

Modelling and Forecasting of Greenhouse Gas Emissions by the Energy Sector in Kenya Using Autoregressive Integrated Moving Average (ARIMA) Models

Michael Mbaria Chege

Department of Mathematics, University of Nairobi, Nairobi, Kenya

Email: mbariamyk@gmail.com

How to cite this paper: Chege, M.M. (2024) Modelling and Forecasting of Greenhouse Gas Emissions by the Energy Sector in Kenya Using Autoregressive Integrated Moving Average (ARIMA) Models. *Open Journal of Statistics*, 14, 667-676.
<https://doi.org/10.4236/ojs.2024.146029>

Received: October 6, 2024

Accepted: November 2, 2024

Published: November 5, 2024

Copyright © 2024 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The energy sector is the second largest emitter of greenhouse (GHG) gases in Kenya, emitting about 31.2% of GHG emissions in the country. The aim of this study was to model Kenya's GHG emissions by the energy sector using ARIMA models for forecasting future values. The data used for the study was that of Kenya's GHG emissions by the energy sector for the period starting from 1970 to 2022 obtained for the International Monetary Fund (IMF) database that was split into training and testing sets using the 80/20 rule for modelling purposes. The best specification for the ARIMA model was identified using Akaike Information Criterion (AIC), root mean squared error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE) and mean absolute scaled error (MASE). ARIMA (1, 1, 1) was identified as the best model for modelling Kenya's GHG emissions and forecasting future values. Using this model, Kenya's GHG emissions by the energy sector were forecasted to increase to a value of about 43.13 million metric tons of carbon dioxide equivalents by 2030. The study, therefore, recommends that Kenya should accelerate the adjustment of industry structure and improve the efficient use of energy, optimize the energy structure and accelerate development and promotion of energy-efficient products to reduce the emission of GHGs by the country's energy sector.

Keywords

Greenhouse Gases, Energy Sector, Autoregressive Moving Averages Models

1. Introduction

Global warming continues to be experienced across the world in both developed

and developing countries. The Earth's temperature has been raising by as many as 0.36°F every 10 years since 1982, reaching the highest level in 2023 [1]. For this particular year, the global temperature was 2.12°F above the 20th-century average of 57°F and 2.43°F above the pre-industrial average. The increasing concentration of greenhouse gases (GHG) in the atmosphere experienced since the industrial age is what could have led to raising global temperatures and changes in climate patterns. With the doubling of carbon dioxide (CO₂) concentration in the atmosphere, global mean temperature is expected to rise by 3°C to 4°C [2]. GHGs trap heat near the Earth's surface and are crucial for maintaining the Earth's habitable temperature [3]. Excessive emissions of GHGs, however, contribute to GHG overburden, potentially disrupting the Earth's carbon cycle and contributing to global warming. This is because GHGs such as CO₂ are highly permeable to visible light from the sun and highly absorbent to long-wave radiation reflected from the earth [4].

China is among the world's top emitters of GHGs, emitting around 10 billion metric tons annually, accounting for 28.8% of global emissions [4]. Kenya is a relatively low emitter of GHGs, emitting less than 0.1% of the global GHG emissions, but GHG emissions in the country have more than doubled since 1995 [5]. In 2013, total GHG emissions in Kenya were as many as 60.2 million metric tons of CO₂ equivalent, representing 0.13% of global GHG emissions. The agricultural sector emitted 62.8% of the total GHG emissions, the energy sector emitted 31.2% of the total GHG emissions, the industrial processes sector emitted 4.6% of the total GHG emissions and the waste sector emitted 1.4% of the total GHG emissions [6]. Although the country is not among the ten largest emitters of GHGs in the atmosphere, it has a goal of reducing GHG emissions by 30% relative to business-as-usual levels by 2030 outlined in its Intended Nationally Determined Contribution (INDC) [5]. The present study aimed to effectively model Kenya's GHG emissions by the energy sector for forecasting future values as accurately as possible using Autoregressive Integrated Moving Averages (ARIMA) models. This is intended to identify how GHG emissions by the energy sector in Kenya are likely to progress in the future as the country continues to use renewable energy as an alternative source of energy.

