

A Mixed Model Approach to Forecasting CO₂ Emissions from Sudan's Energy Sector

Maria H. Mohamed^{1,2}, Altaiyb Omer Ahmed Mohammed¹, Mohammedelameen Eissa Qurashi¹, Mubarak H. Elhafian¹

¹Department of Statistics, Faculty of Science, Sudan University of Science & Technology, Khartoum, Sudan

²College of Business Administration, University of Hail, Hail, Saudi Arabia

Email: marhamidmohamed@gmail.com

How to cite this paper: Mohamed, M.H., Mohammed, A.O.A., Qurashi, M.E. and Elhafian, M.H. (2026) A Mixed Model Approach to Forecasting CO₂ Emissions from Sudan's Energy Sector. *Journal of Data Analysis and Information Processing*, **14**, 278-287.

<https://doi.org/10.4236/jdaip.2026.142014>

Received: July 15, 2025

Accepted: May 27, 2026

Published: May 30, 2026

Copyright © 2026 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 study models annual (1970-2023) percentages of energy-related CO₂ emissions in Sudan using ARIMA for the mean process and ARCH for conditional variance. After differencing to attain stationarity, several ARIMA ranks are compared; residual heteroskedasticity motivates an ARCH extension. The authors finally adopt an ARCH (2) specification and generate 10-year forecasts. They conclude that the ARCH (2) model better captures volatility and yields more reliable predictions than the tested ARIMA alternatives.

Keywords

ARCH, ARIMA, Dioxide Emissions, Energy Sector

1. Introduction

The energy sector is one of the most important economic sectors. It contributes significantly to the economic development of countries. With the large population growth and the population trend towards urbanization, the need for energy has increased [1]. Energy plays a pivotal role in poverty reduction, women's empowerment, sustainable development, and public health. Managing energy demand and reducing greenhouse gas emissions are among the most important challenges facing many countries [2]. The energy sector in Sudan is facing several major problems, including power outages and weak energy infrastructure [3]. There are some environmental impacts resulting from obtaining energy sources, such as deforestation and land degradation resulting from obtaining firewood. This leads to environmental degradation and energy scarcity [4]. In general, energy sources are divided into two types: conventional energy (biomass, petroleum products, and electricity) and non-conventional energy (solar energy, wind energy, hydropower,

etc.). Sudan enjoys a relative abundance of sunlight, solar radiation, moderate wind speeds, hydropower, and biomass energy [5].

2. Method

The ARCH (p) model is the first model of conditional autoregressive heteroskedasticity, where changes in volatility over time can be modeled [6]. The autoregressive conditional heteroscedasticity model was proposed by Engle in 1982 for use in modeling the heteroscedasticity of a time series [7]. Therefore, it is important to consider the fact that the conditional variance may be significantly influenced by the squared values of the residual series from previous periods [8]. This allows us to clarify the conditional heteroskedasticity in the data series $(\varepsilon_{t-1}^2, \varepsilon_{t-2}^2, \varepsilon_{t-p}^2)$ and explain the persistence of volatility within it [9].

The ARCH (p) model can be defined according to the following formula [10]:

$$\varepsilon_t = \sqrt{\alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_p \varepsilon_{t-p}^2} \quad (1)$$

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_p \varepsilon_{t-p}^2 \quad (2)$$

$$\varepsilon_t = z_t \sqrt{h_t} \quad (3)$$

where,

$$\alpha_0 > 0, \alpha_i \geq 0 \quad i = 1, \dots, p$$

z_t : They are independent random variables that follow a standard normal distribution for the time series ε_t and it has the following properties:

$$z_t \cong iid(0,1)$$

$$E(z_t^2) = \sigma_z^2, E(z_t) = 0$$

In general

z_t represents a set of independent random variables with a standard normal distribution.

h_t represents a positive linear function of the squares of past observations for ε_t , or

$$(\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-p})$$

These models are characterized by having a mean equal to zero, with variances that are non-constant and conditional on the past. In this way, a regression model with errors following the ARCH model has been introduced [11].

This model and its various developments are considered an important means of describing change over time [12].

