PAPER • OPEN ACCESS

Forecasting patient admission in orthopedic clinic at a hospital in Kuantan using autoregressive integrated moving average (ARIMA) models

To cite this article: B Mohamed and M Mohamad 2020 *J. Phys.: Conf. Ser.* **1529** 052090

View the article online for updates and enhancements.

Forecasting patient admission in orthopedic clinic at a hospital in Kuantan using autoregressive integrated moving average (ARIMA) models

B Mohamed[1] and M Mohamad[2]

[1]Faculty of Business and Management, Widad University College, BIM Point, Bandar Indera Mahkota, 25200 Kuantan, Pahang
[2]Kulliyah of Nursing, International Islamic University Malaysia, Jalan Sultan Ahmad Shah, Bandar Indera Mahkota, 25200 Kuantan, Pahang

Abstract. This study is an attempt to examine empirically the best ARIMA model for forecasting. The monthly time series data routinely-collected at Orthopedic clinic from January 2013 until June 2018 have been used for this purpose. At first the stationarity condition of the data series is observed by ACF and PACF plots, then checked using the Ljung-Box-Pierce Q-statistic. It has been found that the monthly time series data of the Orthopedic clinic are stationary. The best ARIMA model has been selected by using the MAPE. To select the best ARIMA model the data split into two periods, viz. estimation period and validation period. The model for which the values of MAPE are smallest is considered as the best model. Hence, ARIMA (1, 0, 0) is found as the best model for forecasting the Orthopedic clinic data series. The out of sample forecast by using ARIMA (1, 0, 0) model indicated a fluctuation of monthly orthopedic patients demand, from lowest was 294 and the highest was 299 patients that could receive treatment from the clinic in a month.

1. Introduction

As the economy and society continue to develop, effective hospital management has become increasingly important. Hospitals are facing many problems such as the aging of our domestic population, pandemic diseases like H1N1, and rising operational and maintenance costs. The healthcare sector in Malaysia has been expanded over the past few years. Accordingly, The private hospital is expected to continue to grow as a result of private insurance benefits and increased per-capital income [1]. Adding to this is the aging of the population and the rising wealthy class in Malaysia, the variety of medical insurance products available and the health-conscious society lead to strong market growth in private hospital sector [2].

Another factor that boosts the expansion of private hospital is medical tourism. Malaysia has been rated as the world's fourth best medical tourism destination [3]. Malaysia attracts medical tourists because the availability of good healthcare facilities at low cost and together with its favorable exchange rate, political and economic stability and high rate of literacy. Furthermore, the Malaysian government supporting medical tourism as part of the country's current 5-year development blueprint by incorporating in the Eleventh Malaysia Plan, which covers from 2016 to 2020 [4].

Hence, to be able to plan effectively, it is vital for a private hospital to anticipate future. A private hospital's ability to predict hospital demand can have a significant impact on both customer service and financial performance [5]. The accurate prediction would facilitate, at a micro level, for example,

scheduling nursing and support personnel, and at the macro level for financial and strategic planning for the hospital. Forecasting can also be a great help in controlling resources for the hospital, thereby improving efficiency, reducing costs and increasing profitability.

Consequently, there may be unexpected fluctuations in demand for a hospital service. The gradual increase in the number of patients starts putting pressure on a hospital's available staff and strain facility capacities of a hospital. This capacity crisis has resulted in significant operational gaps, including an increase in patients waiting to be admitted into wards, and long wait time for laboratory, radiology, and other diagnostic services. The implications of this situation would impact the overwhelmed quality of medical staff's diagnosis and prognosis, resulting in increased work dissatisfaction [5]. If this variance can be reduced, higher efficiency can be reached by the hospital. Projection of demand will allow hospital management to plan for these changes and prevent unintended mistake. This research aims to establish parsimony and active models that could forecast the number of patients present on a monthly basis, thereby helping the hospital's management for a better planning.

