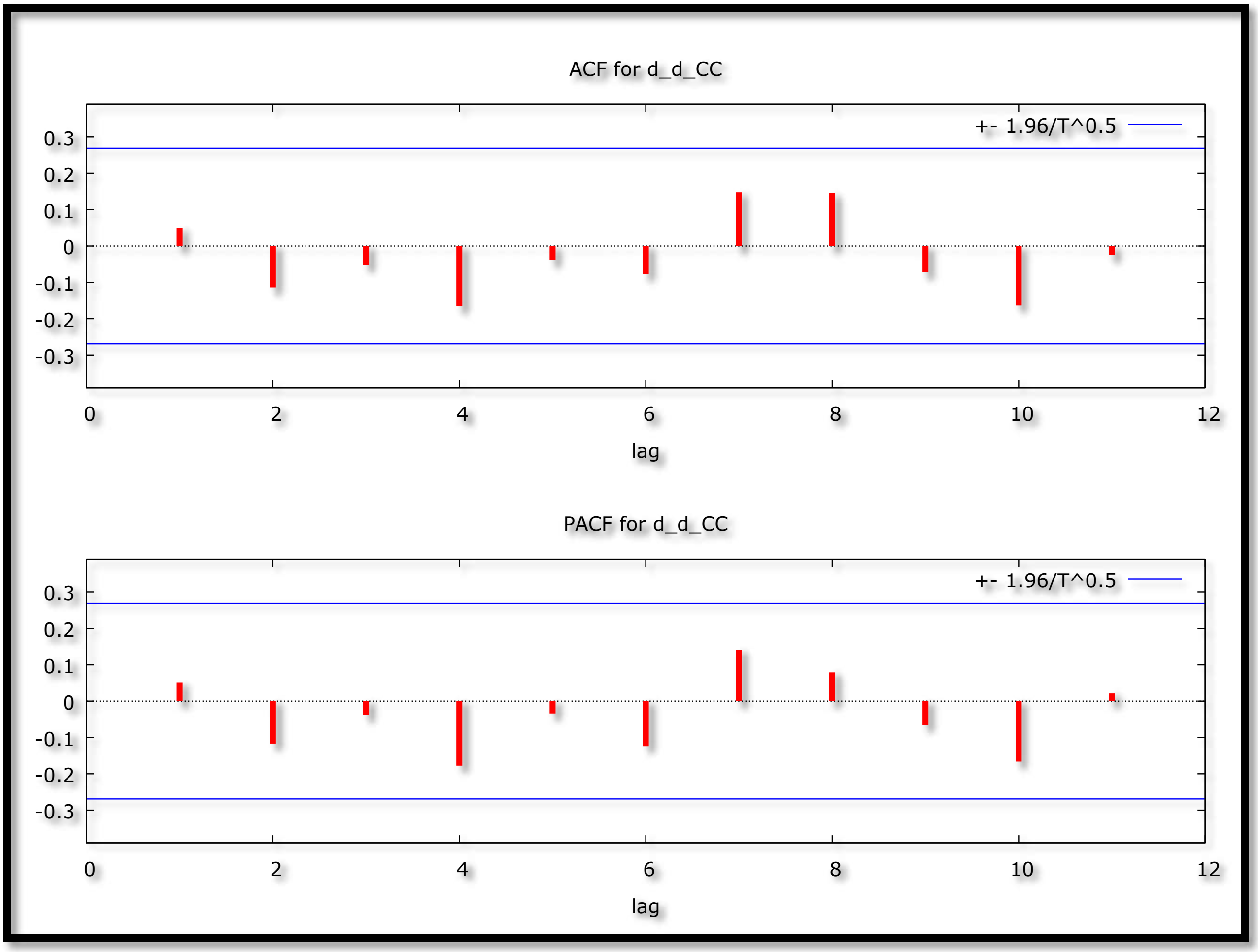


Table 7: 2nd Difference-intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
CC	-6.619111	0.0000	-3.562669	@1%	Stationary
			-2.918778	@5%	Stationary
			-2.597285	@10%	Stationary

Table 8: 2nd Difference-trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
CC	-6.621143	0.0000	-4.144584	@1%	Stationary
			-3.498692	@5%	Stationary
			-3.178578	@10%	Stationary

Table 9: 2nd Difference-without intercept and trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
CC	-6.689658	0.0000	-2.610192	@1%	Stationary
			-1.947248	@5%	Stationary
			-1.612797	@10%	Stationary

Figure 4 and tables 7 – 9 show that the CC series is an I (2) variable.