3. Results and Discussion

The data used in this research represents a time series of (54) observations for a number of variables related to carbon dioxide emissions in Sudan for the period from 1970 to 2023, including energy, which were obtained from the official website of the World Bank. The percentage of carbon dioxide emissions from the en-

ergy sector is 7.4% million tons out of the total emissions across all sectors.

Figure 1 shows that the behavior of the chain is non-linear and tends to be exponential with no stationarity in the variable. To test the stationarity, the unit root test is used. The unit root test is used to examine the properties of the chain and to ensure the stationarity of the series, and to determine the rank of integration and the rank of differences using the Dickey-Fuller test.

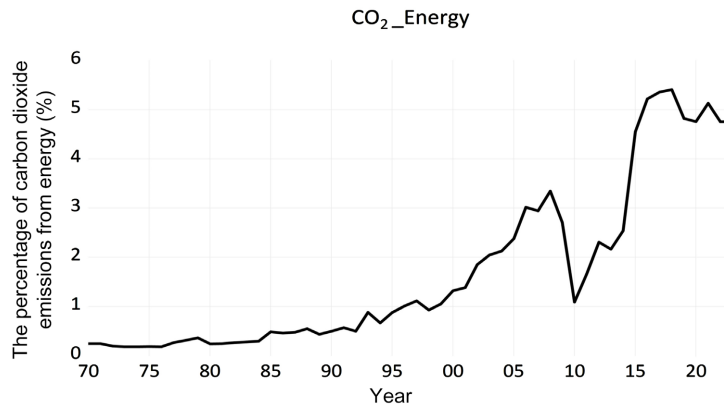


Figure 1. Percentage of carbon dioxide emissions from energy.

Table 1 shows the result of the Dickey-Fuller expanded test to test the stationarity of the series of the percentage of carbon dioxide emissions emitted from energy. The test was carried out at level under three specifications: intercept and trend, intercept only, and without intercept and trend. The results show that the series is stationary for the intercept and trend, but not for the trend and without. In general, we say that the series is non-stationary (the series must stabilize in all its stages).

Table 1. Dickey-fuller test.

Dickey-Fuller test	level		
	Intercept	Intercept and trend	without
t	-0.252776	-2.731415	0.800537
sig	0.9245	0.2289	0.8825
decision	Non significant	Non significant	Non significant
stationary	Non stationary	Non stationary	Non stationary

Since the series was non-stationary, it is necessary to take the first difference and re-test the extended Dickey-Fuller to see if the series is stationary at the first difference or not.

After taking the first difference (**Table 2**), all the results for the series were stationary under all specifications: intercept and trend, intercept only, and without intercept and trend.

Table 2. Dickey-Fuller test after taking the first difference.

Dickey-Fuller test	First Difference		
	intercept	intercept and trend	without
t	-5.873174	-5.462677	-5.745744
sig	0.0000	0.0003	0.0000
decision	significant	significant	significant
stationarity	stationary	stationary	stationary

After we have confirmed that the series is stationary, the next step is to test the autocorrelation and partial autocorrelation of the series to determine the rank of the model.

Figure 2 helps us to verify the stationarity of the series, determine the rank of the model and discover trends. It is clear that the limits of the values in each of the autocorrelation do not exceed one, but in partial autocorrelation it can be 2, so we propose to test the model of the different ARIMA rank for values 0 - 1 - 2 alternately to have 8 models from which to choose the best using the common methods (the lowest value for each of the Akaike criterion, the Bais criterion, the average criterion of absolute error, root mean of error squares, and the largest value of the coefficient of determination).

