International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2023; SP-8(3): 574-582

https://www.mathsjournal.com
Received: 07-05-2023
Accepted: 11-06-2023

Bheemanna
Research Scholar, Department of Statistics, Karnataka University's Karnataka Arts College, Dharwad, Karnataka, India

MN Megeri
Professor, Department of Statistics, Karnataka University's Karnataka Arts College, Dharwad, Karnataka, India

Corresponding Author:
MN Megeri
Professor, Department of Statistics, Karnataka University's Karnataka Arts College, Dharwad, Karnataka, India

Prediction of India's demographic and economic variables using the neural network auto-regression model

Bheemanna and MN Megeri

DOI: https://doi.org/10.22271/maths.2023.v8.i4Sh.1121

Abstract
Forecasting demographic and economic variables is an essential component of research that will help society and the government plan for the best or worst in the future. The data on demographic variables are collected from the Census of India and SRS publications, and economic variables are gathered from the Economic Survey of India from 1971 to 2020. The goal of this research is to forecast demographic and economic factors using the NNAR approach. Because statistical approaches such as the least RMSE training and testing values are used in the process of identifying this method, this research is expected to contribute to the neural network method coupled with the statistical method. The results of this study should be able to predict accurate demographic and economic characteristics. A Neural Network Auto-regression (NNAR) model is used to predict demographic and economic variables for the next ten years, with the best forecasting model being the NNAR (4,4), (4,4), (4,4), (11,6), (10,6), (10,6), (10,6), (5,6), (10,6), (6,4) models. The study's findings show that, except for GDP, all of the selected variables fit the NNAR model well and a comparison shows that the rural population is best fitted when using Mean Absolute Percentage Error (MAPE) when compared to the entire set of demographic and economic variables. The Rural population is the best-fitting model of the three populations; under-five mortality is well-fitting among vital rates; and age dependency ratio is the best forecasting in economic variables using mean absolute percentage error (MAPE).

Keywords: Artificial neural network, economic variables, population dynamics

1. Introduction

Time series forecasting is an important field of forecasting in which historical observations of the same variable are gathered and examined to construct a model that explains the underlying relationship. The model is then utilized to forecast the future time series. This modeling method is especially beneficial when insufficient information about the underlying data generation process is available or when no suitable explanatory model exists that tie the prediction variable to other explanatory factors. Over the last several decades, much work has been put into developing and improving time series forecasting models.

Artificial neural networks (ANNs) are some of the most accurate and commonly used forecasting models, with several applications in forecasting social, economic, engineering, foreign currency, and stock market problems, and others. Given the benefits of artificial neural networks, it is not unexpected that this technology has inspired widespread interest in time series forecasting. Artificial neural networks (ANNs) have recently been extensively researched and employed in time series forecasting. The main benefit of neural networks is their capacity to do flexible nonlinear modeling. There is no need to define an exact model form while using ANNs. Rather, the model is developed adaptively based on the features offered by the data. This data-driven technique is suited for many empirical data sets when there is no theoretical guidance to propose an acceptable data generation procedure.

Artificial neural networks are a feasible alternative to some standard time series models (Chen, Yang, Dong, & Abraham, 2005; Giordano, La Rocca, &Perna, 2007; Jain & Kumar, 2007) [2, 4, 6]. Lapedes and Farber (1987) [7] describe the first attempt to use artificial neural networks to represent nonlinear time series. De Groot and Wurtz (1991) [3] provide a thorough examination of univariate time series forecasting using feed-forward neural networks for two benchmarks of nonlinear time series. An empirical investigation on multivariate time series forecasting using artificial neural networks is conducted by Chakraborty, Mehrotra, Mohan, and Ranka (1992) [1].

G Peter Zhang (2003) [13] developed a hybrid technique that integrates both ARIMA models and ANNs to make use of the distinctive strengths of the ARIMA and ANN models in linear and nonlinear modeling. According to Mehdi Khashei and Mehdi Bijari (2010) [10], ARIMA models are used in a unique hybrid model of Artificial Neural Networks to provide a more accurate forecasting model than Artificial Neural Networks. According to Lutfiani Safitri, Sri Mardiyati, and Hendrisman Rahim (2018) [8], the Lee-Certer model is used by the actuarial community to anticipate death rates. The Lee-Carter model, which is underpinned by a neural network, will be used to anticipate mortality in Indonesia. The daily price of gold is affected by the daily price of gold from a day ago to 24 periods ago using NNAR (25,13) model by Mohamad, Sujito, and Sigit's (2020) [11].

The goal of this research is to forecast demographic and economic factors using the NNAR approach. Because statistical approaches such as the least RMSE training and testing values are used in the process of identifying this method, this research is expected to contribute to the NN method coupled with the statistical method. Hyndman created this NNAR model in 2018 using the open-source R program by Hyndman. R. J and G. Athanasopoulos (2018) [5]. Using 50 observations, the results of this study should be able to predict accurate demographic and economic characteristics. Vijayalakshmi G, K Pushpanjali and A Mohan Babuan (2023) [12] states that an appropriate model must be effectively designed for the production of wheat data in the state of Andhra Pradesh taken over the period of 1982 to 2022 using ARIMA and Neural Network Autoregressive (NNAR) models and it shows higher production values in NNAR (1, 1) Model while it shows lesser production in ARIMA (0, 1, 1) model.

In this study, an attempt is made to study the model and forecast of demographic and economic variables using the ANN model. The first section contains the introduction, the second section contains the data collection and research methodology, the third section contains the results and discussion, and the final section contains the conclusions.

2. Methods and Materials

The total population, urban population, and rural population were taken from the Census, Registrar General of India, while GDP data was gathered from the Economic Survey of India. The Sample Registration System (SRS) publication provided the Birth rate, Death rate, IMR, TFR, under-five mortality rate, Life expectancy at birth, and Age Dependency Ratio.

The forecasting approach is based on an investigation of the relationship pattern between variables to be evaluated with time variables, which are periodic series. There are two types of data forecasting methodologies based on their nature: qualitative and quantitative. Time series models and causal models are the two sorts of quantitative approaches by S. Makridakis, S. C. Wheelwright, and R.J. Hyndman (1998) [9].

The time series approach has seen fast progress, not just for probabilistic (statistical) models, but also for non-probabilistic models like NN. Forecasting NN models include ELM, Multilayer Perceptron (MLP), NNAR, and others.

2.1 Artificial Neural Network

The Artificial Neural Network (ANN) is a technique designed to mimic human neural networks. This approach is used to assist humans in a variety of sectors, including image processing, sensor instruments in homes, workplaces, and hospitals, quantitative forecasting methods, and others. ANN is made up of inputs, processes, and outputs by Hyndman. R. J and G. Athanasopoulos (2018) [5].

