Advances and Applications in Mathematical Sciences
Volume 22, Issue 3, January 2023, Pages 769-778

SUPERIORITY OF GMDH NEURAL NETWORK MODEL OVER HOLT'S AND ARIMA MODELS FOR THE FUTURE PREDICTION OF GENERAL FERTILITY RATE: A CASE STUDY OF INDIA

MINAKSHI MISHRA[1], ANUJ KUMAR* and RATAN KUMAR THAKUR[1]

[1]*Department of Statistics
School of Physical and Decision Sciences
Babasaheb Bhimrao Ambedkar University
(A Central University)
Lucknow-226025, India

Abstract

This study compared the group method of data handling neural network (GMDH-NN) model with the double exponential smoothing (DES) or Holt's and autoregressive integrated moving average (ARIMA) models for the future prediction of GFR for India. For the present study, time series data concerning about general fertility rate (GFR) was collected from 1995 to 2020. The best-fitting model for the future prediction of GFR over the next 20 years was developed using three different models: Holt's, ARIMA and GMDH-NN. As a way to select a model, the coefficient of determination (R2) and the following forecasting errors were taken into consideration: mean absolute error (MAE), mean absolute percentage error (MAPE), mean squared error (MSE) and root mean square error (RMSE). The research shows that the GMDH-NN model outperforms two other models, the Holts and ARIMA models, in terms of performance and accuracy with the lowest values of forecasting error.

1. Introduction

Fertility is one of a population's most important demographic characteristics [4, 17, 20]. The level of childbearing in a population is measured by fertility rates. They are crucial in defining a population's growth rate as well as its age structure [21]. Some fertility measurements are accurate "rates," in the sense that they represent the "risk" of having a child

2020 Mathematics Subject Classification: 68T07.

Keywords: General Fertility Rate, GMDH, Holt's, ARIMA.

*Corresponding author; anujkmsmath2016@gmail.com

Received July 18, 2022; Accepted September 29, 2022

during a specific time frame. The crude birth rate, the general fertility rate and age-specific fertility rates are examples of such measurements. Other fertility measurements, such as cohort completed fertility and the total fertility rate, are not accurate but do capture the level of childbearing in a community [5]. Fertility prediction plays a vital role due to its relevance in predicting and controlling population growth [18]. It is a major factor in biological substitution and the maintenance of mankind's development, as well as having a significant impact on a country's socioeconomic situation. As a result, fertility remains a critical factor in population predictions. GFR is the most common indicator used to estimate fertility and evaluate population increase.

The most widely used measure of fertility is the GFR which is defined as the number of resident live births for a specified geographic area (nation, state, county, etc.) during a specified period (usually a calendar year) divided by the female population aged 15-49 years (usually estimated for a mid-year) and the resulting fraction multiplied by a 1,000 [4]. The GFR is perhaps the most commonly used overall fertility measure because it frequently matches readily accessible numerator and denominator data in a large age range that spans most of the female reproductive years and it reflects the population at the highest risk of delivering birth [17]. According to the latest statistical report 2020, the GFR by residence in India is 67 which include an urban GFR of 53.7 and a rural GFR of 73.7.

Several researchers considered different statistical and time series techniques for the future prediction of fertility rates [10, 6, 22, 8, 19, 2, 20] and mortality rates [1]. Some researchers performed a comparative analysis of the GMDH-NN technique with ARIMA and Holt's techniques [1, 14] and they concluded that the predictive strength of the GMDH-NN technique is found to be superior to other predictive models.

The main aim of this work is future predictions of the GFR for India using the best-fitted model. Future prediction using the GMDH-NN technique has now gained increased attention in the present era and has wide application in many real-life problems.

2. Materials and Methods

The present investigation was carried out on secondary data of GFR for India and the data has been gathered for the period 1995 to 2020 from the source of indiastat.com. The data analysis has been performed using the MS-Excel, *R* and GMDH-NN shell Software.