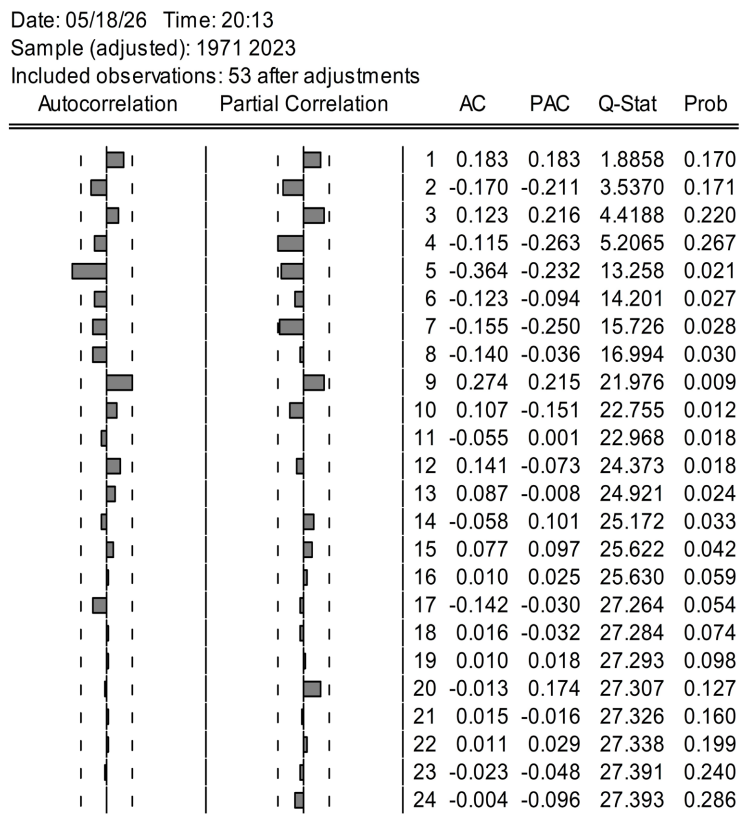


Figure 2. Autocorrelation-partial autocorrelation of the series.

Table 3 shows the criteria for determining the best model. firstly we must make sure that the model is significant and then make sure that the estimated parameters are significant so that after that we have the right to choose the best model from **Table 4** we note that all models in red were insignificant except only model ARIMA (0, 1, 1) ARIMA and model (0, 1, 2) ARIMA and model (1, 1, 1) ARIMA and by comparing them it is clear that the model (0, 1, 1) ARIMA s the best, After the model has been selected and the estimated parameters have been determined, the estimation stage, the model is examined and then the forecasting process comes.

Table 3. ARIMA proposed models for energy emission variable.

Model	Model Evaluation Criteria				
	AIC	BIC	MAPE	RMSE	R ²
ARIMA (0, 1, 1)	2.871	-1.615	18.273	0.430	0.937
ARIMA (0, 1, 2)	3.162	-1.566	18.431	0.424	0.940
ARIMA (1, 1, 0)	1.372	-1.576	17.688	0.438	0.935
ARIMA (2, 1, 0)	2.283	-1.514	17.886	0.435	0.937
ARIMA (1, 1, 1)	1.316	-1.565	18.540	0.424	0.940
ARIMA (1, 1, 2)	1.396	-1.474	18.452	0.428	0.940
ARIMA (2, 1, 1)	1.399	-1.477	18.404	0.427	0.940
ARIMA (2, 1, 2)	2.309	-1.379	18.532	0.432	0.940

Table 4. ARIMA model parameters.

		Estimate	SE	t	Sig.
No Transformation	Difference	1			
	MA Lag 1	-0.366	0.133	-2.756	0.008

After all possible models were identified and we made sure of the moral models with moral parameters, and then the best model was chosen, depending on the methods of differentiation known according to the criteria in the table above, the model was built through the series data and the following model was obtained:

Table 4 contains the parameters of the best model, where it was explained that ARIMA (0, 1, 1). Through **Table 3**, we find that the parameters of the model are significant, which indicates the importance of having these parameters in the model.

After finding the best model, it must be examined and make ensure that all statistical assumptions are met. This will be done using the ways of: drawing random errors with real values, drawing subjective correlations and partial autocorrelations for errors, and determining the value of the Q-Stat test.

Test the randomness of the residual:

After the residual series has been plotted and compared with the actual values, as shown in **Figure 3**, the randomness of the residuals must be examined using the Q-Stat test. In addition, the autocorrelation and partial autocorrelation functions of the residuals should be plotted to ensure that all spikes fall within the confidence limits.



Figure 3. Plotting residuals with actual value.

