

# Forecasting for the Imported Weight of Equipment to Cargo of Sulaimani International Airport

Handreen Abdulla Tahir<sup>1</sup>, Renas Abubaker Ahmed<sup>2</sup>, Aras Jalal Mhamad<sup>3,4</sup>

<sup>1</sup>Department of Statistics and computer, College of Commerce, University of Sulaimani, Sulaymaniyah City, Iraq. <sup>2,3</sup>Department of Statistic and Informatics, College of Administration and Economics-University of Sulaimani, Sulaymaniyah City, Iraq. <sup>4</sup>Accounting Department, College of Administration and Economics, University of Human Development,

Sulaymaniyah City, Iraq.

aras.mhamd@univsul.edu.iq<sup>3,4</sup>

# Abstract

The problem of a weight of imported equipment in the airport with their effects on economic situation is one of the most important problems that challenges faced in the airports of the region especially international Sulaymaniyah airport. This study aims to analyze the time series of weight of imported equipment in international Sulaymaniyah airport for the period between (Jan; 2010 to Nov; 2017) using the modern style in analyzing the time series which is (Box-Jenkins) method for the accuracy and flexibility it has in addition to its high efficiency in analyzing the time series. In this study, we are interested in forecasting the weight of imported equipment of international Sulaymaniyah airport using Box- Jenkins method. The study found that the fit and efficient model is shown according to smallest measurements (AIC, RMSE, MAPE and MAE) is the seasonal model of lag 4 (SARIMA (1,1,1)x(2,0,0)4). According to the results of SARIMA (1,1,1)x(2,0,0)4, the amounts of the weight of imported equipment of international Sulaymaniyah airport have been forecasted for the period from Nov; 2017 to Oct; 2018 (the forecasting was done for 12 months). Finally, we recommend decision makers and the interested people to adapt to formulate strategic plan depend mainly on the scientific method in forecasting of the monthly weight of imported equipment since there is a real problem facing international Sulaymaniyah airport throughout the upcoming years.

Keywords: Time Series Analysis, SARIMA model, Forecasting.

الملخص

تعد مشكلة ثقل المعدات المستوردة عبر المطار وتأثيرها على الوضع الاقتصادي من أهم المشكلات التي تواجهها المطارات وخاصة مطار السليمانية الدولي. تهدف هذه الدراسة إلى تحليل السلاسل الزمنية لوزن المعدات المستوردة في مطار السليمانية الدولي للفترة من (كانون الثاني ؛ 2010 إلى تشرين الثاني؛ 2017) باستخدام الأسلوب الحديث في تحليل السلسلة الزمنية و التي هي (بوكس-جنكس) لأن هذه الطريقة لديها – دقة و كفائه العالية في تحليل السلاسل الزمنية. في هذه الدراسة، نحن مهتمون بالتنبؤ بوزن المعدات المستوردة من مطار السليمانية الدولي باستخدام طريقة (بوكس-جنكس)، بعد تحليل الأحصائي في هذه الدراسة تم تحديد النموذج الملائم بالأعتماد على المقايس المقارنة الأحصائية (AIC و RMSE و MAP و MAE) وهي النموذج الموسمي ذي إزاحة الزمنية 4. وفقًا لنتائج 2,0,0)×(SARIMA(1,1,1)، فقد تم التنبؤ بمقدار أوزان المعدات المستوردة من مطار السليمانية الدولي للفترة الزمنية تشرين الثاني ؛ 2017 إلى تشرين الأول ؛ 2018. أخيرًا ، نوصى جهات المعنية والأشخاص المهتمين بالتكيف لصياغة الخطة الإستراتيجية التي تعتمد بشكل أساسي على الطريقة العلمية في التنبؤ بالوزن الشهري للمعدات المستوردة، وانتباه للمشاكل الحقيقية التي قد تواجه المطار السليمانية الدولي خلال السنوات المقبلة.