ARIMA models that are simple and require only endogenous variables without the need for other exogenous variables have been used severally for modelling emissions of GHGs to forecast future values. Ning *et al.* [4] used ARIMA models to model annual emissions of CO₂ in China for forecasting future values. The data used was for the period from 1997 to 2017, obtained from the China Carbon Emissions Database, and analyzed using EViews. Specific ARIMA model identified as the best for modelling the emission of CO₂ in China was ARIMA (2, 2, 0) model. Using this particular model, CO₂ emissions were predicted for 2018, 2019 and 2020 in Beijing, Henan, Guangdong and Zhejiang. Rahman & Hasan [7], on the other hand, used ARIMA models to model annual CO₂ emissions in Bangladesh for forecasting future values. The data used was for the period starting from 1972

to 2015 analyzed using R. Specific ARIMA model identified as being the best in the modelling of CO₂ emission in Bangladesh to forecast future values was ARIMA (0, 2, 1) model. Using this particular model, CO₂ emission in Bangladesh was forecasted to reach 83.947 metric tons in 2016, 89.905 metric tons in 2017 and 96.286 metric tons in 2018.

2. Materials and Methods

2.1. Data

This particular study used annual data on GHGs emissions from the energy sector in Kenya for the period starting from 1970 to 2022 obtained from the International Monetary Fund database. Observations in the data were 53 without any missing value. The data was visually examined for stationarity using time series, autocorrelation function (ACF) and partial autocorrelation function (PACF) plots before being empirically assessed for stationarity using the Augmented Dickey-Fuller (ADF) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests. Non-stationary time series was differenced until stationarity was achieved. For modelling purposes, the sample was split into training set having 80% of the observations ($n = 42$) and testing set having 20% of the observations ($n = 11$).

Collection of data for the IMF database is through direct reporting of official statistics by countries to the IMF statistics department or through the Fund's area departments that collect data from country authorities or from commercial sources in the course of their regular bilateral surveillance or IMF-program-related activities [8]. There is a fixed calendar that authorities in all countries of the world need to follow when submitting data to the IMF statistics department that is usually monthly or quarterly, but IMF statistics department sometimes collects information directly from official websites for some variables. Frameworks used for collecting and structuring data on anthropogenic GHG emissions are the System of Environmental Economic Accounts (SEEA) and the UN Framework Convention on Climate Change (UNFCCC) inventory for GHG emissions [9]. These two frameworks follow a direct recording principle, which means that emissions are recorded at the level of processes or industries where they are released and estimate emissions directly through emissions monitoring or indirectly through the use of emission factors. The two frameworks also report GHG emissions in metric tons of CO₂ equivalents as the amount of CO₂ emissions having the same global warming potential as one metric ton of a particular GHG.

IMF's statistics department performs quality checks on the submitted data that include tests for compliance with established formats, examinations for outliers and broad cross-sector consistency checks intended to identify large discrepancies across the datasets [8]. IMF's statistics department also updates the data frequently after being initially uploaded to address data inadequacies identified in the course of policy discussions between IMF, its mission teams and country authorities with updates originating solely from official sources. In some cases, IMF's missions spend substantial share of their time in the field collecting and

double-checking aspects of the data through tasks such as verifying data in the primary sources and checking the accuracy of basic calculations and their consistency with methodological standards [9]. This makes reliability of the GHG emissions data in the IMF database to be adequate, which enhances credibility of findings obtained from this particular study.

2.2. Model

Specific univariate time series models used to model Kenya's GHG emissions by the energy sector were the ARIMA models. This group of univariate time series models explain a time series variable using its past values and/or lagged values of the stochastic error terms [10]. The models are called ARIMA models because of containing an autoregressive (AR) part capturing autocorrelation in a time series variable using its values for the previous periods, an order of differencing (I) part indicating the number of times that the time series variable needs to be differences to become stationary and a moving averages (MA) part capturing autocorrelation in a time series variable using lagged values of the stochastic error terms. A general mathematical presentation of ARIMA (p, d, q) model estimated for a time series variable y_t like the one on Kenya's GHG emissions by the energy sector is as given by the equation below. In the equation, Δ^d is the order of differencing needed to make the non-stationary time series analyzed stationary, $\beta_i, i = 1, 2, \dots, p$ are coefficients of the AR part of the model and $\gamma_j, j = 1, 2, \dots, q$ are coefficients of the MA part of the model.

