

Forecasting the Number of Public Health Medical Beds in China: Based on ARIMA Model

Xinyi Liu^{1,a,*}

¹*Department of Gardening, Hainan University, Haikou, China*
a. 20213007006@hainanu.edu.cn

**corresponding author*

Abstract: This study employs the ARIMA model to forecast hospital bed availability in China from 1994 to 2023, with a focus on predicting bed numbers for the subsequent five years (2024–2028). The analysis encompasses two principal phases: the first phase involves forecasting future bed numbers based on historical data, and the second phase evaluates the accuracy of these forecasts by comparing them with actual data from 2021 to 2023. Using the ARIMA(0,2,0) model, the study demonstrates a robust and reliable forecast, revealing a steady increase in hospital bed counts. The accuracy of the model's predictions is high, confirming its effectiveness for short-term forecasting and underscoring the importance of proactive healthcare infrastructure planning. Additionally, the study highlights potential areas for future research, including the integration of additional variables, exploration of alternative modeling approaches, extension of forecast horizons, and sensitivity analyses to enhance predictive precision. These contributions aim to assist healthcare planners and policymakers in addressing the evolving needs of China's healthcare system and ensuring adequate medical service provision.

Keywords: Number of Beds Forecast, ARIMA, Time Series, Chinese Medicine.

1. Introduction

China's healthcare system faces significant challenges, primarily due to a persistent shortage of hospital beds. This shortage has been exacerbated by an aging population and increasing healthcare demands, impacting the quality of patient care and the efficient allocation of medical resources [1]. Due to the uneven geographical distribution of hospital beds and healthcare professionals, the inefficiency in utilizing medical resources is further exacerbated [2]. Accurate forecasting of future hospital bed requirements is therefore crucial for effective healthcare planning and policy-making.

Time series analysis plays a vital role in statistics, involving the examination of data collected sequentially at uniform intervals to study its characteristics and predict future values based on these characteristics. Forecasting is essential across various domains, including finance and industry [3]. The Autoregressive and Moving Average (ARMA) model, pioneered by Yule, Slutsky, Walker, and Yaglom, is a fundamental method for analyzing time series data. The Autoregressive Integrated Moving Average (ARIMA) model extends the ARMA framework by transforming non-stationary data into a stationary format before modeling. ARIMA models are widely used for forecasting linear time series data and are often referred to as Box-Jenkins models after their popularizers, Box and

Jenkins. The ARIMA approach, including the general transfer function model discussed by Box and Tiao, offers substantial flexibility for model identification, parameter estimation, and forecasting [4].

Recent advancements in ARIMA applications demonstrate its continued relevance and versatility. For instance, ARIMA models have been utilized to predict financial market trends, where accurate forecasts can significantly impact investment decisions. The model's efficacy in stock price forecasting underscores its importance for financial analysts and investors [5]. Furthermore, ARIMA's adaptability to incorporate additional time series as input variables, known as ARIMAX or dynamic regression models [6], enhances its predictive capabilities.

In the context of healthcare, ARIMA models have been employed to forecast various aspects of medical resources, including hospital bed availability. This study focuses on forecasting the number of hospital beds in China for the next five years (2024–2028) using historical data from 1994 to 2023. Additionally, the accuracy of the ARIMA model's predictions for the years 2021–2023 is evaluated by comparing them with actual data. The results indicate a continued increase in the number of hospital beds, emphasizing the need for proactive healthcare planning and resource allocation. This finding highlights the value of ARIMA models for accurate short-term forecasting and suggests that similar approaches could be beneficial for other healthcare forecasting applications.

2. Data

This study utilizes publicly available data from the National Bureau of Statistics of China, specifically the number of hospital beds (measured in ten thousand units) from 1994 to 2023. The data is sourced from the "Health Institutions" section of the annual statistical reports. After importing the data into the R environment, this paper carried out several preprocessing steps to ensure the data was suitable for time series analysis.

2.1. Data Cleaning

The year values in the dataset contained the character "year," which was removed to convert the 'Year' column into numeric format. Similarly, the 'Beds' column, representing the number of hospital beds, was converted into numeric format.

2.2. Data Formatting

After cleaning, the data was transformed into a time series object starting from 1994, with a frequency of 1 to reflect the annual nature of the data.

2.3. Missing Data Handling

This paper examined the dataset for any missing values. Since no missing data was found, no further imputation or adjustments were necessary. The cleaned and formatted data was used for the subsequent time series analysis and model development. Figure 1 shows a time series diagram of the pre-processed data.