# يوخته

كيّشەى ھاوردەكردنى كەئوپەل ئەرپىگەى فرۆكەخانەى نيۆدەوئەتى و كارىگەرى ئەسەر بارى ئابوورى يەكيّكە ئەو كيّشانەى كە فرۆكەخانەكانى ناوچەكە رووبەروى دەبنەوە بەتاييەتى فرۆكەخانەى نيۆدەوئەتى سليّمانى، ئامانجى ئەم تويّژينەوەيە شيكاريكردنى برى كيّشى كەل وپەلى ھاووردەيە ئەرپىگەى فرۆكەخانەى نيۆدەوئەتى سليّمانى بەبەكارھيّنانى تيۆرى (بۆكس-جنكس)، ھەئبژاردنى ئەم تيۆريە بۆ شيكاريكردن و خەملائدنى باشترين چەماوە دەگەريّتەوە بۆ وردى و چوستى و بيّخەوشى تيۆريەكە، ئەم تويّژينەوەيەدا تويّژەران گرينگيانداوە بە پيّشبينى كردنى كيشى كەل و پەلى ھاووردەكراو بۆ شارى سليّمانى ئە بەيكارھيّنانى تيۆريەكە، ئەم تويّژينەوەيەدا تويّژەران گرينگيانداوە بە پيّشبينى كردنى كيشى كەل و پەلى ھاووردەكراو بۆ شارى سليّمانى ئە ريّگەى فرۆكەخانەى نيۆدوەئەتى سليّمانى، ئەدواى شيكاريكردن باشترين چەماوە دياريكرا كە بريتيە ئە (1, 1, 1) SARIMA بە پشتبەستن بەچەند پيۆەريكى بەراووردى ئامارى وەك (يزانى گرينگيانداوە بە دياريكرا كە بريتيە ئە (1, 1, 1) كەردى باشترين ئەچەند پيۆەريكى بەراووردى ئامارى وەك (كزانى 2018) دىاريكرا كە بريتيە ئە (1, 1, 1) كەردى بەكرەن يەلەپەت بەچەند پيۆەريكى بەراووردى ئامارى وەك (كرانى 2018) دىاريكرا كە بريتيە ئە (1, 1, 1) كەرلەر بۇ دوانزە مانگى داھاتوو كەئە (سەرما وەرزى 2017) تا گەلا ريّزانى 2018). ئەسەر بنەماى چەماوەى دياريكراو پيڭبينى كرا بۆ دوانزە مانگى داھاتوو كەئە (سەرما وەرزى 2017) تا گەلا ريّزانى 2018). ئەكۆتايى دا ئەسەر بنەماى چەماوەى دياريكراو پيشبينى كرا بۆ دوانزە مانگى داھاتوو كەئە (سەرما وەرزى 2017) تا گەلا ريّزانى 2018). ئەكۆتايى دا ئەمەدى بەرەرى بەرىرەيكان دەكەين كە ھاوردەيدىنى كەل و پەل بەپيتى پيويستى بىت ئە چوارچيۆەي ستانداردى جىھانى وە ھەروھا ئامادەكارى بكريت بۆدەو نەخوازراوانەى كە رووبەروى فرۆكەخانەى ئيۆدەئەتى سيّىانى دەبنەوە.

# 1. Introduction

Economic problems have an effect on our region (Kurdistan) especially Sulaymaniyah province such as another country in the world, in our region the airports are one of the most important centers which improve some of these problems according to their incomes, in addition, the statistical tools could be analyzed these problems especially when time is a significant factor in them. Time series analysis is one of the powerful statistical tools that is used to forecasting the weight of imported equipment which are causes of changing the economic situation in our region, thus these weight of imported equipment are reasons for improving economic situation at Sulaymaniyah governorate.

# **1.1 Research Problem**

The region of Kurdistan is in an economic problem, therefor we want to study the effect of imported weight of equipment through international airport of Sulaimani.

# **1.2 Research Hypothesis**

The hypothesis of the study is as follow:

Dose the imported weight of equipment affects on economic situation in Sulaymaniyah governorate.

# 1.3 Objective of the Study

The main objective of this study is to forecasting the weight of imported equipment and determining their economic effects according to their weights by using seasonal autoregressive moving average model,

# 2. Literature Review

John A. D. aston, David F. Findley, Tucker S. Mcelroy, Kellie C. Wills and Donald E. K. Martin (18 Oct. 2007) in their study they focused on the widely used Box-Jenkins model, they shown how the class of SARIMA models with SMA factor can be parsimoniously generalized to model time series with heteroskedastic seasonal frequency components. Their frequency-specific models decompose this factor by associating one moving average coefficient with a proper subset of the seasonal frequencies 1, 2, 3, 4, 5 and 6 cycles per year and a second coefficient with the complementary subset. A generalization of Akaike's AIC is presented to determine these subsets. Properties of seasonal adjustment filters and adjustments obtained from the new models were examined as are forecasts.

Elangbam Haridev Singh (9 sep. 2013) to forecast demand of international tourism to Bhutan used selecting appropriate model both ARIMA as well as exponential smoothing, firstly the data has been tested to achieve the stationary the researcher found out that the series is stationary at level. The best and adequate model that fitted was SARIMA (0,1,1) (1,1,1) model to the future demand of tourism is forecast.



