Mathematics and Statistics 12(1): 99-104, 2024 http://www.hrpub.org
DOI: 10.13189/ms.2024.120112

The ARCH Model for Analyzing and Forecasting Temperature Data

Ali Sadig Mohommed Bager*

College of Administration and Economics, Al-Muthanna University, Iraq

Received September 16, 2023; Revised December 22, 2023; Accepted January 3, 2024

Cite This Paper in the following Citation Styles

***(a)**: [1] Ali Sadig Mohommed Bager, "The ARCH Model for Analyzing and Forecasting Temperature Data," Mathematics and Statistics, Vol.12, No.1, pp. 99-104, 2024. DOI: 10.13189/ms.2024.120112*

***(b)**: Ali Sadig Mohommed Bager (2024). The ARCH Model for Analyzing and Forecasting Temperature Data. Mathematics and Statistics, 12(1), 99-104. DOI: 10.13189/ms.2024.120112*

Abstract The chaotic nature of the earth's atmosphere and the significant impact of weather on various fields necessitate accurate weather forecasting. Time series analysis plays a crucial role in predicting future values based on past data. The Autoregressive Conditional Heteroscedasticity (ARCH) model is widely used for forecasting, especially in the field of temperature analysis. This study focuses on the ARCH model for analyzing and forecasting temperature changes. The ARCH model is selected based on its ability to capture the regular variations in the predictability of meteorological variables. The methodology section explains the ARCH model and various statistical tests used, such as the heteroscedasticity test (ARCH test), Jarque-Bera test, and Augmented Dickey-Fuller test (ADF). A sample study is conducted on monthly average temperature data from Athenry, Ireland, over a period of four years. The study utilizes the ARCH model to calculate temperature series volatility and assesses the model's performance using goodness-of-fit measures and predictive accuracy. The results show that the ARCH model successfully predicts temperature changes for three years, as indicated by the forecasted temperature series. The statistical performance of the ARCH model is evaluated using in-sample and out-of-sample analyses, demonstrating its effectiveness in capturing temperature variations. The study highlights the importance of time series forecasting and the significant impact of the ARCH model in temperature analysis.

Keywords ARCH Model, Temperature Forecasting, Time Series Analysis, Heteroscedasticity Test, Jarque-bera Test, Augmented Dickey-Fuller Test

1 Introduction

Forecasting is one of the most steadily researched areas because the weather significantly impacts humans' daily lives. Attemperators above normal levels have significant impacts on both the environment and humanity. In terms of environmental impacts, rising temperatures can result in the melting of glaciers, leading to accelerated sea level rise. In turn, it can harm coastal areas and low-lying regions [1]. Additionally, ecosystems can suffer degradation due to increased temperatures. Moreover, rising temperatures can also decrease freshwater resources, posing water availability and management challenges. Health impacts are another concern associated with higher temperatures. Increased temperatures can lead to various health problems, including an elevated risk of diseases. These health risks can have severe consequences for individuals and communities. Furthermore, studying rising temperatures helps estimate the economic costs of climate change. By understanding the potential economic impacts, strategies can be developed to adapt to these challenges. The weather has a significant impact on a variety of fields. Because there is uncertainty in every weather forecast, the forecast must be accompanied by a measure of forecast uncertainty so that users can plan for the different possible outcomes [1]. Forecasts of weather variables are essential due to the numerous variables influencing weather risk. Atmospheric models create the ensemble predictions for a weather variable and consist of numerous future scenarios. One objective of time series analysis is to anticipate the future values of time series data. These approaches are used to describe and analyze phenomena and to predict the future of these phenomena based on data that represent the phenomenon's past. It includes identifying measures to mitigate the adverse effects of rising temperatures and implementing strategies to build resilience in affected sectors. It is essential to study the issue of rising temperatures and its impacts on the environment, human health, and the economy. The models' requirements must be satisfied to express the phenomenon they reflect. Forecasting is also considered the estimation of the future value of the data under investigation based on historical and current data. Using statistical approaches to plan and make assumptions about future events is crucial [1].