2. Literature review

Forecasting has been an important research topic for decades to forecast patients receiving treatment in hospitals. Predictive models are an essential part of the forecasting process. The model building process should take into account the relationships between individual variables and the system as a whole must be checked to ensure the model's forecast capability. This research used ARIMA (autoregressive integrated moving average) model to predict patient admission as it is richer with information-related time series than other similar methods [6]. The finding of relevant studies are briefly described below.

The method of forecasting is used to predict patient admission at Hospital Medicine (HM) in the United States and discover that the approaches predict HM admission accurately [7]. Accordingly, the ARIMA model is appropriate for forecasting the emergency patient admissions at the hospital in Athens [8]. Meanwhile, a study comparing ARIMA and artificial intelligence methods to predict routine hospital admissions due to circulatory and respiratory problems in Madrid, results suggesting that ARIMA model approaches can solve the problem properly [9].

Reis and Mandl developed ARIMA models to provide a foundation for real-time surveillance and bioterrorism detection in Children's Hospital Boston [10]. They forecast overall hospital visits and respiratory-related patient visits at the pediatric emergency department. The results showed that overall visits and respiratory visits are best fitted by ARIMA(2,0,1) and ARIMA(1,0,1) processes respectively. The mean absolute percentage error (MAPE) of the overall visits model was 9.37% and respiratory model was 27.54%. The ARIMA model of overall visits was able to forecast 7 day-long abnormal visit patterns.

A hospital could be able to plan elective admissions properly through accurately predicting emergency patient admissions. Subsequently, Jones et al. developed several ARIMA models of daily hospital bed occupancy from emergency admissions and the number of emergency admissions of Bromley Hospitals NHS Trust [11]. Their models are able to forecast beds occupancy with good accuracy. Champion et al. study supports the finding [12]. They developed ARIMA models to forecast monthly emergency patient arrivals to the emergency department of a hospital in regional Victoria, Australia. They found that ARIMA (0, 1, 1) model is fairly adaptable to the type of data. Thus, they concluded that time series analysis is a useful tool for forecasting emergency department admissions. Abraham et al. further confirmed the findings [13]. They compared several ARIMA models for forecasting daily inpatient admissions and occupancy to Royal Melbourne Hospital, Victoria, Australia. They found that the models are able to forecast emergency occupancy up to seven days ahead with reasonable accuracy.

In epidemic disease forecasting, Promprou et al. applied ARIMA models to forecast dengue hemorrhagic fever (DHF) cases in Southern Thailand [14]. They established ARIMA (1, 0, 1) model and found that the models are very useful in forecasting the DHF in Southern Thailand. Li et al.

supports this finding, suggesting that the models of ARIMA fit the fluctuations in the incidence of hemorrhagic fever with renal syndrome in China[15]. In the meantime, Permanasari et al. prediction number of malaria incidence in human by using Seasonal Autoregressive Integrated Moving Average (SARIMA) and found that the forecasting model able to support and provide a good prediction of the incidence [16]. As such, they recommended that the models may offer the potential for improving planning, control and prevention of the diseases that can be utilized by the health authority.

3. Data and methodology

3.1 Data
For the purpose of time series modelling in this study, data from the Orthopedic clinic of the a private hospital in Kuantan were used. The data were based on the count of patients receiving treatments at the orthopedic clinics and obtained from the admission record books of the clinic. The monthly data from first 60 observations (January of 2013 to December of 2017) were used as training data on which the models were estimated. The subsequent 6 observations (January to June 2018) were kept as hold out sample to verify the accuracy of the forecasting models.

3.2 Methodology
The data from the clinics are collected in the sequence of time, therefore modelling with the time series technique is possible. Particularly, the Box-Jenkins approach to autoregressive integrated moving average or ARIMA (p, d, q) models, where p is the auto-regression (AR), d is the differencing or integration (I) and q is moving-average (MA). All three are based on the simple concept of random disturbances or shocks. ARIMA models are usually formulated with the premise of constant variance in the error term [11]. For ARIMA(p,d,q) model, the AR(p) must be stationary and MA(q) must be invertible [17, 18]. ARIMA (p, d, q) is popularly applied in empirical studies among researchers in the health care industry [15, 10, 19].