Adhistya Erna Permanasari, Indriana Hidayah and Isna Alfi Bustoni (8 Oct. 2013) studied the usefulness of forecasting method in predicting the number of disease incidence was important, they used SARIMA model to forecast the number of Malaria incidence. The dataset for model development was collected from time series data of Malaria occurrences in United States obtained from a study published by Centers for Disease Control and Prevention (CDC). From the results they detected that the best model was SARIMA (0,1,1)(1,1,1)12. The model achieved 21,6% for Mean Absolute Percentage Error (MAPE). It indicated the capability of final model to closely represent and made prediction based on the Malaria historical dataset.

Ette H. Etuk (21 Aug 2014) used SARIMA model to forecast the Monthly rainfall in the Gezira irrigation scheme of Sudan, for this purpose he used a time series sample during (1971 to 2000). He used two tests to test the stationary of the data series under study the results showed that the series is stationary according to Augmented Dickey Fuller (ADF) Test but according to Correlogram in not stationary, a seasonal (i.e. 12-point) differencing yields a stationary series on ADF test and correlogram grounds. The best fitted model was SARIMA((0,0,0)x(0,1,1)12 that adjudged the most adequate depending on AIC value which was minimum, this may be used as a basis for rainfall forecasting for planning purposes in the region.

Daniel Eni and Fola J. Adeyeye (22 Jul. 2015) they obtained historical data of rainfall in Warri Town for the period 2003-2012 for the purpose of model identification and those of 2013 for forecast validation of the identified model. To select the best model for forecasting they used each of sum squared error, Akaike's Information Criterion (AIC) and Schwartz's Bayesian Criterion (SBC), according to these mentioned measures they preferred SARIMA(1, 1, 1) (0, 1, 1) as an adequate and best model to forecast the rainfall in Warri Town foe future periods in time.

Biljana Petrevska and Goce Delcev (25 Jul 2016) they used autoregressive integrated moving average model to forecast the number of international tourism demand focusing on the case of F.Y.R. Macedonia, For this purpose, the Box–Jenkins methodology have been applied and several alternative specifications were tested in the modeling of original time series and international tourist arrivals recorded in the period 1956–2013. Through the results of standard indicators for accuracy testing, the best model was ARIMA(1,1,1) for forecasting, then according to the forecasted values the researchers found out that the number of international tourism will increase by 13.9% in 2018.

### 3. Materials and Methods

# 3.1 Time series

A time series is a sequence of data variables, which is consisting of successive observations on a quantifiable variable(s), that is making an over a time interval [6]. Usually, the observations are chronological and taken at regular intervals (days, months, years). Time series data are also often seen naturally in many field areas including; (Economics, Finance, Environmental, and Medicine) Time series can be represented as a set of observations  $X_t$ , each one being recorded at a specific time t [6], and written as:

 $\{X_1, X_2, ..., X_t\}$  or  $\{X_T\}$ , where T = 1, 2,...t

and Xt is the value of X at time t, then the goal is to create a model of the form:

 $X_t = f(X_{t-1}, X_{t-2}, \dots, X_{t-n}) + e_t$  (2.1)

Where Xt-1 is Xt variable for values of lag 1 that is the previous observations value, Xt-2 is the Xt variable for values of lag 2 means two observations value ago, etc., and it represents noise value which doesn't follow a pattern of forecasting. The Xt value is usually highly correlated with Xt-cycle if a time series is following a pattern repeating, where the cycle was an observations number in a regular cycle [8].



### **3.2 Time series Analysis**

Time series data occurrences are becoming extremely valuable to the operations and development of modern organizations. Financial institutions. Likewise, public and private institutions are using time series data to manage and project the loads on the networks. More and more time series are used in this type of investigation and hundreds of thousands of time series that contain valuable economic and financial information are nowadays available both on and off-line [8].

Time series analysis accounts for the fact that data points taken over time may have an internal structure (such as autocorrelation, trend or seasonal variation) that should be accounted for. As defined earlier, time series analysis comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data. It involves the use of techniques for drawing inferences from time series data. Note however that one other main purpose for analyzing time series is forecasting. Forecasting is the application of a model to predict future values based on previously observed time series values [8].

### **3.3 Component of Time Series**

A basic step in choosing an appropriate model and forecasting procedure to a time series is to consider the type of data patterns exhibited from the time series plots. Traditional methods of time series analysis are mainly concerned with decomposing the variation in series <sup>[11]</sup>. The sources of variation in terms of patterns in the time series data are mostly classified into four main components. The components are:

### 3.3.1 Trend

A time series presents increase or decrease trend for a specific period of month, developing in technological process, huge-scale shift in consumer demands, etc. The upward or downward in the time series movements is said to be Secular trend <sup>[4]</sup>.

