The Nexus of Sustainability

Volume 1, Issue 1, 2024

TIME SERIES FORECASTING OF DAILY EMERGENCY DEPARTMENT ARRIVALS

Musa Khan

MPhil Scholar, Department of Statistics, University of Peshawar, Pakistan

musa4khanu@yahoo.com

Keywords
Emergency Department; Overcrowding; Time series forecasting

Article History
Received: 29 October, 2024
Accepted: 11 December, 2024
Published: 31 December, 2024

Corresponding Author: *
Musa Khan

Abstract
Rising population pressures have led to overcrowding in hospital Emergency Departments (ED), impacting patient care and resource management. This study forecasts daily ED arrivals at District Headquarters (DHQ) Hospital Charsadda using data from January 2024 to January 2025. A seasonal ARIMA (SARIMA) model—specifically SARIMA (1,0,1)(2,1,3)$_7$—was identified as the best fit based on AIC, BIC, MAE, and MAPE criteria. Model diagnostics confirmed its validity, with an MAE of 13 and MAPE of 2.15, demonstrating high accuracy. These results suggest SARIMA is effective for predicting daily ED demand, enabling better staff scheduling and resource allocation to mitigate overcrowding.

INTRODUCTION

The Emergency Department is a primary source of admission to the hospital and is considered an essential part of the healthcare system. A healthcare system operates as a sequence of processes in which the patients enter, wait for service, receive treatment, and then leave. Meanwhile, each patient follows a specific pathway quantified by the ED management before getting the final treatment. During this time, the patients may require various resources such as stretchers, patient beds, specialized equipment, and staff capacity (Jean 2020). The primary role of the health care system is to manage patients' flow efficiently while ensuring safe and effective treatment. However, the demand for healthcare services is increasing due to population growth. The rising number of patients in EDs has become a global concern, often leading to overcrowding (Choudhury, 2019). The American College of Emergency Physicians defines hospital overcrowding as a condition that happens when the demand for medical services exceeds the available resources and health care system capacity (Hoot and Aronsky, 2008). This situation negatively impacts the healthcare system, affecting patients' satisfaction, quality of care, nursing efficiency, length of stay, and diagnosis accuracy. It may also contribute to adverse events (J. Bergs et al. 2014). Managing patients' flow in a hospital, particularly ED arrivals, is a challenging task. Therefore, several researchers are exploring new strategies to assist healthcare administration in improving patient flow management. To address this issue, our study aims to forecast daily patient arrivals in the ED of District Headquarters (DHQ) Hospital Charsadda. Specifically, we seek to predict the total daily number of patient arrivals based on 12 months of historical data obtained from the DHQ Hospital Charsadda, Khyber Pakhtunkhwa (KPK), Pakistan. The insight gained from this study can help ED administration plan resources more effectively and improve overall patient care.