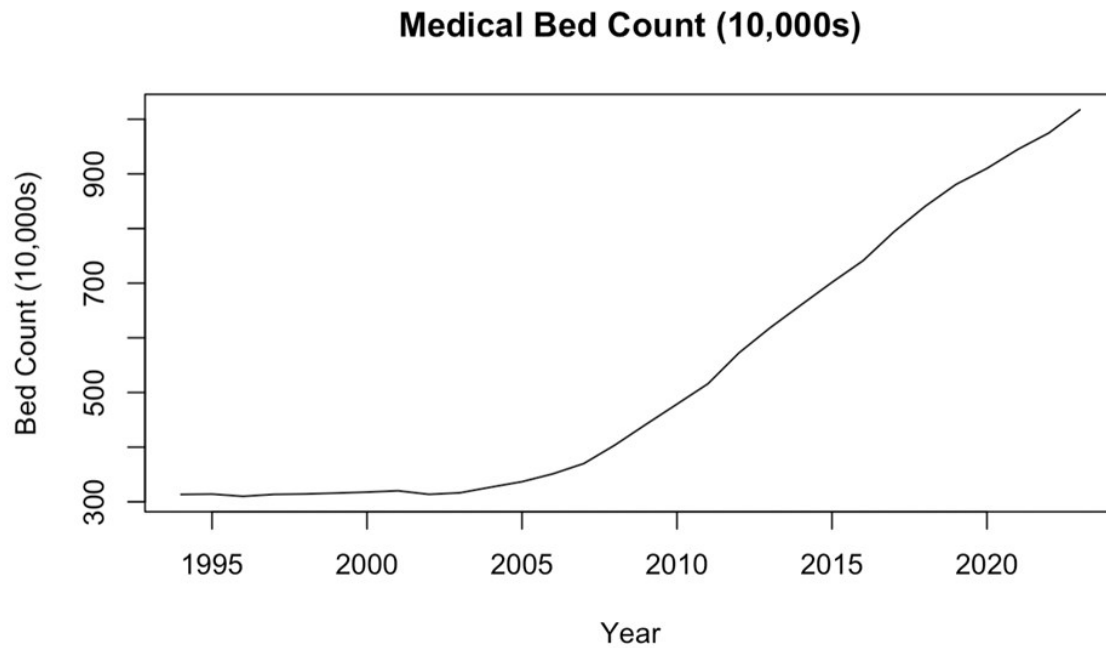


Figure 1: Time series of the number of medical beds in China from 1994 to 2023. (Picture Credit: Original)

3. Methodology

This section outlines the methodological framework employed to forecast hospital bed availability based on time series data. The ARIMA (AutoRegressive Integrated Moving Average) model was selected due to its robustness in handling non-seasonal time series data that exhibit stochastic trends. The methodology encompasses several key stages, including model selection, stationarity testing, parameter estimation, and validation of forecast accuracy. By following a structured approach involving preliminary diagnostics, differencing to achieve stationarity, and model fitting using the `auto.arima` function, the study identifies $ARIMA(0,2,0)$ with drift as the optimal model for forecasting. Subsequently, residual diagnostics and error metrics are utilized to validate the adequacy of the model, ensuring accurate and reliable forecasts. Additionally, a split-sample validation approach was applied, enabling evaluation of the model's predictive performance through comparison of predicted values against actual data for the test period (2021–2023).

3.1. Model Selection and Validation

This study employed an ARIMA model to forecast hospital bed availability based on time series data. ARIMA models are widely used for non-seasonal time series data that exhibit trends and stochastic patterns. The ARIMA model is characterized by three parameters: autoregressive (p), differencing (d), and moving average (q), denoted as $ARIMA(p,d,q)$.

3.1.1. Preliminary Diagnostics

Initially, this paper examined the autocorrelation (ACF) and partial autocorrelation (PACF) plots to identify potential AR (autoregressive) and MA (moving average) terms in the data. According to Figure 2 and Figure 3, the original data series is significantly correlated and does not meet the requirements for the ARIMA model.

Autocorrelation Function (ACF)