### 3.3.2 Seasonal Variation

In a time series the variations of seasonal are a short-term fluctuation that occurs in a year periodically, then it repeats year after year. The customs of people or conditions are responsible for a pattern of repetitive of seasonal variations as a major factor. In general, seasonality is defined as a pattern that repeats itself over fixed intervals of time <sup>[4]</sup>.

# 3.3.3 Cyclical Variations

Cyclical variations are recurrent upward movements in a time series but the cycle period is greater than a year. And, the seasonal variations are regular when comparison with these variations. According to length and size there are different types of varying cycles. The ups and downs in business activities are the effects of cyclical variation<sup>[4]</sup>.

### 3.3.4 Irregular Variation

The short duration fluctuations in a time series that is called Irregular variation, means that it happened unsystematically variation at occurrence; these variations are also referred to as residual variations, cyclical and seasonal variations. Irregular fluctuations results due to the occurrence of unforeseen events such as floods, earthquakes, wars, famines, etc. <sup>[4]</sup>.



### 3.4 Stationary and Non-stationary Series:

Stationary series vary around a constant mean level, neither decreasing nor increasing systematically over time, with constant variance. Non-stationary series have systematic trends, such as linear, quadratic, and so on. A non-stationary series that can be made stationary by differencing is called "non-stationary in the homogenous sense." Stationarity is used as a tool in time series analysis, where the raw data are often transformed to become stationary. For example, economic data are often seasonal or dependent on a non-stationary price level. Using non-stationary time series produces unreliable and spurious results and leads to poor understanding and forecasting. The solution to the problem is to transform the time series data so that it becomes stationary. If the non-stationary process is a random walk with or without a drift, it is transformed to stationary process by differencing.

Differencing the scores is the easiest way to make a non-stationary mean stationary (flat). The number of times you have to difference the scores to make the process stationary determines the value of d. If d=0, the model is already stationary and has no trend. When the series is differenced once, d=1 and linear trend is removed. When the difference is then differenced, d=2 and both linear and quadratic trend are removed. For non-stationary series, d values of 1 or 2 are usually adequate to make the mean stationary. If the time series data analysed exhibits a deterministic trend, the spurious results can be avoided by detrending. Sometimes the non-stationary series may combine a stochastic and deterministic trend at the same time and to avoid obtaining misleading results both differencing and detrending should be applied, as differencing will remove the trend in the variance and detrending will remove the deterministic trend <sup>[10]</sup>.

A stationary process has the property that the mean, variance and autocorrelation structure do not change over time. Stationarity can be defined in precise mathematical terms as:

- 1. The mean  $\mu(t) = E(\gamma(t))$
- 2. The variance  $\sigma^2$  (t) = Var(y(t)) =  $\gamma(0)$

There are two kinds of stationary:

# 3.4.1 Non Stationary around Variance

In the case fluctuation of time series about the contrast and this discrepancy is not fixed, it means that the series is stationary about the contrast, and there are transfers to convert the string non stationary to a series of stationary, including the conversion logarithmic and transfers of power and the square root of the absence of stationary, about the contrast non-fixed and turn it into a series of fixed and stationary contrast by applying the following formula:

$$X\tau = \begin{cases} X_{t}^{\lambda} & \lambda \neq 0\\ \lambda & \lambda = 0\\ X_{t} & \lambda = 0 \end{cases}$$

Where: Xt: the original series



### 3.4.2 Non-Stationary around the Mean

The basic conditional in being a stationary time series about mean and middle hard as the changes that occur in the qualities and characteristics of chains with time makes it unstable so you must remove the property not stability, of these chains are used difference method (Difference) to convert the string unstable to a series stable in terms of time difference and take her first and be in the following format:

$$\Delta X = X_{t_t} - X_{t_{-1}}$$

$$W_t = \Delta X_t = X_t - X_{t-1}$$

Where:

 $\Delta$ : the difference factor

Wt: the new series

Xt: the original series

### 3.5 Box-Jenkins Models:

This is a methodology that George-Box and Gwilyn Jenkins at 1970 applied to time series data. Box and Jenkins popularized an approach that combines the moving average and the autoregressive approaches <sup>[3]</sup>. A Box-Jenkins model explains that the time series is stationary or not. Box and Jenkins is recommended that the non-stationary differencing one or more times series to obtain stationarity, with the "I" standing for "Integrated" of an ARIMA model. A Box-Jenkins methodology is a powerful approach to the solution of many time series analysis problems <sup>[18]</sup>. This methodology depends on parts of procedure which is [autoregressive (AR), moving average (MA) and autoregressive moving average (ARMA)] that can be explained as follow:

# 3.5.1 Autoregressive (AR) model

By using backshift operator equation (2.6) can be rewrite as follow:

Where:  $\phi_p(B_p) = (1 - \phi_1 B_1 - \dots - \phi_p B^p)$  $X_t$ : is the origin series.