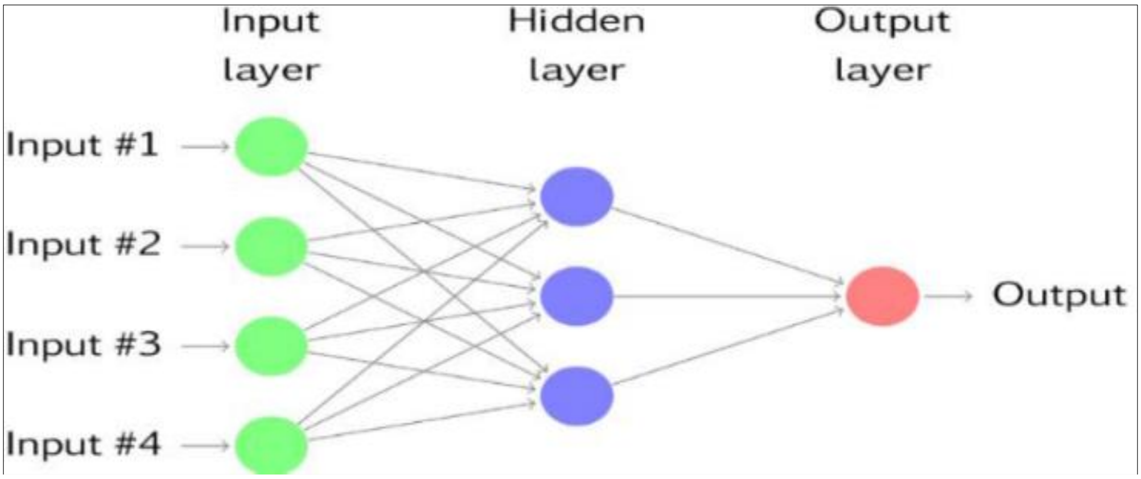


Fig 1: A neural network with three hidden neurons in one hidden layer and four inputs.

The multi-hidden layer ANN model is distinguished by the fact that it hides more than one hidden layer between the input and output layers. We will employ a hidden layer model in this work, one hidden layer with three neurons and one output, as illustrated in Figure 1 by Hyndman. R. J and G. Athanasopoulos, (2018) [5].

2.2 Backpropagation Model

The backpropagation model is an algorithmic model in supervised learning artificial neural networks. This technique is widely employed in predicting models using the ANN algorithm. Training or learning using backpropagation contains three stages: feedforward (forward feed) from the input pattern, counting errors from the learning process, and modifying the weights. In this backpropagation model, the input each nodes is a weighted linear combination. The outputs of the weighted linear combination are changed using a nonlinear function to form the output of this ANN. This linear combination function is denoted by Hyndman. R. J and G. Athanasopoulos, (2018) [5]:

$$z_j = b_j + \sum_{i=1}^{4} w_{i,j} x_i \quad (1)$$

Where
z_j is the sum function of the unit bias to j on the hidden layer
b_j is a weight in the bias unit to j
$w_{i,j}$ is weight of the layer i bias to j
x_i is the network input to i
It has a binary sigmoid activation function, which is a nonlinear function, as shown in Equation (2).

$$s(z) = \frac{1}{1 + e^{-z}} \quad (2)$$

In the equation, the nonlinear binary sigmoid function is a component of the linear combination function (1). In a single-layer network model, this binary sigmoid function is one of the functions for the backpropagation method by Hyndman. R. J and G. Athanasopoulos (2018) [5].

2.3 Neural Network Auto-regression (NNAR)
The NNAR model is an ANN, where the input layer is one variable input with lag 1, lag 2, and so on models until lag to p, thus the name ANN Auto-regressive (NNAR). Hyndman and Athanasopoulos presented NNAR in 2018 with the application R package programmer statistics in the "predict" package with the net function and R Pubs by R studio. This approach is only applicable to feed-forward networks in a single hidden layer and is represented as NNAR (p, k), where p represents lag-p as input and k as hidden layer nodes by Hyndman. R. J and G. Athanasopoulos (2018) [5]. The NNAR approach utilizes a unique hidden layer, as shown in Figure 1, and a nonlinear function, as shown in Equation (1) to give weight and provide output from ANN. We primarily examine feed-forward networks with just one hidden layer, and to denote that the hidden layer contains p-lagged inputs and k nodes, we use the term NNAR (p, k). For instance, an NNAR (9, 5) model consists of a neural network with five neurons in the hidden layer and nine neurons utilized to forecast the output y_t.

2.4 Performance Evaluation
Forecasting accuracy may be calculated to determine how accurately forecasting is performed. The mean absolute percentage error (MAPE) and root mean square error (RMSE) were applied in this study to calculate forecasting accuracy (RMSE). Equation (4-5) (Hyndman, R.J., and G. Athanasopoulos (2018) [5] is used to calculate the two accuracy levels.

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{e_i}{y_i} \right| \quad (4)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} {e_i}^2}{n}} \quad (5)$$

Where n denotes the number of actual values and e_i represents the error, which means actual values minus predicted values.

3. Result and Discussion
This study examines the forecasting of the demographic and economic variables from 1971 to 2020, which is 50 years of observation, and the analysis of the demographic and economic variables for prediction for the next ten years.

Table 1: Model selection for demographic and economic variables

		RMSE	
Variables	**NNAR**	**training**	**testing**
Total population	NNAR (2,4)	199225.5	27730066.1
	NNAR (3,4)	205843.8	28524711.5
	NNAR (4,4)	191675.1	20910852.3
Urban population	NNAR (2,4)	94564.92	18360369.4
	NNAR (3,4)	72141.08	15582940.5
	NNAR (4,4)	67929.6	18990193.2
Rural population	NNAR (2,4)	147010.2	15346464
	NNAR (3,4)	133805.3	11177977.1
	NNAR (4,4)	108203.2	9379160
Birth rate	NNAR (9,6)	0.0297	0.2006
	NNAR(10,6)	0.0287	0.2090
	NNAR(11,6)	0.0239	0.2303
Death rate	NNAR(8,6)	0.0547	0.6423
	NNAR(9,6)	0.0324	0.5232
	NNAR(10,6)	0.3164	0.4563
IMR	NNAR(8,6)	0.2419	11.3309
	NNAR(9,6)	0.2355	6.0530
	NNAR(10,6)	0.2267	8.3574
TFR	NNAR(3,6)	0.0523	0.0667
	NNAR(4,6)	0.0463	0.1189
	NNAR(5,6)	0.0295	0.2447
Under 5 mortality rate	NNAR(8,6)	0.0734	3.1902
	NNAR(9,6)	0.0706	2.6701
	NNAR(10,6)	0.0716	2.6214
Life expectancy at birth	NNAR(4,4)	0.1508	0.2553
	NNAR(5,4)	0.1533	0.9540
	NNAR(6,4)	0.1479	1.0614
GDP	NNAR(2,1)	42463.43	6518081.12

