Journal of Sciences and Management Research 2600-738X

USING AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA) MODELS TO PREDICT THE NUMBER OF PATIENT ADMISSION IN PAEDIATRIC CLINIC AT A PRIVATE HOSPITAL IN KUANTAN.

Bahari Mohamed
Faculty of Business and Management
Widad University College

Meriati Mohamad
Kulliyyah of Nursing
International Islamic University, Malaysia

ABSTRACT

This study present autoregressive integrated moving average (ARIMA) models to forecast monthly patient demand for Paediatric clinic at a private hospital in Kuantan. The ARIMA model developed hold potential for providing operational decision support in the hospital. The forecasting success attained for the Paediatric clinic could aid managers to make capacity and advance planning in the wards and hospital. The ARIMA model was developed from time series data routinely-collected at Paediatric clinic. The study evaluated patient demand at Paediatric clinic by using time series data collected from year 2012 until year 2017. Analyses of time series data of Paediatric clinic produce ARIMA (2, 0, 2) model of monthly data. The ARIMA (2, 0, 2) give rise to MAPE of 11.988 percent respectively, therefore ARIMA (2, 0, 2) model was selected for modelling and forecasting paediatric patient demand based on the lowest MAPE values. The out of sample forecast by using ARIMA (2, 0, 2) model indicated a fluctuation of monthly paediatric patients demand, being the lowest was 325 and the highest was 400 patients that could receive treatment from the clinic in a month. The forecasting models then could be extended to other clinics.
Keywords: Hospital, Paediatric clinic, Forecasting, ARIMA model.

1. BACKGROUND

The effective management of hospital has become increasingly important as Malaysia economic and society continue to develop. Hospitals face many challenges as our domestic population aging, diseases pandemic (e.g. H1N1), and cost of operating and maintenance increasing. The government runs around 145 public hospitals which are financed by taxes and other public revenues and these facilities are supported by more than 217 large private healthcare service providers, as well as around 6442 medical clinics (Inside Malaysia, 2012). The healthcare sector in Malaysia has been expanding at 8 percent to 12 percent annually over the past few years (James, *et al.*, 2009). Accordingly, the private healthcare provider is expected to grow further due to private insurance benefits and rising per-capital income

(Peterson, 2008). These private hospitals are normally equipped with the latest diagnostic and imaging facilities. However, majority of private hospitals facilities are in urban areas.

The ageing of the population a growing affluent class in Malaysia, the range of medical insurance products available and health-conscious society provides a strong market growth for private healthcare providers (Ng, 2008). Coupled with the awareness of cardiovascular and hypertension as the number one killer disease in Malaysia, the demand for private healthcare service is growing tremendously (Inside Malaysia, 2012).

Another factor that boosts the expansion of private healthcare service is medical tourism. Malaysia has been rated as the world's third best medical tourism destination. Malaysia attracts medical tourists because the availability of good healthcare facilities at low cost (Chia, 2009) and together with its favorable exchanges rate, political and economic stability and high rate of literacy. The Association of Private Hospitals of Malaysia (APHM) has projected that number of foreigners seeking treatment in Malaysia for 2009 is expected to be in the region of 625,000, compared with the estimated 501,000 in 2008. The association had previously forecast for the medical tourism sector to contribute around RM580 million by treating more than 849,000 medical tourists by 2010 (Chong, 2009).

Therefore, to be able to plan effectively, it is vital for any organization to anticipate future. Hence, the ability of a private hospital to predict demand is very valuable. Forecasting the demand of hospital services can have a significant effect on both customer service and financial result (Kadri *et al.*, 2014). Thus, accurate prediction would facilitate, at a micro level for example scheduling nursing and support personnel, and at macro level for financial and strategic planning for the hospital. Forecasting also can be a great help to healthcare service providers to manage capacity, thus could improve efficiency, reduce costs and increase profitability.

The purpose of this study is to develop univariate time series models for the monthly volume of patients attending paediatric clinic at a private hospital in Kuantan. Thus, the results can be used for forecasting the demand for paediatric treatments in the hospital at least in the short term. Accordingly, the forecasting models can be a great help to the hospital's management in two important ways. First, staff can be scheduled in accordance with peak demand in the clinics or the hospital as a whole. For example, if patient arrival for treatment tends to increase at a particular time in a week, additional staff can be regularly scheduled to handle this demand. Second, regular demand can be predicted months in advance. Thus, if the demand increases beyond the initially predicted, additional staff strength is required.

The hospital may have three types of patients, that is, elective, urgent and emergency patients. As such, the forecasting results and approach in this study could be applied to improve resource allocation and strategic planning. The forecasting results can be applied to staffing schedule, patients waiting time improvement (especially, schedule elective patients) and forecasting beds occupancy.

2. PROBLEM TO BE SOLVED

Variations in demand for a healthcare service can occur unpredictably. The steady number of patients increase begins to put pressure on the available workforce and strain facility capacity of a hospital. This capacity crisis resulted in significant operation bottlenecks, including an increase in patients waiting to be admitted into wards, and long wait time for laboratory, radiology, and other diagnostic services. The consequences of this situation will affect the quality of treatment and prognosis by medical staff that is overloaded thus leading to decrease in job satisfaction (Kadri *et al*., 2014). If this variability can be reduced, then the hospital can achieve higher efficiency. Demand forecasting can help hospital management prepare for these variations and avoid unintentional mistake. This study attempts to develop parsimony and effective models that could predict the numbers of patients present monthly, therefore, could assist the hospital's management for a better planning.