2.1 Holt's Linear Exponential Smoothing Model

Charles Holt developed a version of exponential smoothing that can be used to forecast a time series with a linear trend. Holt's linear exponential smoothing is also known as double exponential smoothing [11, 16]. Forecasts for Holt's linear exponential smoothing model are obtained using two smoothing parameters (α and β) and the three equations are given below:

Level smoothing equation: $L_t \alpha Y_t + (1-\alpha)(L_{t-1} + b_{t-1})$ (1)

Trend smoothing equation: $b_t = \beta(L_{t-1} + L_{t-1}) + (1-\beta)b_{t-1}$ (2)

Forecasting equation: $F_{t+m} = L_t + mb_t$ (3)

$\alpha =$ smoothing constant level component $(0 < \alpha < 1)$

$\beta =$ smoothing constant trend component $(0 < \beta < 1)$

$m =$ number of periods ahead to be forecast.

2.2 Autoregressive Integrated Moving Average (ARIMA).

In time series forecasting, the Box and Jenkins-developed ARIMA model is commonly employed [22]. Also referred to as the Box-Jenkin's technique. The primary goals of fitting an ARIMA model are to properly forecast future values and to identify the time series stochastic process [3, 15]. These techniques can also be applied in other contexts where creating models for discrete time series and dynamic systems is necessary. In lead periods or seasonal series with a significant random component, this approach performs poorly.

Originally George Box and Gwilym Jenkins have studied ARIMA model extensively during 1968 and their names has been used frequently with general ARIMA process applied for time series analysis, validation, forecasting and control. However, the stochastic model of the series is used to determine the optimal forecast of future values of a time series. A stochastic process can either be stationary or non-stationary. The ARIMA models refer only to a stationary time series but most of the time series are non-stationary. The first stage of Box-Jenkins model is for reducing nonstationary series to a stationary series by taking first or second order differences. The Box-Jenkins's method is named after the statisticians George Box and Gwilym Jenkins, applies autoregressive moving average ARMA model to stationary data i.e. ARIMA models to find the best fit of a time-series model to past values of a time series in time series analysis. The block diagram of Box-Jenkins Methodology is shown in figure 1.

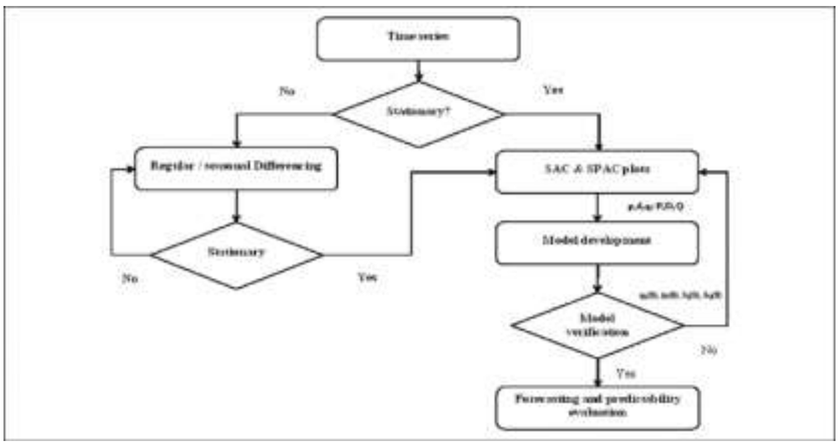

Figure 1. ARIMA model development process.

ARIMA is a statistical model used for forecasting time series data. The ARIMA equation is a regression-type equation in which the independent variables are lags of the dependent variable and/or lags of the forecast errors. The equation of the ARIMA model is given as:

$$\widehat{Y'(y)} = \mu + \varphi_1 * y'_{(t-p)} + \ldots + \theta_1 * y'_{(1-p)} + \theta_1 * \varepsilon_{(t-1)} + \ldots + \theta_q * \varepsilon_{(t-q)} + \varepsilon_t \qquad (4)$$

There are three components in the equation (figure 2): Auto-regressive (AR)- The time series is regressed with its previous values i.e. $y_{(t-1)}$, $y_{(t-2)}$ etc. The order of the lag is denoted as p, Integration: The time series uses differencing to make it stationary. The order of the difference is denoted as d

and Moving Average (MA): The time series is regressed with residuals of the past observations i.e. error $\varepsilon_{(t-1)}$, error $\varepsilon_{(t-2)}$, etc. The order of the error lag is denoted as q. In the equation (5), $\widehat{Y'(t)}$ is the differenced series, ϕ_1 is the coefficient of the first AR term, p is the order of the AR term, θ_1 is the coefficient of the first MA term, q is the order of the MA term and ε_t is the error.