$$\Delta^d y_t = \sum_{i=1}^p \beta_i \Delta^d y_{t-i} + \sum_{j=1}^q \gamma_j u_{t-j} + u_t \quad (1)$$

A certain particular ARIMA model was selected as the best model for modelling Kenya's GHG emissions by the energy sector because of having smallest values of root mean squared error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), mean absolute scaled error (MASE) and Akaike information criterion (AIC). AIC was used instead of Bayesian Information Criterion (BIC) because of being more appropriate in finding the best model for predicting future observations of a univariate time series variable like the one of Kenya's GHG emissions by the energy sector [11]. RMSE, MAE, MAPE and MASE, on the other hand, were used because of their ability to measure the in-sample or out-sample predictive and forecast accuracy of a model when compared to other models considered [12].

Initial ARIMA model considered was identified using patterns that were evident in the ACF and PACF plots of the stationary time series on Kenya's GHG emissions by the energy sector. Other ARIMA models considered were obtained by adding AR or MA terms to the initial ARIMA model considered to ensure that residuals of the models were independent and approximately normally distributed with a variance that is constant and achieve the best fit possible and the highest forecasting accuracy possible. Residuals of the models were assessed for independence using Ljung-Box test, variance that was constant using ARCH test and normality

using the Shapiro-Wilk test. All statistical analyses were conducted using R program.

3. Results and Discussion

For the period investigated, Kenya's GHG emissions by the energy sector measured in million metric tons of CO₂ equivalent ranged from 7.37 to 34.24 ($M = 17.21$, $SD = 8.27$) with a distribution that was skewed positively (Skewness = 0.70). A natural logarithm transformation was applied to the time series variable to reduce the skewness. **Figure 1** is the time series plot, ACF plot and PACF plot obtained for the variable after the transformation. For the period investigated, Kenya's GHG emissions by the energy sector had an increasing trend making the variable not to be stationary. Results obtained from the ADF test for stationarity indicated that the time series variable on Kenya's GHG emissions by the energy sector was not stationary because of having a unit root (ADF = -2.016, $p = 0.568$). Results obtained from the KPSS test for stationarity, on the other hand, indicated that the time series variable was also non-stationary because of having a deterministic trend (KPSS = 1.408, $p = 0.010$).