2.1 Time series and hospital admissions

Forecasting patients receiving treatment at hospital has been an important research topic for decades. Many previous studies have been focused on predicting hospital demand capacity using the time-series models (e.g. Jones *et al*., 2002), network flow models (e.g. Akcali, *et al*., 2006), logistic regression models (e.g. Littig and Isken, 2007), and queuing model (e.g. Kao and Tung, 1981), etc. Forecasting process must consider the relationships among individual variables and the model must be validated as a whole to ensure capability of the model's forecast. Thus, combining right data with right model is very essential to obtain the better results (Jain, 2005/2006).

Time-series analysis has been applied previously to health care and particularly to hospital demand capacity. A time series is a sequence of historical data recorded over time, with a consistency in the activity and the method of the measurement (Wang, 2008). The main characteristic of time series modelling is that it only takes into consideration the relationship between the historical data at time $t(y_t)$ from the past observations $y_1, y_2, ...y_n$ (Bowerman *et al*., 2005). The advantage of time-series analysis is simplicity and effectiveness, and attractive for practical applications (Kadri *et al*., 2014). The finding of relevant researches is briefly described below.

Reis and Mandl (2003) built autoregressive integrated moving average (ARIMA) models to forecast overall visits and respiratory-related visit at paediatric emergency department in Children's Hospital Boston. The main objective of this study is to develop automated surveillance systems that can detect abnormality in disease patterns as an early signal of a bioterrorist attack. The results showed that overall visits and respiratory visits are best fitted by ARIMA(2,0,1) and ARIMA(1,0,1) with the mean absolute percentage error (MAPE) of overall visits model is 9.37% and 27.54% respectively. The ARIMA model of overall visits is able to forecast 7 day-long abnormal visit patterns. Thus, the model can provide a foundation for real-time surveillance and bioterrorism detection.

Jones *et al*. (2002) described several ARIMA models of daily hospital beds occupancy from emergency admissions and number of emergency admissions at Bromley Hospitals NHS Trust and developed models that are able to forecast beds occupancy with good accuracy. Accordingly, Abraham *et al*. (2009) developed several ARIMA models to forecast daily inpatient admissions and occupancy in Royal Melbourne Hospital, Victoria, Australia. They found that the models are able to forecast emergency occupancy up to seven days ahead with reasonable accuracy but emergency admissions are unpredictable. ARIMA model also performed better in forecasting cholera cases in Beira, Mozambique (Bergh *et al*., 2008).

The time series analysis is also a useful tool for forecasting emergency department (ED) demand. Kadri *et al*. (2014) used ARIMA models to forecast emergency patient arrivals to the ED of hospitals in France. They found that the ARIMA models are suitable to predict the number of ED visits. The ARIMA models also suitable to forecast dengue haemorrhagic fever (DHF). Promprou *et al*. (2006) established ARIMA (1, 0, 1) model and found that the models are very useful in forecasting the DHF in Southern Thailand.

Chang *et al*. (2004) applied ARIMA models to determine whether the severe acute respiratory syndrome (SARS) epidemic was significantly connected with the rate of medical service utilization in Taiwan. They found that the models are able to detect the changes in medical service utilization. Meanwhile, Earnest *et al*. (2005) successfully applied ARIMA (1, 0, 3) model to forecast the number of bed occupied in a tertiary hospital during the SARS outbreak in Singapore. Accordingly, ARIMA model applied to historical hemorrhagic fever with renal syndrome (HFRS) data is accurately forecasting the HRFS incidence in China (Liu *et al*., 2011). Thus, accurate forecasting of the hospital or clinic admission is possible using ARIMA model.

3. METHODS

3.1 Data

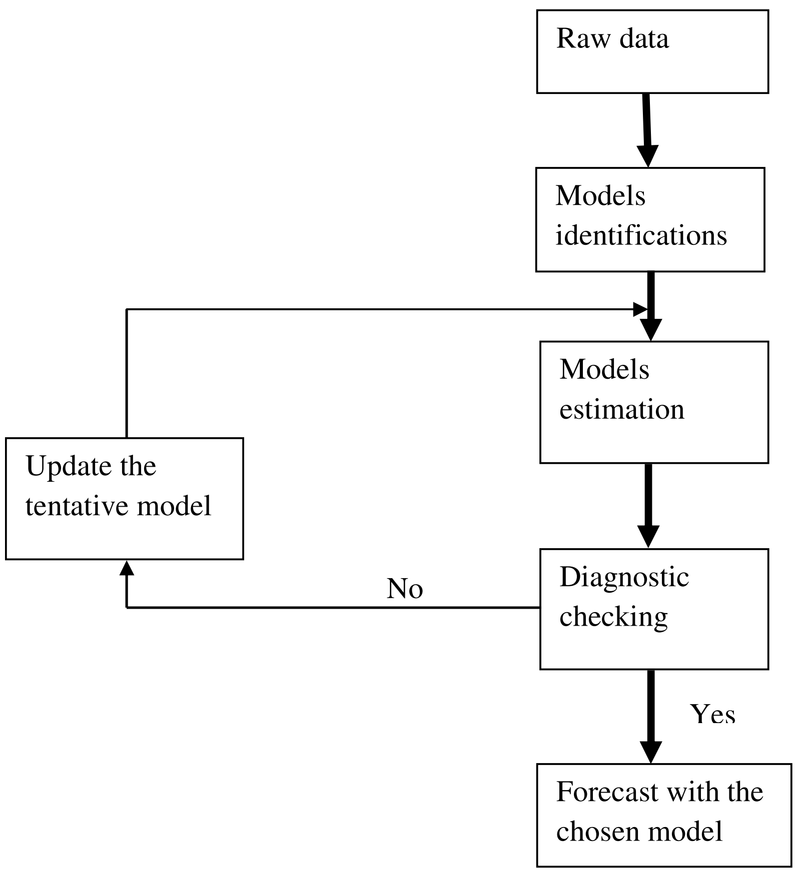
Data from the Paediatric clinic of the hospital were used as the basis for this study. The monthly data from January of 2012 to December of 2017 were used as training data on which the forecasting models were trained (estimating model parameters) base on the count of patients receiving treatments at the paediatric clinics. The monthly data from January to June of the year 2017 were kept as hold out sample to verify the accuracy of the forecasting models. The data were related to patients treated by the consultant doctors at the clinics only.
The data for this study were obtained from the admission record books of the clinic. The data include the date of the patients received treatment from the clinic.