The authors in [2] used high-frequency climate data and three types of neural networks to investigate the accuracy of forecasting temperatures with hourly input data versus less frequent daily inputs. The convolutional neural network outperformed the other models, and the effect of state-of-the-art weather forecasting techniques was evaluated when combined with the convolutional neural network.

In [3], the authors modeled and forecasted parameters such as maximum and minimum temperature and morning and evening relative humidity using parametric models such as ARIMA and GARCH. The study data came from the Hoshangabad district of Madhya Pradesh from January 1996 to November 2019, and the AIC and BIC criteria were used to select the best models. It was found that the ARIMA-GARCH models were the most suitable for forecasting the parameters studied.

The researchers in [4] seek autoregressive integrated moving average (ARIMA), one of the most often used linear models in time series forecasting. As a result, the autoregressive integrated moving average (ARIMA) is one of the most widely used linear models in time series forecasting and the only model that can be utilized in any forecasting situation. The authors of [5] assess the growth rate's performance utilizing time series production. It has been identified as the most directly influential factor, and weedicide as the most indirectly significant element in Jaipur district and Rajasthan throughout time.

The authors of [6] investigated the month-ahead temperature forecasting in Jerusalem, Palestine, using ARIMA and GARCH modeling techniques. In addition, the authors of [7] utilized the AugmentedDickey–Fuller test to extend beyond the linear AR() process assumption. Its limiting distribution is derived when the error process has a continuous spectral density and is strictly positive.

This study [8] compares the performance of different statistical and machine learning models (ARIMA, KNN, SVR, and LSTM) for time series forecasting. The evaluation of the models was based on metrics such as MSE, MAE, Median AE, and RMSE. The results show that KNN has the best performance, notably in the medium and long terms. ARIMA is most appropriate for univariate small data sets, whereas Machine Learning algorithms show better short-term predictions.

The study of [9] proposed an ARIMAX model that accounts for external factors such as temperature, rainfall, and relative humidity to improve the accuracy of the forecast. It could lead to better decision-making and control of cocoa black pod disease. Furthermore, the authors of [10] compared three fuzzy models to 21 years of agricultural crop yield data from the GB Pant University of Agriculture and Technology Pantnagar, India, to judge accuracy. The efficiency of the models was judged through Mean Square Error and average percentage error. The results of the study will help predict large-scale production.

This paper attempts to determine the optimal model for predicting temperature changes and to forecast temperature utilizing the ARCH model. The ARCH model has been regarded as a criterion. In addition, the study focuses on forecasting a monthly average temperature series for Athens from 2020 to 2023 using the ARCH model and NCSS software. The ARCH model was used to calculate the temperature series' volatility. The goodness of fit and the predictive models are evaluated using an enhanced Dickey-Fuller and Heteroskedasticity test.

The rest of the paper is organized as follows: Section 2 presents materials and methodology. Section 3 presents the Estimation Auto regression conditionally helerodastisy ARCH model followed by reports results in Section 4. The paper concludes in Section 5.

2 Materials and Methodology

In this section, the Autoregression condition heterroscedastic model (ARCH) has been described in detail.

2.1 Autoregression condition heterroscedastic model (ARCH)

In 1982, Engle first proposed the autoregressive conditionally heterroscedastic model ARCH. The ARCH model assumes that $\{\beta_t\}$ is a stationary process satisfying [3]:

$$\beta_t = \theta + \varepsilon \tag{1}$$

$$\varepsilon = \alpha_t + \delta_t \tag{2}$$

Where: the $\delta_t \sim N(0,1)$

$$\varphi_t^2 = \propto_0 + \sum_{i=1}^{p} \propto_i \delta_{t-i}^2 \tag{3}$$

the $\propto_0 > 0$ and $\propto_1 \geq 0$ for $i > 0$ present the parameter model

2.2 Test ARCH (Heteroscedasticity Test)

The heteroscedasticity test uses stream flow series as a test of serial independence applied to some mode's serially uncorrelated fitting error to test the time series errors by calculating the autocorrelation coefficients for the residuals and writing the test hypothesis in the following [11]:

$H_0 : \partial_1 = \partial_2 = \cdots = \partial_k = \ldots \partial_{\mathrm{m}} = 0$
$\partial = 1, 2, \ldots \ldots m$
$H_1 : \partial_k \neq 0 \quad k = 1, 2, \ldots \ldots m$
Then, the formal as the following:

$$\phi_{(n)} = n(n+1) \sum_{k=1}^{m} \frac{\partial_k^2}{n-k} - \chi_{m-p}^2 \tag{4}$$

Where:
r: represents the sample size
m: represents the number of lags included in the test
p : represents the number of parameter estimators in model
∂_k^2 : represents the population autocorrelation function of the squared time series.