	NNAR(3,1)	42771.5	6940832
	NNAR(4,1)	42365.5	6985030.3
Age dependency ratio	NNAR(2,1)	0.3802	3.3752
	NNAR(3,1)	0.3710	3.3863
	NNAR(4,1)	0.3723	3.4656

3.1 Data training
To make the ANN model work properly, the data training must be completed. First, 40 data points from 50-year demographic and economic variables were used to create training data samples using R programming. The training model with the lowest RMSE value should be selected as the best. Following that, models with the lowest RMSE training values will be chosen. The selection of the best NNAR model among three NNAR models is based on RMSE in forecasting the year's population. Table 1 displays the results of the calculations.

3.2 Data Testing
Testing data was drawn from the last ten years of selected demographic and economic variables data, as the training and testing data should be balanced in number because it is a sample taken to create and validate the model in R programming. The criteria used to select the model tested are the same as those used in the training model, which are the smallest error, which is RMSE. The results of model selection in the training model and testing model, which are shown in Table 1.

The selected NNAR model represents the entire population as NNAR (4, 4), which denotes a neural network with four neurons in the single hidden layer and four past observations utilized to forecast the output. In addition, the remaining the selected variables are the same. We chose one ANNR model among the three since it has the lowest RMSE will be utilized for predicting.

3.3 Forecasting of the demographic variables
Demographic variables are used to define the nature of samples prepared from populations.

Table 2: Forecasts of the demographic variables from 2021 to 2030

Year	Total Population	Urban Population	Rural Population	Birth rate	Death rate	IMR	TFR	U5M	Life expectancy
2021	1412187698	483450099	928710004	19.3	6.0	28.9	2.0	29.9	70.2
2022	1432030444	493900108	937761262	19.1	5.9	28.3	2.0	27.9	70.3
2023	1451318497	503810393	946486840	18.9	5.9	27.9	1.9	26.1	70.4
2024	1469866455	513154134	954961511	18.8	5.8	27.5	1.9	24.6	70.5
2025	1487474121	521800180	962978740	18.6	5.7	27.1	1.9	23.2	70.5
2026	1504373905	529663043	970736721	18.4	5.6	26.8	1.9	21.9	70.6
2027	1520227854	536685560	978082324	18.2	5.6	26.6	1.9	20.7	70.6
2028	1534975341	542840946	985038266	18.1	5.5	26.5	1.8	19.5	70.7
2029	1548589843	548136715	991552311	17.9	5.4	26.3	1.8	18.5	70.7
2030	1561082167	552609488	997659376	17.8	5.3	26.1	1.8	17.4	70.8

Note: IMR- Infant Mortality rate, TFR- Total fertility rate, U5M- Under-5 mortality rate

Table 2 shows that over the following five years, India's population as a whole is projected to increase by 1.49 billion, from 1.39 billion in 2020 to 1.56 billion in 2030, reaching an average of about 0.17 billion (170 million). The percentage distribution of total population shows 12% increase is observed for period of ten years and annually 1.2%. The urban population was 0.47 billion in 2020 and is expected increase 0.08 billion (80 million) in ten years and it is percentage of increase from initially period to finally period of forecasting is 17% and annually it is increases by 1.7%. The population of rural areas was 0.92 billion in 2020, and it is expected that it would rise by 0.07 billion (70 million) in ten years. The percentage increase from the initial period to the final period of forecasting is 7%, and it rises by 0.7% yearly. Figure 2(a), (b), and (c) show that populations' actual values for the black line and training values for the blue line are best fitted for the NNAR (4, 4) model, and after ten forecasted values, the red line is showing an increasing trend, but last year may has been slow decreases due to the total fertility rate being less than two per woman in Table 1, massive urbanization, family planning, people's high standard of living, late family planning, and late marriage.

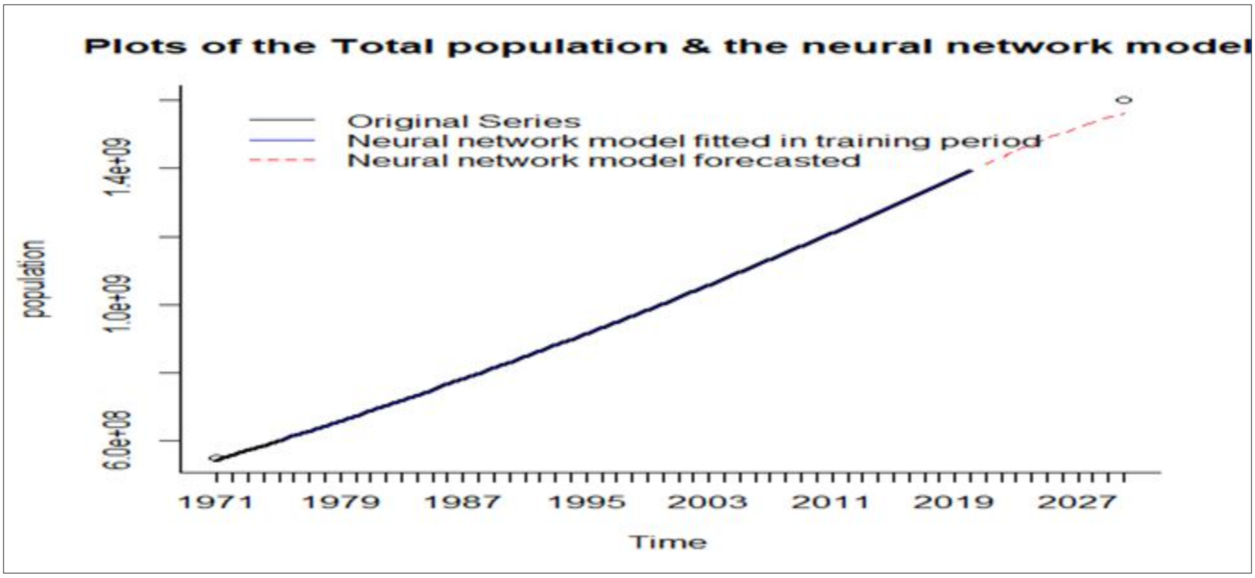
India had a birth rate of 19.6 per 1,000 people in 2020. By 2025, that rate was predicted to decline to 18.6 per 1,000 people, which would represent a 5.1% decrease during the next nearly five years, and by 2030, it would have fallen to 17.8 per 1,000 people, a 9.1% decrease from 2020. The most significant factors influencing fertility decline are late marriage, family planning, female education, urbanization, methods of contraception, as well as medical care.

