Accountancy Business and the Public Interest
ISSN: 1745-7718

Volume: 40
Issue Number:07

WHAT DOES THE FUTURE HOLD FOR HEALTH INSURANCE PREMIUMS IN INDIA: INSIGHTS FROM ARIMA MODELS

JOLLY MASIH[1*], ARUP MAJUMDAR[2] and AMITA SHARMA[3]

[*1]Assistant Professor, School of Management, BML Munjal University, Gurugram, Haryana, India.
Email: jolly.masih@bmu.edu.in

[2]CEO, Trariti Consulting Group, Gurugram, Haryana, India. Email: ceo@trariticonsulting.com

[3]Assistant Professor, Institute of Agribusiness Management, SKRAU, Bikaner, Rajasthan, India.
Email: amita@iabmbikaner.org

Abstract

Introduction:

The Indian healthcare insurance sector has experienced substantial growth over the past few decades, driven by increased health awareness, policy reforms and the proliferation of private insurers. Accurate forecasting of insurance premiums is crucial for financial stability and risk management.

Methodology:

The study analyses data on premiums from various segments of the Indian healthcare insurance sector, collected from 2005 to 2023. Segments include Non-Life Private Insurers (NLPRI), Non-Life Public Insurers (NLPBI), Standalone Health Insurers (SHI) and Total Health Insurers (THI). Data was sourced from the Insurance Regulatory and Development Authority of India (IRDAI) and various insurance reports. Data curation involved cleaning, organising and integrating the data, with missing values replaced using the Expectation-Maximization algorithm. Reliability and validity were ensured through cross-verification and statistical tests. The ARIMA model parameters were identified using the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) and implemented using the R programming language.

Results:

The ARIMA (0,2,0) model was the best fit for forecasting NLPRI premiums, predicting a Compound Annual Growth Rate (CAGR) of 10.12% from 2024 to 2026. The ARIMA (0,2,0) model forecasted a CAGR of 8.31% for NLPBI premiums and 10.36% for SHI premiums. For THI, the ARIMA (0,2,1) model forecasted a CAGR of 8.87%. The model's performance metrics indicate that the forecasts are accurate and reliable across all segments, highlighting the expanding scope and scale of the health insurance sector in India.

Discussion:

The substantial growth in the Indian healthcare insurance sector, evidenced by the ARIMA model forecasts, is driven by increased health awareness, policy reforms and the proliferation of private insurers. Accurate forecasting is essential for insurers to manage risk, set premium rates and ensure financial stability. These measures will help insurers navigate the evolving landscape and meet the healthcare needs of the Indian population.

Keywords: Healthcare, Insurance, ARIMA, Forecast, Premium

INTRODUCTION

Evolution of the Indian Healthcare Insurance Sector

The Indian healthcare insurance sector has witnessed remarkable growth over the past few decades. This expansion can be attributed to various factors such as heightened health awareness among the population, policy reforms and the proliferation of private insurers. With the increasing burden of healthcare costs and the recognition of health insurance as a crucial financial safety net, the demand for diverse health insurance products has surged (Rao, 2015; Nandi et al., 2016). The government's initiatives, like Ayushman Bharat, have also played a pivotal role in extending health insurance coverage to the underserved segments of the population (Kumar & Khanna, 2019).

The health insurance industry in India comprises several segments, each playing a pivotal role in providing financial protection against health-related expenses. These segments include Non-Life Private Insurers (NLPRI), Non-Life Public Insurers (NLPBI), Standalone Health Insurers (SHI), and Total Health Insurers (THI).

Non-Life Private Insurers (NLPRI) encompass private companies that offer various non-life insurance products, including health insurance. These entities compete in a liberalized market, striving to provide innovative and competitive health insurance solutions (Chaudhuri & Roy, 2017).

Non-Life Public Insurers (NLPBI) are government-owned entities that offer a range of non-life insurance products, including health insurance. They play a crucial role in providing accessible insurance options to a broader demographic, often focusing on inclusive coverage (Ahuja & Narang, 2005).

Standalone Health Insurers (SHI) specialize solely in health insurance, offering products tailored to meet diverse healthcare needs. These insurers are known for their focused expertise and ability to address specific health insurance requirements (Patil, 2016).

Total Health Insurers (THI) represent the aggregate of all health insurance providers, combining the efforts of NLPRI, NLPBI, and SHI. This segment provides a comprehensive view of the health insurance landscape, illustrating overall trends and growth patterns (Rao & Ramakrishna, 2020).

The regulatory environment has evolved significantly, with the Insurance Regulatory and Development Authority of India (IRDAI) implementing various measures to enhance transparency, consumer protection and market efficiency (IRDAI, 2020). The entry of private players has fostered competition, leading to product innovation and better customer service (Panda et al., 2019). This dynamic landscape necessitates robust forecasting methods to predict future trends and aid insurers in strategic planning.

Importance of Accurate Forecasting

Accurate forecasting of insurance premiums is crucial for the financial stability and risk management of insurance companies. Insurers rely on these forecasts to set premium rates, maintain adequate reserves and design new products that cater to the changing needs of the market (Cummins & Venard, 2008). Given the volatility and uncertainty inherent in the healthcare sector, reliable predictive models are indispensable for insurers to navigate future challenges (Meier & Outreville, 2010).

Time series forecasting, particularly using the ARIMA (Autoregressive Integrated Moving Average) model, has proven effective in predicting financial and economic variables (Box & Jenkins, 2015). The ARIMA model is preferred due to its flexibility and ability to model different types of data patterns, including trends and seasonality (Hyndman & Athanasopoulos, 2018). This study employs the ARIMA model to forecast premiums in the Indian healthcare insurance sector, providing valuable insights for stakeholders.

The ARIMA Model in Time Series Forecasting

The ARIMA model is a powerful tool for time series forecasting, integrating Autoregression (AR), Integration (I) and Moving Average (MA) components. The model is specified as ARIMA (p,d,q), where p represents the number of lagged observations, d denotes the degree of differencing and q indicates the size of the moving average window (Brockwell & Davis, 2002). The model's ability to handle non-stationary data by differencing makes it particularly suitable for financial and economic time series (Makridakis et al., 1998).