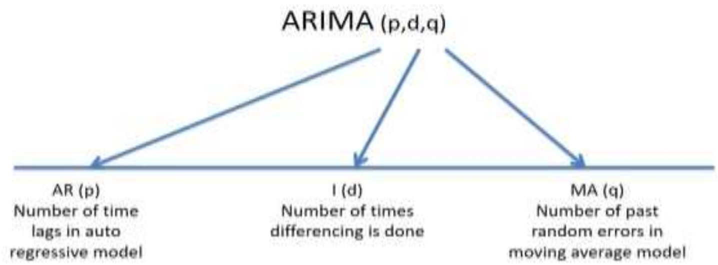

Figure 2. Components of the ARIMA Model.

2.3 Group Method of Data Handling Neural Network (GMDH-NN)

Ivakhnenko [13] was the first to utilize the GMDH method to describe complicated systems, which contained a collection of data with many inputs and a single output. The GMDH network's major objective is to build a function for a feed-forward network using a second-degree transfer function. The algorithm of GMDH automatically determines the appropriate model structure with the number of layers, neurons inside the hidden layers and the input variables. The mapping between the input and output variables done through a GMDH-NN is a nonlinear function called the Volterra series [7], in the form of equation (5). The Volterra series as a one-input variable second-degree polynomial is analyzed using equation (6).

$$\hat{y} = a_0 + \sum_{i=1}^{m} a_i x_i + \sum_{j=1}^{m} a_{ij} x_i x_j + \sum_{i=1}^{m} \sum_{k=1}^{m} a_{ijk} x_i x_j x_k + \ldots \quad (5)$$

where x represents the input variable, a_i represents coefficients, y represents the output variable and m represents the number of observations of input variable.

$$G(x_i) = a_0 + a_1 x_i + a_2 x_i^2 \tag{6}$$

The aim of the GMDH algorithm is to find the a_i unknown coefficients in the Volterra series. The a_i coefficients are solved with regression methods for xi input variable [10, 12]. A schematic representation of the GMDH-NN model development process is depicted in figure 3.

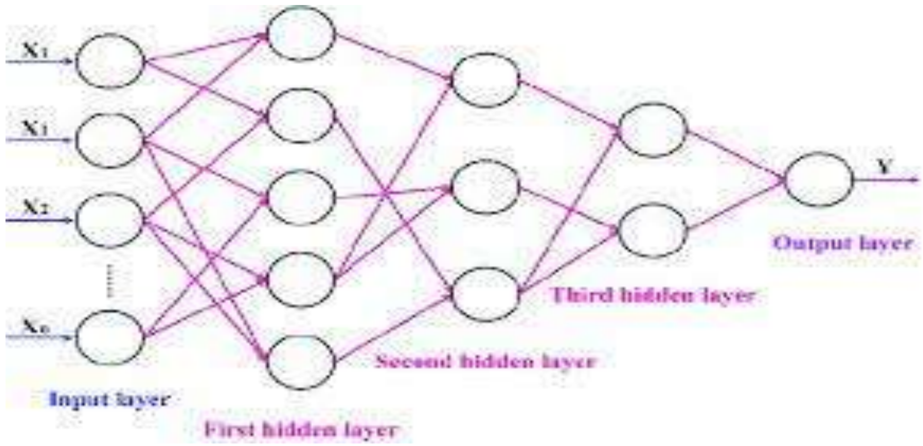

Figure 3. GMDH-NN model development process.

3. Results and Discussion

In this section, a comparative analysis has been carried out to depict the predictive strength of the GMDH-NN model over ARIMA and Holt's models for future prediction of the GFR using different model selection criteria's (Table 1). The outcome revealed that GMDH-NN technique is significantly better than the other considered techniques on the basis of lowest values of forecast errors namely, MAE MAPE, MSE and RMSE.

Table 1. Comparison of Various Models based on Model selection Criteria's

Model selection criteria's	GMDH-ANN	ARIMA	Holt's method
Best parameters	Training set= 80%, Testing set= 20%	$p = 1, d = 1, q = 2$	$\alpha = 0.29, \beta = 1.00$

MAE	0.418	1.636	1.740
MAPE	0.485	1.765	1.853
MSE	0.303	6.429	5.916
RMSE	0.551	2.535	2.432
R2	0.999	0.966	0.967

From figure 5, predicted values of GFR for 1995 to 2020 were close to the actual values of GFR using GMDH-NN model as compared to Holt's and ARIMA models where as future predicted values of GFR were observed from 2021 to 2040 for the three models (figure 6).