In 2020, India's death rate was 6.0 per 1,000 people, with a forecast drop to 5.7 per 1,000 people in the next half-decade, a 5% decline, and 5.3 per 1,000 people by the end of the decade, an 11.6% decrease from 2020. The IMR in India is declined by 3 children per 1000 people for the period of 5 year from 2020 to 2025 and it is percentage of decline is 9.6. In the later period this decline is slow that is one child for the period 5 years (from 2025 to 2030). It was expected that by 2030, the rate would have decreased to 26.1 deaths per 1,000 kids, which represents a 13% decrease from the decade before it, from 27.1 deaths per 1,000 children in the next five years, a 9.6% decrease from 2020.

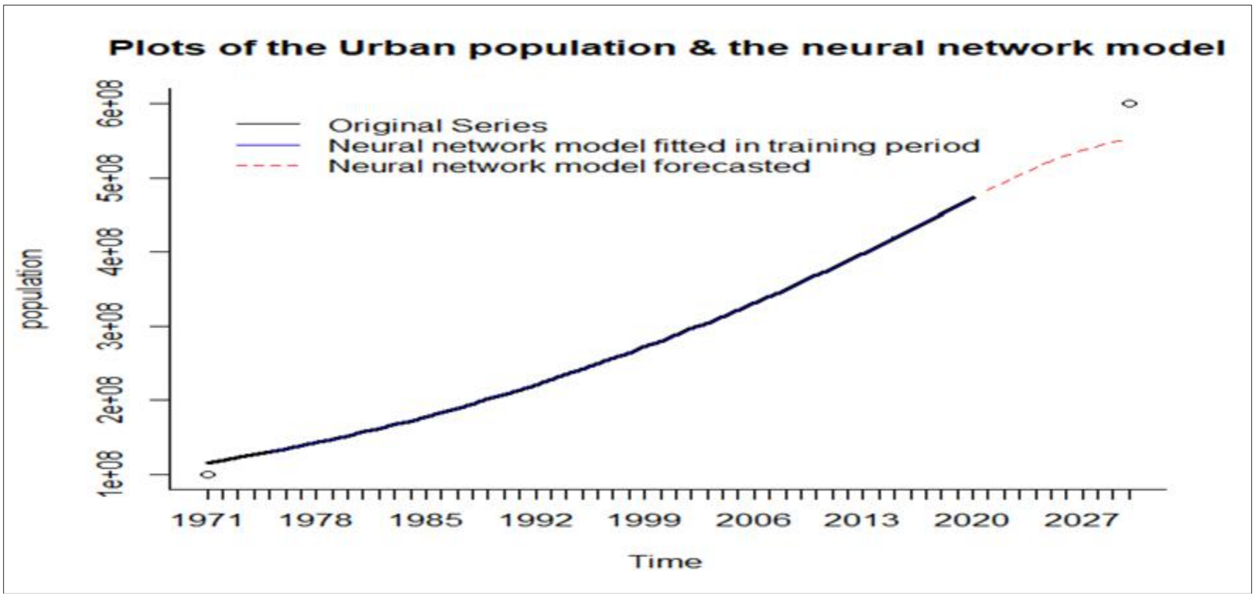
The TFR is 2 children per women in 2020 and at the end of prediction period these 1.8 (2 children) per women. This decline is hardly 10 percent from initial period to end period of forecasting. Due to planning, chemical food, late marriage, stress, business planning, and packaged food, this rate drops by fewer than two per woman.

Under-five mortality rate in India for the period 2020 to 2030 was declined by 14.4 children. For the first 5 year of forecasting Under-five mortality is declined by 8.6 children per 1000 live births and next 5 year (2025 to 2030) its decline is 6 children per 1000 live births. Figure 2 shows that the death rate, IMR, TFR, and under-five mortality rate are decreasing, with the actual values shown in black and expected values shown in blue; both are well-fitted NNAR models, and the next forecasted NNAR model is best fit for the red line. Because of health facilities and female education, these rate factors are on the decline.

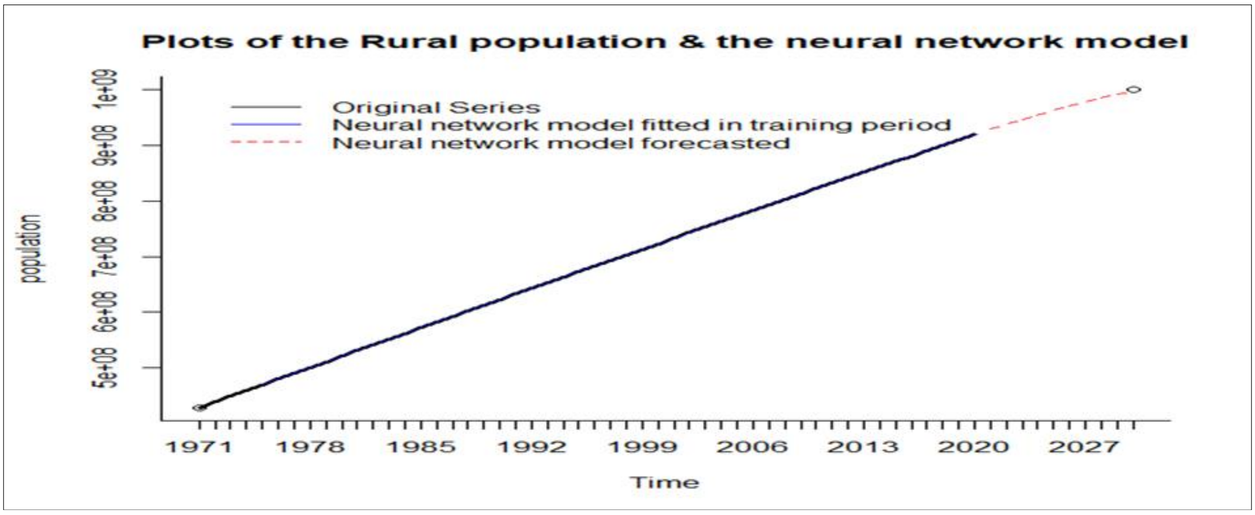
In India, the average life expectancy at birth was 70 years in 2020, increased to 70.5 years in 2025 (a 0.7% increase over the next five years), and is predicted to reach 70.8 years by 2030 (a 1.1% increase over the earlier ten years). A higher return on human capital is produced by longer life expectancies, which encourage more spending on health and education, boost economic growth, and lower mortality.