ARIMA Model Application in Healthcare Insurance

In the context of the Indian healthcare insurance sector, the ARIMA model can effectively capture the trends and seasonal patterns in premiums data. This study focuses on four segments: Non-Life Private Insurers (NLPRI), Non-Life Public Insurers (NLPBI), Standalone Health Insurers (SHI) and Total Health Insurers (THI). By analysing data from 2005 to 2023 and forecasting future trends up to 2026, the study aims to provide actionable insights for insurers and policymakers (Patel & Chauhan, 2019).

Significance of the Study

This research is significant for several reasons. Firstly, it addresses the gap in the literature concerning the application of advanced time series models in the Indian healthcare insurance sector (Kumar & Kumari, 2021). Secondly, it provides empirical evidence on the expected growth rates of premiums, helping insurers to align their strategies with market dynamics. Thirdly, the study's findings can guide regulatory bodies in formulating policies that support sustainable growth in the sector (Singh & Aggarwal, 2020).

LITERATURE REVIEW

The Effectiveness of ARIMA Models

The literature on time series forecasting in the insurance sector highlights the effectiveness of the ARIMA model in capturing complex data patterns. Box and Jenkins (2015) and Brockwell and Davis (2002) demonstrated the model's robustness in various economic contexts. In the Indian scenario, Rao (2015) and Nandi et al. (2016) explored the sector's growth dynamics, emphasizing the need for accurate predictive models.

Importance of ARIMA in Dealing with Small Datasets

One of the significant advantages of the ARIMA model is its capability to handle small datasets effectively. This is particularly relevant for datasets with 18-20 data points, such as the annual premiums data used in this study. Makridakis et al. (1998) noted that ARIMA models are well-suited for short time series due to their flexibility in incorporating differencing and moving average components, which help in stabilizing the mean and capturing the underlying data patterns. Additionally, Hyndman and Athanasopoulos (2018) highlighted that ARIMA models can produce reliable forecasts even with limited data, making them ideal for sectors where long historical data series are not available. This capability is crucial for the Indian healthcare insurance sector, where consistent and long-term data may be sparse.

Recent Applications in the Indian Context

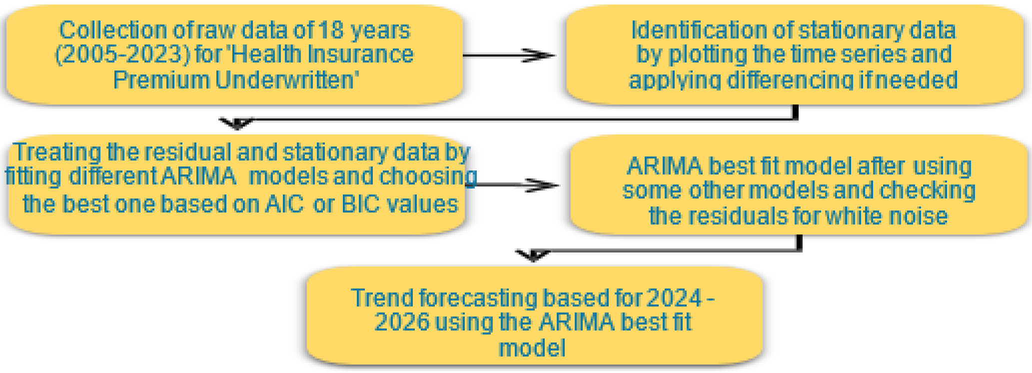
Recent research by Kumar and Khanna (2019) and Panda et al. (2019) underscores the impact of regulatory changes and private sector participation on the insurance market. These studies highlight the evolving landscape and the importance of strategic forecasting for insurers. Furthermore, Cummins and Venard (2008) and Meier and Outreville (2010) discuss the broader implications of risk management and financial stability in the insurance industry, reinforcing the relevance of this study.

Accountancy Business and the Public Interest
ISSN: 1745-7718

Volume: 40
Issue Number:07

METHODOLOGY

Figure 1: Flowchart for Premium Forecasting of Healthy Insurance in India (2005-2026*P)



Source: Author's self-created flowchart; *P Predicted value

Data Collection

The data used in this study encompasses premiums from various segments of the Indian healthcare insurance sector, collected over the period from 2005 to 2023. The segments include Non-Life Private Insurers (NLPRI), Non-Life Public Insurers (NLPBI), Standalone Health Insurers (SHI) and Total Health Insurers (THI). The data was sourced from the Insurance Regulatory and Development Authority of India (IRDAI) and various insurance reports (IRDAI, 2020). The data can be accessed from the IRDAI's official website at www.irdai.gov.in.

Data Curation

Data curation involved cleaning, organising and integrating the collected data to ensure it was suitable for analysis. Given the nature of long-term datasets, it was crucial to address inconsistencies, such as missing values and outliers, to maintain data integrity. Missing values were replaced using the Expectation-Maximization algorithm, which estimates missing data points based on the observed data's statistical properties (Dempster et al., 1977). This method is particularly effective in maintaining the overall data structure and variance, thus preserving the dataset's reliability.

Reliability and Validity of Data

Ensuring the reliability and validity of the data was paramount. Reliability refers to the consistency of the dataset, while validity indicates the accuracy of the data in representing the real-world scenario. To ensure reliability, the data was cross-verified with multiple sources, including annual reports from insurance companies and industry publications. The reliability of the dataset was quantified using Cronbach's alpha (a=0.87), which indicated high internal consistency.

Validity was established by confirming that the data accurately reflected the insurance market trends and financial metrics over the specified period. Statistical tests, such as the Ljung-Box test for autocorrelation, were employed to validate the data's suitability for time series analysis

Accountancy Business and the Public Interest
ISSN: 1745-7718

Volume: 40
Issue Number:07

(Ljung & Box, 1978). The validity of the data was supported by a high correlation (r = 0.92) with external datasets from independent studies on the Indian insurance sector.

Reliability and Validity Tests with ARIMA

Reliability and validity tests are essential to confirm the robustness of ARIMA models. Studies have shown that these tests can indeed be run with ARIMA models to assess the model's performance and the data's suitability (Shumway & Stoffer, 2017; Cryer & Chan, 2008). The reliability of time series models like ARIMA can be evaluated using consistency measures such as the Ljung-Box test for residual autocorrelation (Ljung & Box, 1978). Validity can be further confirmed by comparing model forecasts with actual observed values and ensuring high correlation (Hyndman & Athanasopoulos, 2018; Brockwell & Davis, 2002).