 $a_t$ : is white noise,  $a_t \sim N(0, \sigma_a^2)$ 

 $\phi_p$ : is the estimated PACF.



To find Variance-Covariance the equation (2.6) should be multiplied by  $(X_{t-k})$  and taking expectation so we get:

$$E(X_t X_{t-k}) =$$

$$E(\phi_1 X_{t-1} X_{t-k} + \phi_2 X_{t-2} X_{t-k} + \dots + \phi_p X_{t-p} X_{t-k} + a_t X_{t-k}) \dots (2.7)$$

Note:

$$E(X_t X_{t-k}) = \lambda_k .$$
  
$$E(a_t X_{t-k}) = 0 .$$

Then:

$$\lambda_k = \phi_1 \lambda_{k-1} + \phi_2 \lambda_{k-2} + \ldots + \phi_p \lambda_{k-p} ; K > 0$$
 .....(2.8)

To get the ACF the equation (2.8) should be divided by the variance of the series ( $\Upsilon_0$ ).

 $p_k = \phi_1 p_{k-1} + \phi_2 p_{k-2} + \ldots + \phi_p p_{k-p}$  .....(2.9) Note:

$$\frac{\lambda_k}{\lambda_o} = p_k, \quad \Upsilon_0 = \sigma_X^2$$

Then the PACF for the AR(P) model can be estimated by using Yule-Walker equations

$$p_j = \phi_k p_{j-1} + \phi_{k(k-1)} p_{j-2} + \ldots + \phi_{kk} p_{j-p} \qquad (2.10)$$

### 3.5.2 Moving Average (MA) Model:

The order of moving average model depends on the number of significant ACF and  $- \Theta_1$  is the coefficient of dependency of observations (Xt) on the error term et and the previous error term at-1, the MA(q) model can be write as follow <sup>[10]</sup>:

Equation (2.11) can be rewrite with back shift operator as follow:  $X_t = O(B)a_t$ 

Where:

$$\bigcirc(\mathbb{B}) = 1 - \bigcirc_1 \mathbb{B} - \dots - \bigcirc_q \mathbb{B}^q$$

The Var-Cov of MA(q) model is:

$$\Upsilon k = \begin{matrix} \sigma_a^2 \left( -\Theta + \Theta_1 \Theta_{k+1} + \dots + \Theta_q \Theta_{q-k} \right) \\ 0 \end{matrix}; K = 1, 2, \dots, q \\ ; K > q \end{matrix} \qquad \dots \dots (2.13)$$

Note:

 $\bigcirc_0 = 1.$ And the ACF is:



$$Pk = \begin{array}{c} \frac{-\odot_{k} + \odot_{1} \odot_{k+1} + \ldots + \odot_{q} \odot_{q-k}}{1 + \odot_{1}^{2} + \odot_{2}^{2} + \cdots + \odot_{q}^{2}} \\ 0 \end{array}; k = 1, 2, \dots, q \\ ; k > q \end{array}$$
.....(2.15)

#### 3.5.3 Autoregressive Moving Average Model (ARMA)

There is large family of models which is named "Autoregressive-Moving Average Models" and abbreviated by ARMA. Many of researchers in different application fields prove that ARMA models fits more than other traditional methods for forecasting <sup>[10]</sup>. The ARMA model is a more general model as a mixture of the AR(p) and MA(q) models and it is called an autoregressive moving average model (ARMA) of order (p,q). The ARMA(p,q) is given by <sup>[19]</sup>:

Where:

 $\phi_p(\mathbf{B}) = 1 - \phi_1 \mathbf{B} - \dots - \phi_p \mathbf{B}^p$  $\odot_q(\mathbf{B}) = 1 - \odot_1 \mathbf{B} - \dots - \odot_q \mathbf{B}^q$ 

We write equation (2.16) as:

$$X_t = \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + a_t - \bigcirc_1 a_{t-1} - \ldots - \bigcirc_q a_{t-q} \dots (2.17)$$

#### 3.5.4 The Autoregressive Integrated Moving Average Models (ARIMA)

When the time series data is not stationary, difference operator can be used to remove nonstationary, the time series data after differencing is called adjusted data and the fitted model is called integrated model which is combine both autoregressive and moving average models.