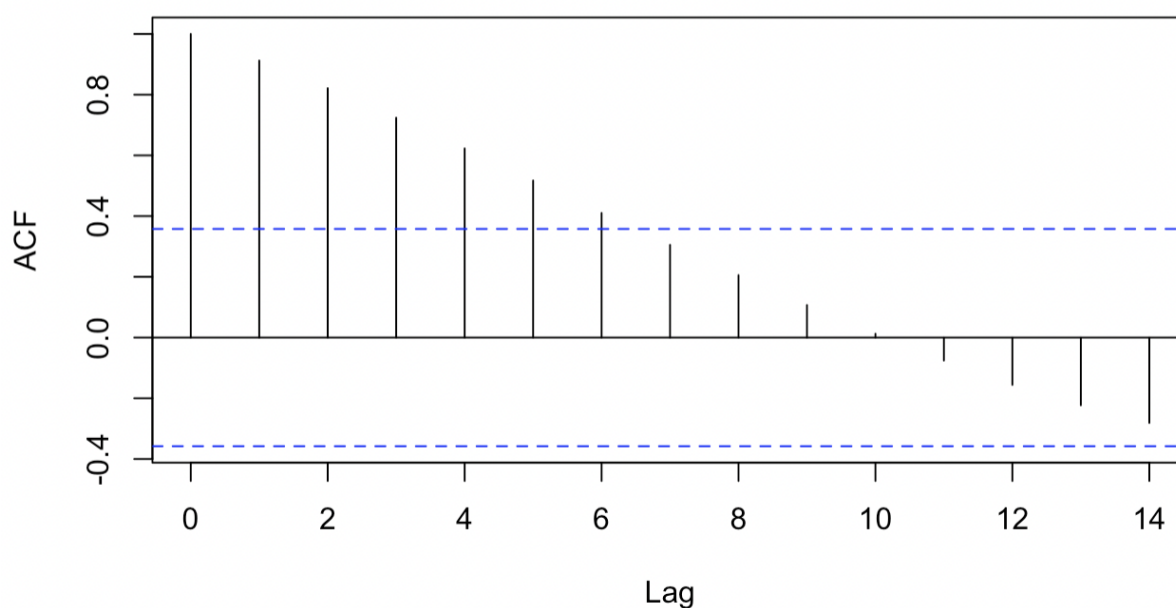


Figure 2: ACF plot of raw data. (Picture Credit: Original)

Partial Autocorrelation Function (PACF)

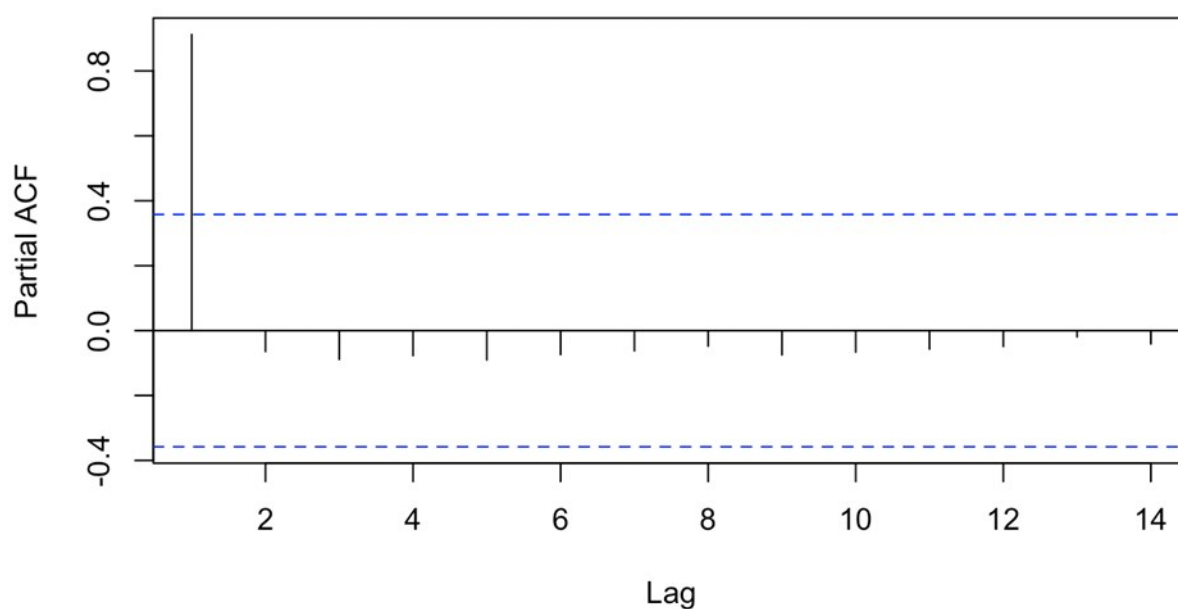


Figure 3: PACF plot of raw data. (Picture Credit: Original)

3.1.2. Stationarity Testing

To assess the stationarity of the data, this paper employed the Augmented Dickey-Fuller (ADF) test [7]. The test yielded a Dickey-Fuller statistic of -2.6493 with a p-value of 0.3232, which is

significantly higher than the typical 0.05 threshold for rejecting the null hypothesis. This indicates that the data is non-stationary, as we fail to reject the null hypothesis that a unit root is present in the series.

