Journal of Agriculture, Biology and Applied Statistics
Vol. 3, No. 1, 2024, pp. 1-9

URL : www.arfjournals.com
https://DOI:10.47509/JABAS.2024.v03i01.01

Forecasting Dengue Cases using Hybrid GARCH Model

S.R. Krishna Priya[1], N. Naranammal[2] and S. Sneha[3]
[1,2,3]Department of Statistics, PSG College of Arts & Science, Coimbatore

To cite this article

S.R. Krishna Priya, N. Naranammal & S. Sneha (2023). Forecasting Dengue Cases using Hybrid GARCH Model. Journal of Agriculture, Biology and Applied Statistics. Vol. 3, No. 1, pp. 1-9. *https://DOI:10.47509/JABAS.2023.v03i01.01*

Abstract: The Dengue incidence has increased over the past ten years in India. Almost half of the world's population, about 4 billion people, live in areas with a risk of Dengue. So, forecasting the Dengue cases will help to take preventive measures to overcome the disease. In this study Dengue cases of India has been forecast using ARIMA and ARIMA-GARCH model. Annual data from year 1996 to 2023 has been used to develop the model. Goodness of fit measures has been used to compare the traditional ARIMA and hybrid ARIMA-GARCH model. From the results, ARIMA (1,1,1)-GARCH (1,1) outperformed the ARIMA (1,1,1) model.

Keywords: Dengue, Stationarity, Heteroskedasticity, volatility, ARIMA model, ARCH and GARCH model.

1. Introduction

Dengue viruses are spread to people through the bite of an infected Aedes species mosquito. It is a vector-borne disease that is major public health threat globally. Dengue virus was isolated in Japan in 1943, the first epidemic of clinical dengue-like illness was recorded in Chennao in 1790 (Gupta *et al.*, 2012). It is estimated that 50 to 500 million people worldwide are infected with Dengue. Every year nearly 10000 to 20000 people are facing death worldwide due to Dengue (Othman *et al.*, 2022). Many researchers have performed the analysis on dengue incidence for many regions (Hsu *et al.*, 2017; Choudhury *et al.*, 2008; Patsaraporn and Somboonsak, 2019).

Previously, studies have been carried out using ARIMA and GARCH model in traffic modelling (Sun *et al.*, 2006), oil price fluctuation (Xiang, 2022), agricultural price forecasting (Bhardwaj *et al.*, 2014).

This paper is an attempt to forecast Dengue cases in India using ARIMA and ARIMA-GARCH models.

Article History Received : 28 December 2023; Revised : 20 January 2024; Accepted : 26 January 2024; Published : 30 March 2024

2. Materials and Methods

2.1. Data Description

The data used for this study is annual data of Dengue cases in India from 1996 to 2023. Data has been collected from website India stat. To analyse and forecast the dengue cases ARIMA and ARIMA-GARCH model have been used.

2.2. Stationarity Test

Unit root test, tests whether the time series data is stationary or not. The hypothesis for the unit root test is given as follows;

H_0: *The time series data is non stationary*

H_1: *The time series data is stationary*

The most commonly used unit root test is Augmented Dickey Fuller (ADF) test.

2.3. ARIMA Model

The ARIMA (auto regressive integrated moving average) model is a time series forecasting method that combines the auto regression (AR), differencing(I), moving average components (MA). The three main parameters in ARIMA are P (order of auto regression), d (degree of differencing) and q (order of moving average).

The ARIMA model is quite similar to the ARMA model other than the fact that it includes one more factor known as Integrated (I) i.e. differencing which stands for I in the ARIMA model. So, in short ARIMA model is a combination of a number of differences already applied on the model in order to make it stationary, the number of previous lags along with residuals errors in order to forecast future values.

2.4. ARCH Model

ARCH is an Autoregressive model with Conditional Heteroskedasticity. The ARCH process introduced by Engle (1982) explicitly recognizes the difference between the unconditional and the conditional variance allowing the latter to change over time as a function of past errors. ARCH model is a statistical model used to analyze volatility. The components of ARCH model are,

- Autoregressive: The current value can be expressed as a function of the previous values i.e. they are correlated
- Conditional: This informs that the variance is based on past errors.
- Heteroskedasticity: This implies the series displays unusual variance (Kumar, 2020).