2.3 Jarqe –Bera Test

This test is based on calculating the difference between the Skewness and Kutosis coefficients of the series under study compared with the Skewness and Kutosis coefficients for the Normal distribution. It is calculated according to the following formula :

$$JB = \frac{N-K}{6} \left[S^2 + \frac{1}{4}(K-3)^2 \right] \tag{5}$$

Where S represents the Skewness, it can be calculated as the following [12]:

$$S = \frac{1}{N} \sum_{i-1}^{N} \left(\frac{y_i - \bar{y}}{\widehat{\sigma}} \right)^3 \tag{6}$$

Where Z represents the Kurtosis, it can be calculated as the following [11]:

$$Z = \frac{1}{N} \sum_{i-1}^{N} \left(\frac{y_i - \bar{y}}{\widehat{\sigma}} \right)^4 \tag{7}$$

2.4 Augmented Dickey-Fuller Test (A.D.F.)

The unit root in the time series analysis will be tested using the Augmented Dickey-Fuller Test (A.D.F.). It is the most commonly used test for analyzing the stationary of a series. We can use the following function to find the A.D.F. test :

$$\Delta y_t = \vartheta + \beta_t + \lambda y_{t-1} + \sum_{i=1}^{j} \theta_i \Delta y_{t-i} + e_t \tag{8}$$

Where :
y_t :represents the value of the time series at the time.
Δ : illustrates the first difference of time series.
$\vartheta, \beta, \lambda$: represent the parameter estimator.
Then :

$$DF_t = \frac{\hat{y}}{se\hat{y}} \tag{9}$$

The null hypothesis of the A.D.F. test is $\mathrm{H}_0 : \mathrm{y} = 0$ means the data needs to be differenced to make it stationary versus the alternative hypothesis of $\mathrm{H}_1 : \mathrm{y} < 0$ means the data is stationary and does not need to be differenced [13].

3 Estimation Auto regression conditionally helerodastisy ARCH model

The regression model will be defined as :

$$Y_t = B_0 + B_1 + B_2 X_{2t} + B_3 X_{3t} + \cdots + e_t \tag{10}$$

Where: $e_t \sim N\left(0, \sigma^2\right)$
Then the variance: $\operatorname{Var}\left(e_t\right) = \sigma_i^2$
The conditional variance is as:

$$\begin{aligned} \operatorname{Var}\left(X_{(t)} \mid x(t-1)^{(x)}\right) &= EX_{(t)^2} \mid X_{(t-1)} \\ &= E\left(\sigma\left(t^2\right) \varsigma\left(t^2\right) \mid X_{(t-1)}\right) \\ &= \omega + \alpha X_{(t-1)^2} \end{aligned} \tag{11}$$

If the value of $X_{(t-1)} = x(t-1)$ knows Normal distribution $\sim N\left(0, \sigma^2\right)$

$$\sigma^2(t) = \omega + \alpha X_{(t-1)^2}$$

In other formal :

$$f_{x(t) \mid X(t-1)^{(x)}} = \frac{1}{\sigma(t)} \varphi\left(\frac{x}{\sigma(t)}\right) \tag{12}$$