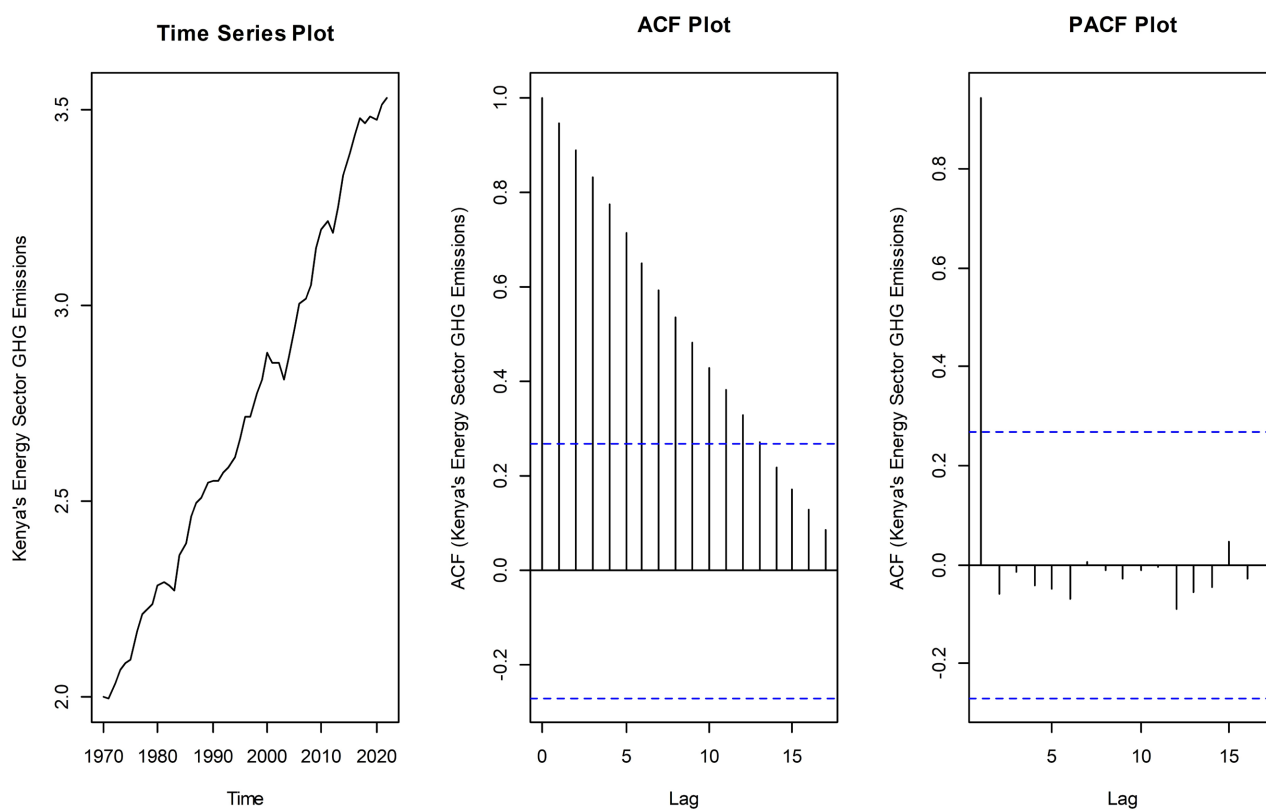


Figure 1. Time series, ACF and PACF plots of Kenya's energy sector GHG emissions.

Figure 2 is the time series plot, ACF plot and PACF plot obtained for the first differences of Kenya's GHG emissions by the energy sector. For the period investigated, first differences of Kenya's GHG emissions by the energy sector do not

appear to have a trend that could make the variable not to be stationary. Results obtained from the ADF test for stationarity confirmed that first difference of Kenya's GHG emissions by the energy sector was stationary because of not having a unit root ($ADF = -3.640$, $p = 0.038$). Results obtained from the KPSS test for stationarity, on the other hand, confirmed that the first difference of Kenya's GHG emissions by the energy sector was stationary because of not having a deterministic trend ($KPSS = 0.081$, $p = 0.100$).