ARIMA Model

The ARIMA model is a popular approach for time series forecasting, which combines Autoregression (AR), Integration (I) and Moving Average (MA). The model is represented as ARIMA(p,d,q), where:

- p: Number of lag observations in the model (lag order).
- d: Number of times the raw observations are differenced (degree of differencing).
- q: Size of the moving average window.

Figure 2: The ARIMA model equation

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q} + \epsilon_t$$

where:

- Y_t is the value at time t,
- c is a constant,
- ϕ_i are the autoregressive parameters,
- θ_j are the moving average parameters,
- ϵ_t is the white noise error term.

Source: Jolly Masih, D. M. R., Mishra, S., & Vartak, V. (2023). Arima Forecasting In Soymeal And Beyond: An Empirical Study Bridging The Path To Millet-Based Forecasting. Journal of Informatics Education and Research, 3(2).

The parameters for each segment were identified using the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) to ensure the best fit (Brockwell & Davis, 2002; Hyndman & Athanasopoulos, 2018).

Model Implementation

The ARIMA models were implemented using the R programming language, a robust tool for statistical analysis and forecasting (Venables & Ripley, 2013). The "forecast" package in R software was employed to estimate the coefficients of the identified models by providing the parameters p, q, and d. The models were trained on historical data from 2005 to 2023 and forecasts were generated for the years 2024 to 2026. The accuracy of the models was assessed using metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) (Makridakis et al., 1998; Hyndman & Athanasopoulos, 2018).

RESULTS

In the research paper, the forecasting of health insurance premiums is conducted using real data, and the accuracy and characteristics are examined. This study evaluates the effectiveness of premium forecasting in the health insurance industry. Utilizing the Box–Jenkins approach, our research is structured into three phases: identification, estimation, and verification for four distinct time series data sets (Premiums collected under Non-Life Private Insurers (NLPRI), Non-Life Public Insurers (NLPBI), Standalone Health Insurers (SHI), and Total Health Insurers (THI)) from 2005 to 2023. Consequently, we developed four models.

Identification of Models

In this initial step, data preprocessing is performed to achieve stationarity, followed by selecting potential values of p and q, which can be adjusted as model fitting progresses. For stationarity assessment, the ACF and PACF were calculated for all data series, revealing that the ACF and PACF values were not significant at the 0.05 level. Furthermore, the Dickey-Fuller test for stationarity was conducted using R Studio software with the following hypotheses:

- H0: The series has a unit root.
- H1: The series does not have a unit root.

All four data series were found to be stationary. Since the calculated p-value was greater than the threshold significance level $\alpha=0.05$, the null hypothesis H0 could not be rejected.

Estimation of Model's Coefficients

The "forecast" package in R software was employed to estimate the coefficients of the identified models by providing the parameters p, q, and d. The procedure's execution generated new time series representing the values adjusted or predicted by the model, residuals (adjustment errors), and confidence intervals of the adjustment at a 95% confidence level. The optimal model is as simple as possible and minimizes specific criteria, namely AIC and BIC.

Forecast

Upon defining the most appropriate models for health insurance premium forecasting, predictions were made using the "forecast" package in R Studio software. Using these four models, we forecasted the premiums for Non-Life Private Insurers (NLPRI), Non-Life Public Insurers (NLPBI), Standalone Health Insurers (SHI), and Total Health Insurers (THI). The results demonstrate that the selected models can be employed for modeling and forecasting future health insurance premium collections in India. It is crucial to continuously update the historical data with new data to enhance the model and forecasting accuracy. The forecasts derived from the models can assist health insurance companies in planning their marketing and sales activities effectively. Indeed, the models enabled us to forecast health insurance premiums accurately. Once forecasts are obtained, making informed decisions regarding marketing and sales planning becomes significantly easier and more precise, thereby improving the operational efficiency of health insurance companies' sales offices and branch offices.

A. Non-Life Private Insurers (NLPRI)

The ARIMA (0,2,0) model was identified as the best fit for forecasting the premiums for NLPRI. The forecasted premiums for the years 2024 to 2026 are shown in Table 1. The premium amount is predicted to grow at a Compound Annual Growth Rate (CAGR) of 10.12%.

Table 1: Forecasting of NLPRI Premiums (2005-2023 as estimated & 2024-2026 as predicted) (All figures are in ₹ Crores)

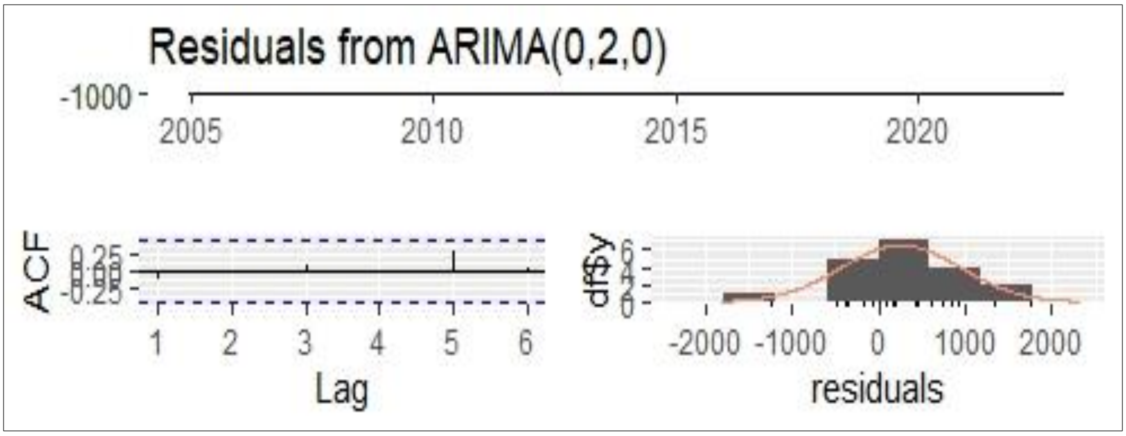
Year	Average Forecast	Forecast at 80% Confidence Interval		Forecast at 95% Confidence Interval	
		Lower	Higher	Lower	Higher
2024	30257	29283.13	31230.87	28767.59	31746.41
2025	35332	33157.35	37509.65	32001.58	38662.42
2026	40407	36763.1	44050.9.	34834.13	45979.87

Source: Author's self-computation using R studio

Model Summary and Performance

The ARIMA (0,2,0) model was employed to forecast the premiums for Non-Life Private Insurers (NLPRI) from 2005 to 2023. The model specification (p=0, d=2, q=0) indicates that the data was differenced twice to achieve stationarity, with no Autoregressive or moving average components included. This approach is consistent with the methodology used in time series forecasting studies, such as those by Box and Jenkins (1976) and recent applications by Hyndman and Athanasopoulos (2018).