3.1.3. Differencing

Using the `'ndiffs'` function in R, this paper determined that the dataset required second-order differencing ($d=2$) to achieve stationarity. According to Figure 4, it can be seen that the data's stationarity has improved, and the differenced data fluctuates around a constant near zero.

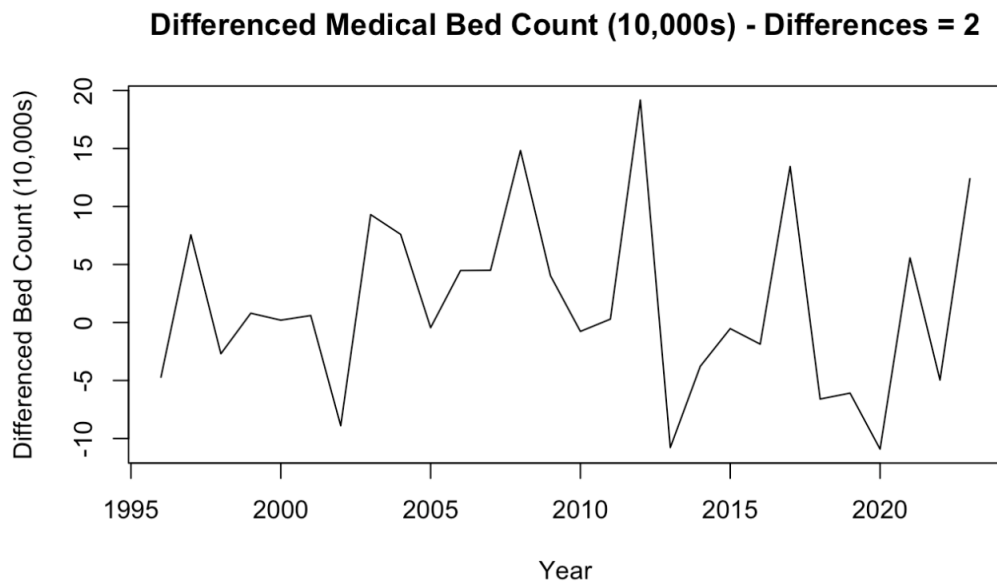


Figure 4: Differenced Medical Bed Count (10,000s) - Differences = 2. (Picture Credit: Original)

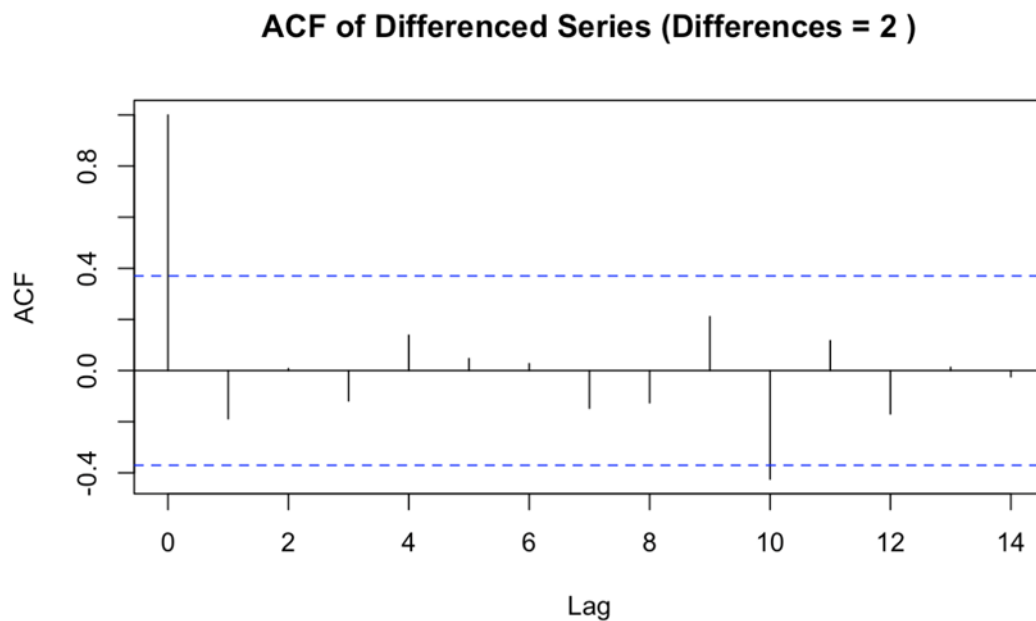


Figure 5: ACF of Differenced Series (Differences = 2). (Picture Credit: Original)