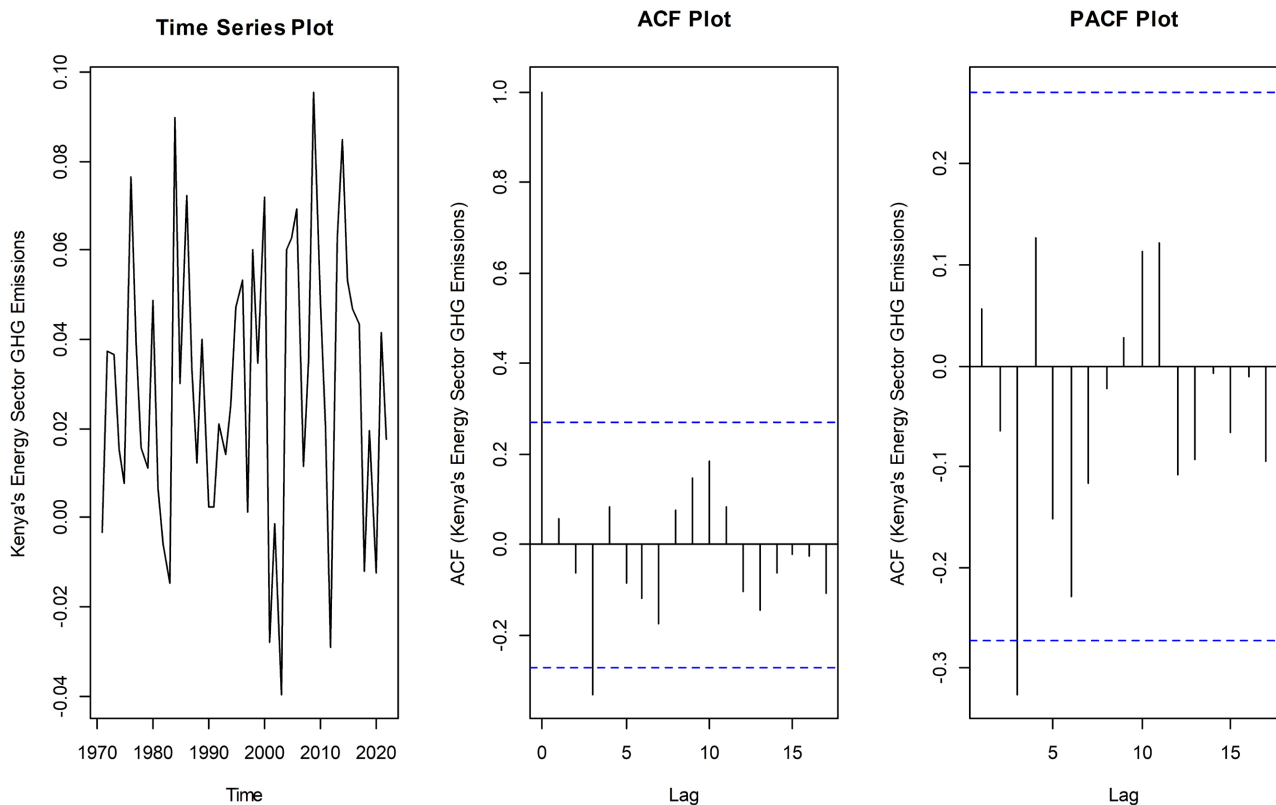


Figure 2. Time series, ACF and PACF plots of first differences of Kenya's energy sector ghg emissions.

ACF plot with significant spikes at lag 0 and lag 3 and PACF plot with significant spike only at lag 3 for first differences of Kenya's GHG emissions by the energy sector. This does not show any pure AR or MA process. As a starting point, ARIMA (0, 1, 1) was considered because MA part appeared to dominate in the long run as indicated by a decaying PACF and an ACF that cuts off at lag 3. Other ARIMA models considered were ARIMA (1, 1, 1), ARIMA (3, 1, 1), ARIMA (1, 1, 3), ARIMA (3, 1, 3), ARIMA (1, 1, 4) and ARIMA (3, 1, 4) because of a possible need to include AR terms and other higher order MA terms in the modelling of Kenya's GHG emissions by the energy sector to make the residuals obtained to be independent and normally distributed with a variance that is constant.

Table 1 contains the findings obtained for the analyses conducted for assessing the best specification for the ARIMA model.

The best ARIMA model for modelling Kenya's GHG emissions by the energy

sector among the ARIMA models considered is ARIMA (1, 1, 1) because of not only having the smallest AIC (AIC = -158.50) indicating an in-sample model fit that is better than that of the other models considered, but also the smallest RMSE (RMSE = 0.047), MAE (MAE = 0.039), MAPE (MAPE = 1.151), MASE (MASE = 1.142), indicating an out-sample forecasting accuracy that is higher than that of the other models considered. ARIMA (1, 1, 1) model also had residuals that were independent (Ljung-Box $\chi^2(1) = 0.027, p = 0.870$) and approximately normally distributed (S-W = 0.987, $p = 0.916$) with a variance that is constant (ARCH LM Test $\chi^2(12) = 9.277, p = 0.679$), which indicates applicability and stability of the model for modelling GHG emissions by the energy sector in Kenya to forecast future values.

Table 1. RMSE, MAE, MAPE, MASE and AIC of the ARIMA models considered.

Models	Model Selection Criteria				
	RMSE	MAE	MAPE	MASE	AIC
ARIMA (0, 1, 1)	0.220	0.199	5.743	5.812	-144.85
ARIMA (1, 1, 1)	0.047	0.039	1.151	1.142	-158.50
ARIMA (3, 1, 1)	0.169	0.151	4.367	4.420	-149.04
ARIMA (1, 1, 3)	0.230	0.208	5.993	6.067	-147.01
ARIMA (3, 1, 3)	0.048	0.040	1.174	1.165	-154.37
ARIMA (1, 1, 4)	0.183	0.162	4.671	4.733	149.61
ARIMA (3, 1, 4)	0.051	0.043	1.266	1.260	152.65