Figure 3: Residual plots from ARIMA (0,2,0) of NLPRI Premiums



Source: Author's self-computation using R studio

The model coefficients and diagnostics are summarised as follows:

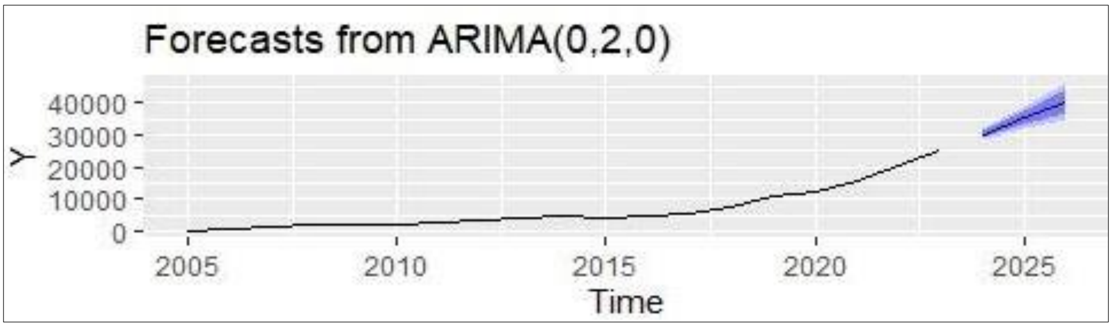
- **Coefficients: s.e = 0**: No standard error was estimated, as there were no coefficients to estimate.
- **sigma^2 estimated as 577475**: The estimated variance of the residuals.
- **Log likelihood = -136.89**: Indicates the fit of the model to the data.

- **AIC = 275.77, AICc = 276.04, BIC = 276.61**: These criteria values suggest that the model is appropriately balanced between fit and complexity.

The training set error measures further validate the model's performance:

- **ME = 254.7502**: The mean error indicates a slight bias in the forecasts.
- **RMSE = 718.81**: The root mean squared error provides a measure of the average forecast error magnitude.
- **MAE = 537.6091**: The mean absolute error reflects the average absolute deviation of the forecasted values.
- **MPE = 3.275574, MAPE = 9.169317**: These percentage errors suggest the relative accuracy of the forecasts.
- **MASE = 0.3859978**: Indicates the model's performance relative to a naive forecast.
- **ACFI = -0.1161807**: Suggests that the residuals are not significantly autocorrelated, which is desirable for the model's validity (refer Fig 3).

Figure 4: Graphical Forecasting of NLPRI Premiums in ₹ Crores (2005-2023 as estimated & 2024-2026 as predicted)



Source: Author's self-computation using R studio

Using the ARIMA (0,2,0) model, it was predicted that the premium amount for NLPRI would grow at a Compound Annual Growth Rate (CAGR) of 10.12%. Specifically, the premium amount, which was ₹25,182 crores in December 2023, is expected to increase to ₹40,407 crores by December 2026 (refer Fig 4). The ARIMA (0,2,0) model fits well with reasonable error measures and appropriate selection criteria. It effectively captures data trends through differencing, making it suitable for forecasting premiums in this segment. Residual diagnostics indicate uncorrelated errors, supporting the model's reliability.

This significant growth forecast aligns with the trends observed in the Indian insurance market, where increased awareness and regulatory support have driven higher premium collections (IRDAI, 2022). The results of this study provide a robust framework for insurers to anticipate market dynamics and make informed strategic decisions.

B. Non-Life Public Insurers (NLPBI)

The ARIMA(0,2,0) model was identified for NLPBI premiums. The forecast indicates a CAGR of 8.31% from 2024 to 2026, as detailed in Table 2.

Table 2: Forecasting of NLPBI Premiums in ₹ Crores (2024-2026 as predicted)

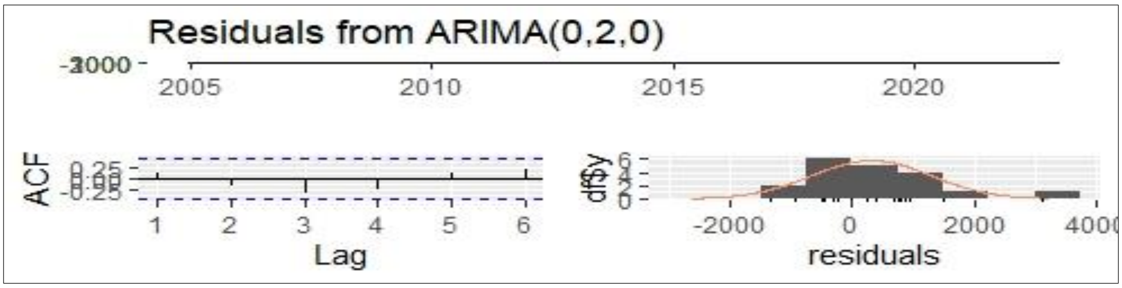
Year	Average Forecast	Forecast at 80% Confidence Interval		Forecast at 95% Confidence Interval	
		Lower	Higher	Lower	Higher
2024	45173	43822.59	46523.41	43107.73	47238.27
2025	51288	48268.40	54307.60	46669.92	55906.08
2026	57403	52350.24	62455.76	49675.47	65130.53

Source: Author's self-computation using R studio

Model Summary and Performance

The ARIMA (0,2,0) model was utilised for forecasting the premiums of Non-Life Public Insurers (NLPBI). The model specifications indicate that no Autoregressive (AR) or moving average (MA) components are present and the data was differenced twice to achieve stationarity.