3.2 Method of analysis

The data from the clinics can be considered as a collection of observations made in 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 that become popularly applied in empirical studies among researchers in health care industry (e.g., Channouf *et al*., 2006; Reis and Mandl, 2003).

The Box-Jenkins approach to time series model building is a method of finding a suitable ARIMA model that adequately represents the given set of data. The method is an iterative, which may go through the process many times before arriving at suitable model. That is, the methodology is designed to arrive at ARIMA models through a three interactive stages procedure based on models identifications, models estimation, and diagnostic checking or model validation, and then utilize the models for forecasting (Bowerman *et al*., 2005; Wang, 2008; Wilson and Keating, 2009).

Figure 1: The Box-Jenkins Methodology flowchart



Source: Adapted from Wilson and Keating, 2009

Box-Jenkins methodology

Figure 1 shows Box-Jenkins process flowchart. The raw data is examined to determine either the time series stationary or non-stationary. In Box-Jenkins methodology, the non-stationary series must be transformed to stationary series to model. The first step is model identification. In identification process, it is very important to exam the behaviour of the autocorrelation function (ACF) and the partial autocorrelation function (PACF) from the given data of time series. ACF plainly mean as measure of significance of correlations between the current observation and the past observations, and to ascertain how far back in time are they correlated, i.e., how many time lags they correlated. And PACF values are the coefficients of a linear regression of the time series using its lagged values as independent variables. The values of ACF and PACF both will fall between – 1 and + 1 if the time series is stationary (Wang, 2008). The ACF and PACF provide a useful measure of the degree of dependence among the values of a time series at different times. As such, they play an important role in forecasting future values of the series in terms of past and present values. A pair of correlograms (ACF and PACF) is used to indentify an appropriate ARIMA model for the given set of data. The correlogram is a graph showing the time series ACF or PACF values against the lag k. Several important information of the given time series can be acquired by examining a correlogram. For example, the times series observed could be considered stationary if the ACF of the time series "cut of fairly quickly" or "dies down fairly quickly", and then the order of ARIMA model can be indentified i.e., the values of p and q can be identified by observing the pair of correlograms. The general rule in this identification process as follows:

(i) If ACF abruptly stop after q spikes, then the appropriate model is MA(q) type.
(ii) If PACF abruptly stop after p spikes, then the appropriate model is an AR(p) type.
(iii) If neither function falls off abruptly, but both die down extremely slow, the appropriate model is an ARMA (p,q) type (Bowerman *et al*., 2005; Wilson and Keating, 2009).

The second step is model estimation. Specifically, parameters of the model are tentatively identified. Box-Jenkins methodology required that ARIMA model must satisfy the condition of stationary and invertible. An AR model of order p, with parameters $\phi_1.......\phi_p$ is stationary if and only if the combination value of parameters less than 1. The MA model of order q, with parameters $\theta_1.........\theta_q$ is invertible if and only if the value of parameters is less than 1. For ARIMA(p,d,q) model, the AR(p) must be stationary and MA(q) must be invertible (Bowerman *et al*., 2005; Wang, 2008). In this study SPSS 12.0 package is used in estimation process.

The third step is diagnostic checking process or model validation. That is to identify the correct parameters of the model has been chosen. If the model is "good", then the residuals are expected to be random and close to zero. There are several ways of checking if a model is good. The common approach is to examine the residues. In this step, ACF behaviour of the residuals produced by estimation process is examined. For a good model, the residual time series should be close to an independent and identically distributed (iid) zero-mean white noise (random series with normally and independently distributed). As such, about 95% of the sample autocorrelations should fall between the bound $\pm\, {}^{2}\!/_{\sqrt{n}}$ (further explain in data analysis and modelling section). Instead of checking to see whether each sample autocorrelation coefficient falls within the bounds, it is also possible to carry out what is called Ljung–Box Q statistic test. If correct model is obtained, the residual should be normally distributed and uncorrelated. Therefore, the autocorrelation of the residuals should be small. Thus Ljung-Box Q statistic should be small. The model also would have a smaller standard error and larger *p*-value corresponding to Ljung-Box Q statistic. Repeating the process of estimation is necessary if the diagnostic checking is unsuccessful in obtaining appropriate model. Therefore, the process is loop through many times until the correct parameters of the model identified. The model then can be used in forecasting process.