   A brief description of the time series models and definitions used in this study are as following paragraph. In practice, most of the time series are non-stationary. According to Bowerman et al. Box-Jenkins methodology required time series use in forecasting to be stationary, i.e., stationarity implying that the time series is invariant with respect to time and the mean is constant through time [17]. As such, non-stationary time series must be transformed into stationary time series values. The transformation process can be done through differencing or integration.
The non-seasonal mixed ARIMA (p,q) model is applied when data show evidence of non-stationarity and can be expressed in equation form of

$$Z_t = \partial + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + ... + \phi Z_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - .... - \theta_q a_{t-q} \qquad (1)$$

Where Z_t is a realization of the time-series, ϕ and θ are parameters of the model and a is an Independent and Identically Distributed (IID) error term with mean of zero and constant variance.

   The Autoregression (ARIMA) model $Z_t = \partial + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + ... + \phi Z_{t-p} + a_t$ is non-seasonal autoregressive model of order p. In an autoregressive (AR) process, each value in a series at time t depends only on its previous values and on a random noise. It is a linear function of the previous value or values. That is, in order to make a forecast one needs to know the p previous values. AR assumes that the future values of an examined variable may be approximated and forecasted by its own previous values. That is, its past behaviors may suggest important information regarding its near future dynamics [20]. In a first-order autoregressive process, only the single preceding value is used; in a second-order process, the two preceding values are used, and so on. These processes are commonly indicated by the notation AR (n) or ARIMA (n, 0, 0), where the number in parentheses indicates the order.

The Moving-average (ARIMA) model $Z_t = \partial + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - .... - \theta_q a_{t-q}$ is non-seasonal moving average model of order q. The moving-average (MA) component of an ARIMA model tries to predict future values of the series based on deviations from the series mean observed for previous values. In a moving-average process, each value is determined by the weighted average of the current disturbance and one or more previous disturbances. The order of the moving-average process specifies how many previous disturbances are averaged into the new value. In the standard notation, an MA (n) or ARIMA (0,0,n) process uses n previous disturbances along with the current one.

The time series model building technique of Box-Jenkins is a method of finding a suitable ARIMA model that adequately represents the data set. The approach is an iterative that can go through the process several times before finding the suitable model. That is, the methodology is designed to arrive at ARIMA models through a three interactive stage procedure based on model identifications, model estimation, and diagnostic checking or model validation, and then utilizes the models for forecasting [see; 17, 18, 21].

The common method used to measure the accuracy of the forecast is the mean absolute percentage error (MAPE). A MAPE of 0% indicated a perfect fit of the model in the training data [10, 21]. The mean absolute percentage error is calculated as:
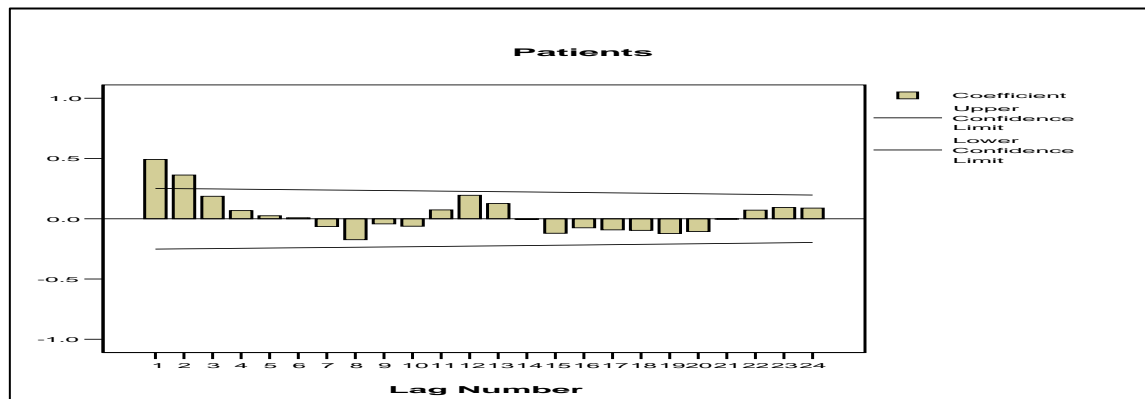
$$MAPE = \frac{\sum \left| (A_t - F_t)/A_t \right|}{n},$$