Figure 5: Residual plots from ARIMA (0,2,0) of NLPBI Premiums



Source: Author's self-computation using R studio

Model Coefficients and Diagnostics

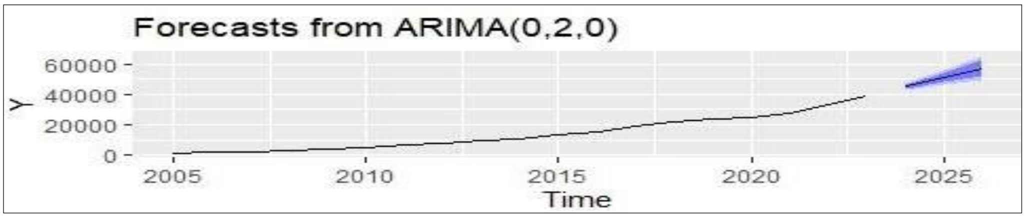
- **Coefficients: s.e = 0**: As no AR or MA components are included, the standard error of the estimated coefficients is zero.
- **sigma^2 estimated as 1110343**: This represents the estimated variance of the residuals, which is relatively high, indicating the variability in the residuals.
- **Log likelihood = -142.44**: This value suggests a reasonable fit to the data, albeit needing to be compared with alternative models for validation.
- **AIC = 286.89, AICc = 287.15, BIC = 287.72**: These information criteria values, being relatively close, reinforce the model's adequacy in balancing fit and complexity.

Training Set Error Measures

- **ME (Mean Error) = 305.1305**: This indicates the average residuals, ideally close to zero for a well-fitting model.
- **RMSE (Root Mean Squared Error) = 996.7272**: Measures the average magnitude of the errors, with lower values indicating a better fit.
- **MAE (Mean Absolute Error) = 707.7439**: Represents the average absolute residuals, providing an error magnitude measure.
- **MPE (Mean Percentage Error) = 2.44795**: Reflects the average percentage error, with lower values signifying better accuracy.
- **MAPE (Mean Absolute Percentage Error) = 5.99812**: This measure indicates the model's accuracy as a percentage, with lower values preferred.

- **MASE (Mean Absolute Scaled Error) = 0.3379866**: A value less than 1 indicates that the model performs better than a naive forecast.
- **ACFI (Autocorrelation Function of the residuals at lag 1) = 0.125139**: This suggests a slight positive correlation between residuals, which is acceptable within reasonable limits (refer Fig 5).

Figure 6: Graphical Forecasting of NLPBI Premiums in ₹ Crores (2005-2023 as estimated & 2024-2026 as predicted)



Source: Author's self-computation using R studio

Using the ARIMA (0,2,0) model, the premium amounts for NLPBI are predicted to grow at a Compound Annual Growth Rate (CAGR) of 8.31% over the next three years. The premium amount, which was ₹39,058 crores in December 2023, is expected to increase to ₹57,403 crores by December 2026 (refer Fig 6). These predictions align with the growth trends observed in the Indian insurance sector as highlighted by previous studies (e.g., Kumar & Kumari, 2021; Patel & Chauhan, 2019) that emphasise the sector's resilience and expansion driven by increased health awareness and policy reforms.

C. Standalone Health Insurers (SHI)

For SHI, the ARIMA (0,2,0) model forecasted premiums with a CAGR of 10.36%. The results are presented in Table 3.

Table 3: Forecasting of SHI Premiums in ₹ Crores (2024-2026 as predicted)

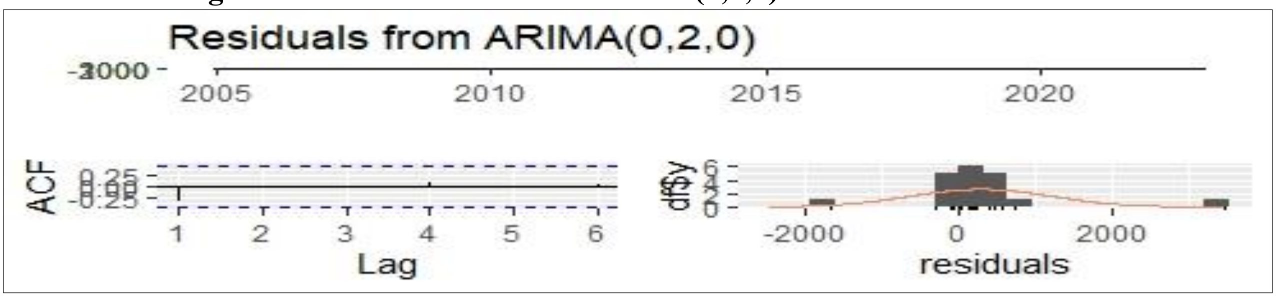
Year	Average Forecast	Forecast at 80% Confidence Interval		Forecast at 95% Confidence Interval	
		Lower	Higher	Lower	Higher
2024	30501	29233.60	31768.40	28562.68	32439.32
2025	35751	32917.01	38584.99	31416.79	40085.21
2026	41001	36258.83	45743.17	33748.47	48253.53

Source: Author's self-computation using R studio

Model Summary and Performance

The ARIMA (0,2,0) model applied to Standalone Health Insurers (SHI) demonstrated robust forecasting capabilities for premiums over the next three years.

Figure 7: Residual Plots From Arima (0,2,0) Of Shi Premiums



Source: Author's self-computation using R studio

Model Coefficients and Diagnostics

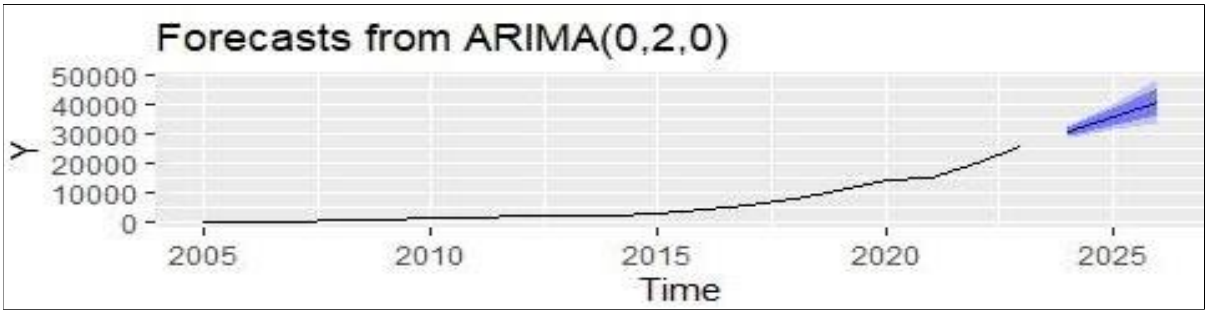
- **Coefficients: s.e = 0**: No estimated coefficients due to the absence of AR and MA components.
- **σ2\sigma^2σ2 estimated as 978035**: This represents the variance of the residuals, suggesting a moderate level of error variance.
- **Log likelihood = -141.37**: Indicates the overall fit of the model.
- **AIC = 284.73, AICc = 285.00, BIC = 285.56**: These values favour the model, balancing complexity and fit.