In recent years, many research studies have focused on forecasting Hourly, Daily, and Monthly patient volumes in the ED. EDs are among the most frequently utilized providers of critical care in the healthcare sector, making their study crucial for enhancing hospital management and the entire healthcare system. Sun et al. (2009) applied the ARIMA model to forecast daily admission of patients to a regional hospital in Singapore. They concluded that time series modeling is a valuable tool for predicting ED workload, assisting in staff roster and resource allocation. Kam et al. (2010) analyzed two years of daily data collected from a Korean hospital and applied three models to forecast the daily number of visitors to the ED. Based on the residual analysis, Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Mean Absolute Percentage Error (MAPE), they found that the multivariate SARIMA model provided a best fit for forecasting the daily numbers of patients visiting the ED compared to univariate SARIMA and the Average method. Bergs et al. (2013) employed an automated exponential smoothing to predict the monthly ED visits in a Belgian hospital, using five years of monthly data collected from 4 Belgian hospitals. Marcilio et al. (2013) developed three different models the Generalized Linear Model (GLM), the Generalized Estimating Equation (GEE), and SARIMA to forecast daily counts of patients visiting the ED of a hospital in Sao Paulo, Brazil. They concluded that the forecasting accuracy depends on the model type and the length of the time horizon being predicted, recommending SARIMA for forecasting the daily number of patients seeking ED. Shahid et al. (2016) used the Box Jenkins Methodology to forecast monthly patient volume coming for ultrasound examination in the Radiology Department of Fatima Memorial Hospital (FMH) Lahore, Pakistan. Their results indicated that the ARIMA is the best fit model for short and long-run forecasting and can be implemented in other hospital departments for future scheduling and administration. Similarly, Ibrahim et al. (2016) applied the SARIMA model to forecast the patients' incoming at the Outpatients Medical Laboratory of Mayo Hospital Lahore. Zhou et al. (2018) compared three models: Single SARIMA, Nonlinear Autoregressive Neural Network NARNN and a hybrid SARIMA-NARNN to forecast the daily and monthly ED. Their finding reveal that the hybrid SARIMA-NARNN performed best for monthly forecasts, while NANN was superior for daily prediction. Choudhury (2019) applied the Box-Jenkins ARIMA model to predict hourly ED arrivals. He concluded that the ARIMA is the most suitable model to provide the most accurate hourly forecasting. Arum et al. (2020) evaluated numerous forecasting models on patient arrival data and found that the multivariate Vector Autoregressive Moving Average (VARMA) outperformed the univariate ARMA and Holt-Winters model based on the error matrices such as MAR, RMSE, and MAPE. Whitt and Zhang (2019) determined that SARIMA was the most suitable and effective model for forecasting ED arrivals among other models, including the Artificial Neural Network. Recently, Kamboh et al. (2024) used the ARIMA model to forecast the patients' admissions at Liaquat University of Medical and Health Sciences (LUMHS) Hospital in Jamshoro, Pakistan. Their study recommended ARIMA as the best model for managing patient volumes and predicting future trends in hospital admissions with high precision.

1. MATERIALS AND METHODS

The time series data on the daily number of patient arrivals to the ED of DHQ Hospital Charsadda (Operating 24 hours a day, 7 days a week) for the period Jan-2024 to Jan-2025 was used for analysis, model fitting, and forecasting. The hospital, located in the main city of Charsadda, serves more than 30,000 patients monthly. The software used for the data analysis in this study is Gretl 2021. The Seasonal ARIMA model was employed to forecast the daily patient arrivals at the ED. The Autoregressive Integrated Moving Average (ARIMA) model was primarily developed by Box and Jenkins in 1976. Later, they extended it to incorporate seasonality, presenting the SARIMA model. This model applies seasonal differencing of an appropriate order to eliminate non-stationarity caused by seasonality in a time series. The seasonal difference of the first order is the difference between the current observation and the corresponding observation from the previous season, that is $Z_t = Y_t - Y_{t-s}$. For daily time series, S=7; for monthly time series, S=12; and for quarterly time series, S=4. Generally, the SARIMA model is represented as SARIMA (p, d, q) (P, D, Q) where p represents autoregressive terms, d non-seasonal difference, and q for moving average terms, while P stands for seasonal autoregressive terms, Q stand for seasonal moving average terms, and D is the order of seasonality difference. The multiplicative form of the SARIMA model in terms of the back shift operator is given below:

$$\phi_{AR}(B)\varphi_{SAR}(B)(1-B)^d(1-B^s)^D Y_t = \theta_{MA}(B)\vartheta_{SMA}(B^S)E_t \quad (1)$$

$$\theta_{AR}(B) = \left(1-\theta_1 B - \theta_2 B^2 - \ldots - \theta_p B^p\right) \quad (2)$$

$$\phi_{SAR}(B) = \left(1-\phi_1 B^s - \phi_2 B^{2s} - \ldots - \phi_p B^{Ps}\right) \quad (3)$$

$$\varphi_{MA}(B) = \left(1+\varphi_1 B + \varphi_2 B^2 + \ldots + \varphi_p B^q\right) \quad (4)$$

$$\vartheta_{SMA}(B^S) = \left(1+\vartheta_1 B^s + \vartheta_2 B^{2s} + \ldots + \vartheta_Q B^{Qs}\right) \quad (5)$$

In these above equations, ϕ_{AR}, and φ_{SAR} represent non-seasonal and seasonal autoregressive terms, while θ_{MA}, and ϑ_{SMA} are the non-seasonal and seasonal moving average operators, respectively. The parameter *s* represents the number of periods per season. (Adhikari & Agrawal 2013).

1.1. The Box-Jenkins Methodology

Fitting an ARIMA model to a time series involves a four-step process known as the Box-Jenkins methodology. It is a popular technique for fitting a time series model and forecasting. The steps involved are outlined below.