(2)

Where, A_t is actual value in period t, F_t is forecast value in period t, and n is the number of periods used in the calculation.
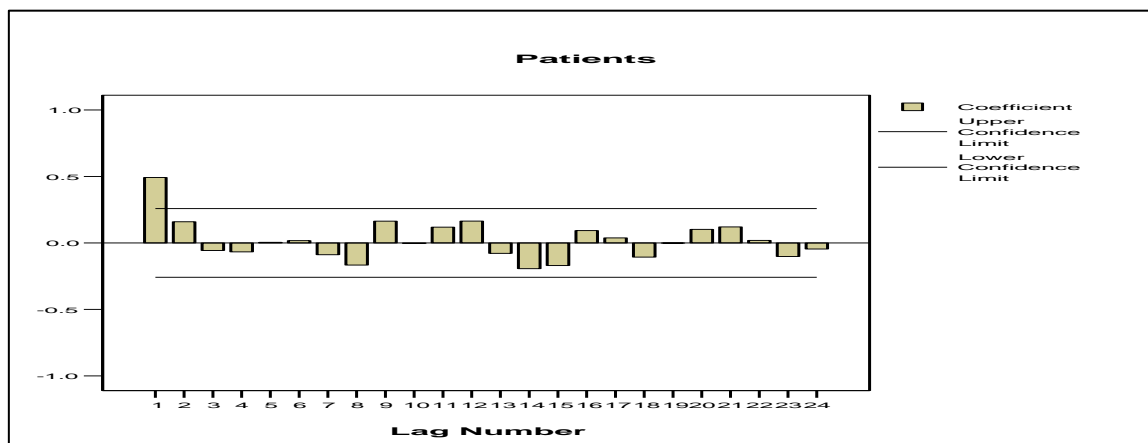
4. Results
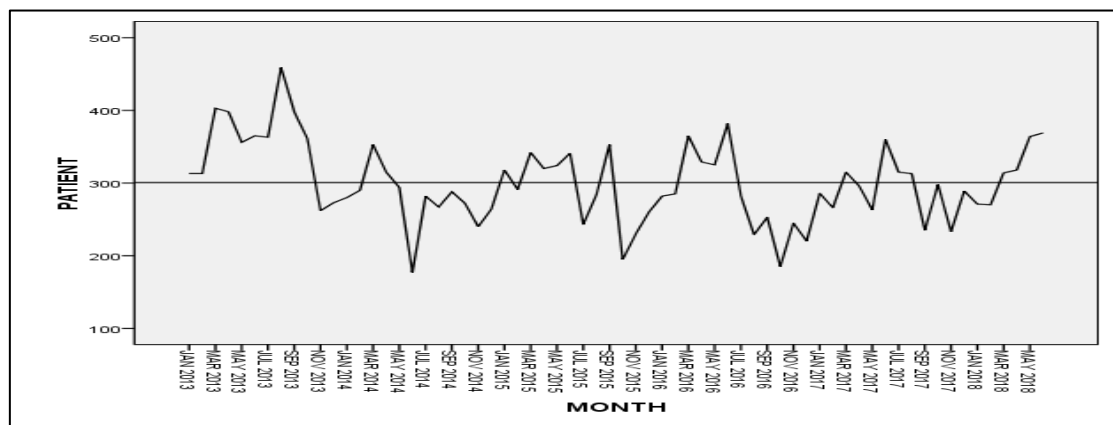
Figure 1: Plot of monthly orthopedic patient demand of original values