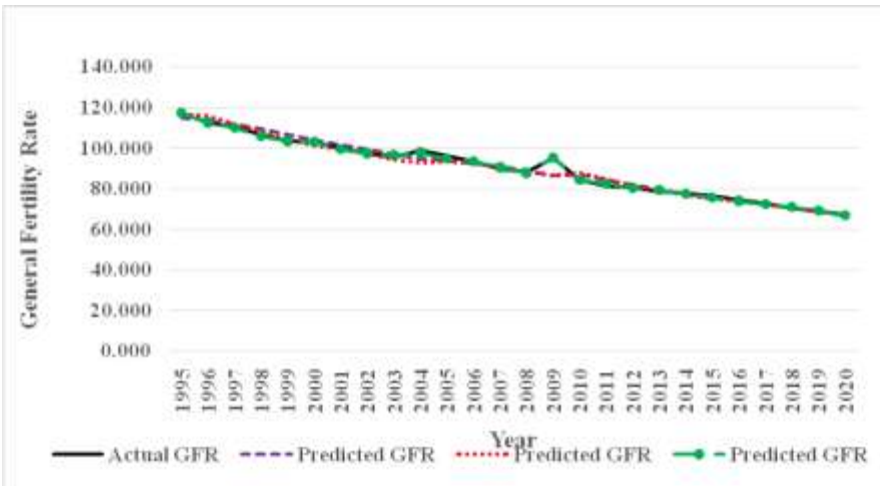

Figure 5. Comparison of Holt, ARIMA and GMDH-NN models for the prediction of GFR.

The future prediction of GFR at 95% prediction interval (PI) using Holt and ARIMA models were very close for 2021 to 2040, however, GMDH-NN is different (figure 6). Future predicted values of GFR using GMDH-NNmodel (95% PI: 29.265-30.735) were more superior as depicted the difference of lowest and highest values become less as compared to Holt's (95% PI:21.403 - 37680) and ARIMA (95% PI: 15.959 - 42.299) models.

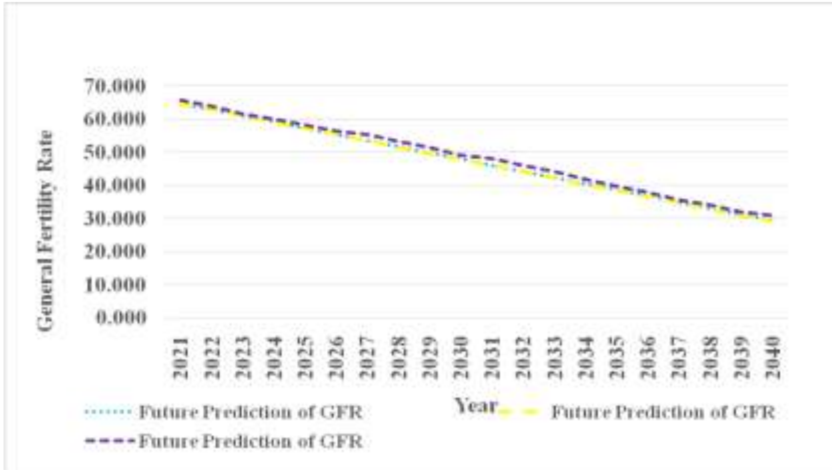

Figure 6. Comparison of Holt, ARIMA and GMDH-NN models for the future prediction of GFR.

4. Conclusion

This study reveals that the prediction of GFR using GMDH-NN technique was more appropriate than other considered conventional techniques for India. Also, it does not require complicated assumptions needed for traditional time series models. The ARIMA regression and Holt methods might not be suitable for long-term forecasting of GFR for India. Therefore, GMDH-NN might be more suitable for data with non-linear, such as GFR. GMDH-NN increases forecasting accuracy of different fertility rates to inform the government of India will be able to allocate future resources and plan for children's services.

Acknowledgment

The authors are highly thankful to the Editor-in-Chief for his immense support on time and helped in revising the manuscript with efficient reviewers' comments.