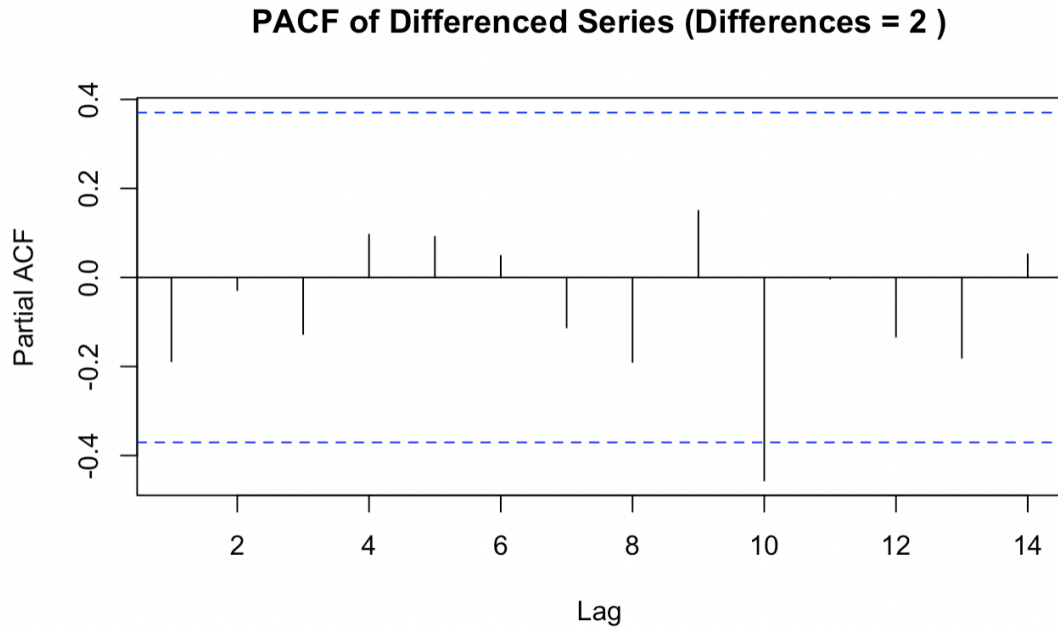


Figure 6: PACF of Differenced Series (Differences = 2). (Picture Credit: Original)

According to Figure 5 and Figure 6, it can be seen that the data after difference does not have significant autocorrelation.

Given the non-stationarity, it was necessary to apply differencing to the data to achieve stationarity. After applying the second-order differencing, the Box-Ljung test was performed to assess the residuals' white noise characteristics. The test produced a p-value of 0.1598 (with a chi-square statistic of 14.299 and 10 degrees of freedom), suggesting that the residuals are not significantly different from white noise, thus supporting the adequacy of the differencing process.

3.1.4. Model Selection

This paper used the 'auto.arima' function to automatically select the best-fitting ARIMA model, which identified ARIMA(0,2,0) with drift as the optimal model based on the lowest AICc score.

Table 1: Error Metrics for ARIMA(0,2,0) Model.

ME	1.381
RMSE	7.486
MAE	5.614
MPE	0.358%
MAPE	1.123%
ACF1	-0.180

According to Table 1, these error metrics, along with the log-likelihood of -97.06, and information criteria values (AIC = 196.12, AICc = 196.27, BIC = 197.45), further corroborate that the ARIMA(0,2,0) model is a good fit for the differenced series.

3.2. Parameter Estimation and Model Fitting

After selecting the ARIMA(0,2,0) model, this paper proceeded to fit the model to the time series data.

Table 2: Model Parameters for ARIMA(0,2,0).

Sigma^2	60.04
Log Likelihood	-97.06
AIC	196.12
AICc	196.27
BIC	197.45

According to Table 2, these results confirmed that the model was a suitable fit for the data. Using this model, this paper generated forecasts for the years 2024–2028.

3.3. Model Diagnostics and Adjustment

Following the initial model fitting, this paper conducted a series of diagnostic tests to ensure the model's accuracy. The residuals from the fitted model were analyzed through residual plots and ACF/PACF charts to verify they resembled white noise. Additionally, the Ljung-Box test confirmed that the residuals exhibited no significant autocorrelation, indicating the model was adequate. Initially, this paper considered an ARIMA(0,0,0) model, but this led to unrealistic forecasts, as the differencing produced constant zero predictions. To resolve this, drift was allowed in the ARIMA(0,2,0) model, which captured the upward trend in the data and provided more reliable forecasts.

3.4. Forecasting with Training Data and Model Evaluation

To validate the forecasting performance, this paper split the dataset into a training set (1994–2020) and a test set (2021–2023). The ARIMA(0,2,0) model was fitted to the training data, and according to Table 3 and Table 4, the following results were obtained.