Figure 2: ACF of monthly orthopedic patient original values

Figure 3: PACF of monthly orthopedic patient original values

Figure 1 shows the time plot of the original values of time series seem to fluctuate around a constant mean, and, therefore, could be considered stationary. The graph also shows that is no seasonality and no trend in the data. Furthermore, figure 2 of the ACF of the series shows that the series dies down fairly quickly after lag two through damped sine-wave fashion. As such, the series is considered stationary. Since the time series is stationary without being difference or integration, the optimal degree of integration is determined to be zero ($d = 0$). By comparing the pair of correlograms in figure 2 and 3, we could tentatively determine the order of ARIMA models. Figure 3 shows that the PACF behaviour cut off fairly quickly after lag one as compare to figure 2 of ACF. Therefore, the autoregressive of order p is a possible ARIMA model. The single spike in the PACF strongly suggests that an AR (1) model is appropriate for monthly orthopedic patient demand. Consequently, AR (1) is then tested to determine adequate ARIMA (p,d,q) model from the available training data.

From the analysis, finally, ARIMA (1, 0, 0) model (first order Autoregressive) is recognized to generate the appropriate results to represent monthly orthopedic demand. The following Table 1 shows the estimated parameters for ARIMA (1,0,0) model of monthly orthopedic demand.

**Table 1**: Estimated parameters of ARIMA (1,0,0) model

Parameter	CONSTANT	ϕ
Estimate	299.02857	.48505
Std. Error	12.171544	.114030
T-RATIO	24.567841	4.253733

APPROX. PR.	.00000000	.00007757

4.1. Stationarity condition of ARIMA (1, 0, 0) model

By observing the Table 1 and following Bowerman *et al*., the stationary of the estimated model could be evaluated [17]. That is an AR (1) or ARIMA (1,0,0) model the coefficients ϕ, must be less than 1. The Table 1 shows ϕ is 0.48505 which is less than 1. As such the model fulfills the stationarity condition and model was considered stable. Furthermore, the coefficients ϕ has t-ratio greater than 2 with p-value less than 5 percent ($p<.05$).