Notationally, all AR(p) and MA(q) models can be figured as ARIMA(1, 0, 0) this tells there is no differencing and no MA part. The general of ARIMA is written as ARIMA (p, d, q) where p is the order of the AR part, d is the degree of differencing and q is the order of the MA part <sup>[11]</sup>.

$$W_t = \nabla^d X_t = (1 - B) dX_t$$

The general ARIMA process is of the form

#### 3.6 Seasonal Time Series Models

Seasonal time series refers to similar or almost similar patterns which a time series appears to follow during corresponding months of successive years. The movement is popularly repeating events which takes place annually or quarterly. The plot of the series against time checks for non-seasonal or seasonal changes that reveal non-stationarity. Peaks in a series for every 12 months



would indicate annual seasonality whereas peaks in a series of every 3 months would indicate quarterly seasonality. In the seasonal model the significant ACF and PACF patterns are determining the order of seasonality. The number of times per year which seasonal variations occur also matters. If the seasonality is annual, the seasonal variation ACF spikes are heightened patterns at seasonal lags over and above the regular non-seasonal variation once per year. If the seasonality is quarterly, there will be prominent ACF spikes four times per year <sup>[5]</sup>.

### 3.7 Seasonal Autoregressive Integrated Moving Average (Seasonal ARIMA) Models

The seasonal difference is an important tool in non-stationary seasonal modeling. The difference of seasonality of period s for the series  $\{X_t\}$  is symbol by  $\nabla_s X_t$  and is expressed as:

$$\nabla_s \mathbf{X}_t = \mathbf{X}_t - \mathbf{X}_{t-s} \qquad \qquad (2.20)$$

For a time series of size (n), the difference of seasonality will be n-s which is s data values are lost because of difference seasonality. In a non-stationary seasonal model, a time series  $\{X_t\}$  is called a multiplicative seasonal ARIMA model with non-seasonal orders p, d and q, seasonal orders P, D, and Q, and seasonal period s if the differenced series expressed as below:

Satisfies an ARMA (p×q)(P×Q)s model with seasonal period s. We say that  $\{X_t\}$  is an ARIMA (p,d,q)(P,D,Q)s model with seasonal period s. The seasonal part of a seasonal ARIMA model has the same structure as the non-seasonal part: It may have an AR factor, an MA factor, and/or an order of differencing. Box and Jenkins have generalized the ARIMA model to deal with seasonality and defined a general multiplicative seasonal ARIMA model in the form:

$$\phi(B)(1 - B^4)X_t = \Theta(B)(B^4)a_t$$
 ......(2.22)

Where B denotes the backward shift operator,  $\Phi$ ,  $\phi$ ,  $\Theta$ , and  $\odot$  are polynomials of order p, P, q, and Q respectively and  $a_t$  is the purely random process with mean zero and constant variance  $\sigma_a^2$ . Consider a Seasonal ARIMA (0, 1, 1) × (0, 1, 1)4 model for instance. The model specification is:

$$(1 - B^4)(1 - B)X_t = (1 - O_1B)(1 - O_1B^4)a_t$$
......(2.23)

By expansion, we have:

$$X_t = X_{t-1} + X_{t-4} - X_{t-5} + a_t - \bigcirc_1 a_{t-1} - \bigcirc_1 a_{t-4} + \bigcirc_1 \bigcirc_1 a_{t-4}$$

The estimates of the parameters of the model in (2.23) may be obtained using the method of maximum likelihood <sup>[5], [15]</sup>.

### 3.8 Time series analysis steps:

There are several stages to build any model of Box - Jenkins to forecast and can represent the stages of the scheme follows (1):





#### **3.9 Augmented Dickey-Fuller test**

Augmented Dickey-Fuller tests the hypothesis which is stated that the series is not stationary, can be tested in regression equation <sup>[9]</sup>.

Where a random walk,  $a_t = a_{t-1} + a\varepsilon_t$  is allowed <sup>[9]</sup>.

#### **3.10Autocorrelation Function (ACF)**

The autocorrelation function measures the degree of correlation between neighboring observations in a time series. The autocorrelation coefficient is estimated from sample observation using the formula <sup>[10]</sup>:



Thus, the autocorrelation function at lag k is defined as:

$$p_k = \frac{\lambda_k}{\lambda_0}$$
 ,  $k = 0, \pm 1, \pm 2, ...$ 

#### 3.11 Partial Autocorrelation Function (PACF)

The partial autocorrelation function at lag k is the correlation between  $X_t$  and  $X_{t-k}$  after removing the effect of the intervening variables  $X_{t-1}$ ,  $X_{t-2}$ , ....,  $X_{t-k+1}$  which locate within (t, t-k) period, partial autocorrelation function will be donated by  $\phi \kappa \kappa$ , PACF is calculated by iteration <sup>[17]</sup>.