Table 3: Training set model parameters for ARIMA(0,2,0).

Sigma^2	58.87
Log Likelihood	-86.41
AIC	174.83
AICc	175
BIC	176.05

Table 4: Training Set Error Measures for ARIMA(0,2,0).

ME	1.053
RMSE	7.383
MAE	5.388
MPE	0.349
MAPE	1.162
MASE	0.227
ACF1	-0.099

3.5. Forecast Accuracy Evaluation

The next step was to assess the accuracy of the model's forecasts by comparing the predicted values for 2021–2023 with the actual data. The evaluation process involved the following steps:

3.5.1. Forecast Comparison

This paper compared the model's predictions for the years 2021–2023 with the observed hospital bed counts, and the errors between predicted and actual values were calculated.

3.5.2. Error Calculation

This paper computed the forecast error for each year using the formula $1 - (\text{err} / \text{ActualVal})$, where "err" represents the absolute difference between the forecast and actual values.

3.5.3. Accuracy Percentage

Forecast accuracy was expressed as a percentage by multiplying the accuracy value by 100, with the average accuracy exceeding 90%.

3.5.4. Standard Deviation

The standard deviation of the accuracy percentages was computed to evaluate the variability of the forecasts, confirming that the predictions were stable and consistent.

4. Results and Discussion

This study aimed to forecast the number of hospital beds in China using an ARIMA model, covering the period from 1994 to 2023. The research consists of two primary steps: first, this paper forecasted the number of beds for the next five years (2024–2028) based on the 1994–2023 data, and second, this paper evaluated the accuracy of the model's forecasts for 2021–2023 by comparing them with actual data based on a model trained on the 1994–2020 data.

4.1. Forecast Results

Based on the ARIMA(0,2,0) model applied to the data from 1994 to 2023, the forecasted number of hospital beds for the next five years (2024–2028) is as follows, according to Table 5, along with the corresponding 80% and 95% confidence intervals.

Table 5: Forecast of number of hospital beds in the next five years (2024-2028).

Year	Forecasted Value	80% CI Lower Bound	80% CI Upper Bound	95% CI Lower Bound	95% CI Upper Bound
2024	1059.75	1049.82	1069.68	1044.563	1074.937
2025	1102.13	1079.925	1124.335	1068.171	1136.089
2026	1144.51	1107.355	1181.665	1087.686	1201.334
2027	1186.89	1132.5	1241.28	1103.708	1270.072
2028	1229.27	1155.626	1302.914	1116.641	1341.899

These forecast results suggest a steady increase in the number of hospital beds over the next five years, with a growing trend continuing from the historical data. The confidence intervals provide a range of potential outcomes, reflecting the uncertainty associated with the forecast. Figure 7 shows the bed situation predicted by the model for the next five years.

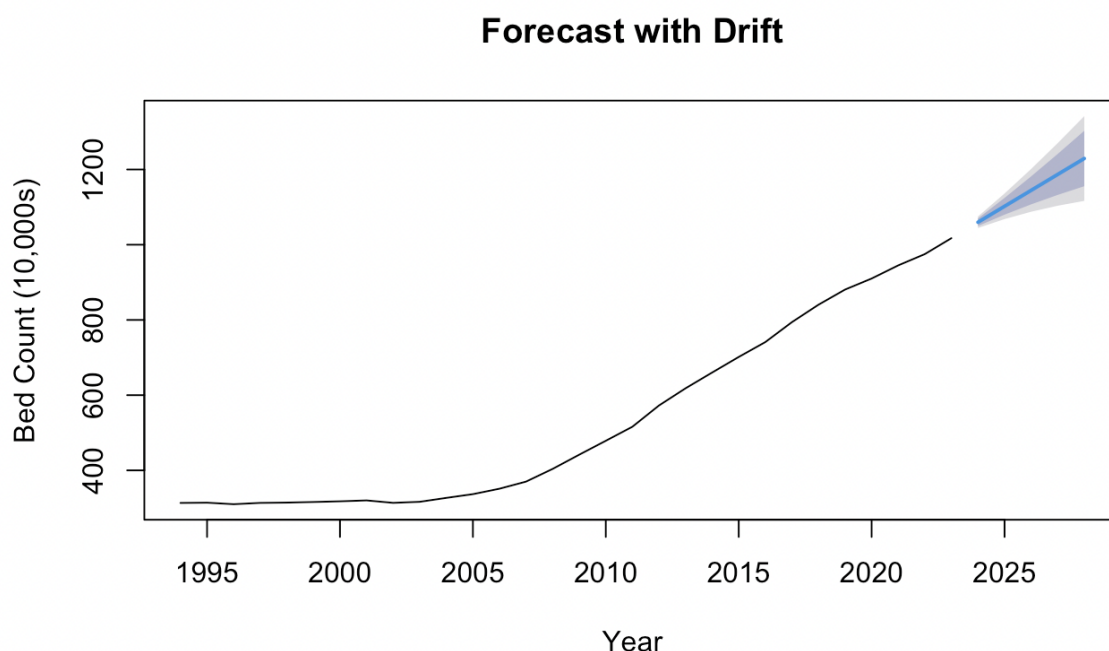


Figure 7: Forecast results for the next five years (2024-2028). (Picture Credit: Original)