1.1.1. Identification

It is the first step in the Box Jenkins methodology to determine the order of the time series model, that is: the order of the Autoregressive (AR) term p order of Moving Average (MA) components q, the order of Seasonal AR terms P, and order of Seasonal MA terms Q and the order of differencing (seasonal/non-seasonal) d/D. The difference is used to eliminate the trend in the time series records. However, a suitable seasonal order difference is used to remove or reduce the seasonal pattern observed in the series. The differentiation continues until the series achieves stationarity.

To determine the appropriate model: If the Autocorrelation Function (ACF) decays exponentially and the Partial Autocorrelation Function (PACF) cuts off after a few lags, then a pure AR model is to be fitted; on the other hand, if PACF decay exponentially while ACF cut off after a few lags, a pure MA model is appropriate. If both PACF and ACF decay geometrically, an ARMA model is recommended.

1.1.2. Process for examination Stationarity

To check the stationarity in a time series, graphical techniques such as a time series plot (e.g., histogram) and a plot of the ACF and PACF (also called a Correlogram) are used. The Historigram tells us about the tendency of data, while the Correlogram helps identify the order of the ARIMA model and also shows the performance of a series. Additionally, statistical tests, called unit root tests, are formally used to check the statistical significance of stationarity in a time series. The popular tests are the Augmented Dickey-Fuller (ADF) **test** and the Dickey-Hasza-Fuller (DHF) **test** (for seasonal data).

1.1.3. Estimation of Parameter

Once the model is identified, the next step is parameter estimation. This is generally done using the methods of Maximum Likelihood and Least Square Estimation. In this study, the maximum likelihood estimation method is used for parameter estimation. To ensure the best model is selected, some competing models in the neighborhood of the estimated model are evaluated. The best fit model is determined using selection criteria such as the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC). The best-fit model with the lowest AIC or BIC and error matrices is considered optimal. The criterion (AIC/BIC) serves as a goodness-of-fit measure, assisting to select the most suitable model among competing substitutes.

1.1.4. Diagnostic Check

The third step of this procedure is diagnosing the model or verifying the adequacy of the proposed model. The general way to check the adequacy of the model is to check the dependency and normality of the residual term obtained from the model. If the model's residual does not exhibit any dependency and white noise, then the model is said to be suitable for the analysis or forecasting. The tests generally used for residual analysis are the Box-Pierce test, Ljung box test, and $\chi 2$ test of normality. Additionally, the ACF and PACF plot (Correlogram) of the residuals are used to check for autocorrelation in the residual terms. To test for residual normality, the Jarque-Bera test is used, which follows chi-square distribution with two degrees of freedom.

1.1.5. Forecasting

Once the model passes all the diagnostic checks, it is considered adequate for forecasting. The final and fourth step of the BJ Methodology is to generate a forecast for future values based on the estimated model.

2. RESULTS AND DISCUSSION

A time series plot for the patients' data was created (shown in Fig.1) to examine the pattern and tendencies in data. The ACF and PACF plots (the Correlogram) were generated before differencing the series, as shown in Fig.2. Both the time series and Correlogram plots indicate that the series was non-stationary as a slight increasing trend and a little repetition of the same shape were observed. Most of the ACF spikes in Fig.2 exceed the 95% confidence interval, further