4.2. Diagnostic checking of ARIMA (1, 0, 0) model
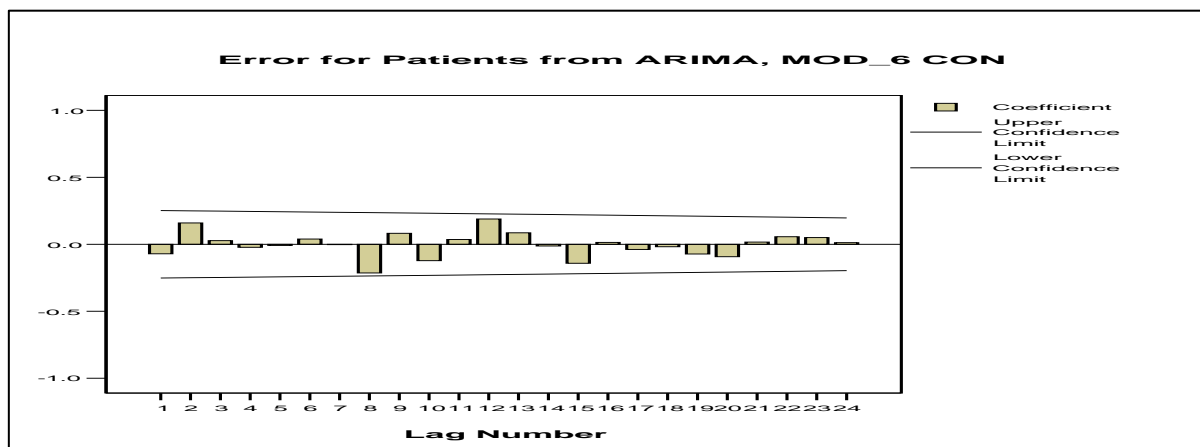
Figure 4: ACF of residuals from ARIMA (1, 0, 0) model
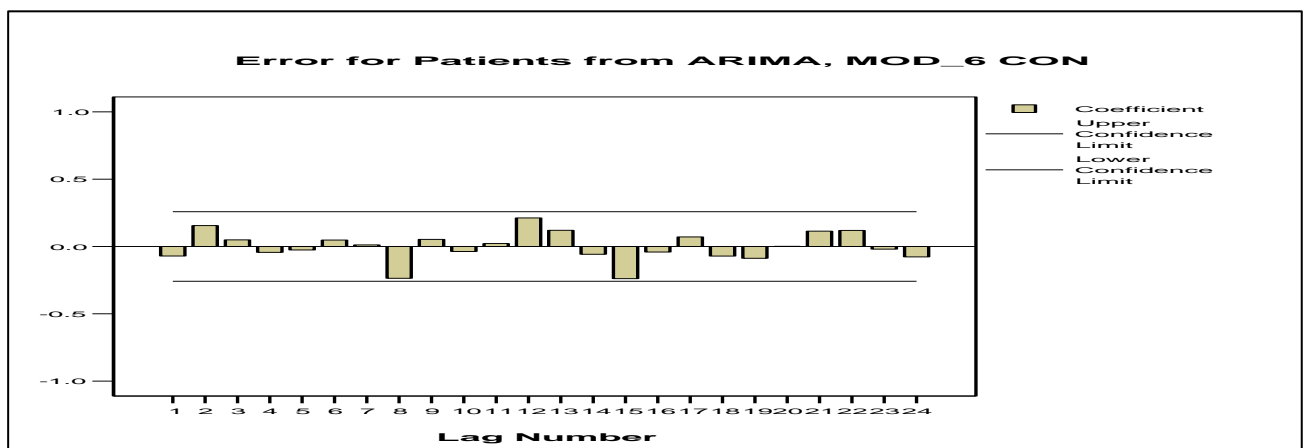
Figure 5: PACF of residuals from ARIMA (1, 0, 0) model

```
                               Auto-  Stand.

       Lag  Corr.   Err.  -1  -.75  -.5 -.25   0   .25  .5   .75   1   Box-Ljung  Prob.

                          □⇩⇩⇩⇩⇧⇩⇩⇩⇩⇧⇩⇩⇩⇩⇧⇩⇩⇩⇩⇧⇩⇩⇩⇩⇧⇩⇩⇩⇩⇧⇩⇩⇩⇩⇧⇩⇩⇩⇩□

        1  -.071   .126                .   *⇔   .                      .315      .574

        2   .159   .125                .   ⇔*** .                     1.946      .378

        3   .027   .124                .   ⇔*   .                     1.994      .574
```

```
4   -.022   .123          .    *    .              2.027   .731

5   -.009   .122          .    *    .              2.032   .845

6    .039   .120          .   ⇔*    .              2.137   .907

7   -.002   .119          .    *    .              2.137   .952

8   -.214   .118          .****⇔    .              5.409   .713

9    .082   .117          .   ⇔**   .              5.899   .750

10  -.122   .116          .  **⇔    .              7.010   .725

11   .035   .115          .   ⇔*    .              7.105   .790

12   .189   .114          .   ⇔****.              9.872   .627

13   .086   .112          .   ⇔**  .             10.456   .656

14  -.013   .111          .    *    .             10.468   .727

15  -.142   .110          .***⇔     .             12.130   .669

16   .014   .109          .    *    .             12.147   .734

17  -.038   .108          .   *⇔    .             12.274   .783

18  -.018   .106          .    *    .             12.303   .831

19  -.072   .105          .   *⇔    .             12.777   .850

20  -.093   .104          .  **⇔    .             13.577   .851

21   .017   .102          .    *    .             13.603   .886

22   .056   .101          .   ⇔*   .             13.914   .905

23   .050   .100          .   ⇔*   .             14.167   .922

24   .012   .098          .    *    .             14.183   .942
```

Figure 6: ACF of error from ARIMA (1, 0, 0) model

Diagnostic checking was performed to verify the ARIMA (1, 0, 0) model adequacy.Figure 4 and 5 show correlogram of the ACF and PACF of the residuals from ARIMA(1,0,0) model respectively. Figure 6 shows the autocorrelation plot of residuals of the ARIMA (1,0,0) model. By observing figure 4 and 5, we could conclude that ACF and PACF of the residuals lay within the boundary lines, indicating that overall the residual time series approximates zero mean white noise behaviour. From figure 6, we could observe that the autocorrelation of the residuals is small. Thus Ljung-Box Q statistic values are relatively small. The residuals of ARIMA (1, 0, 0) model also have a smaller standard errors and larger p-values (i.e., $> .05$) corresponding to Ljung-Box Q statistic [see; 17]. As such, we could conclude that the ARIMA (1, 0, 0) model is adequate and could be used in forecasting.