Where: $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{\left(\frac{-x^2}{2}\right)}$
Consequently, the maximum likelihood method is the joint conditional density as :

$$\begin{aligned} & L(\omega, \alpha) = f\left|x_{(n)}, x_{(n-1)}, \ldots\ldots\ldots, x_{(1)}\right| x_{(0)} \mid \\ & = \frac{1}{\sigma(n)} \varphi\left(\frac{x_{(n)}}{\sigma(n)}\right) x \ldots x \frac{1}{\sigma(1)} \varphi\left(\frac{x(1)}{\sigma(1)}\right) \\ & = \frac{1}{(2\pi)^{n/2} \sigma(n) \ldots\ldots \sigma(1)} e^{\frac{-\left(\sum_{t=1}^{n} x(t)^2\right.}{2\left(\sigma(t)^2\right)}} \end{aligned} \tag{13}$$

After that, we solve the logarithm we find :

$$\begin{aligned} L(\omega, \alpha) = \left(-\frac{1}{2}\right)(N \log(2\pi)) - \frac{\sum_{t=1}^{w} \log \sigma_t^2}{2} \\ - \frac{1}{2} \sum_{t=1}^{w} \frac{x(t)^2}{\sigma_t^2} \end{aligned} \tag{14}$$

4 Applied

This section includes an applied study on selecting a model of appropriate volatility forecasting for the data under analysis.

4.1 Sample Study

Athenry is located 25 kilometers (16 miles) east of Galway City in County Galway, Ireland. Athenry lies 12 miles (19 kilometers) north-northwest of Loughrea. It was an island in the North Atlantic Ocean in northwestern Europe (Land; 98.2 percent, Water; 1.8 percent). The climate of Athnary is moderate, humid, and varied, with few temperature extremes. The weather in Athnary is classified as temperate oceanic. January and February are the coldest months of the year, with average daily air temperatures between 4 and 7 °C. The warmest months are July and August, with mean daily temperatures of 14 to 16 °C and maximums ranging from 17 to 18 °C along the shore to 19 to 20 °C inland. May and June are the sunniest months, with an average of seven hours of sunshine daily. The data under consideration constitute a semi-periodic time series of monthly average temperatures from January 1, 2020, to December 31, 2023, with 36 observations. The graph and Figure below illustrate a semi-periodic time series, as monthly average temperatures change with the seasons. Met Eireann, Ireland's National Meteorological Service, is consulted via its website for the monthly temperature averages:https://www.met.ie/climate/available-data/monthly-data#top

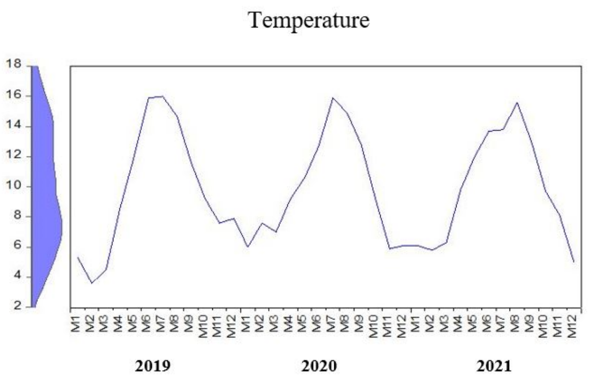

Figure 1. Temperature averages

This data type in Figure 1 is commonly used in climate research, weather forecasting, and other fields where temperature data analysis is essential. Figure 1 includes monthly temperatures, typically representing the recorded temperatures for each month over a specific period, with the vertical axis representing the temperatures and the horizontal axis representing the months by year. Figure 1 allows for the organization and analysis of temperature data over time. Patterns, trends, and comparisons can be observed and studied by plotting the temperatures on the vertical axis and aligning them with the corresponding months on the horizontal axis.

Table 1. Descriptive statistic

Mean	9.83
Median	9.2
Maximum	16
Minimum	3.6
Std.Dev	3.7455
Skewness	0.2011
Kutosis	1.785
Jarque-Bera	2.46
Probability	0.293

Table 1 presents the main summary statistics. The mean of this series was 9.83 and a median of 9.2, with a standard deviation of 3.745. We also note that the max value in the time series was 16.00, and the min value of 3.6. Furthermore, The skewness coefficient was 0.2011, which is a positive value, and this indicates that the series is pointed on the right side and Skewness is positive as well; the coefficient of kutosis coefficient was 1.785, which is a value less than 3, and this indicates that the data distribution is Kurtosis. Finally, the Jarque-Bera value indicates that it is distributed according to the Normal distribution at the 5% significance level.