The residual randomness test is used to ensure that the model has exhausted all patterns in the data, leaving no unexplained temporal relationships. One of the tests used to examine the randomness of residues is the Ljung-Box test under the hypothesis of random residues (no autocorrelation). **Table 5** shows that the residuals in this model are not random (Ljung-Box = 31.682, sig = 0.016).

Table 5. Model statistics.

Model	Number of Predictors	Model Fit Statistics		Ljung-Box Q (18)		Number of Outliers
		Stationary R-squared		Statistics	DF Sig.	
energy-Model_1	0	0.075		31.682	17 0.016	0

From **Figure 4** we note that the residuals of the series is nonstationary and the values of autocorrelation and partial autocorrelation are all outside the limits, which indicates that the model chosen which indicates that the model that was chosen cannot be relied upon in the prediction process in this way. A transformation or other appropriate procedure must be performed, or this variable must be excluded. So we use ARCH models.

The test results are significant for both the Fisher test and the Lagrange multiplier, and all coefficients are significant as shown in **Table 6**, indicating that the ARCH (1) effect was confirmed. However, it was not sufficient or fully appropriate, possibly because the variance-error relationship requires more than one lag, in addition to the low coefficient of determination. Therefore, the ARCH (1) model cannot be relied upon. Therefore, we will test ARCH (2).

Date: 05/19/26 Time: 01:19
 Sample: 1970 2023
 Q-statistic probabilities adjusted for 1 ARMA term

Figure 4

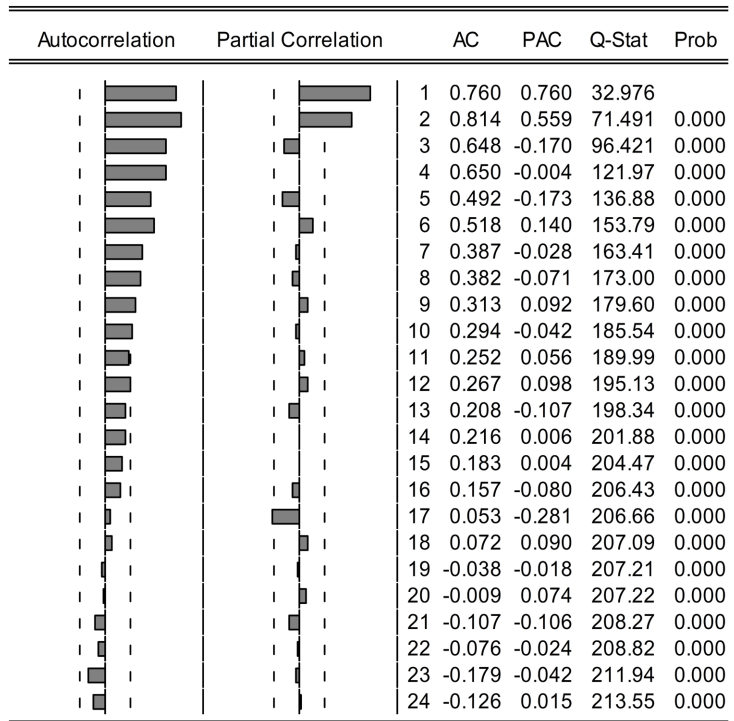


Figure 4. Auto correlation and partial auto correlation for the residuals.

Table 6. Heteroskedasticity test ARCH (1).

F-statistics	10.28621	Prob. F(1, 51)	0.0023	
Obs*R-SQUARED	8.895464	Prob Chi-Square (1)	0.0029	
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.524038	0.199429	2.627696	0.0113
RESID ² (-1)	0.435541	0.135800	3.207213	0.0023
R-squared	0.167839	Mean dependent var		0.886343
Adjusted R-squared	0.151522	S. D. dependent var		1.298925
S. E. of regression	1.196477	Akaike info criterion		3.233645
Sum squared resid	73.00940	Schwarz criterion		3.307996
Log likelihood	-83.69160	Hannan-Quinn criterion.		3.262237
F-statistic	10.28621	Durbin-Watson stat		2.424982
Prob (F-statistic)	0.002316			

The test was significant for both the Fisher test statistic and the Lagrange multiplier (see Table 7). However, the coefficients of the constant term and the first error term were not statistically significant. Therefore, the parameters were estimated using the maximum likelihood method.