4.2. Forecast Accuracy Assessment

To evaluate the accuracy of the model's predictions for 2021–2023, this paper compared the forecasted values with the actual data for those years. The following table summarizes the results.

Table 6: Forecast of number of hospital beds in the last three years (2021-2023).

Year	Point Forecast	80% CI Lower Bound	80% CI Upper Bound	95% CI Lower Bound	95% CI Upper Bound
2021	939.44	929.6071	949.2729	924.4019	954.4781
2022	968.81	946.823	990.797	935.1838	1002.4362
2023	998.18	961.3888	1034.9712	941.9127	1054.4473

Table 7: Comparison of the predicted and actual number of beds for 2021-2023.

Year	Actual	Forecast	Error	Accuracy	Accuracy_Percentage
2021	945.01	939.44	5.57	0.9941059	99.41059
2022	974.99	968.81	6.18	0.9936615	99.36615
2023	1017.37	998.18	19.19	0.9811376	98.11376

According to Table 6 and Table 7, the model shows high accuracy in predicting the number of hospital beds for 2021–2023, with accuracy percentages of 99.41%, 99.37%, and 98.11%, respectively. These results indicate that the ARIMA(0,2,0) model performs well for short-term forecasts. Figure 8 compares the difference between the predicted and true values.

Forecast with Drift (2021-2023)

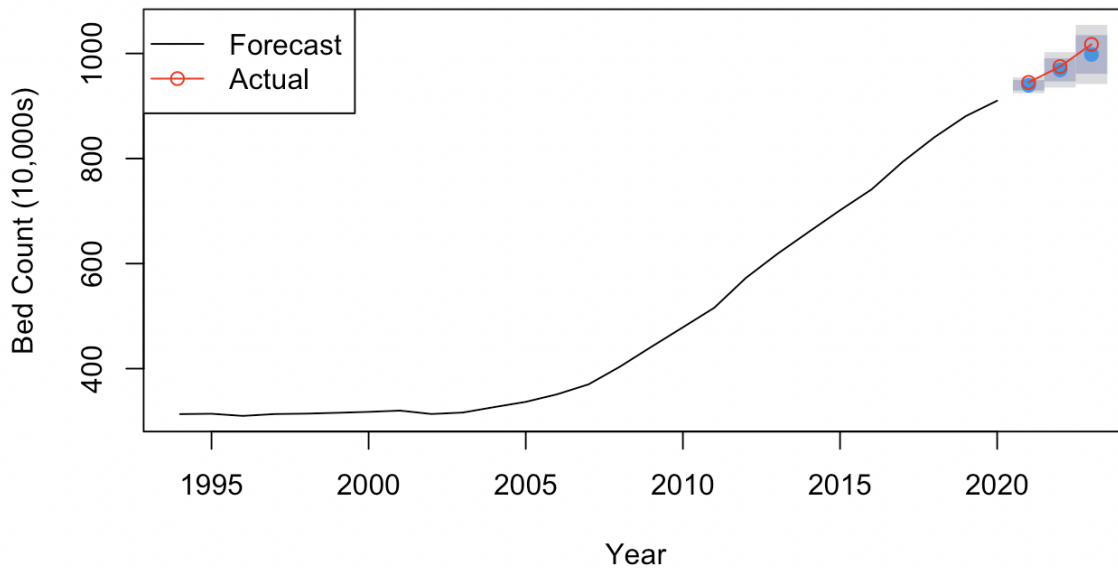


Figure 8: Line chart comparison of predicted and true values.

4.3. Standard Deviation and Accuracy Evaluation

To further assess the reliability of the model's accuracy, this paper calculated the standard deviation of the forecast accuracy percentages for the years 2021–2023. The average forecast accuracy percentage across these three years is 98.96%, and the standard deviation of the accuracy percentage is 0.7362%. This low standard deviation indicates that the model's forecasts are not only accurate but also highly consistent, with minimal variability in performance over time.