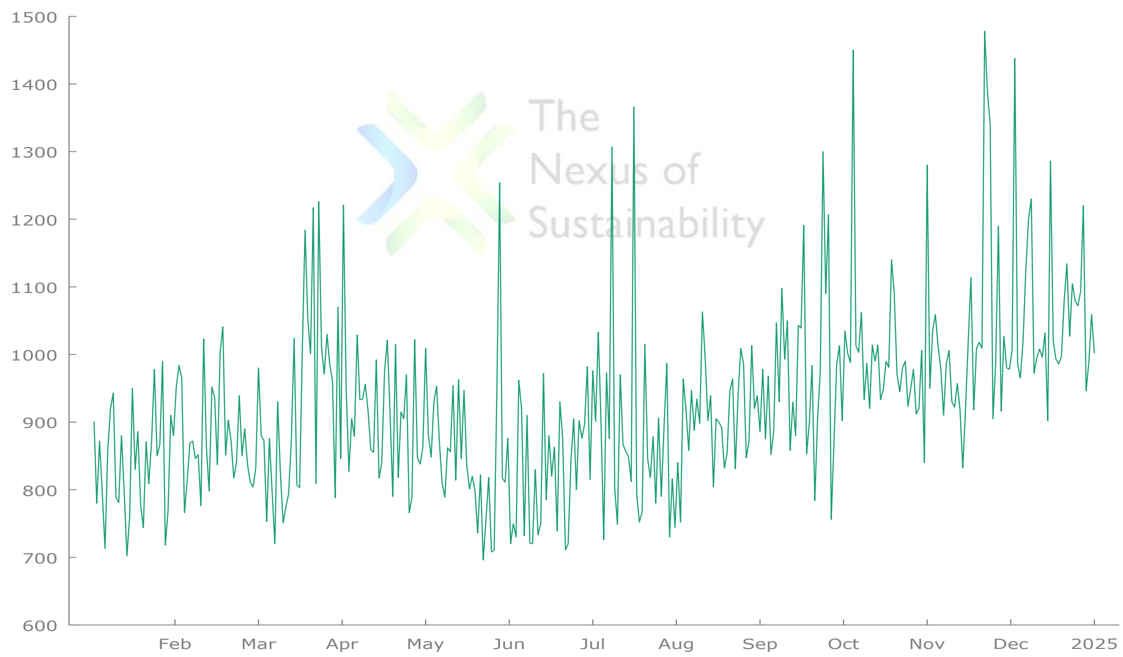

Fig.1 Time series plot before difference

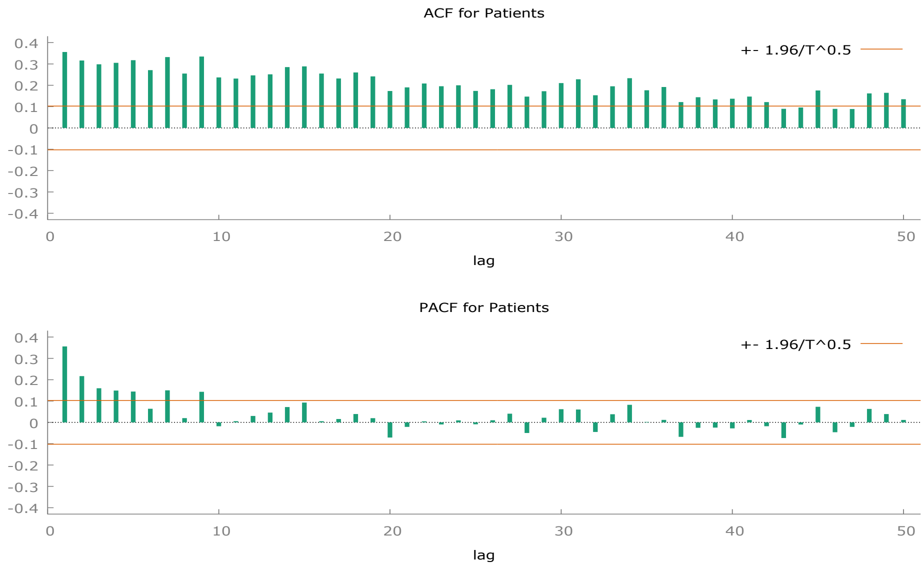

Fig.2 Correlogram before seasonal differencing

confirming non-stationarity. The Dickey Hasza Fuller (DHF) unit root test results presented in Table 1 also confirm that the series was non-stationary before differencing. To achieve stationarity, a seasonal difference was applied. Fig.3 represents the time series plot after the first seasonal difference (D=1), showing that the series became stationary. In Fig. 4, the Correlogram and the DHF test result after the seasonal difference shown in Table 1 confirm this transformation.