Table 2. Augmented Dickey-Fuller test statistic

		t-Statistic	Prob
Test critical values	**Augmented Dickey-Fuller test statistic**	-4.837042	0.0005
	1% level	-3.679322	
	5% level	-2.967767	
	10% level	-2.622989	

In Table 2, we proved the series is stationary using the (Augment Dickey-Fuller) test. The results showed the rejection of the Null hypothesis, which states the unit root of a time series at a significance level less than 0.05. We use the ARCH test for the residual selected from the test to check the presence of heteroscedasticity.

Through Table 3, It is clear that the value of the F-statistic is less than 0.05. According to this result, we conclude that the residues are volatility by the presence of the ARCH effect, meaning the heteroscedasticity effect.

According to the outputs of the ARCH model in Table 4, the constant term (C) is statistically significant in both the mean and variance equations. It suggests that the constant significantly impacts the average temperature and the volatility of temperature fluctuations. In the variance equation, the term RESID(-1) is also statistically significant at indices returns. It implies that the volatility of temperature, or the risk associated with temperature fluctuations, is influenced by past square residual terms. Additionally, the ARCH (-1) term is statistically significant, indicating that the past temperature volatility significantly impacts the current volatility. These findings highlight the importance of considering both the mean and variance equations in the ARCH model for accurate temperature forecasting and risk assessment. However, without further information on the specific data and methodology used in the study, providing a more detailed analysis or interpretation of the results is challenging.

Table 3. Heteroscedasticity test ARCH

F-statistic	P-value
36.812	0.0001

Table 4. Estimator of ARCH model

variable	Std.Error	Coefficient	p-value	Z-statistic
C	0.006615	0.900195	0.0000	136.0938
temp	0.001104	0.906518	0.0000	821.3575
Variance Equation				
C	0.8929	0.4190	2.1309	0.0331
RESID (-1)	0.4325	0.0360	12.0111	0.0000

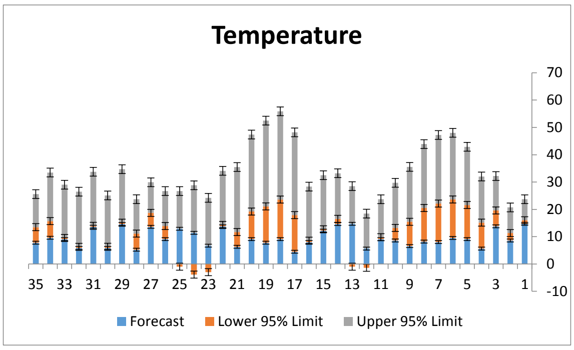

Figure 2. Forecast (lower & upper 95%limit)

Table 4 shows that after checking the model's efficiency, it was found that the appropriateness of the ARCH model, which was determined by the time series, could extract the predictive values for three years. Figure 2 shows the monthly temperature (measured in Celsius degree (C)) for each month vs the predicted values represented by 95% prediction intervals for a return period of 3 years.

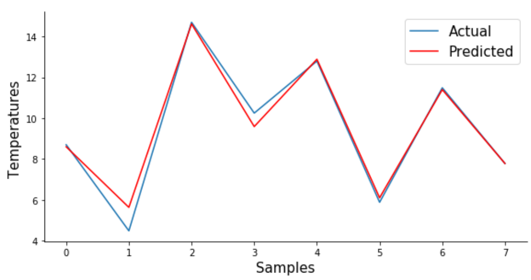

Figure 3. Prediction (forecast) of temperature rate

Based on the information provided, it seems that Figure 3 displays a prediction forecast of temperature rates. The red line represents the predicted values, while the blue line represents the actual temperature values. The convergence of the predicted values to the actual value indicates the strength of the prediction model. If the predicted values closely align with the actual temperature values, the model is accurate and reliable in forecasting temperature rates. This convergence demonstrates the effectiveness of the prediction model in capturing and predicting temperature patterns. It is important to note that the accuracy of the prediction model can vary depending on various factors, such as the quality and quantity of data used, the complexity of the model, and the specific characteristics of the predicted temperature patterns. Regular evaluation and validation of the model's performance against actual data are essential to ensure its reliability.