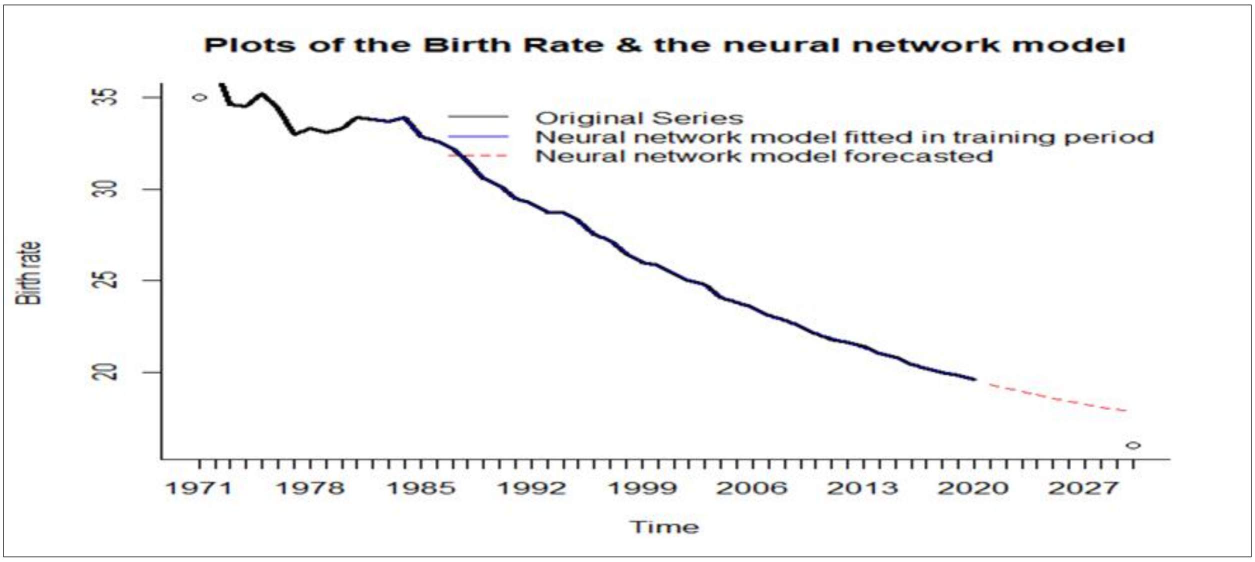
(a)



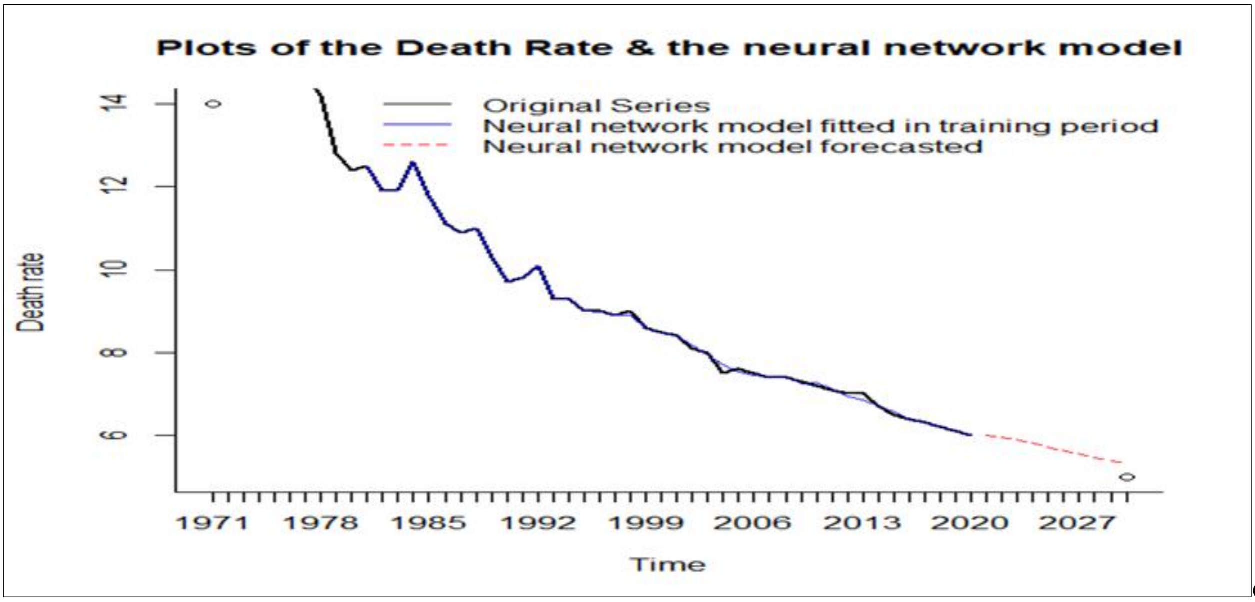
(b)



(c)