3.3 The ARIMA Model

The primary objective of time series analysis is to study the dynamic or the mechanism that generates the data. Then, forecast future values of the series base on the analysis (Chin and Fan, 2005; Billings and Yang, 2006). The most common technique of forecasting for time series data is by using autoregressive integrated moving-average (ARIMA) modelling (Jones *et al.*, 2002). ARIMA models were developed by Box and Jenkins (Wilson and Keating, 2009). There are three basic components to an ARIMA model: auto-regression (AR), differencing or integration (I), and moving-average (MA). All three are based on the simple concept of random disturbances or shocks. Between two observations in a series, a disturbance occurs that somehow affects the level of the series. These disturbances can be mathematically described by ARIMA models. Each of the three types of processes has its own characteristic way of responding to a random disturbance. In general form, an ARIMA model is typically expressed as:

ARIMA (*p*,*d*,*q*), where *p* is the order of auto-regression, *d* is the order of differencing (or integration), and *q* is the order of moving-average involved (SPSS Trends 13.0 manual).
The non-seasonal mixed ARIMA (*p,q*) model can be expressed in equation form of

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

or $$Z_t = \sum_{i=1}^{p} \phi_i Z_{t-i} + a_t + \sum_{i=1}^{q} \theta_i a_{t-i}$$

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. ARIMA models are usually formulated with the premise of constant variance in error term (Jones *et al.*, 2002).

Differencing (ARIMA)

First, let discuss the differencing or integration process of ARIMA. The differencing or integration is denoted by "I", the second component of ARIMA. In practice, most of the time series are non-stationary. The characteristics of the observed values of the time series change over time (Pindyck and Rubinfeld, 1991), i.e., the time series experience periods of high volatility followed by periods of relative serenity. 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. 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 symbol "I" denote that the time series has been transformed into a stationary time series.

Therefore, an integrated series can be considered by looking at the changes, or differences, from one observation to the next. Let consider *n* values $y_1, y_2, ...y_n$ of a time series, if the *n* values fluctuate with constant variation, e.g., the difference from one observation to the next is often small. Thus, the time series can be considered stationary. This stationarity, of the differences is highly desirable from a statistical point of view.

For example, the standard form for integrated models (of first differences), or models that need to be differenced, is I(1) or ARIMA(0,1,0), i.e., first difference of the time series values $y_1, y_2, ...y_n$ are $z_t = y_t - y_{t-1}$, where $t = 2, ..., n$. Sometime second differences are required in order to produce stationary time series values; such models are termed I(2) or ARIMA(0,2,0). The second differences of the time series values are $z_t = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2})$ or $z_t = y_t - 2y_{t-1} + y_{t-2}$ for $t = 3,4,......,n$.

Differencing beyond the second or third order is rare. Usually, when a series exhibits such extreme trends, it is not stationary due to a non-constant variance. Applying a log or square root transformation to the series before estimating the model will generally removes the trend from the data, i.e., stabilize the variance (SPSS Trends 13.0 manual; Bowerman *et al.*, 2005).

Auto-regression (ARIMA)

The 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 behaviours may suggest important information regarding its near future dynamics (Chin and Fan, 2005). 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. For example, non-seasonal autoregressive model of order 1 is denoted by notation AR (1) or ARIMA (1, 0, 0) and the process has the functional form of $Z_t = \phi_1 Z_{t-1} + a_t$ or Value Z_t = Coefficient ϕ * Value Z_{t-1} plus disturbance a_t. Where: *Value* Z_t is the value of the series at time *t*. The *Coefficient* ϕ is a value that indicates how strongly each value depends on the preceding value. The sign and magnitude of the coefficient are directly related to the sign and magnitude of the partial autocorrelation at lag 1. When the coefficient is greater than –1 and less than +1, the influence of earlier observations dies out exponentially. The coefficient ϕ is an unknown parameter that must be estimated from sample data. The *Disturbance* or *random shock* a_t is the value that assumed to have been randomly selected from a normal distribution that has mean zero and a variance that is the same for each and every time period *t*. The random shocks $a_1, a_2, a_3 .....$ in different time periods are assumed to be statistically independent of each other.

Conceptually, an autoregressive process is one with a “recall capacity,” in that each value is correlated with all preceding values. In an AR (1) process, the current value is a function of the preceding value, which is a function of the one preceding it, and so on. Thus, effect of shock or disturbance in an autoregressive process is diminishing as time passes. In practical term, AR processes are more useful for modelling longer-term effects (SPSS Trends 13.0 manual; Bowerman *et al*., 2005).

Moving-average (ARIMA)

The 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. For example, non-seasonal moving average model of order 1, denoted by MA (1) or ARIMA (0, 0, 1) has the functional form: $Z_t = a_t - \theta_1 a_{t-1}$ or Value Z_t = disturbance_t - Coefficient * $\text{disturbance}_{t-1.}$ Where: *Value* Z_t is the value of the series at time *t*. *Coefficient* θ is a term that indicates how strongly each value depends on the preceding disturbance terms. The sign and magnitude of the coefficient are directly related to the sign and magnitude of the autocorrelation at lag 1 and *disturbance* is the chance error associated with the series value at time *t*. Each value in a moving-average series is a weighted average of the most recent random disturbances. Thus, in a moving-average process, a disturbance affects the system for a finite number of periods (the order of the moving-average) and then abruptly ceases to affect it. In practical terms, MA processes are more useful for modelling short-term fluctuations (SPSS Trends 13.0 manual; Bowerman *et al*., 2005).