The ARCH effect is caused by autocorrelation of heteroskedasticity observed over different periods of time. It means volatility is present in the data.

The mean equation of ARCH model is given by,

$$y_t = \beta_1 + \beta_2 X_{2t} + \cdots + \beta_k X_{kt} + \mu_t \tag{1}$$

The variance equation of ARCH model is given by,

$$\text{Var (u)} = \sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \cdots + \alpha_p u_{t-p}^2 \tag{2}$$

2.5. GARCH Model

Generalized Autoregressive Conditional Heteroskedasticity (GARCH) is an extension of the ARCH model that incorporates a **moving average component** together with the **autoregressive component.** Bollerslev (1986) generalized the ARCH model and introduced the GARCH model. GARCH is the ARMA equivalent of ARCH, which only has an autoregressive component. GARCH models permit a wider range of behavior more persistent volatility. (Kumar, 2020). The equation is given by

$$\sigma_t^2 = \alpha_0 + \Sigma_{i=1}^{q} \ \alpha i \in_{t-i}^2 + \Sigma_{j=1}^{p} \beta j \sigma_{t-j}^2 \tag{3}$$

2.6. Goodness of Fit Measures

Goodness of fit measures is used to evaluate the developed models.

2.6.1. AIC and BIC

$$\text{AIC} = -2\log L + 2n \tag{6}$$

$$\text{BIC} = -\log L + n \log T \tag{7}$$

where,
L is likelihood function; n is number of hyper parameters and T is the total number of observations. Less value of AIC and BIC indicates the better model.

2.6.2. Root Mean Squared Error

$$\text{RMSE} = \frac{\sqrt{\Sigma_{i=1}^{n} \ (yi - \hat{y}i)^2}}{n} \tag{8}$$

2.6.3. Mean Squared Error

$$\text{MSE} = \frac{1}{n} \Sigma_{i=1}^{n} \ (yi - \hat{y}i^2) \tag{9}$$

2.6.4. Mean Absolute Percent Error

$$\text{MAPE} = \frac{1}{n} \Sigma_{i=1}^{n} \left| \frac{yi - \hat{y}i}{yi} \right| \, | \tag{10}$$

where,

n - number of observations y_i – actual value and $\hat{y}i$ – predicted value

3. Results and Discussion

The time series plot of the Dengue cases from the year 1996 to 2023 is represented in the figure 1.

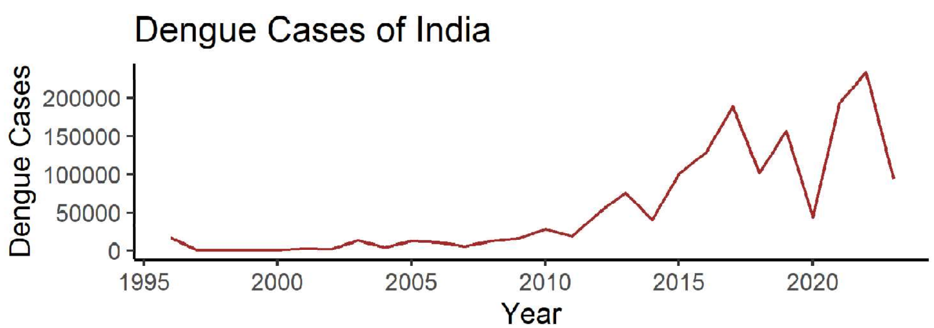


Figure 1: Time Series Plot of Dengue Cases

As the graph indicates that, the minimum number cases recorded is 650 in the year 2000 and the maximum number of cases recorded is 233251 in the year 2022. From the year 1996 to 2023 over all 1553583 cases have been recorded in India.

The above series exhibits the properties of a non-stationary series. Also, it is seen that there is an upward trend which indicates that our time series data is indeed non-stationary. To further check its stationarity, unit root test has been used.

3.1. Augmented Dickey Fuller Test

Unit roots are the reason for non-stationarity. A time series data is said to be stationary if a change in time does not influence the change in the distribution's form. The statistical power of these tests is low. The most commonly used test to check the stationarity is Augmented Dickey Fuller test.