4.3. Forecasting monthly patients demand at Orthopaedic clinic
The main objective of this study is forecasting. Forecasting is to predict future values of a time series. Time series models developed are used to forecast future demand in the clinic. However, we need to evaluate the accuracy of the forecasting models over certain periods of time in order to identify the best model, i.e., the model that has little error as possible. The common method that usually employs is the mean absolute percentage error (MAPE). MAPE is used to measure of quality of fit that measure

the percentage of the deviation between forecasted and observed values of a given time series. A MAPE of 0% indicated a perfect fit of the model in the training data [10, 21].

Table 2: Forecast monthly patients demand at Orthopaedic clinic of ARIMA (1, 0, 0) model

Period (months)	Actual (Patients)	Forecast	Error	Error%	Absolute Error%
Jan 2018	271	294.16418	-23.164	-8.548	8.548
Feb 2018	270	296.66908	-26.669	-9.877	9.877
Mar 2018	314	297.88409	16.116	5.132	5.132
Apr2018	318	298.47344	19.527	6.141	6.141
May 2018	364	298.75930	65.241	17.923	17.923
Jun 2018	369	298.89796	70.102	18.998	18.998
July 2018		298.96521	**MAPE**		**11.103**
Aug 2018		298.99784			
Sept 2018		299.01366			
Oct 2018		299.02134			
Nov 2018		299.02506			
Dec 2018		299.02687			
Jan 2019		299.02774			
Feb 2019		299.02817			
Mar 2019		299.02837			
Apr 2019		299.02847			
May 2019		299.02852			
Jun 2019		299.02854			

Table 2 shows the actual monthly patients demand of Orthopaedic clinic from January 2018 to June 2018, forecasted patients demand from July 2018 to June 2019, and corresponding errors of forecasting for the period from January 2018 to June 2018. From table 2 it appears that the absolute percentage errors are high for May 2018 and June 2018. The forecast indicates increasing trend and moving toward constant of monthly orthopedic patients demand, being the lowest is 294 and the highest is 299 patients could receive treatment per month. The table is also presented that the ARIMA (1, 0, 0) model yielded a MAPE of 11.103 percent.

5. Discussion and conclusion

This study considers a number of ARIMA models during estimating and forecasting of patients demand at orthopedic clinics. The models were developed based on monthly time series data. Finally the best model was selected. Still, it is very essential that forecast needs to be updated as and when more data becomes available. Obviously, the forecasting models act as useful components of the healthcare system. However, the model should not be judged totally based on technical criteria such as MAPE, but the application of the model in real situation such as ability to predict patient's present in the clinic practically accurate should be given due consideration.

The ARIMA models are a useful tool for analyzing time series data and then make use of the model in forecasting. Forecasting offers the potential for improving planning in hospital service. The ARIMA

(1,0,0) model could be very useful for the Orthopedic Surgeon. These models are also useful for the managers and staff that engaged in resource planning. Nurse Manager could reliably anticipate the number of beds required in the near future and prepare in advance for patients admission and staff as required. That is, staffing and scheduling adequate numbers of staff to meet predicted patient care needs. Thus, providing effective and efficient nursing care. And elective operation can be performed as schedule. Therefore, hospital beds are fully utilized most of the time.

Reference

[1]   Peterson, C. A. (2008). Healthcare growth, *Malaysia Business*, May 1, p.4

[2]   Ng, S. (2008). Here's to our health, *Malaysian Business*, Nov. 16, p. 29

[3]   Ab Rahman N, Sivasampu S, Noh KM, Khoo EM. Health profiles of foreigners attending primary care clinics in Malaysia. *BMC health services research*. 2016 Dec;**16**(1):197.