Therefore  $\Phi_{kj} = \Phi_{k-1,j} - \Phi_{kk} \Phi_{k-1,k-1}$  ,  $j = 1,2,\ldots,k-1$ 

#### 3.12 Model Selection Criteria

#### 3.12.1 Akaike Information Criterion AIC

Akaike Information Criterion AIC<sup>[2]</sup> is defined as

$$AIC = -2 \log L + 2p$$
 ......(2.24)

where L is the maximized likelihood function and p is the number of effective parameters. The best model is the one with the smallest AIC. The likelihood function part reflects the goodness of fit of the model to the data, while 2p is described as a penalty. Since L generally increases with p, AIC reaches the minimum at a certain p. AIC is based on the information theory.

#### 3.12.2 Mean Absolute Percentage Error (MAPE)

The MAPE also is said to be mean absolute percentage deviation (MAPD) that is an accurate measure of a method for constructing time series fitted model, the accuracy in time series processes is expressed by:

Where  $X_t$  is the actual value and  $\hat{X}_t$  is the forecast value <sup>[14]</sup>.



### 3.12.3 Mean Absolute Error (MAE)

The MAE is mathematically expressed by:

$$MAE = \frac{1}{n} \sum_{t=1}^{n} |e_t|$$
 .......(2.26)

Where *it* is the error term and n is the number of forecasting <sup>[18]</sup>.

### 3.12.4 Root Mean Square Error (RMSE)

The RMSE is expressed by:

Where:  $X_i$  = actual value

 $\hat{X}_i$ = forecasteed value

N = number of forecasted time period <sup>[16]</sup>.

### 3.13 Estimating the Parameters of an ARMA Model

Iterative method can be used to estimate the parameters of the ARMA model. At every point sum square residuals should be calculated of suitable grid of the parameter values, and the sufficient values are given minimum sum of squared residuals. For an ARMA (1,1) the model is given by

Given N observation  $X_1, X_2, ..., X_N$ , we guess values for  $\mu$ ,  $\phi_1$ ,  $\odot_1$ , set  $a_0 = 0$  and  $Y_0 = 0$  and then calculate the residuals recursively by

$$a_{1} = X_{1} - \mu$$
$$a_{2} = X_{2} - \mu - \phi_{1}(X_{1} - \mu) - \bigcirc_{1}$$
$$a_{N} = X_{N} - \mu - \phi_{1}(X_{1} - \mu) - \bigcirc_{1} a_{N-1}$$

The residual sum of squares  $\sum_{t=1}^{N} a_t^2$  is calculated. Then other values of  $\mu$ ,  $\phi_1$ ,  $\odot_1$ , are tried until the minimum residual sum of squares is found <sup>[7], [12]</sup>.

Note: It has been found that most of the stationary time series occurring in practices can be fitted by AR(1), AR(2), MA(1), MA(2), ARMA(1,1) or white noise models that are customarily needed in practice <sup>[10]</sup>.



### 3.14 Models Forecasts

The main goal of constructing a model for a time series is to make future forecasteions for a given series. It also plays a significant role in assessing the forecasts accuracy. The ultimate test of an ARIMA model is power or ability to forecast. In order to obtain a forecast with a minimal errors, there are seven features of a good ARIMA models taken into account <sup>[13], [15]</sup>. First, a good model is parsimonious. That is, it has the smallest number of coefficients which explain the data set. Secondly, a sufficient AR model should not be nonstationary. Thirdly, the MA of the model should be invertible. Fourth, insufficient model the residuals must be independent. Fifth, the distribution of residuals of a good model must be distributed normal<sup>[13]</sup>. From the existing theory of the series up to time t, namely,  $X_1, X_2, X_3, ..., X_{t-1}, X_t$ , we can forecast the value of  $X_{t+h}$ , that will happen h time units ahead. In this case, time t is the forecast origin and his the lead time forecast. This forecast is denoted and estimated as

$$\hat{X}_t(L) = E(X_{t+h}|X_1, X_2, \dots, X_t)......(2.28)$$

Once an adequate and satisfactory model is fitted to the series of interest, forecasts can be generated using the model <sup>[13]</sup>.

The one-step-ahead forecast for time t + 1 is given by:

$$x_{t+1} = \phi_1 x_t + \dots + \phi_p x_{t-p+1} + a_{t+1} - \bigcirc_1 a_t \dots - \bigcirc_q a_{t-q+1} \dots$$
(2.30)

### 4.1 introduction

Decision making is so sensitive when it is based on the forecasting methods, therefore the researcher should be careful during use the forecasting methods, these methods are depending on the number of observation and its run term. Also the method of estimation is important to estimate the parameters of the model in use, in this study maximum likelihood estimation method have been used to estimate the parameter of the best model that is adequate the data under consideration.