Table 1: Result of Stationarity Test

Data	*ADF test*	*p-value*
Level	-2.15	0.22
Differentiated	-5.86	<0.01

The Augmented Dickey-Fuller test shows the p-value as 0.2259 which is greater than 0.05. So, the null hypothesis is accepted which indicated that the data is non-stationary. After differencing one time, p value of ADF test is <0.01 which indicate that the data have become stationary.

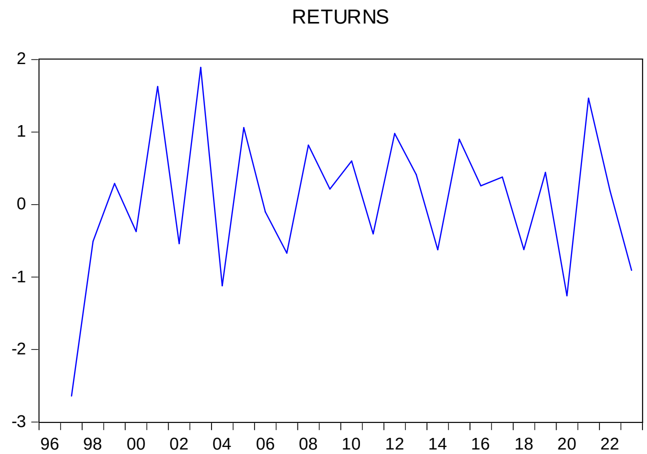


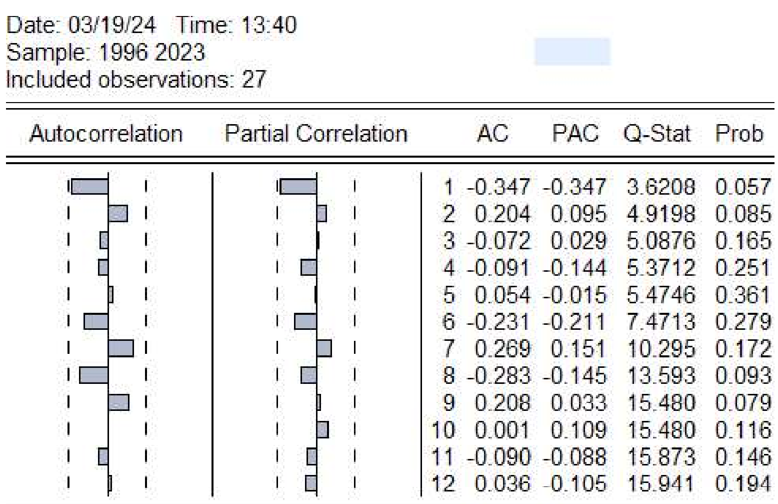
Figure 2: Time Series Plot of Dengue Cases after Differencing

The figure 2, differentiated plot provides more useful information about the data as there is the phenomenon of volatility clustering. There is no evidence of a trend, and the series appears to show a tendency to mean reversion. However, there may be an autocorrelation in the data.

3.2. Results of ARIMA Model

3.2.1. Correlogram of ARIMA

ACF and PACF plot shows the p and q values in the order of ARIMA model and it is presented in figure 3.



	AC	PAC	Q-Stat	Prob
1	-0.347	-0.347	3.6208	0.057
2	0.204	0.095	4.9198	0.085
3	-0.072	0.029	5.0876	0.165
4	-0.091	-0.144	5.3712	0.251
5	0.054	-0.015	5.4746	0.361
6	-0.231	-0.211	7.4713	0.279
7	0.269	0.151	10.295	0.172
8	-0.283	-0.145	13.593	0.093
9	0.208	0.033	15.480	0.079
10	0.001	0.109	15.480	0.116
11	-0.090	-0.088	15.873	0.146
12	0.036	-0.105	15.941	0.194

Figure 3: Correlogram of ARIMA model

From the figure 3, the Correlogram shows that there is a spike in the lag 1 of the plots. So, the best ARIMA model suitable for the data is ARIMA (1,1,1).

The parameter estimates of the ARIMA (1,1,1) model is presented in the Tables 2.