Evaluation of ARIMA models (without a constant)

Table 10

Model	AIC	U	ME	MAE	RMSE	MAPE
ARIMA (1, 2, 1)	**1413.898**	0.86938	3593.7	93912	141630	5.3361
ARIMA (1, 2, 0)	1414.292	0.86744	4490.7	95662	145180	5.4445
ARIMA (0, 2, 1)	1414.248	0.86757	4396.3	95520	145110	5.4436
ARIMA (1, 2, 2)	1414.976	0.89854	25583	96034	140310	5.509
ARIMA (1, 2, 3)	1416.971	0.89826	25465	96008	140300	5.5053
ARIMA (2, 2, 0)	1415.579	0.86758	5624.4	95262	144190	5.4596
ARIMA (3, 2, 0)	1417.464	0.86697	7125.9	95167	144020	5.4932
ARIMA (2, 2, 2)	1415.179	0.8999	25218	94830	137660	5.5796
ARIMA (0, 2, 2)	1415.098	0.87392	7186	96325	143490	5.5182
ARIMA (0, 2, 3)	1416.996	0.87807	11599	96740	143330	5.5901

A model with a lower AIC value is better than the one with a higher AIC value (Nyoni, 2018n). Theil's U must lie between 0 and 1, of which the closer it is to 0, the better the forecast method (Nyoni, 2018l). The study will consider AIC in order to choose the best model for forecasting CO_2 in China. Therefore, for forecasting annual total CO_2 in Zimbabwe, the ARIMA (1, 2, 1) model is selected.

Residual & Stability Tests

ADF Tests of the Residuals of the ARIMA (1, 2, 1) Model

Table 11: Levels-intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
R_t	-6.074221	0.0000	-3.565430	@1%	Stationary
			-2.919952	@5%	Stationary
			-2.597905	@10%	Stationary

Table 12: Levels-trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
R_t	-6.029017	0.0000	-4.148465	@1%	Stationary
			-3.500495	@5%	Stationary

			-3.179617	@10%	Stationary

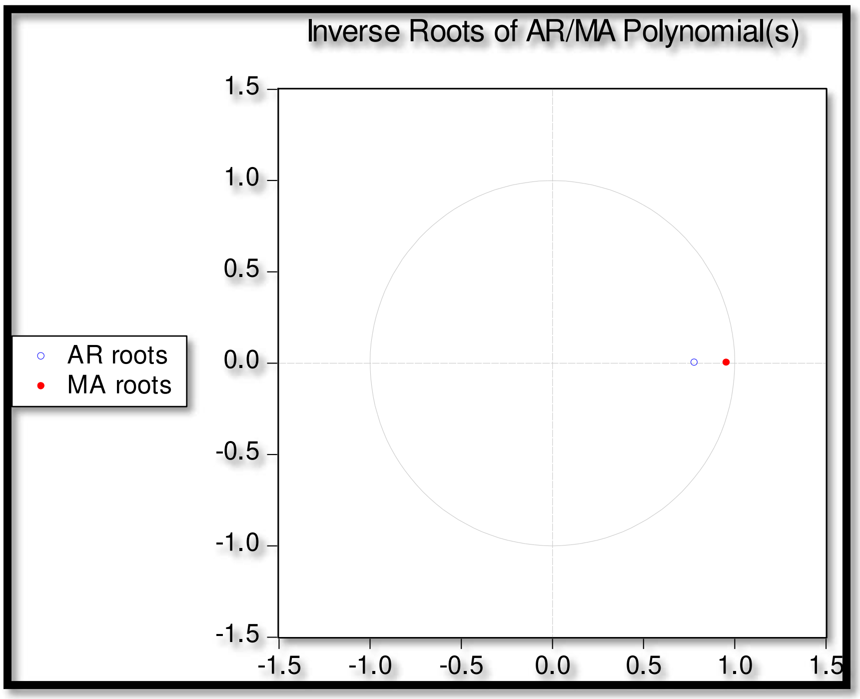
Table 13: without intercept and trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
R_t	-6.095749	0.0000	-2.611094	@1%	Stationary
			-1.947381	@5%	Stationary
			-1.612725	@10%	Stationary

As shown in tables 11 – 13 above, the residuals of the ARIMA (1, 2, 1) model are stationary.

Stability Test of the ARIMA (1, 2, 1) Model

Figure 5



Since the corresponding inverse roots of the characteristic polynomial lie in the unit circle, it illustrates that the chosen best model, the ARIMA (1, 2, 1) model is stable and hence acceptable.

FINDINGS

Descriptive Statistics

Table 14

Description	Statistic
Mean	3098600
Median	2209700
Minimum	433230
Maximum	10292000
Standard deviation	2846100

Skewness	1.3073
Excess kurtosis	0.64279

The mean is positive, i.e. 3098600. The wide gap between the minimum carbon dioxide emission (i.e. 433230) and the maximum carbon dioxide emission (i.e. 10292000) is consistent with the reality that the Chinese carbon dioxide emission series is sharply trending upwards as already shown in figure 1 above. Skewness is 1.3073 and the most essential thing about it is that it is positive, indicating that it is positively skewed and non-symmetric. Kurtosis is 0.64279; indicating that the carbon dioxide emission series is not normally distributed.