5 Conclusions

This study uses the ARCH (Autoregressive Conditional Heteroscedasticity) model to predict temperature for the next three years. The highest temperature recorded was 16 degrees Celsius, and the lowest was 3.1 degrees Celsius. Forecasting and time-series analysis have been critical academic subjects, as accurate forecasting is crucial for various decision-making processes. Researchers are continuously striving to improve the effectiveness of forecasting models. The ARCH model has emerged as one of the most commonly used models in forecasting research and application. The study yielded novel results, as the ARCH model accurately predicted temperature. Regarding in-sample statistical performance, both ARCH models best-predicted temperature averages. The out-of-sample statistical performance also showed good outcomes. The study specifically found a significant impact on the temperature forecast for Athenry City. It is important to note that the specific details and findings of the study were not provided, so further information regarding the methodology, data sources, and specific results would be necessary for a more comprehensive understanding of the study's findings.

REFERENCES

[1] M. K. A. Issa, "Weighted least squares estimation for ar (1) model with incomplete data," *Mathematics and Statistics*, vol. 10, pp. 342–357, 2022.

[2] S. Lee, Y.-S. Lee, and Y. Son, "Forecasting daily temperatures with different time interval data using deep neural networks," *Applied Sciences*, vol. 10, no. 5, p. 1609, 2020.

[3] P. Mishra, P. MISHRA, C. FATIH, G. VANI, J. M. LAVROD, V. JAIN, A. DUBEY, and A. K. CHOUDHARY, "Modeling and forecasting of metrological factors using arch process under different errors distribution specification," *Mausam*, vol. 72, no. 2, pp. 301–312, 2021.

[4] G. M. Naidu, B. R. Reddy, and B. R. Murthy, "Time series forecasting using arima and neural network approaches." *International Journal of Agricultural & Statistical Sciences*, vol. 14, no. 1, 2018.

[5] C. K. Saini, R. Singh, and S. Jain, "A study on growth rate and estimation of direct and indirect effect of some influential time series production factors on total barley (hordeum vulgare) production." *International Journal of Agricultural & Statistical Sciences*, vol. 15, no. 1, 2019.

[6] B. M. Iqelan, "Time series modelling of monthly temperature data of jerusalem/palestine." *MATEMATIKA: Malaysian Journal of Industrial and Applied Mathematics*, pp. 159–176, 2015.

[7] E. Paparoditis and D. N. Politis, "The asymptotic size and power of the augmented dickey–fuller test for a unit root," *Econometric Reviews*, vol. 37, no. 9, pp. 955–973, 2018.

[8] M. Iaousse, Y. Jouilil, M. Bouincha, and D. Mentagui, "A comparative simulation study of classical and machine learning techniques for forecasting time series data." *International Journal of Online & Biomedical Engineering*, vol. 19, no. 8, 2023.

[9] A. Ling, G. Darmesah, K. Chong, and C. Ho, "Application of arimax model to forecast weekly cocoa black pod disease incidence," *Mathematics and Statistics*, vol. 7, no. 4A, pp. 29–40, 2019.

[10] A. K. Rana, "Comparative study on fuzzy models for crop production forecasting," *Mathematics and Statistics*, vol. 8, no. 4, pp. 451–457, 2020.

[11] Y. Jouilil and D. Mentagui, "Comparing the forecasting accuracy metrics of support vector regression and arima algorithms for non-stationary time proces," *Mathematics and Statistics*, vol. 11, no. 2, pp. 294–299, 2023.

[12] A. Hussein and A. S. M. Bager, "The significant variables of effluent constructed wetlands treated domestic wastewater by a subset regression model," *International Journal of Environment and Waste Management*, vol. 32, no. 3, pp. 344–354, 2023.

[13] N. Saini, R. Sharma, R. Arora, and A. K. Bhardwaj, "Time series analysis of trend in production and its impact on price of red chili in india." *International Journal of Agricultural & Statistical Sciences*, vol. 16, no. 1, 2020.