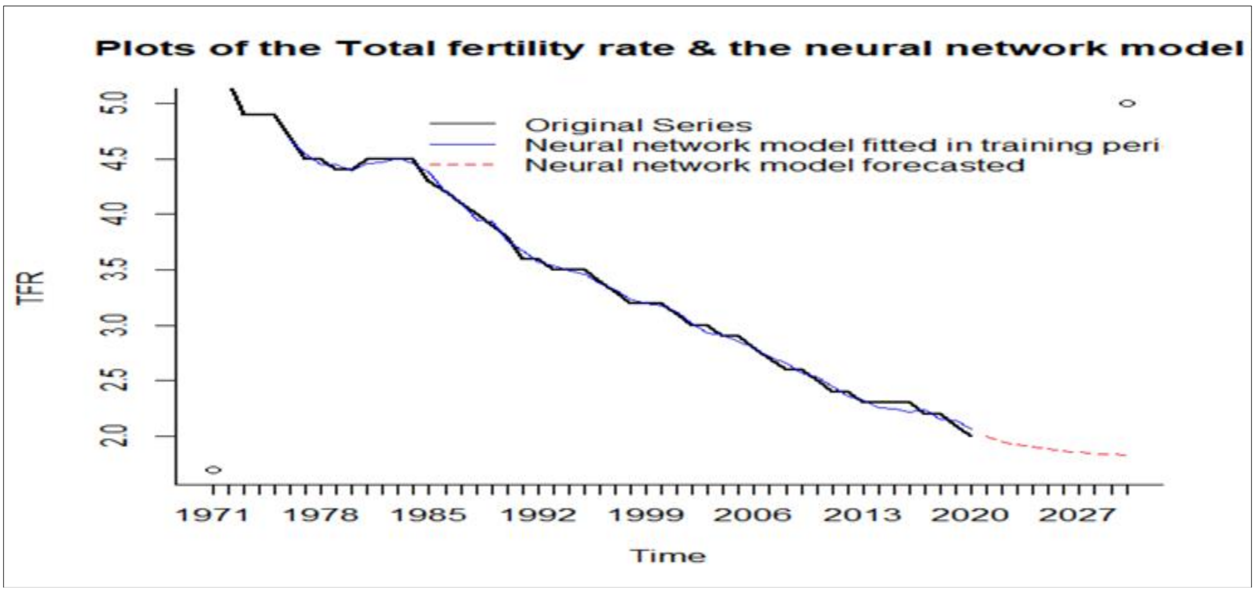

Plots of the Birth Rate & the neural network model
Original Series
Neural network model fitted in training period
Neural network model forecasted
Birth rate
35
30
25
20
1971 1978 1985 1992 1999 2006 2013 2020 2027
Time

(d)

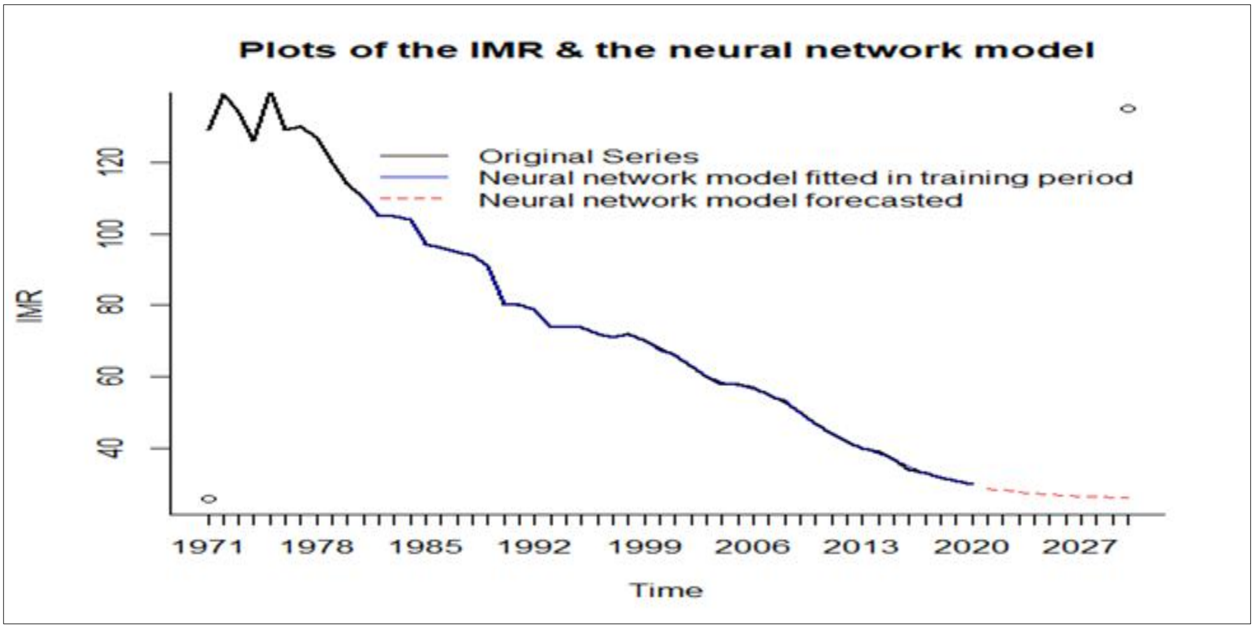

Plots of the Death Rate & the neural network model
Original Series
Neural network model fitted in training period
Neural network model forecasted
Death rate
14
12
10
8
6
1971 1978 1985 1992 1999 2006 2013 2020 2027
Time

e

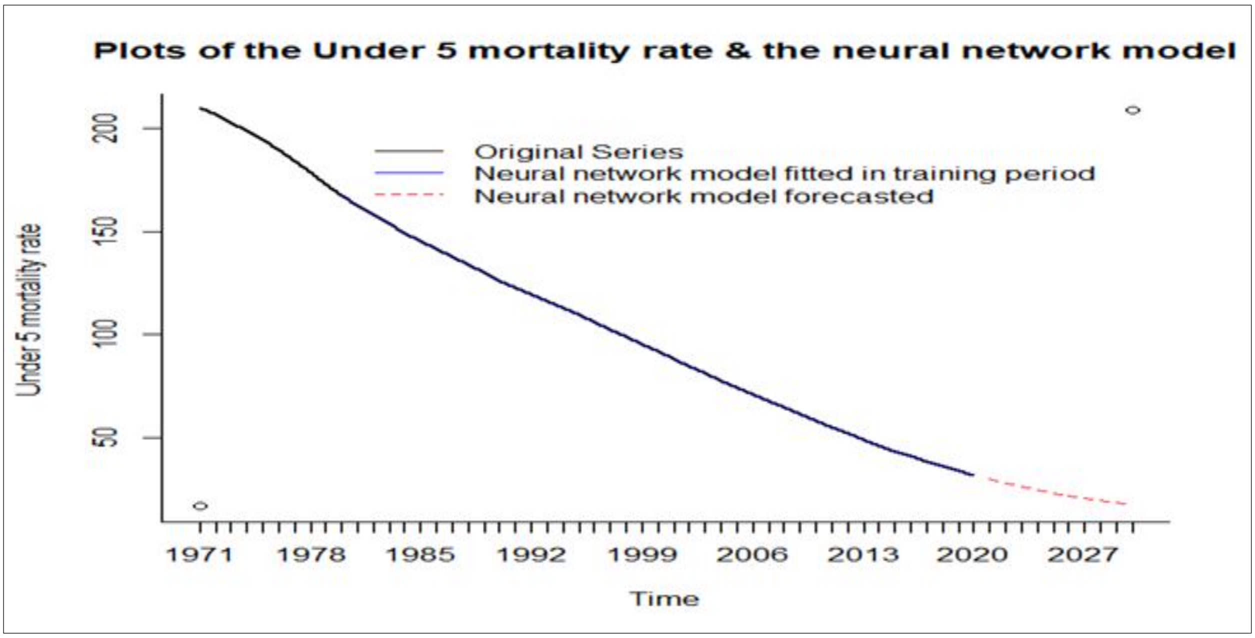
(e)