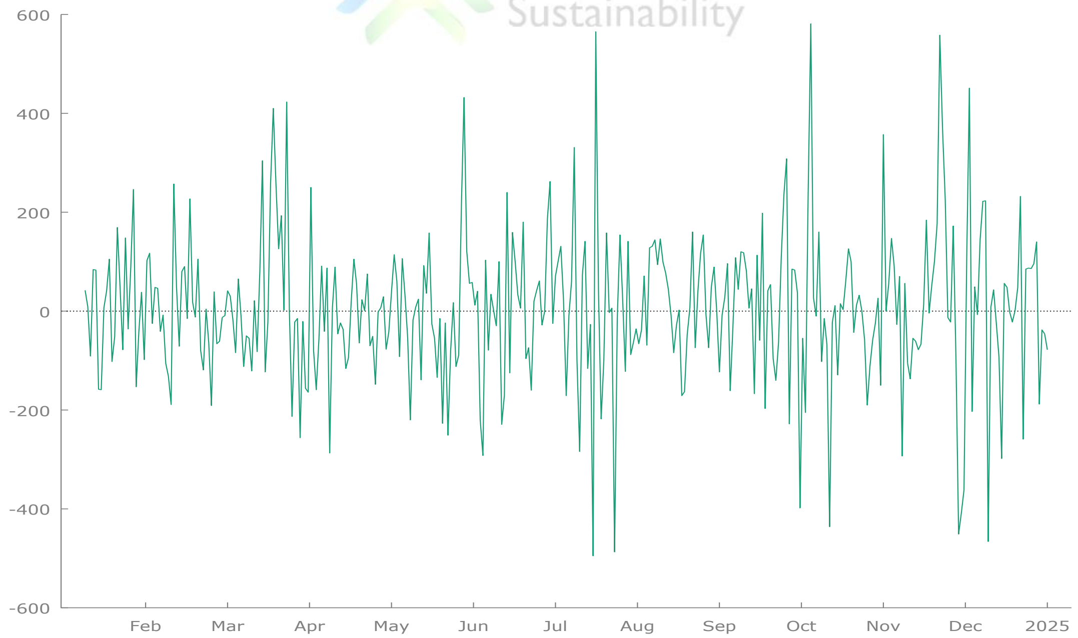

Fig.3 Time series plot after seasonal difference (D=1)

The order of model parameters was identified from the behavior of the Correlogram shown in Fig.4. Since the ACF dying down suggests that q=0, in the same way, the PACF also dying down suggests p = 0.

The ACF cuts off at seasonal lags, as the correlation at lag 7 is significant but becomes non-significant at higher seasonal lags (e.g., lag 14, 21, 28, etc). However, the PACF remains statistically significant at these seasonal lags (e.g., lag 7, lag 14, 21, 28, 35), suggesting that Q = 1 and P = 0. Consequently, the estimated model was SARIMA (0,0,0) $(0,1,1)_7$. Alongside this model, some other models in their neighborhood were also fitted. Table 3 presents these models along with their corresponding Akaike Information Criterion (AIC) and Bayesian Information Criterion BIC values. The best model was selected based on the lowest AIC and BIC values and the highest of the R^2 value. As seen in Table 3, the SARIMA (1,0,1) $(2,1,3)_7$ has the lowest value of AIC and BIC, making it the best model to fit.

Fig.4 The Correlogram after seasonal difference (D=1)

After selecting the appropriate model, the next step was to diagnose it, which was conducted using the Ljung-Box Correlation Test and the residual Correlogram. It is evident from Fig. 5 and Table 2 that the residual term is independent, confirming the adequacy of the model. The model was estimated using the method of Maximum Likelihood Estimation. The parameters of the fitted model SARIMA (1,0,1) $(2,1,3)_7$ are presented in Table 4. The coefficients for the autoregressive and moving average terms are statistically significant, as shown by their low p-values (all near to zero). The large absolute values of the Z-statistics also confirm their significance. These results suggest that the selected model effectively captures the temporal dependencies in the data, making it suitable for forecasting emergency department arrivals. Once the model was validated, it was used for forecasting. The forecasting was completed using the SARIMA (1, 0, 1) $(2, 1, 3)_7$. The forecasting accuracy was assessed using Mean Absolute Percentage Error (MAPE) and Mean Absolute Error (MAE), which were found to be 2.15% and 13, respectively, demonstrating a high level of accuracy in predicting emergency department arrivals.