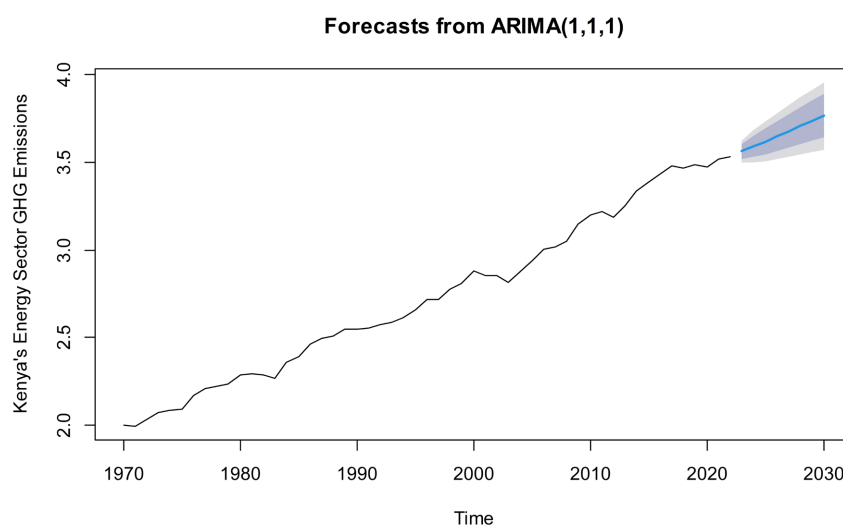


Figure 3. Kenya's GHG emissions by the energy sector forecasted for 2023 to 2030 period.

Figure 3 shows the Kenya's GHG emissions by the energy sector forecasted for the period starting from 2023 to 2030. Kenya's GHG emissions by the energy sector are forecasted to continue increasing for the period starting from 2023 to 2030 to a value of about 35.24 million metric tons of CO₂ equivalent (95% PI [33.10,

37.52]) in 2023 and 43.13 million metric tons of CO₂ equivalent (95% PI = [35.66, 52.15]) in 2030.

4. Discussion and Conclusion

The aim of this study was to assess how Kenya's GHG emissions by the energy sector could be modelled using ARIMA models for forecasting future values as accurately as possible. Findings obtained indicated that ARIMA (1, 1, 1) model was the best model for modelling Kenya's GHG emissions by the energy sector to forecast future values among the ARIMA models considered. Findings obtained also indicated that Kenya's GHG emissions by the energy sector were likely to continue increasing in the future to a value of about 43.13 million metric tons of CO₂ equivalents by 2030 if the current trend continues.

The data used for the study was from a highly credible source and the model used to model it for forecasting future values had not violated any of its assumptions. The model used for modelling the data, however, failed to consider the influence of other factors such as population size, energy consumption, policy changes, technological advancement and economic patterns on GHG emissions by the energy sector in Kenya. GHG emissions by the energy sector in Kenya are likely to increase with the increase in population size, energy consumption and economic activities [13]. This is because GHGs are emitted as individuals undertake their day-to-day social, recreational and economic activities. GHG emissions by the energy sector in Kenya are, however, likely to decrease with policy changes and technological advancement because many inventions made in technologies and changes made in policies are intended to reduce GHG emissions by various sectors in a country [14]. The model identified in this paper is, therefore, appropriate for forecasting Kenya's GHG emissions by the energy sector only under conventional circumstances. In case of special circumstances such as a major breakthrough in new renewable energy sources with low emissions of greenhouse gases, the model is likely to be inaccurate in forecasting Kenya's GHG emissions by the energy sector.

Based on the findings obtained, this particular study concludes that Kenya's GHG emissions by the energy sector can be effectively modelled using the ARIMA (1, 1, 1) model to forecast its future values that are forecasted to continue increasing over years in the future to reach a value of about 43.13 million metric tons of CO₂ equivalent by 2030. The study, therefore, recommends that Kenya should accelerate the adjustment of industry structure and improve the efficient use of energy, optimize the energy structure and accelerate development and promotion of energy-efficient products to reduce emission of GHGs by the country's energy sector. This would enable the country to achieve the goal of reducing GHG emissions by 30% relative to business-as-usual levels by 2030 outlined in its Intended Nationally Determined Contribution (INDC). Further research is, however, necessary for identifying how Kenya's GHG emissions by the energy sector could be modelled for forecasting future values by considering previous values of GHG