Plots of the Total fertility rate & the neural network model
Original Series
Neural network model fitted in training peri
Neural network model forecasted
TFR
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1971 1978 1985 1992 1999 2006 2013 2020 2027
Time

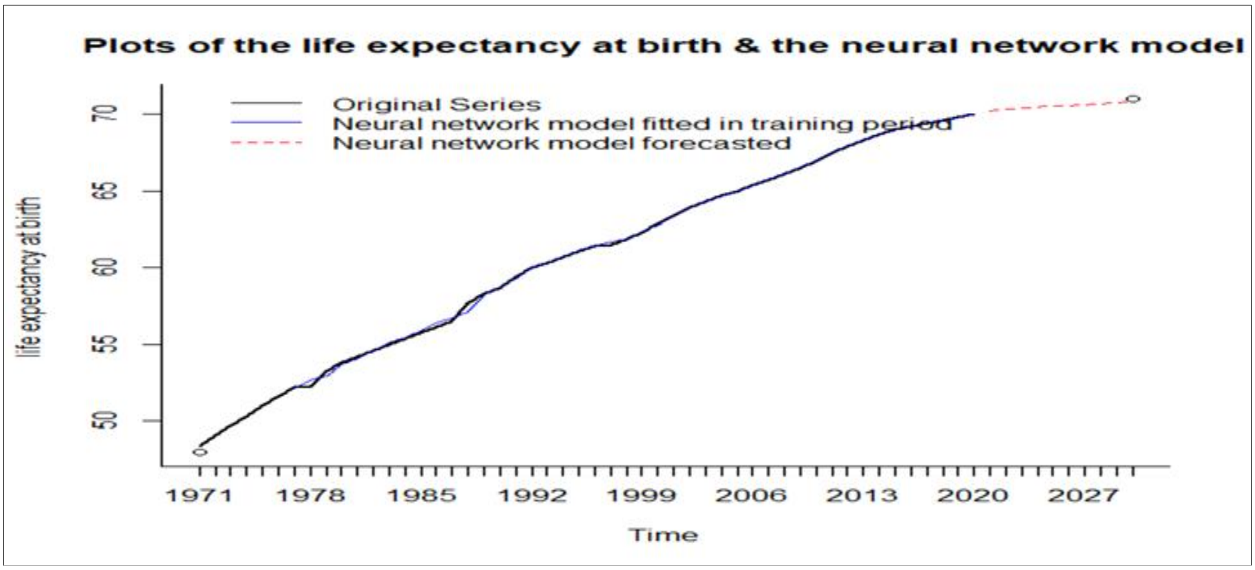
(f)



(g)



(h)



(i)

Fig 2: Forecasted of the demographic variables from 2021 to 2030

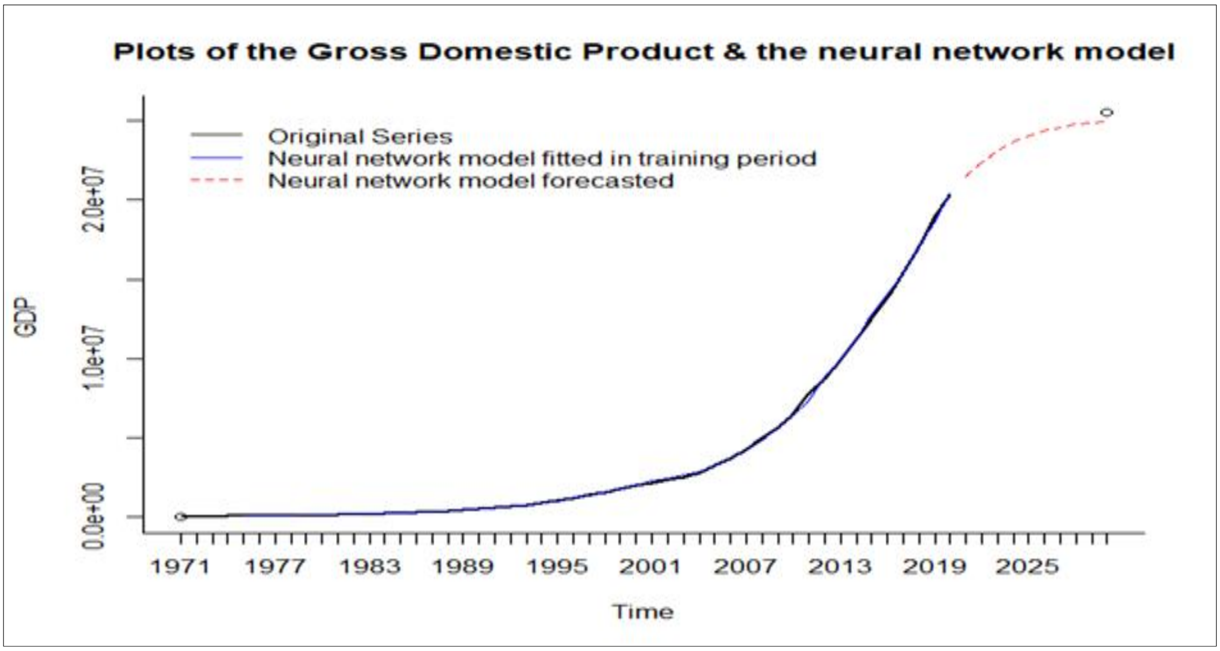
3.3 Forecasting of the Economic variables

An economic variable is any measurement that contributes to analyzing how an economy performs.