Table 7. Heteroskedasticity test ARCH (2).

F-statistics	31.39431	Prob. F(2, 49)	0.0000	
Obs*R-SQUARED	29.20698	Prob Chi-Square (2)	0.0000	
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.184100	0.157250	1.170750	0.2474
RESID ² (-1)	0.130581	0.110729	1.179283	0.2440
RESID ² (-2)	0.732641	0.110757	6.614849	0.0000
R-squared	0.561673	Mean dependent var	0.895991	
Adjusted R-squared	0.543782	S. D. dependent var	1.309718	
S. E. of regression	0.884635	Akaike info criterion	2.648678	
Sum squared resid	38.34637	Schwarz criterion	2.761250	
Log likelihood	-65.86562	Hannan-Quinn criterion.	2.691835	
F-statistic	31.39431	Durbin-Watson stat	1.915192	
Prob (F-statistic)	0.000000			

Based on the model adopted in **Table 8**, the percentage of carbon dioxide emissions emitted from the energy sector in Sudan was predicted for a period of 10 years, starting from 2024 to 2033, as shown below in **Table 9** and **Figure 5**.

Table 8. Estimate the parameters using maximum likelihood.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR (1)	1.381580	0.169689	8.141826	0.0000
AR (2)	-0.407707	0.175620	-2.321529	0.0203
Variance Equation				
C	0.021386	0.004598	4.650960	0.2474
RESID ² (-1)	1.722711	0.559302	3.080111	0.0021
R-squared	0.929877	Mean dependent var	1.754462	
Adjusted R-squared	0.928475	S. D. dependent var	1.715710	
S. E. of regression	0.458854	Akaike info criterion	0.540797	
Sum squared resid	10.52734	Schwarz criterion	0.690892	
Log likelihood	-10.06071	Hannan-Quinn criterion.	0.598340	
Durbin-Watson stat	2.125078			
Inverted AR Roots	0.95		0.43	

Table 9. Forecast.

Model	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033
Forecast	4.8345	4.8345	4.8345	4.8345	4.8345	4.8345	4.8345	4.8345	4.8345	4.8345
UCL	5.6965	6.3157	6.7436	7.0918	7.3931	7.6624	7.9083	8.1359	8.3487	8.5494
LCL	3.9726	3.3534	2.9254	2.5772	2.2760	2.0066	1.7608	1.5332	1.3203	1.1197

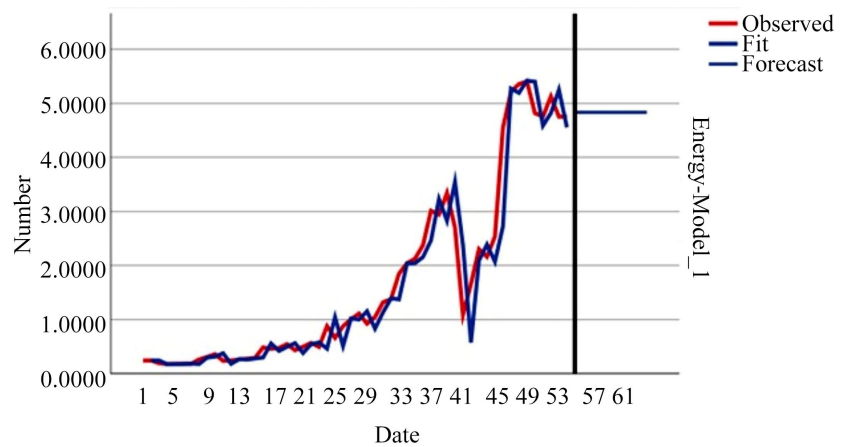


Figure 5. The curve of observed and forecast values.

Figure 5 illustrates the relationship between the observed values, the fitted values, and the forecast resulting from the estimated model. It shows a significant convergence between the observed values and the fitted values.