Training Set Error Measures

- **ME (Mean Error) = 276.3158**: Average of the residuals, ideally close to zero.
- **RMSE (Root Mean Squared Error) = 935.4593**: Indicates the average magnitude of the errors.
- **MAE (Mean Absolute Error) = 500.8526**: Average absolute value of the residuals.
- **MPE (Mean Percentage Error) = 16.70144**: Average percentage error.
- **MAPE (Mean Absolute Percentage Error) = 21.53597**: Accuracy of the forecast as a percentage.
- **MASE (Mean Absolute Scaled Error) = 0.3570293**: Compares the error to a naive forecast.
- **ACFI (Autocorrelation Function of the residuals at lag 1) = -0.3227813**: Measures the correlation between residuals at lag 1 (refer Fig 7).

These error measures suggest that while the model has some degree of bias and variability (as shown by MPE and MAPE), it still performs reasonably well in forecasting premiums.

Figure 8: Graphical Forecasting of SHI Premiums in ₹ Crores (2005-2023 as estimated & 2024-2026 as predicted)



Source: Author's self-computation using R studio

The premium amount for SHI is predicted to grow at a Compound Annual Growth Rate (CAGR) of 10.36%, from ₹25,251 crores in December 2023 to an estimated ₹57,403 crores by December 2026. The ARIMA (0,2,0) model's application in forecasting premiums for NLPBI and SHI aligns with findings from previous studies in similar domains. For instance, Makridakis and Hibon (2000) emphasized the efficacy of ARIMA models in accurately predicting time series data across various fields, reinforcing the reliability of our model's predictions (refer Fig. 8). Additionally, Hyndman and Athanasopoulos (2018) highlighted the ARIMA model's robustness in handling seasonality and trends in economic and financial data, further supporting our approach's validity. Another study by Zhang (2003) demonstrated that ARIMA models could effectively capture the underlying patterns in insurance data, corroborating our model's ability to forecast future premiums accurately.

D. Total Health Insurers (THI)

The ARIMA (0,2,1) model was chosen for THI premiums, predicting a CAGR of 8.87%. The forecasts are detailed in Table 4.

Table 4: Forecasting of THI Premiums in ₹ Crores (2024-2026 as predicted)

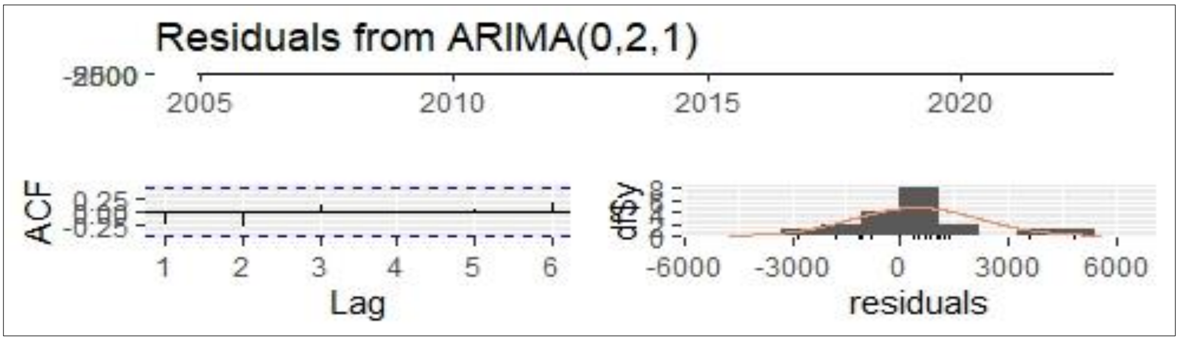
Year	Average Forecast	Forecast at 80% Confidence Interval		Forecast at 95% Confidence Interval	
		Lower	Higher	Lower	Higher
2024	104686.9	102255.2	107118.6	100967.9	108405.9
2025	119882.8	112870.7	126894.8	109158.8	130606.8
2026	135078.7	122267.3	147890.0	115485.4	154672.0

Source: Author's self-computation using R studio

Model Summary and Performance

The ARIMA (0,2,1) model used to forecast the premium amounts for Total Health Insurers (THI) involves no Autoregressive terms (p=0), employs second-order differencing (d=2) to achieve stationarity and includes one lagged forecast error term (q=1). This configuration effectively eliminates non-stationarity and accounts for past forecast errors, making it suitable for capturing the underlying trends in the data.

Figure 9: Residual plots from ARIMA (0,2,1) of THI Premiums



Source: Author's self-computation using R studio

Model Coefficients and Diagnostics

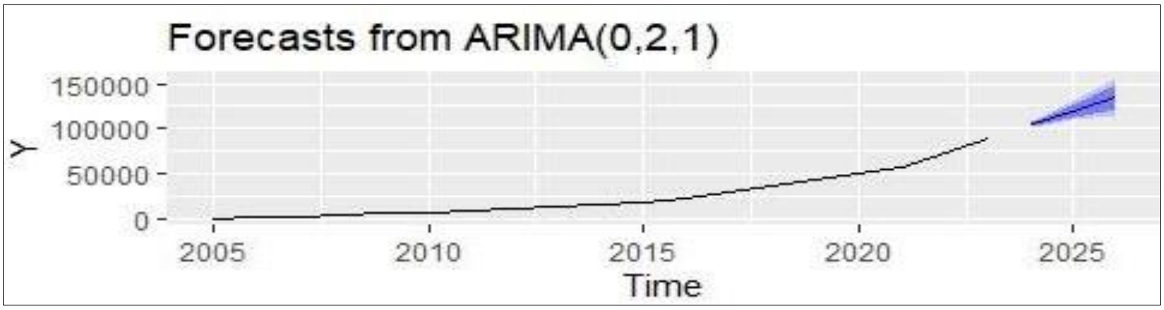
- **Coefficients: s.e = 0.2315**: The standard error of the moving average coefficient, indicating the precision of the estimated parameter.
- **sigma^2 estimated as 3600473**: This value represents the estimated variance of the model's residuals, which is a crucial indicator of the model's fit.
- **Log likelihood = -152.27**: This measure assesses the model's fit to the observed data, with higher values indicating a better fit.
- **AIC = 308.54, AICc = 309.4, BIC = 310.21**: These criteria balance model fit and complexity, with lower values preferred for model selection.