4. DATA ANALYSIS AND MODELING

To fit a time series data into ARIMA models, the data are required to be stationary. Thus, the raw data need to be transformed to stationary first (if required) before next steps of modelling process are performed. In other words, the transformed data or original data have statistical properties that are constant through time, for example the time series has constant mean and variance, has no trend overtime. In order to determine whether the data are stationary or non-stationary, the time series data must first plotted to examine their pattern, i.e., a graph showing the observations against time. To further confirm the stationarity of the time series we can observe the ACF behaviour of the series. If the ACF behaviour shows the series cuts off fairly quickly or dies down fairly quickly then we can consider the time series is stationary. If non-stationary, then the series is further processed to make it stationary. Differencing is an effective way to remove trend and seasonal components in a time series. As the time series is stationary, it could be used for ARIMA modelling purpose. In ARIMA modelling process, inspecting correlogram of ACF and the correlogram of PACF of the time series is very essential. Notice that in these pair of ACF and PACF correlograms the lines that parallel to the x-axis are representing the error bounds for the data. The lines are determined based on $\pm {}^{2}\!/\!\sqrt{n}$, where n represents the number of data. If the value of the ACF and PACF lie within these lines, then the values are considered not significantly different from zero. In other words, the plot shows approximate 95% confidence limits at this value, and the observed values of ACF and PACF which fall outside these limits are considered different from zero at the 5% level. Thus, the nonzero lags (spikes) could be identified as tentative order of ARIMA models (Bowerman *et al*., 2005; Wilson and Keating, 2009). Parameters for these ARIMA models were estimated by the SPSS 12.0 package. Melard's algorithm was used for estimating ARIMA models parameters. Ten iterations were specified with default tolerance of 0.001.