Table 1: DHF Seasonal Unit Root Test

Test before seasonal difference		Test after seasonal difference	
Test statistic value	-0.862907	Test statistic value	-21.601068

Critical value	-2.040000	Critical value	-2.040000
Conclusion	The series is non-stationary	Conclusion	The series is stationary

Table 2. Ljung Box test

Lag order	14	21	28	35	42	49
Test statistic value	6.92299	11.4522	13.6328	20.1549	24.7786	32.7604
P value	0.4369	0.6502	0.8848	0.8588	0.9005	0.8459

Table 3: Models and their respective AIC and BIC values

Model	AIC	BIC	Model	AIC	BIC
SARIMA (1, 0, 1) (1, 1,1)	4442	4468	**SARIMA (1, 0, 1) (2, 1,3)**	**4441**	**4462**
SARIMA (1, 0, 1) (1, 1,2)	4444	4470	SARIMA (2, 0, 1) (1, 1 ,1)	4444	4480
SARIMA (1, 0, 1) (1, 1,3)	4445	4476	SARIMA (2, 0, 2) (1, 1 ,1)	4443	4477
SARIMA (1, 0, 1) (2, 1,1)	4443	4473	SARIMA (2, 0, 2) (2, 1 ,2)	4446	4470
SARIMA (1, 0, 1) (2, 1,2)	4445	4480	SARIMA (2, 0, 2) (2, 1 ,3)	4449	4473

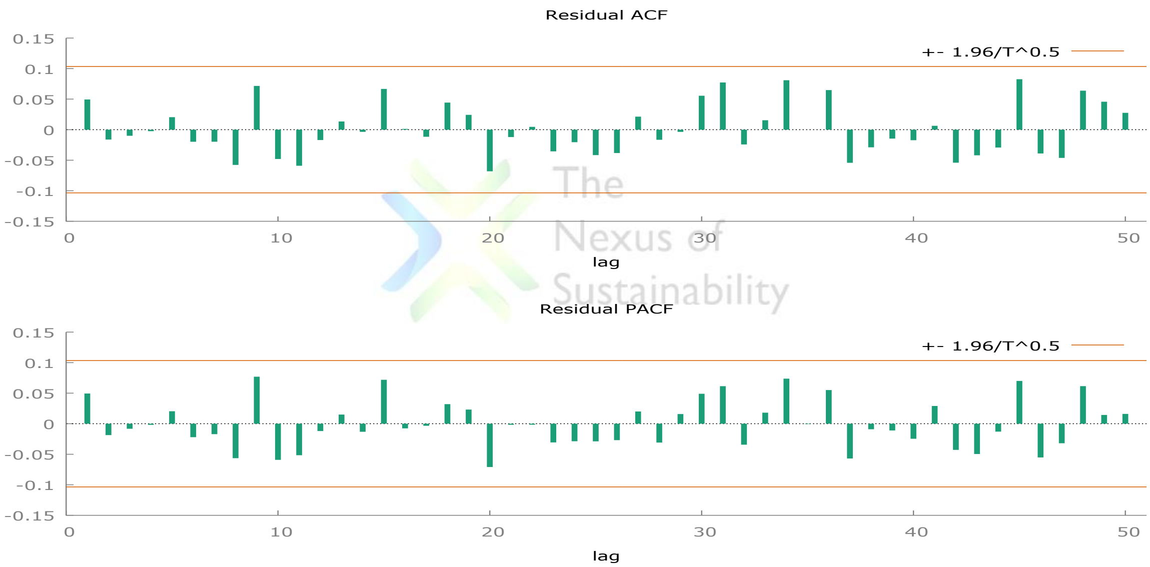

Fig.5 Residual Correlogram

Table 4. SARIMA (1, 0, 1) (2, 1, 3)

Parameter	Coefficient	Std. Error	Z-Value	P-Value
phi_1	0.972276	0.0260159	37.37	1.10e-305
Phi_1	0.536219	0.108339	4.949	7.44e-07
Phi_2	-0.872507	0.125895	-6.93	4.19e-12
theta_1	-0.894982	0.0474411	-18.87	2.21e-79
Theta_1	-1.47287	0.127813	-11.52	1.00e-30
Theta_2	1.38066	0.206085	6.7	2.09e-11
Theta_3	-0.907785	0.127272	-7.133	9.85e-13

3. Conclusion

Patient arrivals with increasing trends in Emergency Departments (EDs) have become a major issue all over the country and worldwide. The rise in patient volume may cause stressful situations, increased healthcare costs, and a decline in the quality of services provided by the ED. To manage this issue effectively, hospital ED management needs to forecast patient arrivals accurately for the future to handle the overcrowding and provide efficient service. The primary objective of this study was to find a best-fit model to predict the daily number of patient arrivals to the ED of QHQ Hospital Charsadda. This objective was achieved using the Box-Jenkins Methodology. Several SARIMA models were fitted, and the best model, SARIMA (1, 0, 1) (2, 1, 3)$_{7,}$ was selected based on the lowest value of the AIC and BIC, along with the higher value of R^2. The findings reveal that the SARIMA model is the best fit for forecasting ED arrivals. This model can be applied to other hospitals with similar ED arrival patterns to estimate daily patient inflows. The result of this study may be useful for upcoming scheduling, medical researchers, hospital ED management, and overall patient care. Therefore, it is suggested that the administration of DHQ hospital Charsadda devise strategies based on our findings to improve public healthcare services efficiently.