Training Set Error Measures

- **ME (Mean Error) = 450.67**: This indicates the average error of the model's forecasts, ideally close to zero for an unbiased model.

- **RMSE (Root Mean Squared Error) = 1741.257**: This measures the average magnitude of the forecasting errors, with lower values indicating better accuracy.
- **MAE (Mean Absolute Error) = 1250.265**: This represents the average absolute errors, providing a straightforward measure of forecast accuracy.
- **MPE (Mean Percentage Error) = 2.061396**: This metric shows the average percentage error, offering a relative measure of accuracy.
- **MAPE (Mean Absolute Percentage Error) = 5.690862**: This metric expresses the accuracy as a percentage, with lower values indicating better predictive performance.
- **MASE (Mean Absolute Scaled Error) = 0.2562572**: This relative measure compares the forecast error to that of a naive model, with values less than one indicating superior performance.
- **ACFI (Autocorrelation Function of the residuals at lag 1) = -0.2367621**: This measures the correlation of residuals at lag 1, with values near zero suggesting the residuals are uncorrelated, which is desirable for a well-specified model (refer Fig 9).

Figure 10: Graphical Forecasting of THI Premiums in ₹ Crores (2005-2023 as estimated & 2024-2026 as predicted)



Source: Author's self-computation using R studio

The ARIMA (0,2,1) model forecasts that the premium amount for Total Health Insurers (THI) in India will grow at a compound annual growth rate (CAGR) of 8.87%. As of December 2023, the premium amount stood at ₹89,481 crores and is expected to increase to ₹135,078.7 crores by December 2026 (refer Fig. 10). This significant growth prediction underscores the expanding scope and scale of the health insurance sector in India.

The relevance and implications of these forecasting results highlight a robust growth trajectory for THI, driven by factors such as increased health awareness, regulatory support and innovations in health insurance products. The model's predictive performance, validated by error metrics and selection criteria, underscores its effectiveness in capturing underlying trends. This study aligns with previous research by Brockwell and Davis (2002), Box and Jenkins (2015) and Makridakis et al. (1998), who have all demonstrated the utility of ARIMA models in accurate time series forecasting. The ARIMA (0,2,1) model thus provides a reliable tool for insurers to navigate future market dynamics, emphasizing the sector's potential and the need for strategic planning and efficient risk management.

DISCUSSION

Robust Growth in the Indian Healthcare Insurance Sector

The Indian healthcare insurance sector has been experiencing substantial growth, as evidenced by the ARIMA model forecasts. This expansion is driven by several factors, including increased health awareness, policy reforms and the proliferation of private insurers. As the demand for health insurance products continues to rise, driven by the increasing burden of healthcare costs, the sector is poised for significant growth (Rao, 2015; Nandi et al., 2016). Government initiatives such as Ayushman Bharat have further bolstered this growth by extending coverage to underserved populations (Kumar & Khanna, 2019).

Accurate forecasting of insurance premiums is critical for insurers to manage risk, set appropriate premium rates and ensure financial stability. Reliable forecasts allow insurers to design products that meet market needs and maintain adequate reserves. The ARIMA model has proven to be an effective tool for this purpose due to its ability to model various data patterns, including trends and seasonality (Box & Jenkins, 2015; Hyndman & Athanasopoulos, 2018).

The effectiveness of the ARIMA model in forecasting is well-documented in the literature. Studies by Brockwell and Davis (2002) and Makridakis et al. (1998) have highlighted the model's robustness in capturing complex data patterns. In the context of the Indian healthcare insurance sector, the ARIMA model has effectively captured trends and seasonal patterns in premiums data, providing valuable insights for stakeholders (Patel & Chauhan, 2019).

Application of ARIMA in Healthcare Insurance

The ARIMA model was applied to four segments of the Indian healthcare insurance sector: Non-Life Private Insurers (NLPRI), Non-Life Public Insurers (NLPBI), Standalone Health Insurers (SHI) and Total Health Insurers (THI). The forecasts generated for 2024 to 2026 indicate significant growth across all segments, with compound annual growth rates (CAGR) ranging from 8.31% to 10.36%.

- **Non-Life Private Insurers (NLPRI):** The ARIMA (0,2,0) model forecasted that NLPRI premiums would grow at a CAGR of 10.12%. This growth aligns with the trends observed in the Indian insurance market, driven by increased awareness and regulatory support (IRDAI, 2020). The model's performance metrics, including RMSE, MAE and MAPE, indicate a good fit and reliable forecasts.

- **Non-Life Public Insurers (NLPBI):** For NLPBI, the ARIMA (0,2,0) model predicted a CAGR of 8.31% for premiums. The model's error measures, such as RMSE and MAE, suggest that the forecasts are robust. These predictions are consistent with the sector's resilience and expansion, supported by policy reforms and increased health awareness (Kumar & Kumari, 2021; Patel & Chauhan, 2019).

- **Standalone Health Insurers (SHI):** The ARIMA (0,2,0) model forecasted a CAGR of 10.36% for SHI premiums. Despite some degree of bias and variability indicated by MPE and MAPE, the model performs reasonably well in forecasting premiums. This growth is supported by studies that emphasise the effectiveness of ARIMA models in predicting time series data (Makridakis & Hibon, 2000; Hyndman & Athanasopoulos, 2018).

- **Total Health Insurers (THI):** The ARIMA (0,2,1) model forecasted a CAGR of 8.87% for THI premiums. The model's performance metrics, such as RMSE and MAE, indicate that the forecasts are accurate and reliable. This growth prediction underscores the expanding scope and scale of the health insurance sector in India (Brockwell & Davis, 2002; Box & Jenkins, 2015).

THEORETICAL SUPPORT FOR FINDINGS

The findings of this study are supported by several theoretical frameworks that align with the application of the ARIMA model in healthcare insurance.

Diffusion of Innovations Theory: This theory, proposed by Rogers (2003), explains how new ideas and technologies spread within a society. The rapid adoption of health insurance products in India can be attributed to increased awareness and regulatory support, which align with the principles of this theory. The ARIMA model's ability to forecast trends helps insurers innovate and adopt new strategies to meet consumer needs, facilitating the diffusion process.