Figure 2: Plot of monthly patient demand at paediatric clinic

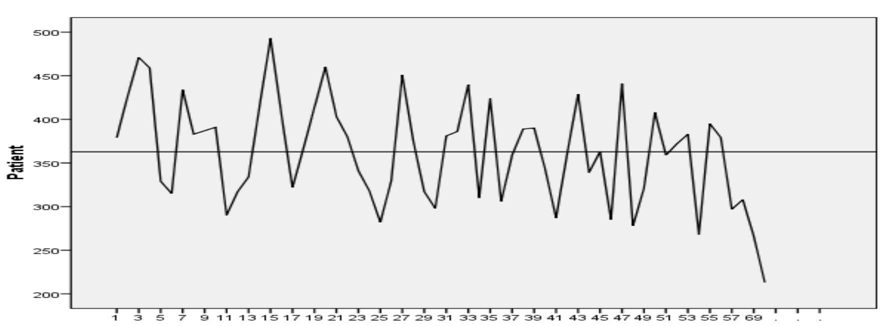


Figure 3: ACF of monthly patient demand at paediatric clinic

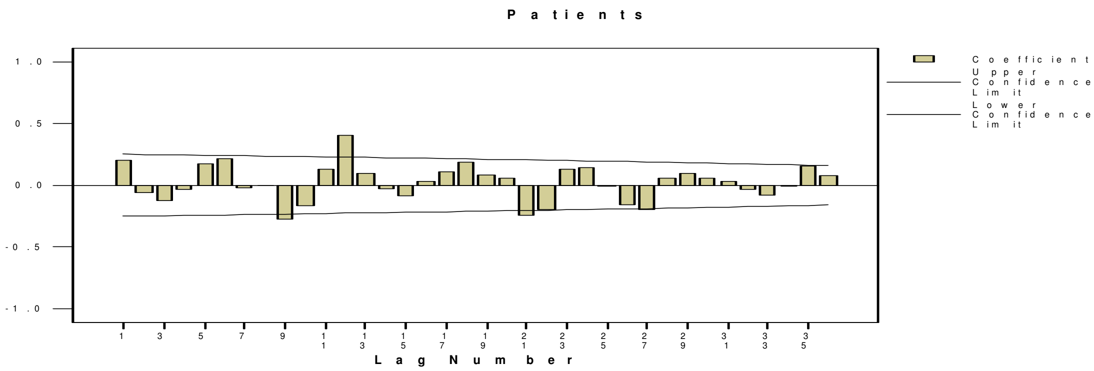


Figure 4: PACF of monthly patient demand at paediatric clinic

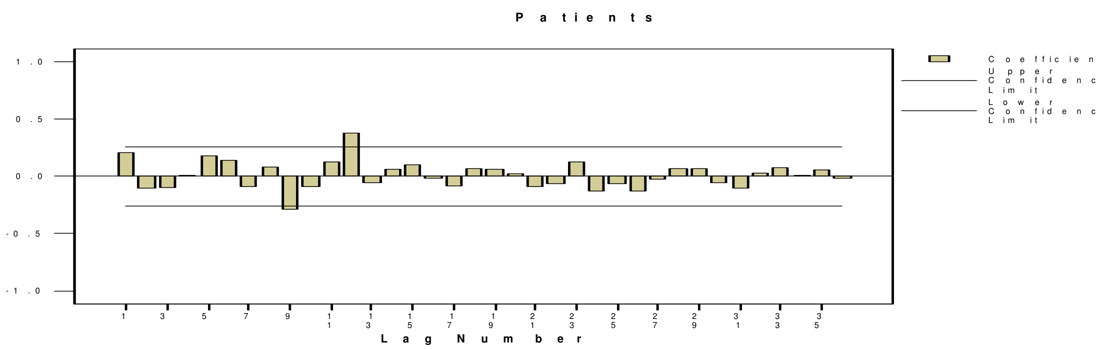


Figure 2 shows the time plot of the time series. Numbers of patients received treatment were relatively stable between 2012 and 2016, but decrease toward the end of 2016 was evident. Average number of patients visiting the paediatric clinic is 362.7667 or approximately 363 patients per month. The time series shows that there is no seasonality and trend in the data as evident from plot in figure 2. Accordingly, from the graphs, the time series can be considered stationary. Furthermore, the ACF behaviour (figure 3) of the series shows that the series dies down fairly quickly. As such, the series was considered stationary. Since the time series was stationary without being difference or integration, the optimal degree of integration was determined to be zero (d = 0). By means of observing the pair of ACF and PACF correlograms (Figure 3 and 4), tentative ARIMA models were determined. Both the correlogram of ACF and PACF for monthly paediatric demand show the characteristic of dies down. Therefore, the mixed autoregressive moving average of order (p,q) was possible ARIMA model. Accordingly, the order of autoregressive (spikes of PACF) would be AR of 9 or 12 and similarly the order of moving average (spikes of ACF) would be MA of 9, 12, 21, or 27. Consequently, various combination of AR and MA orders were then tested to determine adequate ARIMA (p,d,q) model from the available training data.

Finally, ARIMA (2, 0, 2) model (second order Autoregressive combined with second order Moving Average) was found to generate appropriate results to represent monthly paediatric demand. The following Table 1 shows the estimated parameters for ARIMA (2, 0, 2) model of monthly paediatric demand.

Table 1: Estimated parameters of ARIMA (2, 0, 2) model

Parameter	**CONSTANT**	ϕ_1	ϕ_2	θ_1	θ_2
Estimate	366.71822	.98949	-.99852	.95987	-.96816
Std. Error	6.4846995	.0190798	.0101789	.0916222	.1326717
T-RATIO	56.551305	51.860708	-98.097093	10.476362	-7.297387
APPROX. PR.	.0000000	.0000000	.0000000	.0000000	.0000000

Stationary and Invariability condition of ARIMA (2, 0, 2) model

From Table 1, the model shows that the sum of coefficients of AR (2) (i.e., $\phi_1 + \phi_2$) component was less than 1 and also the sum of coefficients of MA (2) (i.e., $\theta_1 + \theta_2$) component was less than 1. As such, the model was fulfilled the requirement of stationary and invertible condition.