emissions by the energy sector and current and previous values of variables such as population size, energy consumption, policy changes, technological advancement and economic patterns. For such a study, variance autoregressive (VAR) models and vector error correction (VEC) models should be considered depending on whether the variables analyzed are co-integrated of order one.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Lindsey, R. and Dahlman, L. (2024) Climate Change: Global Temperature (Understanding Climate). Science & Information for Climate Smart Nation. <http://www.climate.gov/news-features/understanding-climate/climate-change-global-temperature>
- [2] Reddy, V.R., Reddy, K.R. and Acock, B. (1995) Carbon Dioxide and Temperature Interactions on Stem Extension, Node Initiation, and Fruiting in Cotton. *Agriculture, Ecosystems & Environment*, **55**, 17-28. [https://doi.org/10.1016/0167-8809\(95\)00606-s](https://doi.org/10.1016/0167-8809(95)00606-s)
- [3] Li, Z., Gan, B., Li, Z., Zhang, H., Wang, D., Zhang, Y., *et al.* (2023) Kinetic Mechanisms of Methane Hydrate Replacement and Carbon Dioxide Hydrate Reorganization. *Chemical Engineering Journal*, **477**, Article ID: 146973. <https://doi.org/10.1016/j.cej.2023.146973>
- [4] Ning, L., Pei, L. and Li, F. (2021) Forecast of China's Carbon Emissions Based on ARIMA Method. *Discrete Dynamics in Nature and Society*, **2021**, Article ID: 1441942. <https://doi.org/10.1155/2021/1441942>
- [5] World Bank (2023) Climate Action Key to Kenya's Upper-Middle-Income Country Aspirations. World Bank Group. <https://www.worldbank.org/en/news/press-release/2023/11/16/climate-action-key-to-kenya-s-upper-middle-incomeafe-1123-country-aspirations>
- [6] Climate Links (2017) Greenhouse Gas Emissions Factsheet: Kenya. <https://www.climatelinks.org/resources/greenhouse-gas-emissions-factsheet-kenya>
- [7] Rahman, A. and Hasan, M.M. (2017) Modeling and Forecasting of Carbon Dioxide Emissions in Bangladesh Using Autoregressive Integrated Moving Average (ARIMA) Models. *Open Journal of Statistics*, **7**, 560-566. <https://doi.org/10.4236/ojs.2017.74038>
- [8] Jerven, M. (2016) Data and Statistics at the IMF: Quality Assurance for Low-Income Countries. Background Paper BP/16/06; IEO Background Paper, Independent Evaluation Office, International Monetary Fund.
- [9] Astolfi, R., Baptista, N.D.S., Bhanumati, P., de Haan, M., Moll, S., Pegoue, A., Quadrelli, R. and Ribarsky, J. (2023) Quarterly Greenhouse Gas Emissions by Economic Activity. In: Arslanalp, S., Kostial, K. and Quiros-Romero, G., Eds., *Data for a Greener World*, International Monetary Fund, 1-22.
- [10] Schaffer, A.L., Dobbins, T.A. and Pearson, S. (2021) Interrupted Time Series Analysis Using Autoregressive Integrated Moving Average (ARIMA) Models: A Guide for Evaluating Large-Scale Health Interventions. *BMC Medical Research Methodology*, **21**, Article No. 58. <https://doi.org/10.1186/s12874-021-01235-8>
- [11] Chakrabarti, A. and Ghosh, J.K. (2011) AIC, BIC and Recent Advances in Model

- Selection. In: Bandyopadhyay, P.S. and Forster, M.R., Eds., *Philosophy of Statistics*, Elsevier, 583-605. <https://doi.org/10.1016/b978-0-444-51862-0.50018-6>
- [12] Siamba, S., Otieno, A. and Koech, J. (2023) Application of ARIMA, and Hybrid ARIMA Models in Predicting and Forecasting Tuberculosis Incidences among Children in Homa Bay and Turkana Counties, Kenya. *PLOS Digital Health*, **2**, e0000084. <https://doi.org/10.1371/journal.pdig.0000084>
- [13] Ahmed, M., Huan, W., Ali, N., Shafi, A., Ehsan, M., Abdelrahman, K., *et al.* (2023) The Effect of Energy Consumption, Income, and Population Growth on CO₂ Emissions: Evidence from NARDL and Machine Learning Models. *Sustainability*, **15**, Article 11956. <https://doi.org/10.3390/su151511956>
- [14] Carrilho-Nunes, I. and Catalão-Lopes, M. (2022) The Effects of Environmental Policy and Technology Transfer on GHG Emissions: The Case of Portugal. *Structural Change and Economic Dynamics*, **61**, 255-264. <https://doi.org/10.1016/j.strueco.2022.03.001>