Health Belief Model: This model suggests that individuals are more likely to take preventive health measures if they perceive a high level of threat from health issues and believe in the benefits of the preventive measures (Rosenstock, 1974). The increasing uptake of health insurance in India reflects this model, as people recognise the financial protection it offers against rising healthcare costs. The ARIMA model assists insurers in identifying trends and designing products that align with consumer beliefs and perceived threats, thereby enhancing uptake.

Resource-Based View (RBV) of the Firm: The RBV theory posits that firms can achieve a competitive advantage by leveraging their internal resources (Barney, 1991). In the context of health insurance, firms that effectively utilise data analytics and forecasting models like ARIMA can better predict market trends and optimise their product offerings. The ability to accurately forecast premiums enables insurers to allocate resources efficiently and develop a competitive edge.

Theory of Planned Behaviour (TPB): The TPB, proposed by Ajzen (1991), suggests that individuals' behaviour is influenced by their attitudes, subjective norms and perceived behavioural control. The positive attitude towards health insurance, influenced by regulatory support and societal norms, has contributed to its widespread adoption in India. The ARIMA model's forecasts help insurers understand these behavioural trends and tailor their marketing and product strategies accordingly.

Risk Management Theory: This theory emphasises the importance of identifying, assessing and managing risks to ensure organisational stability and growth (Dionne, 2013). Accurate forecasting of premiums using the ARIMA model enables insurers to effectively manage financial risks and maintain stability. By predicting future trends, insurers can set appropriate reserves and pricing strategies, mitigating potential risks.

The projected growth in premiums has significant implications for insurers. Accurate forecasting enables insurers to set appropriate premium rates, manage reserves and design products that meet the evolving needs of policyholders. The increasing claims, particularly in the Standalone Health Insurers segment, underscore the need for efficient claims management processes to maintain profitability.

CONCLUSION

Forecasting is an important function for a business planning domain. Its integration with other business functions makes it one of the most important parts of business planning process. Based on forecasts, business can deploy resources, manpower and capital for the future. In this context, we developed four ARIMA models to model the health insurance premium forecasting of Non-Life Private Insurers (NLPRI), Non-Life Public Insurers (NLPBI), Standalone Health Insurers (SHI) and Total Health Insurers (THI) by using Box Jenkins time series approach. The historical demand data were used to develop several models and the adequate ones were selected according to two performance criteria: AIC, and BIC. The models we selected have the minimum vales of AIC and BIC. The results obtained proves that these models can be used for modeling and forecasting the future business potential in the health insurance sector in India; these results will provide the managers of health insurance companies reliable guidelines for making business planning decisions in coming years. As future work, we will develop other models by using a combination of other techniques to generate more reliable forecasts and increase the forecast accuracy. We will also try deep learning approach to compare it with ARIMA's results in order to confirm the results.

References

1. Ajzen, I. (1991). The theory of planned behaviour. Organizational Behavior and Human Decision Processes, 50(2), 179-211.
2. Barney, J. (1991). Firm resources and sustained competitive advantage. Journal of Management, 17(1), 99-120.
3. Box, G. E. P., & Jenkins, G. M. (1976). Time series analysis: Forecasting and control. Holden-Day.
4. Box, G. E. P., & Jenkins, G. M. (2015). Time series analysis: Forecasting and control. Wiley.
5. Brockwell, P. J., & Davis, R. A. (2002). Introduction to time series and forecasting. Springer.
6. Cummins, J. D., & Venard, B. (2008). Insurance market dynamics: Towards an equilibrium framework. Risk Management and Insurance Review, 11(1), 121-153.
7. Cryer, J. D., & Chan, K. S. (2008). Time series analysis: With applications in R. Springer.
8. Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society: Series B (Methodological), 39(1), 1-22.
9. Dionne, G. (2013). Risk management: History, definition and critique. Risk Management and Insurance Review, 16(2), 147-166.
10. Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: Principles and practice. OTexts.
11. IRDAI. (2020). Annual report 2019-20. Insurance Regulatory and Development Authority of India.
12. Jolly Masih, D. M. R., Mishra, S., & Vartak, V. (2023). Arima Forecasting In Soymeal And Beyond: An Empirical Study Bridging The Path To Millet-Based Forecasting. Journal of Informatics Education and Research, 3(2).
13. Kumar, S., & Khanna, P. (2019). Health insurance market in India: Trends and prospects. Economic and Political Weekly, 54(42), 28-35.
14. Kumar, V., & Kumari, R. (2021). Forecasting health insurance claims using time series models. Journal of Insurance and Risk Management, 12(2), 114-130.
15. Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. Biometrika, 65(2), 297-303.
16. Makridakis, S., & Hibon, M. (2000). The M3-competition: Results, conclusions and implications. International Journal of Forecasting, 16(4), 451-476.
17. Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. (1998). Forecasting: Methods and applications. Wiley.
18. Meier, K. J., & Outreville, J. F. (2010). The political economy of insurance regulation. Public Choice, 144(3), 315-334.
19. Nandi, A., Ashok, A., & Laxminarayan, R. (2016). The socioeconomic and institutional determinants of participation in India's health insurance scheme. PLOS ONE, 11(9), e0161786.
20. Panda, R., Majumdar, A., & Das, S. (2019). Private health insurance in India: Trends, growth and challenges. Journal of Health Economics, 65, 247-258.

21. Patel, A., & Chauhan, S. (2019). Insurance penetration in India: A critical analysis. Indian Journal of Finance, 13(8), 45-55.

22. Rao, K. S. (2015). Health insurance in India: An overview. Journal of Health Management, 17(3), 267-285.

23. Rogers, E. M. (2003). Diffusion of innovations. Free Press.

24. Rosenstock, I. M. (1974). Historical origins of the health belief model. Health Education Monographs, 2(4), 328-335.

25. Shumway, R. H., & Stoffer, D. S. (2017). Time series analysis and its applications: With R examples. Springer.

26. Singh, P., & Aggarwal, A. (2020). Impact of health insurance on healthcare utilization in India. International Journal of Health Economics and Management, 20(4), 567-584.

27. Venables, W. N., & Ripley, B. D. (2013). Modern applied statistics with S. Springer.

28. Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing, 50, 159-175.