Results Presentation[1]

Table 15

ARIMA (1, 2, 1) Model:

$$\Delta^2 CC_{t-1} = -0.729771\Delta^2 CC_{t-1} + 0.914701\mu_{t-1} \ldots\ldots\ldots\ldots\ldots\ldots [5]$$

P: (0.0038) (0.0000)

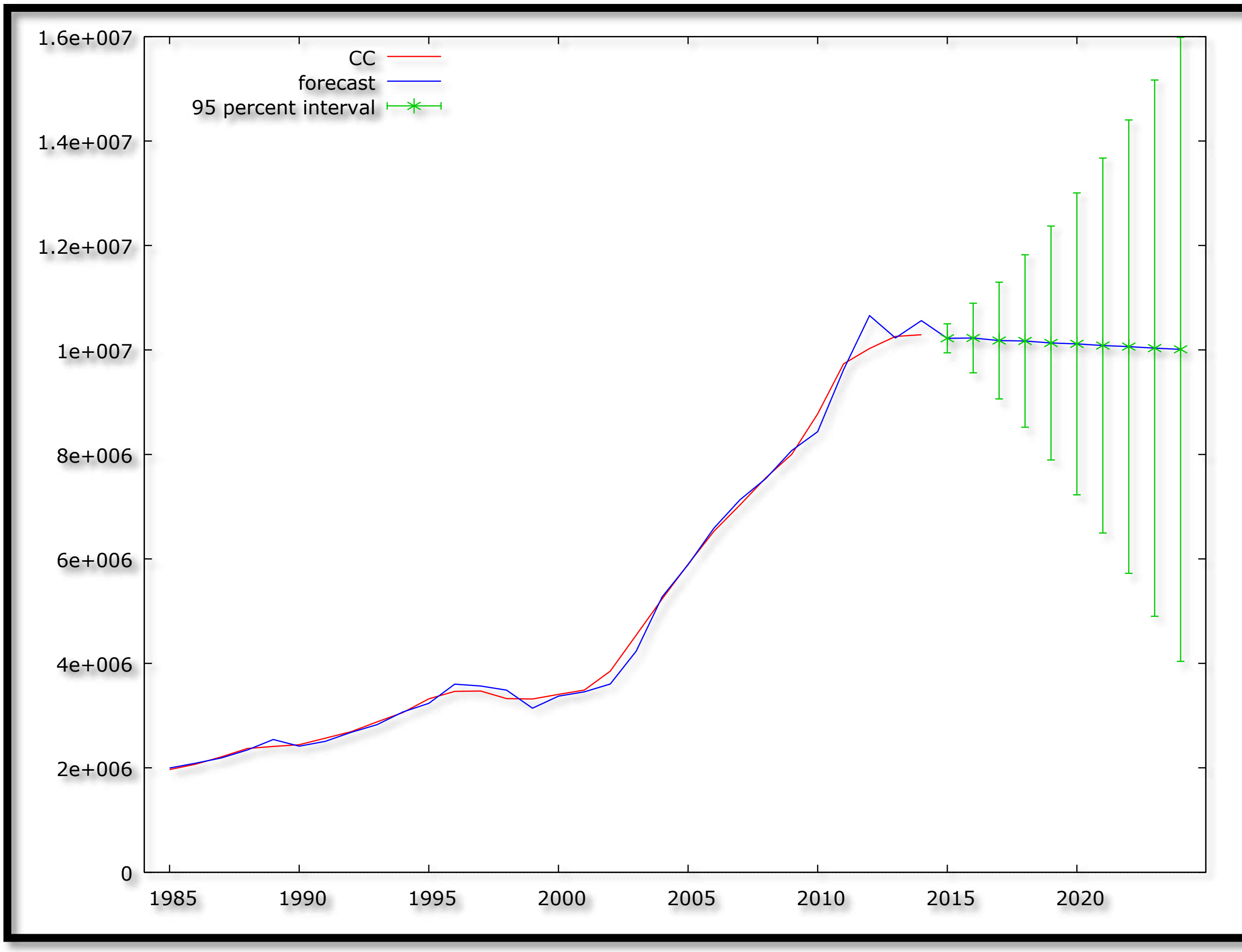
S. E: (0.252021) (0.205541)

Variable	Coefficient	Standard Error	z	p-value
AR (1)	-0.729771	0.252021	-2.896	0.0038***
MA (1)	0.914701	0.205541	4.45	0.0000***

Forecast Graph

Figure 6

[1] *The *, ** and *** means significant at 10%, 5% and 1% levels of significance; respectively.*