4. Conclusion

It was concluded that the best model based on selection methods was the ARIMA (2, 1, 2) model. Upon examining the model, it was found that the residuals of this model were random (Ljung-Box = 8.155, sig = 0.881). Nevertheless, we find that the residuals of the series are unstable, and some values of autocorrelation and partial autocorrelation do not fall within the limits, indicating that the chosen model is inefficient and cannot be relied upon for forecasting. Therefore, it is necessary to find a way to address this issue, which involves selecting other nonlinear models such as ARCH models. After confirming that the test results are significant for both the Fisher test and the Lagrange multiplier, and that all coefficients are significant, it was indicated that the ARCH (1) model cannot be relied upon. Thus, we test ARCH (2), and it was concluded that the latter model represents the data and can be relied upon for forecasting.

Acknowledgements

The researchers express their gratitude to the Sudan University of Science and Technology, represented by the College of Science and the College of Graduate Studies. We also extend our thanks to the University of Hail, represented by the College of Business Administration.

Conflicts of Interest

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

References

- [1] Osobajo, O.A., Otitoju, A., Otitoju, M.A. and Oke, A. (2020) The Impact of Energy Consumption and Economic Growth on Carbon Dioxide Emissions. *Sustainability*,

- 12, Article 7965. <https://doi.org/10.3390/su12197965>
- [2] Adams, S. and Nsiah, C. (2019) Reducing Carbon Dioxide Emissions; Does Renewable Energy Matter? *Science of the Total Environment*, **693**, Article 133288. <https://doi.org/10.1016/j.scitotenv.2019.07.094>
- [3] Abdalla, M. and Qarmout, T. (2023) An Analysis of Sudan's Energy Sector and Its Renewable Energy Potential in a Comparative African Perspective. *International Journal of Environmental Studies*, **80**, 1169-1187. <https://doi.org/10.1080/00207233.2023.2177417>
- [4] Elnourani, M., Elhag, H.S.H., Alasad, W.I. and Bashier, M.N. (2024) Khartoum War's Echoes in Oil and Energy Sectors: Economic and Environmental Implications for Sudan and South Sudan. *Heliyon*, **10**, e34739. <https://doi.org/10.1016/j.heliyon.2024.e34739>
- [5] Omer, A.M. (2007) Renewable Energy Resources for Electricity Generation in Sudan. *Renewable and Sustainable Energy Reviews*, **11**, 1481-1497. <https://doi.org/10.1016/j.rser.2005.12.001>
- [6] Degiannakis, S. and Xekalaki, E. (2004) Autoregressive Conditional Heteroscedasticity (ARCH) Models: A Review. *Quality Technology & Quantitative Management*, **1**, 271-324. <https://doi.org/10.1080/16843703.2004.11673078>
- [7] Cryer, J.D. and Chan, K.S. (2008) *Time Series Analysis: With Applications in R*. Springer.
- [8] Bollerslev, T., Chou, R.Y. and Kroner, K.F. (1992) ARCH Modeling in Finance. *Journal of Econometrics*, **52**, 5-59. [https://doi.org/10.1016/0304-4076\(92\)90064-x](https://doi.org/10.1016/0304-4076(92)90064-x)
- [9] Kumar, R. and Dhankar, R.S. (2010) Empirical Analysis of Conditional Heteroskedasticity in Time Series of Stock Returns and Asymmetric Effect on Volatility. *Global Business Review*, **11**, 21-33. <https://doi.org/10.1177/097215090901100102>
- [10] Bera, A.K. and Higgins, M.L. (1993) ARCH Models: Properties, Estimation and Testing. *Journal of Economic Surveys*, **7**, 305-366. <https://doi.org/10.1111/j.1467-6419.1993.tb00170.x>
- [11] Gouriéroux, C. and Monfort, A. (1992) Qualitative Threshold ARCH Models. *Journal of Econometrics*, **52**, 159-199. [https://doi.org/10.1016/0304-4076\(92\)90069-4](https://doi.org/10.1016/0304-4076(92)90069-4)
- [12] Engle, R. (2002) New Frontiers for Arch Models. *Journal of Applied Econometrics*, **17**, 425-446. <https://doi.org/10.1002/jae.683>