4.4. Impact of Forecast Results

The forecast results indicate a steady increase in the number of hospital beds over the next five years, underscoring the importance of proactive planning and policy-making in the healthcare sector. The reliable predictions provided by the model suggest that policymakers and healthcare planners should anticipate this growth and consider expanding facilities and resources accordingly. This forward-looking approach can help mitigate potential shortages and ensure that adequate medical services are available to meet the rising demand.

The high accuracy of the model's predictions for 2021–2023 further demonstrates its effectiveness in short-term forecasting. This indicates that similar ARIMA models could be valuable tools for other healthcare forecasting applications, such as predicting patient admissions or resource utilization. The model's consistency also highlights the importance of regular updates and recalibration to maintain forecasting accuracy, especially as new data becomes available.

4.5. Recommendations for Future Research

While the ARIMA(0,2,0) model has provided robust forecasts in this study, several avenues for future research could enhance the accuracy and robustness of predictions.

4.5.1. Incorporating Additional Variables

Including related variables such as economic indicators, healthcare policy changes, and demographic shifts may offer a more comprehensive view of the factors influencing hospital bed supply. This could lead to more nuanced models and improved forecasting outcomes.

4.5.2. Exploring Alternative Models

Evaluating other time series models, such as Seasonal ARIMA (SARIMA) or machine learning approaches, may yield improvements, particularly if the data exhibits seasonality or non-linear patterns [8]. Comparing the performance of these models with the ARIMA model could reveal potential enhancements for better accuracy. Additionally, combining the ARIMA model with other models, such as integrating the short-term forecasting capability of ARIMA with the long-term forecasting ability of diffusion models, especially when data is limited, can be more advantageous than using either model alone [9].

4.5.3. Long-Term Forecasting

Extending the forecast horizon beyond five years could provide valuable insights into long-term trends and planning needs. Combining short-term accuracy with long-term projections may offer a more strategic approach to healthcare resource planning.

4.5.4. Sensitivity Analysis

Conducting sensitivity analyses to assess the impact of different parameters and assumptions on forecast results can help evaluate model robustness and identify areas for potential improvement [10].

5. Conclusion

The ARIMA(0,2,0) model effectively forecasts the number of hospital beds in China, indicating a steady increase over the next five years. The model's high accuracy in predicting bed numbers for 2021–2023 underscores its reliability and provides confidence in its short-term forecasts. The forecast results highlight the necessity for proactive healthcare planning and infrastructure development to address the anticipated growth in bed numbers. Policymakers and planners should leverage these predictions to guide decisions on facility expansion and resource allocation. Additionally, the study's findings suggest that similar ARIMA models can be valuable for other healthcare-related forecasting tasks. Future research should explore incorporating additional variables, alternative modeling approaches, and long-term forecasting to enhance predictive accuracy and support more comprehensive healthcare planning. The results and recommendations offered in this study aim to assist in addressing the evolving needs of China's healthcare system and ensuring adequate medical service provision.

References

- [1] Zhuang, Z., Cao, P., Zhao, S., Han, L., He, D., & Yang, L. (2021). The shortage of hospital beds for COVID-19 and non-COVID-19 patients during the lockdown of Wuhan, China. *Annals of Translational Medicine*, 9(3).
- [2] Xiang, Y., He, S., Webster, C., & Yu, M. (2022). Divergent distributions of physicians and healthcare beds in China: Changing patterns, driving forces, and policy implications. *Applied Geography*, 138, 102626.
- [3] Clements, M. P., & Hendry, D. F. (2020). *Forecasting economic time series*. Cambridge University Press.
- [4] Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice*. OTexts.
- [5] Chatfield, C., & Xing, H. (2019). *The analysis of time series: An introduction with R*. Chapman and Hall/CRC.
- [6] Pankratz, A. (2021). *Forecasting with dynamic regression models*. Wiley.

- [7] Worden, K., Iakovidis, I., & Cross, E. J. (2019). *On stationarity and the interpretation of the ADF statistic. In Dynamics of civil structures, volume 2: Proceedings of the 36th IMAC, a conference and exposition on structural dynamics 2018 (pp. 29-38). Springer International Publishing.*
- [8] Dimri, T., Ahmad, S., & Sharif, M. (2020). *Time series analysis of climate variables using seasonal ARIMA approach. Journal of Earth System Science, 129, 1-16.*
- [9] Christodoulos, C., Michalakelis, C., & Varoutas, D. (2010). *Forecasting with limited data: Combining ARIMA and diffusion models. Technological Forecasting and Social Change, 77(4), 558-565.*
- [10] Jamil, R. (2020). *Hydroelectricity consumption forecast for Pakistan using ARIMA modeling and supply-demand analysis for the year 2030. Renewable Energy, 154, 1-10.*