Diagnostic checking of ARIMA (2, 0, 2) model

Figure 5: ACF of residuals of ARIMA (2,0,2) model

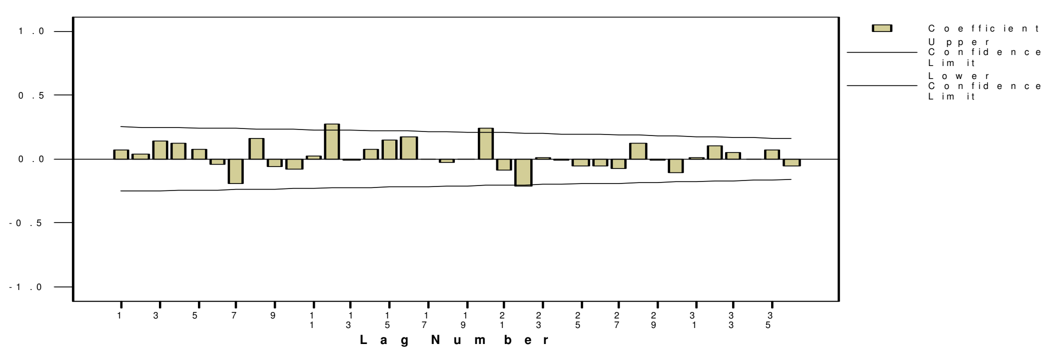


Figure 6: PACF of residuals of ARIMA(2,0,2) model

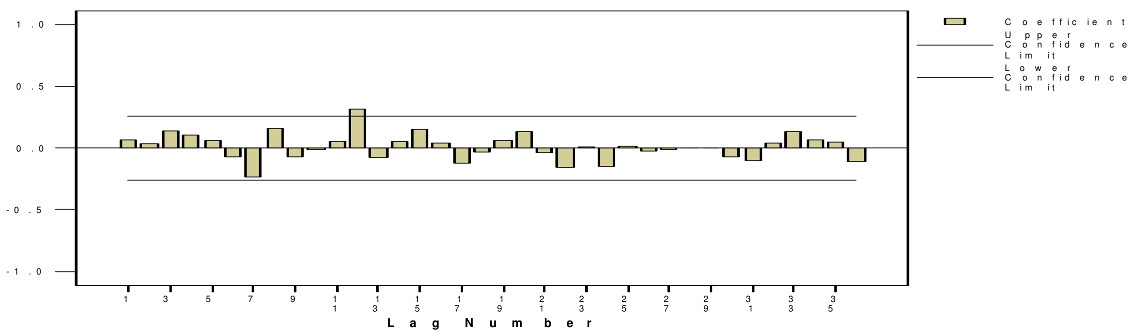


Figure 7: Autocorrelation plot of residuals of ARIMA (2, 0, 2) model

```
 Auto- Stand.

Lag  Corr.   Err. -1   -.75   -.5 -.25    0   .25  .5   .75   1   Box-Ljung  Prob.
                   ¤⇩⇩⇩⇩⇧⇩⇩⇩⇩⇧⇩⇩⇩⇩⇧⇩⇩⇩⇩⇧⇩⇩⇩⇩⇧⇩⇩⇩⇩⇧⇩⇩⇩⇩⇧⇩⇩⇩⇩⇧⇩⇩⇩⇩¤
  1   .068   .126              .    ⇔*    .                       .296   .587
  2   .037   .125              .    ⇔*    .                       .384   .825
  3   .142   .124              .    ⇔***  .                      1.698   .637
  4   .123   .123              .    ⇔**   .                      2.698   .610
  5   .079   .122              .    ⇔**   .                      3.122   .681
  6  -.039   .120              .   *⇔     .                      3.224   .780
  7  -.193   .119              .****⇔     .                      5.848   .558
  8   .159   .118              .    ⇔***  .                      7.667   .467
  9  -.060   .117              .   *⇔     .                      7.931   .541
 10  -.079   .116              .  **⇔     .                      8.393   .590
 11   .025   .115              .    *    .                       8.439   .673
 12   .271   .114              .    ⇔*****                      14.115   .293
 13  -.008   .112               .   *   .                       14.119   .365
 14   .078   .111               .   ⇔** .                       14.610   .405
 15   .151   .110               .   ⇔***.                       16.490   .350
 16   .173   .109               .   ⇔***.                       19.013   .268
 17   .001   .108               .   *   .                       19.013   .328
 18  -.026   .106               .  *⇔   .                       19.075   .387
 19  -.005   .105               .   *   .                       19.077   .452
 20   .239   .104               .   ⇔***.*                      24.376   .226
 21  -.085   .102               . **⇔   .                       25.062   .244
 22  -.209   .101               ****⇔   .                       29.328   .136
 23   .014   .100               .   *   .                       29.348   .169
 24  -.012   .098               .   *   .                       29.362   .207
```

Residual analysis of the ARIMA (2,0,2) model was carried out to verify the model. Figure 5 and 6 shows the ACF and PACF correlogram of the residuals respectively and Figure 7 show autocorrelation plots of residuals. By observing Figure 5 and 6, we could see that ACF and PACF of the residuals roughly lie within the boundary lines except for a few lags (spikes), indicating that overall the residual time series approximates to a zero mean white noise behaviour. From Figure 7, we could observe that the autocorrelation of the residuals were small. Thus Ljung-Box Q statistic values were small. The model also had a smaller standard error and larger p-value (i.e., $> .05$) corresponding to Ljung-Box Q statistic. As such, we could conclude that the model was adequate.

4.1 Forecasting

The main objective in this study is forecasting. Forecasting is to predict future values of a time series. Time series models developed are used to forecast future demand on each 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 employ 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 to the training data (Reis and Mandl, 2003; Wilson and Keating, 2009). The mean absolute percentage error is calculated as: $MAPE = \frac{\sum |(A_t - F_t)/A_t|}{n}$.

Where, A_t = Actual value in period t, F_t = Forecast value in period t, and n = Number of periods used in the calculation. The accuracy of forecasts of patients demand at the clinics was described and presenting in Tables 2. Table 2 shows the actual and forecasted patients demand and MAPE.

Table 2: Forecast monthly patients demand at paediatric clinic from ARIMA (2,0,2) model

Period (month)	**Actual (patients)**	**Forecast**	**Error**	**Error%**	**Absolute Error%**
Jan 2017	337	372.035	-35.035	-10.396	10.396
Feb 2017	417	400.142	16.858	4.042	4.042
Mar 2017	278	390.293	-112.293	-40.393	40.393
Apr 2017	388	352.583	35.417	9.128	9.128
May 2017	343	325.323	17.677	5.154	5.154
Jun 2017	327	336.119	-9.119	-2.813	2.813
July 2017		373.916	**MAPE**		**11.988**
Aug 2017		400.316			
Sept 2017		388.589			
Oct 2017		350.733			
Nov 2017		325.204			
Dec 2017		337.846			
Jan 2018		375.733			
Feb 2018		400.381			
Mar 2018		386.841			
Apr 2018		348.950			
May 2018		325.193			
Jun 2018		339.614			
July 2018		377.482			
Aug 2018		400.339			

Sept 2018		385.055
Oct 2018		347.236
Nov 2018		325.288
Dec 2018		341.417

Table 2 shows actual patients demand from January 2017 to June 2017, forecasted patients demand from July 2017 to December 2018, and corresponding errors of forecasting for the period from January 2017 to June 2017. From Table 2 it was observed that the absolute percentage errors high in March 2017. The forecast indicate a fluctuation of monthly patients demand at paediatric clinic, being the lowest was 325 and the highest was 400 patients receive treatment per month. The Table 2 also present that the ARIMA (2, 0, 2) model yielded a MAPE of 11.988 percent.

5. CONCLUSION AND RECOMMENDATIONS

This study has considered number of ARIMA models during estimating and forecasting of patients demand at paediatric clinic. The models were developed base on monthly time series data. Finally the best model was selected. Still, it is very essential that forecast need to be updated as and when more data becomes available. Obviously, the forecasting models act as useful components to the healthcare system particularly to the Paediatric Clinic. However, the model should not be judged totally base on technical criteria such as MAPE, but the application of the model in real situation such as able to predict patients presentation in the clinic practically accurate should be given due consideration.

5.1 Model selection

Model was selected base on MAPE value. The smallest the MAPE is the better. Evaluating the Table 1 and 2 we could determine the best model to represent Paediatric clinic meant for forecasting patients demand. MAPE for ARIMA (2,0,2) was 11.988 percent. As such, monthly paediatric patients demand is better modelled as an ARIMA (2,0,2) model because of smaller MAPE value. This means that paediatric patient demand may be characterized as the combination of autoregressive and moving average.

5.2 Scope and limitation of study

There are some limitations of time series analysis models and ARIMA model in particular. ARIMA model very much depend on stationary of time series, which some time difficult to ascertain. ARIMA model only capture short-range dependency and the model reflect the behaviour of the particular clinics or hospitals. However, such model could be applied to other clinics or hospitals with little adjustment.