Predicted Annual Total CO_2 in China

Table 16

Year	Prediction	Std. Error	95% Confidence Interval
2015	10222792.25	141439.317	9945576.29 - 10500008.22
2016	10228863.89	339864.283	9562742.14 - 10894985.65
2017	10180052.08	569958.705	9062953.54 - 11297150.61
2018	10171292.63	842156.902	8520695.43 - 11821889.82
2019	10133304.11	1142629.104	7893792.22 - 12372816.00
2020	10116646.13	1474489.022	7226700.75 - 13006591.50
2021	10084421.74	1831390.032	6494963.23 - 13673880.24
2022	10063557.26	2214189.762	5723825.07 - 14403289.45

2023	10034402.65	2619551.498	4900176.05 - 15168629.24
2024	10011297.94	3047465.039	4038376.22 - 15984219.66

Figure 5 (with a forecast range from 2015 – 2024) and table 16, clearly show that China's annual total CO_2 emission is likely to rise over the next decade. With a 95% confidence interval of 4038376.22 kt to 15984219.66 kt and a projected annual total CO_2 emission of 10011297.94 kt by 2024, the chosen ARIMA (1, 2, 1) model is apparently sending warning signals to Environmental Economists in China on the need to continue taking action, especially in light of climate change and global warming.

Policy Implications

a) There is need for continued reduction in consumption of fossil fuels in China.
b) There is need to innovate new and more effective energy saving technologies in China.
c) There is also need to continuously educate the Chinese nation on the essence of lower pollution levels.
d) The Chinese government ought to reduce pollution by implementing policy actions such as increasing tax on the polluting companies, especially those that use fossil fuels in their daily production activities.

CONCLUSION

The study shows that the ARIMA (1, 2, 1) model is not only stable but also the most suitable model to forecast annual total CO_2 in China for the next 10 years. The model predicts that by 2024, China's annual total CO_2 emission will be approximately, 10000000 kt. This is a warning signal to Environmental Economists in China, particularly with regards to climate change and global warming. The results of this study are invaluable for the Chinese government, especially when it comes to medium-term and long-term planning.

REFERENCES

[1] Asteriou, D. & Hall, S. G. (2007). Applied Econometrics: a modern approach, Revised Edition, *Palgrave MacMillan*, New York.

[2] Du Preez, J. & Witt, S. F. (2003). Univariate and multivariate time series forecasting: An application to tourism demand, *International Journal of Forecasting*, 19: 435 – 451.

[3] Goh, C. & Law, R. (2002). Modeling and forecasting tourism demand for arrivals with stochastic non-stationary seasonality and intervention, *Tourism Management*, 23: 499 – 510.

[4] Hossain, A., Islam, M. A., Kamruzzaman, M., Khalek, M. A. & Ali, M. A (2017). Forecasting carbon dioxide emissions in Bangladesh using Box-Jenkins ARIMA models, *Department of Statistics*, University of Rajshahi.

[5] Lotfalipour, M. R., Falahi, M. A. & Bastam, M (2013). Prediction of CO_2 emissions in Iran using Grey and ARIMA models, *International Journal of Energy Economics and Policy*, 3 (3): 229 – 237.

[6] Nyoni, T (2018l). Modeling Forecasting Naira / USD Exchange Rate in Nigeria: a Box – Jenkins ARIMA approach, *University of Munich Library – Munich Personal RePEc Archive (MPRA)*, Paper No. 88622.

[7] Nyoni, T (2018n). Modeling and Forecasting Inflation in Kenya: Recent Insights from ARIMA and GARCH analysis, *Dimorian Review*, 5 (6): 16 – 40.

[8] Nyoni, T. (2018i). Box – Jenkins ARIMA Approach to Predicting net FDI inflows in Zimbabwe, *Munich University Library – Munich Personal RePEc Archive (MPRA)*, Paper No. 87737.

[9] Pao, H., Fu, H., & Tseng, C (2012). Forecasting of CO_2 emissions, energy consumption and economic growth in China using an improved grey model, *Energy*, 40 (2012): 400 – 409.

[10] Pruethsan, S (2017). VARIMAX model to forecast the emission of carbon dioxide from energy consumption in rubber and petroleum industries sectors in Thailand, *Journal of Ecological Engineering*, 18 (3): 112 – 117.

[11] Rahman, A & Hasan, M. M (2017). Modeling and forecasting of carbon dioxide emissions in Bangladesh using Autoregressive Integrated Moving Average (ARIMA) models, *Scientific Research Publishing – Open Journal of Statistics*, 7: 560 – 566.

[12] Song, H., Witt, S. F. & Jensen, T. C. (2003b). Tourism forecasting: accuracy of alternative econometric models, *International Journal of Forecasting*, 19: 123 – 141.

[13] Sun, X (2009). Analyze China's CO_2 emission pattern and forecast its future emission, Masters Thesis, *Nicholas School of Environment*, Duke University.