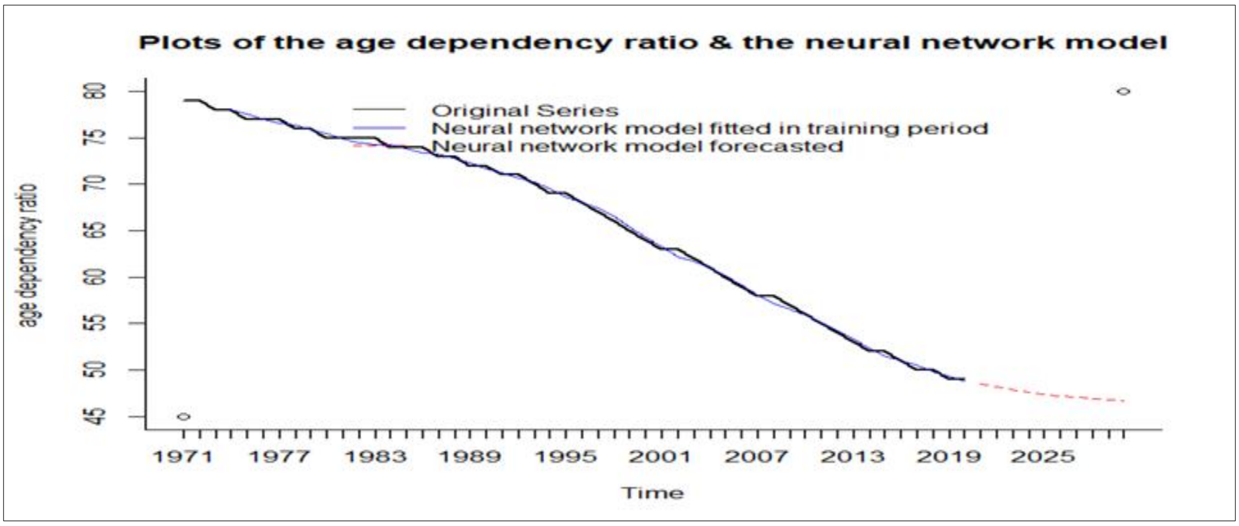
Table 3 shows that India's GDP was 203 trillion in 2020 and that it anticipates rising to 240 trillion in the next half-decade, which would represent an increase of 18.2% from 2020, and 249 trillion in the following decade, which would indicate a rise of 22.6% over the following ten years. In the following ten years, it will increase by 46 trillion.

Table 3: Prediction of the economic variables from 2021 to 2030

Year	GDP	Age Dependency Ratio
2021	21450729	48.5
2022	22348653	48.2
2023	23086679	47.9
2024	23653641	47.6
2025	24071380	47.4
2026	24373610	47.2
2027	24593880	47.1
2028	24754605	46.9
2029	24870889	46.8
2030	24954093	46.7



(a)



(b)

Fig 3: Forecasted of the economic variables from 2021 to 2030

According to predictions, India's age dependency ratio will drop from 49 in 2020 to 47.4 in 2025, a 3.2% decline over the next half-decade, and 46.7 by 2030, a 4.6% decrease over the next ten years. According to the Neural Network Auto-regression model in Figure 3, as the working-age population grows, the age dependence ratio growth rate will begin to decrease over the following ten years. The age-dependency ratio will gradually decline as the GDP growth rate rises.

Table 4: Accuracy values for forecasting

Variables	MAPE
Total population	0.0180
Urban population	0.0357
Rural population	0.0179
Birth rate	0.1306
Death rate	0.4280
IMR	0.4314
TFR	1.1816
Under 5 mortality rate	0.0859
Life expectancy at birth	0.1486
GDP	5.5780
Age dependency ratio	0.4945

The demographic and economic variables are well fitted to the neural network auto-regression model for forecasting. According to Table 4, the NNAR (4,4), (4,4), (4,4), (11,6), (10,6), (10,6), (10,6), (5,6), (10,6), and (6,4) models are the best predicting models. Forecasting accuracy values for demographic variables such as total population, urban population, rural population, birth rate, death rate, IMR, TFR, Under-5 mortality rates, and life expectancy at birth are 0.0180, 0.0357, 0.0180, 0.1306, 0.4280, 0.4314, 1.1816, 0.0859, and 0.1486. The NNAR (4, 1) and (3, 1) models are the best forecasting models for economic variables. The best predicting accuracy values are 5.5780 and 0.4945 for GDP and age dependence ratio, respectively.

4. Conclusion

The study is a successful attempt to predict demographic and economic factors over the next 10 years. The NNAR method can predict demographic and economic variables as well. The NNAR (4,4), (4,4), (4,4), (11,6), (10,6), (10,6), (11,6), (5,6), (10,6), and (6,4) models are the best forecasting models. Among three populations, rural population is the best fitting model; among vital rates, under-five mortality is well-fitting and age dependency ratio is the best fitting economic variable. Among all the selected variables, rural population is the best predictor. India's population growth suggests that it has reached a stabilizing point with a dropping fertility rate. India is taking advantage of the demographic dividend. The birth rate has decreased, but the death rate has also decreased. The Indian population is ageing as life expectancy rises. The urban population is slowly growing, whereas the rural population is slowly declining. India will have a larger working population than in past years. This condition provides enough necessary labour, capital savings owing to a lower dependence ratio, and human capital since women also earn due to a lower number of children. It is also increasing the demand for finished goods as the standard of living rises.

References

1. Chakraborty K, Mehrotra K, Mohan CK, Ranka S. Forecasting the behavior of multivariate time series using neural networks. Neural Networks. 1992;5:961-970.
2. Chen Y, Yang B, Dong J, Abraham A. Time-series forecasting using flexible neural tree model, Information Sciences. 2005;174(3&4):219–235.
3. De Groot C, Wurtz D. Analysis of univariate time series with connectionist nets: A case study of two classical examples. Neurocomputing. 1991;3:177–192.
4. Giordano F, La Rocca M, Perna C. Forecasting nonlinear time series with neural network sieve bootstrap. Computational Statistics and Data Analysis**.** 2007;51:3871-3884.
5. Hyndman RJ, Athanasopoulos G. Forecasting: principles and practice, 2nd edition, OTexts: Melbourne, Australia, c2018.
6. Jain A, Kumar AM. Hybrid neural network models for hydrologic time series forecasting. Applied Soft Computing. 2007;**7**:585–592.
7. Lapedes A, Farber R. Nonlinear signal processing using neural networks: Prediction and system modeling. Technical Report LAUR-87-2662, Los Alamos National Laboratory, Los Alamos, NM, c1987.
8. Lutfiani Safitri, Sri Mardiyati, Hendrisman Rahim. Forecasting the mortality rates of Indonesian population by using neural network, International Conference on Mathematics: Pure, Applied and Computation, IOP Conf. Series: Journal of Physics: Conf. Series; c2018; p.974-012030.
9. Makridakis S, Wheelwright SC, Hyndman RJ. Forecasting: Methods and Applications, John Wiley & Sons, Inc; c1998.
10. Mehdi Khashei, Mehdi Bijari. An artificial neural network (p,d q) model for t me series forecasting, Expert Systems with Applications, 2010;37:479-489.
11. Mohamad As'ad, Sujito, Sigit Setyowibowo. Prediction of Daily Gold Prices Using an Autoregressive Neural Network, Inform: Jurnal Ilmiah Bidang Teknologi Informasidan Komunikasi. 2020;5(2):69-73.
12. Vijayalakshmi G, Pushpanjali K, Mohan Babu A. A comparison of ARIMA & NNAR models for production of rice in the state of Andhra Pradesh, International Journal of Statistics and Applied Mathematics. 2023;8(3)**:**251-257.
13. Zhang GP. Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing. 2003;50:159-175.