[4]   Seow AN, Choong YO, Moorthy K, Chan LM. Intention to visit Malaysia for medical tourism using the antecedents of Theory of Planned Behaviour: A predictive model. *International Journal of Tourism Research*. 2017 May;**19**(3):383-93.

[5]   Kadri F, Harrou F, Chaabane S, Tahon C. Time series modelling and forecasting of emergency department overcrowding. *Journal of medical systems*. 2014 Sep 1;**38**(9):107.

[6]   Jiang S, Yang C, Guo J, Ding Z. ARIMA forecasting of China's coal consumption, price and investment by 2030. Energy Sources, Part B: Economics, Planning, and Policy. 2018 Mar 4;13(3):190-5.

[7]   Kim K, Lee C, O'Leary K, Rosenauer S, Mehrotra S. Predicting patient volumes in hospital medicine: A comparative study of different time series forecasting methods. Northwestern University, Illinois, USA, *Scientific Report*. 2014 Jan 23.

[8]   Boutsioli Z. Measuring unexpected hospital demand: the application of a univariate model to public hospitals in Greece. *Hospital topics*. 2009 Nov 10;**87**(4):14-21.

[9]   Navares R, Díaz J, Linares C, Aznarte JL. Comparing ARIMA and computational intelligence methods to forecast daily hospital admissions due to circulatory and respiratory causes in Madrid. *Stochastic environmental research and risk assessment*. 2018 Oct 1;**32**(10):2849-59.

[10]   Reis BY, Mandl KD. Time series modeling for syndromic surveillance. *BMC Medical Informatics and Decision Making*. 2003 Dec;**3**(1):2.

[11]   Jones SA, Joy MP, Pearson JO. Forecasting demand of emergency care. *Health care management science*. 2002 Nov 1;**5**(4):297-305.

[12]   Champion R, Kinsman LD, Lee GA, Masman KA, May EA, Mills TM, Taylor MD, Thomas PR, Williams RJ. Forecasting emergency department presentations. *Australian Health Review*. 2007;**31**(1):83-90.

[13]   Abraham G, Byrnes GB, Bain CA. Short-term forecasting of emergency inpatient flow. *IEEE Transactions on Information Technology in Biomedicine*. 2009 Feb 24;**13**(3):380-8.

[14]   Promprou S, Jaroensutasinee M, Jaroensutasinee K. Forecasting Dengue Haemorrhagic Fever Cases in Southern Thailand using ARIMA Models.

[15]   Li Q, Guo NN, Han ZY, Zhang YB, Qi SX, Xu YG, Wei YM, Han X, Liu YY. Application of an autoregressive integrated moving average model for predicting the incidence of hemorrhagic fever with renal syndrome. *The American journal of tropical medicine and hygiene*. 2012 Aug 1;**87**(2):364-70.

[16]   Permanasari AE, Hidayah I, Bustoni IA. SARIMA (Seasonal ARIMA) implementation on time series to forecast the number of Malaria incidence. In2013 *International Conference on Information Technology and Electrical Engineering (ICITEE)* 2013 Oct 7 (pp. 203-207). IEEE.

[17]   Bowerman BL, O'Connell RT, Koehler AB. *Forecasting, time series, and regression: An applied approach*. Thomson Brooks/Cole; 2005.

[18]   Wang GC. A guide to Box-Jenkins modeling. *The Journal of Business Forecasting*. 2008 Apr 1;**27**(1):19.

[19]   Anggraeni W, Aristiani L. Using Google Trend data in forecasting number of dengue fever cases with ARIMAX method case study: Surabaya, Indonesia. In2016 *International Conference*

on *Information & Communication Technology and Systems (ICTS)* 2016 Oct 12 (pp. 114-118). IEEE

[20] Chin L, Fan GZ. Autoregressive analysis of Singapore's private residential prices. *Property Management*. 2005 Sep 1;**23**(4):257-70.

[21] Wilson JH, Keating B. *Business forecasting with ForecastX*. McGraw-Hill; 2009.