Table 2: Parameter Estimates of ARIMA (1,1,1) Model

Variable	*Coefficient*
C	0.11
AR(1)	-0.45
MA(1)	-0.06

The equation of ARIMA (1,1,1) model is,

$$Y_t' = 0.1145 - 0.4557 y_{t-1} + 0.4241 e_{t-1} \tag{11}$$

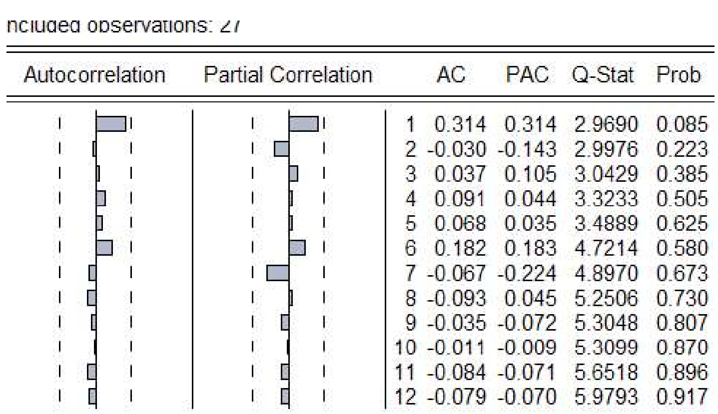
3.3. Results of ARCH Effect

To forecast using ARIMA-GARCH model, the presence of ARCH effects has to be done by checking the residual diagnostics for heteroskedasticity and the results is presented in table 3.

Table 3: Heteroskedasticity test for ARCH effects

Parameter	*Values*
Observed R squared	5.79
Prob.F(1,24)	0.01
Prob.chi.Square(1)	0.01

From the Table 3, the probability of chi-square is <0.05, so it is significant and the probability of residual is < 0.05 which implies that there is heteroskedasticity in the data. To obtain the p and q values for GARCH model, correlogram of squared residuals in presented in figure 4.


ncluded observations: 27

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.314	0.314	2.9690	0.085
		2	-0.030	-0.143	2.9976	0.223
		3	0.037	0.105	3.0429	0.385
		4	0.091	0.044	3.3233	0.505
		5	0.068	0.035	3.4889	0.625
		6	0.182	0.183	4.7214	0.580
		7	-0.067	-0.224	4.8970	0.673
		8	-0.093	0.045	5.2506	0.730
		9	-0.035	-0.072	5.3048	0.807
		10	-0.011	-0.009	5.3099	0.870
		11	-0.084	-0.071	5.6518	0.896
		12	-0.079	-0.070	5.9793	0.917

Figure 4: Correlogram of Squared Residuals

From figure 4, one can see the partial correlation graph for the lags. The resulting values are all significant indicating the presence of ARCH effects. The necessary values are all significant indicating that ARCH model of lag 1 is enough.

3.4. Result of ARIMA-GARCH Model

Having identified the ARIMA (1,1,1) as the better model, now ARIMA-GARCH model has to be estimated. To find out the best ARIMA-GARCH model, different combination of ARIMA-GARCH model have been proposed and AIC and BIC values are presented in table 4.

Table 4: AIC and BIC of different ARIMA-GARCH Models

Models	*AIC*	*BIC*
ARIMA (1,1,1)- GARCH (1,1)	2.41	2.71
ARIMA (1,1,1)- GARCH (1,2)	2.58	2.91
ARIMA (1,1,1)- GARCH (1,3)	2.67	3.05
ARIMA (1,1,1)- GARCH (2,1)	2.61	2.94
ARIMA (1,1,1)- GARCH (2,2)	2.76	2.87
ARIMA (1,1,1)- GARCH (2,3)	2.79	3.22
ARIMA (1,1,1)- GARCH (3,1)	2.70	3.09
ARIMA (1,1,1)- GARCH (3,2)	2.75	3.19
ARIMA (1,1,1)- GARCH (3,3)	2.83	3.31

From table 4, comparing the AIC and BIC values of the different models, ARIMA (1,1,1)-GARCH (1,1,) has the lowest AIC, BIC values which indicated that the ARIMA (1,1,1)-GARCH (1,1) is the best model for forecasting Dengue cases.