4. REFERENCES

Rizk, J. (2020). Coxian phase-type modeling and time series forecasting of patient length of stay and hospital overcrowding. PhD thesis, University of Limerick.

Choudhury, A. (2019). Hourly forecasting of emergency department arrivals: time series analysis. arXiv preprint arXiv:1901.02714.c

Hoot, N. R., & Aronsky, D. (2008). Systematic review of emergency department crowding: causes, effects, and solutions. Annals of emergency medicine, 52(2), 126-136.

Bergs, J., Heerinckx, P., & Verelst, S. (2014). Knowing what to expect, forecasting monthly emergency department visits: A time-series analysis. International emergency nursing, 22(2), 112–115. https://doi.org/10.1016/j.ienj.2013.08.001

Jones, S. S., Evans, R. S., Allen, T. L., Thomas, A., Haug, P. J., Welch, S. J., & Snow, G. L. (2009). A multivariate time series approach to modeling and forecasting demand in the emergency department. Journal of biomedical informatics, 42(1), 123-139.

Kam, H. J., Sung, J. O., & Park, R. W. (2010). Prediction of Daily Patient Numbers for a Regional Emergency Medical Center using Time Series Analysis. Healthcare informatics research, 16(3), 158–165. https://doi.org/10.4258/hir.2010.16.3.158

Bergs, J., Heerinckx, P., & Verelst, S. (2014). Knowing what to expect, forecasting monthly emergency department visits: A time-series analysis. International emergency nursing, 22(2), 112–115. https://doi.org/10.1016/j.ienj.2013.08.001

Marcilio, I., Hajat, S., & Gouveia, N. (2013). Forecasting daily emergency department visits using calendar variables and ambient temperature readings. Academic emergency medicine: Official Journal of the Society for Academic Emergency Medicine, 20(8), 769–777.

Shahid, K., Manzoor, T., & Ibrahim, M. (2016). FORECASTING OF MONTHLY PATIENT VOLIUME AT RADIOLOGY DEPARTMENT COMING FOR ULTRASOUND: A TIME SERIES APPROCH. Journal of University Medical & Dental College, 7(3), 22-27.

Ibrahim, M., Jamil, M., & Akbar, S. (2016). The fitting of the SARIMA model on peads patients coming at outpatient's medical laboratory (OPML), Mayo hospital, Lahore. Journal of University Medical & Dental College, 7(1), 12-18.

Zhou, L., Zhao, P., Wu, D., Cheng, C., & Huang, H. (2018). Time series model for forecasting the number of new admission inpatients. BMC medical informatics and decision making, 18(1), 1-11.

Choudhury, A. (2019). Hourly forecasting of emergency department arrivals: time series analysis. arXiv preprint arXiv:1901.02714.c

Arum, K. C., Ugwu, C. L. J., Ugwuowo, F. I., Oranye, H. E., & Agu, P. I. (2020). Time Series Model Comparative Approaches in Analyzing Accident and Emergency

Departmental Demand in Eastern Nigeria. Afr. J. Math. Stat. Stud, 3, 55-62.

Whitt, W., & Zhang, X. (2019). Forecasting arrivals and occupancy levels in an emergency department. Operations Research for Health Care, 21, 1-18.

Kamboh, S., Talpur, M. G. H., Chand, N. K., Zardari, L. A., Shaikh, A. W., & Kamboh, S. A. (2024). Time Series Modeling and Forecasting of the Patients' Inflow and Admission in the Hospitals: A cases study of LUMHS Hospital Jamshoro Pakistan. VFAST Transactions on Mathematics, 12(1), 311-322.

Adhikari, R., & Agrawal, R. K. (2013). An introductory study on time series modeling and forecasting. arXiv preprint arXiv:1302.6613.