References

[1] D. A. Adeyinka and N. Muhajarine, Time series prediction of under-five mortality rates for Nigeria: comparative analysis of artificial neural networks, Holt-Winters exponential smoothing and autoregressive integrated moving average models, BMC Medical Research Methodology 20(1) (2020), 1-11.

[2] N. Arbu, A. Lim and C. Me-ead, Forecasting the Patterns and Trends Age-Specific Fertility Rate in South Asia, Turkish Journal of Computer and Mathematics Education 12(8) (2021), 2595-2600.

[3] D. Asteriou and S. G. Hall, ARIMA models and Box-Jenkins Methodology: Applied Econometrics, New York, Palgrave Macmillan, (2011).

[4] A. A. Bhende and T. Kanitkar, Principles of Population Studies, Himalaya Publishing House, New Delhi, India, (1997).

[5] J. Bongaarts, Trends in fertility and fertility preferences in sub-Saharan Africa: the roles of education and family planning programs, Genus 76(1) (2020), 1-15.

[6] B. M. Devkota, Estimation of total fertility rate and birth averted due to contraception: Regression Approach, Nepal Population Journal 18(17) (2018), 105-112.

[7] I. Ebtehaj, H. Bonakdari, A. H. Zaji, H. Azimi and F. Khoshbin, GMDH-type neural network approach for modeling the discharge coefficient of rectangular sharp-crested side weirs, Engineering Science and Technology, an International Journal 18(4) (2015), 746-757.

[8] A. A. Elhag and H. Abu-Zinadah, Forecasting under applying Machine Learning and Statistical Models, Thermal Science 24 (Supplement 1) (2020), 131-137.

[9] A. F. Fagbamigbe and A. S. Adebowale, Current and Predicted Fertility using Poisson Regression Model: Evidence from 2008 Nigerian Demographic Health Survey, African Journal of Reproductive Health 18(1) (2014), 71-83.

[10] S. J. Farlow, The GMDH Algorithm in SJ Farlow, ed., Self-organizing methods in modelling, GMDH type algorithms, (1984).

[11] C. C. Holt, Forecasting seasonals and trends by exponentially weighted moving averages, International Journal of Forecasting, Elsevier 20(1) (2004), 5-10.

[12] H. Iba, H. Degaris and T. Sato, A numerical approach to genetic programming for system identification, Evolutionary computation 3(4) (1995), 417-452.

[13] A. G. Ivakhnenko, The group method of data of handling: A rival of the method of stochastic approximation, Soviet Automation Control 1(3) (1968), 43-55.

[14] D. Li, M. R. Moghaddam, M. Monjezi, D. J. Armaghani and Mehrdanesh, Development of a Group Method of Data Handling Technique to Forecast Iron Ore Price, Applied Sciences 10(7) (2020), 364.

[15] Z. Liu, Z. Zhu, J. Gao and C. Xu, Forecast methods for Time Series Data: A Survey, IEEE Access 9 (2021), 91896-91912.

[16] A. Nazim and A. Afthanorhan, A comparison between single exponential smoothing (SES), double exponential smoothing (DES), holt's (brown) and adaptive response rate exponential smoothing (ARRES) techniques in forecasting Malaysia population, Global Journal of Mathematical Analysis 2(4) (2014), 276-280.

[17] K. B. Pathak and F. Ram, Techniques of Demographic Analysis, Himalaya Publishing House, New Delhi, India, (2016).

[18] A. Radkar, Indian Fertility Transition, Journal of Health Management 22(3) (2020), 413-423.

[19] H. L. Shang and H. Booth, Synergy in fertility forecasting: improving forecast accuracy through model averaging, Genus 76(1) (2020), 1-23.

[20] S. Singh, C. Shekhar, A. Bankole, R. Acharya, S. Audam and T. Akinade, Key drivers of fertility levels and differentials in India, at the national, state and population subgroup levels, 2015-2016: An application of Bongaarts' proximate determinants model, Plos One, 17(2) (2022), 1-39.

[21] A. K. Tiwari and S. Mishra, Fertility pattern in India and estimation of total fertility rate, International Journal of Agricultural and Statistical Sciences 17(Supplement 1) (2021), 2279-2289.

[22] P. K. Tripathi, R. K. Mishra and S. K. Upadhyay, Bayes and classical prediction of total fertility rate of India using autoregressive integrated moving average model, Journal of Statistics Applications and Probability 7(2) (2018), 233-244.