Table 5: Estimate values of ARIMA (1,1,1)-GARCH (1,1)

Mean Equation *Parameter*	*Coefficient*	*Standard Error*	*P-Value*
Constant	0.2211	0.035458	0.0000
Ar (1)	0.01508	0.275183	0.9563
Ma (1)	-0.91138	0.144398	0.0000
Variance Equation			
α_0	-0.025931	0.308223	0.9330
β_1	0.424069	9.255408	0.9635

The equation of ARIMA (1,1,1)-GARCH (1,1) model is,

$$\sigma_t^2 = 0.2754 - 0.0259\epsilon_{t-1}^2 + 0.4241\sigma_{t-1}^2 \tag{12}$$

3.5. Goodness of Fit Measures of the ARIMA and ARIMA-GARCH Model

Table 6: Goodness of Fit Measures of the Proposed Models

Model	*AIC*	*BIC*	*RMSE*	*MAE*	*MAPE*
ARIMA (1,1,1)	2.89	3.08	0.94	0.79	119.93
ARIMA (1,1,1)-GARCH (1,1)	2.41	2.71	0.82	0.67	94.41

From the table 6, it is clear that ARIMA (1,1,1)-GARCH (1,1) performed better than ARIMA (1,1,1) for forecasting Dengue cases of India.

5. Conclusion

Forecasting Dengue cases improves the understanding of the risk associated with this disease and it provides accurate predictions for better public health planning and resource allocation. In the present study, ARIMA and ARIMA-GARCH models have been developed for forecasting dengue incidence. When comparing various combination of ARIMA-GARCH models, ARIMA (1,1,1)-GARCH (1,1) had lower AIC and BIC values indicates that the model performs better than other models. Based on the goodness of fit measure the ARIMA-GARCH model outperforms the traditional ARIMA model. ARIMA (1,1,1)-GARCH (1,1) provides valuable insights into the dynamics of dengue outbreak allowing for better preparedness and allocation of resources for prevention and control efforts.

References

Othman, M., Indawati, R., Suleiman, H.A., Qomaruddin, M.B. and Sokkalingam, R. (2022). Model Forecasting Development for Dengue Fever Incidence in Surabaya City Using Time Series Analysis. processes. 10(11).

Mutheneni, S.R., Morse, A.P., Caminade, C. and Upadhyayula., S.A. (2017). Dengue burden in India: recent trends and importance of climatic parameters. *Emerg Microbes Infect.* 6(8):e70.

Gupta, N., Srivastava, S., Jain, A and Chaturvedi UC. (2012). Dengue in India. *Indian J Med Res.* 136(3): 373-90.

Singh, P.S., and Chaturved, H.K. (2022). A retrospective study of environmental predictors of dengue in Delhi from 2015 to 2018 using the generalized linear model. *Sci Rep.* 12(1): 8109.

Nayak, S.D.P.M and Narayan, KA. (2019). Forecasting Dengue Fever Incidence Using ARIMA Analysis. *International Journal of Collaborative Research on Internal Medicine & Public Health.* 11.

Zhou, B., He,D. and sun, Z. (2006). Traffic modeling and prediction using ARIMA/GARCH model. Modeling and Simulation Tools for Emerging Telecommunication Networks:101-121.

Xiang, Y. (2022). Using ARIMA-GARCH Model to analyze fluctuation law of international oil price. *Mathematical Problem in Engineering.* 2022.

Bhardwaj,S.P., Paul, R.K., Singh, D.R. and Singh, K.N. (2014).An empirical investigation of ARIMA and GARCH models in agricultural price forecasting. *Economic Affairs.* 59(3): 415-428.

Somboonsak, P. (2019). Forecasting dengue fever epidemics using ARIMA model. AICCC 2019: 144-150.

Naher, S., Rabbi, F., Hossain, M.M., Banik, R., Pervez, S. and Boitchi, A.K. (2022). Forecasting the incidence of dengue in Bangladesh- Application of time series model. *Health Sci Rep.* 5(4): e666

Bhatnagar, S., Lal, V., Gupta, S.D. and Gupta, O.P. (2012). Forecasting incidence of dengue in Rajasthan, using time series analyses. *Indian J Public Health.* 56(4): 281-5.