The data were recorded by the hospital staff, as such the data may contain bias that very difficult to identify. The lack of control over the data recorded may increase the risk of misinterpreting of finding and therefore drawing invalid conclusions. Furthermore, the time series analysis does not consider any intervention influence (e.g., diseases pandemic) during data recording process.

5.3 Discussion and conclusion

The ARIMA models are a useful tool for analysing time series data and then make use of the model in forecasting. Forecasting offers the potential for improve planning in healthcare service provider. The ARIMA (2,0,2) model could be very useful for the paediatrician. These models are also useful for the hospital's managers and staff that engaged in resource planning. Nurse Manager could reliably anticipate the number of beds required in near future and prepare in advance for patients admission and staff as required. Thus, provide effective and efficient nursing care. And elective operation can be performed as schedule and hospital beds are fully utilized most the time. Therefore, modelling and forecasting the patient volume provides useful information for hospital and health care authorities. The information is useful in allocating resources, scheduling staff, and planning future expansion.

REFERENCES

Abraham, G., Byrnes, G. B., Bain, C. A. (2009). Short-term forecasting of emergency inpatient flow, *IEEE Transactions on Information Technology in Biomedicine, 13*(3), 380-388.

Akcali, E., Cote, M. J., Lin, C. (2006). A network flow approach to optimizing hospital bed capacity decisions, *Health Care Management Science, 9*, 391-404.

Bergh, F. V. D., Holloway, J. P., Pienaar, M., Koen, R., Elphinstone, C. D., Woodborne, S. (2008). A comparison of various modelling approaches applied to Cholera case data, *ORION 24*(1), 17-36.

Billings, D., Yang, J. S. (2006). Application of the ARIMA models to urban roadway travel time prediction – A case study, *IEEE International conference on System, Man, and Cybernetics 8*(11), 2529-2534.

Bowerman, B. L., O' Connell, R. T., Koehler, A. B. (2005). *Forecasting, Time Series, and Regression: An applied approach*, Thomson Learning, Inc.

Chang, H. J., Huang, N., Lee, C. H., Hsu, Y. J., et al. (2004). The impact of the SARS epidemic on the utilization of medical services: SARS and the fear of SARS, *American Journal of Public Health 94*(4), 562-564.

Channouf, N., L'Ecuyer, P., Ingolfsson, A., Avramidis, A. N. (2007). The application of forecasting techniques to modelling emergency medical system calls in Calgary, Alberta, *Health Care Management Science 10*, 25-45.

Chia, S. (2009). How sustainable is our healthcare system? *Starbiz*, Monday 19 October.

Chin, L., Fan, G. Z. (2005). Autoregressive analysis of Singapore's private residential prices, *Property Management 23*(3/4), 257-270.

Chong, Y. (2009). Health is wealth, *Malaysian Business*, Mar. 1, p. 20

Earnest, A., Chen, M.I., Ng, D. and Sin, L.Y. (2005), Using autoregressive integrated moving average (ARIMA) models to predict and monitor the number of beds occupied during a SARS outbreak in a tertiary hospital in Singapore, *BMC Health Services Research*, 5:36.

Inside Malaysia, (2012), Healthcare -economic transformation programme - PEMANDU**,** etp.pemandu.gov.my/upload/Inside%20Investor%20-%20Healthcare.pdf (online), downloaded January 2018.

Jain, C. L. (2005/2006). Benchmarking forecasting models, *The Journal of Business Forecasting, 24*(4), 9-12.

James, S., Chong, Y., Abdullah, A. (2009). An eye on 2009 - Sectors to watch, *Malaysian Business*, Jan 1, p. 29

Jones, S. A., Joy, M. K., Pearson, J. (2002). Forecasting demand of emergency care, *Health Care Management Science 5*(40), 297-305.

Kadri, F., Harrou, F., Chaabane, S., and Tahon, C. (2014), Time series modelling and forecasting of emergency department overcrowding, *Journal of Medical Systems*, 38, 107-126.

Kao, E. P. C., Tung, G. G. (1981). Bed allocation in public health care delivery system, *Management Science, 27*(5), 507-520.

Littig, S. J., Isken, M. W. (2007). Short term hospital occupancy prediction, *Health Care Managment Scienc*e, 10, 47-66.

Liu, Q., Liu, X., Jiang, B. and Yang, W. (2011). Forecasting incidence of hemorrhagic fever with renal syndrome in China using ARIMA model, *BMC Infectious Diseases*, 11: 218.

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

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

Pindyck, R. S., Rubinfeld, D. L. (1991). *Econometric model & economic forecasts*, 3rd Edition, McGraw-Hill

Promprou, S., Jaroensutasinee, M., Jaroensutasinee, K. (2006). Forecasting dengue haemorrhagic fever cases in Southern Thailand using ARIMA models, *Dengue Bulletin 30*, 99-106.

Reis, B. Y., Mandl, K. D. (2003). Time series modelling for syndromic surveillance, *BMC Medical Informatics and Decision Making 3*(2), 1-11.

SPSS Trends 13.0 manual, java.mcmaster.ca/**SPSS**.manual/**SPSS**%20Trends%2013.0.pdf (online), download Jan12, 2018.

Wang, G. C. S. (2008). A guide to Box-Jenkins modelling, *The Journal of Business Forecast, spring,* 19-28.

Wilson, J. H., Keating, B. (2009). *Business forecasting with Forecast X^{TM}*, 6th edition, McGraw Hill