### 4.2 Data Description

The dataset used for the analysis in this study came from the international Sulaymaniyah airport of Sulaymaniyah city which is contained one variable and deals with monthly weight of imported equipment since January 2010 up to November 2017 as shown in appendix. These data are measurements of the weight of imported equipment in Sulaymaniyah city from international Sulaymaniyah airport.

# 4.3 Applications

The time series plots are display observations on the y-axis against equally spaced time intervals on the x-axis. They are used to evaluate patterns, knowledge of the general trend and behaviors in data over time. The time series plot of monthly weight of imported equipment in Sulaymaniyah city is displayed in Figure 1 below:



Time Series Plot for Weight of imported equipment



Figure 1: Monthly plot of time series weight of imported equipment in Sulaymaniyah city

Figure 1 indicates that the data of time series is not random. The plot shows consistent pattern of short-term changes for data which indicates the existence of seasonal fluctuations. This series varies randomly over time and there is seasonal fluctuations. For further testing of the stationery of the time series, we applied Augmented Dickey-Fuller test for monthly weight of imported equipment in Sulaymaniyah international airport. The augmented Dickey-Fuller tests the hypotheses which is stated that the series is non-stationary series. Table 1 shows the results of ADF of the data of the time series of weight of imported equipment.

**Table 1:** Dickey-Fuller test for monthly weight of imported equipment in Sulaymaniyah

 international airport

Test	t-Statistic	P-value
ADF	0.647249	0.9903

Table (1) explain that the p-value of the Dickey-Fuller test equals 0.9903 and it is greater than 0.05. This result indicates that the data of time series of monthly weight of imported equipment is not random, and demonstrates this results by the examination of the autocorrelation and partial autocorrelation functions as shown below.









Estimated Autocorrelations for Weight of imported equipment



**Figure 3:** Partial Autocorrelation Function for the monthly weight of imported equipment in Sulaymaniyah international airport

All the above results and plots support that the data of time series is not random at the level, which needs some treatments to be transformed to a random series. Therefore, we used many transformations and we found that the most suitable transformation is by differencing the series. We note that the time series for the first-differenced series in Figure 4 indicates that the series is stationary.



Time Series Plot for adjusted Weight of imported equipment

**Figure 4**: Time series plot of the first difference of monthly weight of imported equipment in Sulaymaniyah city

**Table 2:** Dickey-Fuller test of the first differences for monthly weight of imported equipment inSulaymaniyah international airport

Test	t-Statistic	P-value
ADF	-6.592854	0.000



Table (2) explain that the p-value of the Dickey-Fuller test equals 0.000 and it is less than 0.05. This result indicates that the non-stationarity hypotheses of the differenced monthly weight of imported equipment in Sulaymaniyah international airport is rejected and this demonstrates by estimating the autocorrelation and partial autocorrelation function (ACF and PACF) for the first-differenced series in Figure 5 and 6

Estimated Autocorrelations for adjusted Weight of imported equipment



**Figure 5:** Autocorrelation Function for the first-differenced series of the monthly weight of imported equipment in Sulaymaniyah international airport.



Estimated Partial Autocorrelations for adjusted Weight of imported equipment

**Figure 6:** Partial Autocorrelation Function for the first-differenced series of the monthly weight of imported equipment in Sulaymaniyah international airport.

The results above demonstrates the success of differencing of the time series data of the monthly weight of imported equipment in Sulaymaniyah international airport. Thus, the series became stationarity.

# 4.4 Model Identification

This section shows how we determine the order of the seasonal ARIMA model. We computed all relevant criteria to determine good SARIMA model of weight of imported equipment. Those are



the ACF and PACF in addition to RMSE, MAE, MAPE, and AIC criteria. To take a decision must be scanning all the plots of ACF and PACF coefficients of the series as shown in the figure (5 and 6) respectively, It can be seen from the AC coefficients and the PAC coefficients of the time series data that it is necessary to measure the changes in seasonal during identifying and model estimating. Weight of imported equipment data is monthly and according to the identification criteria, the following models have been examined and estimated as shown in table (3) below. The best seasonal model is chosen through the RMSE, MAE, MAPE and AIC criteria if it shows the lowest values of these criteria as it is shown in table (3).

**Table 3**: SARIMA Models Criteria for the monthly Weight of imported equipment in Sulaymaniyah

 international airport

N A a d a l				MDE	A1C
woder	RIVISE	IVIAE	WAPE	IVIPE	AIC
	0.074400	0 500070	2 5 2 0 5 5	0.450070	0 750440
ARIMA(1,1,1)X(0,0,0)4	0.6/1193	0.508973	2.52955	0.456078	-0.753442
ARIMA(2,1,1)x(0,0,0)4	0.672002	0.508289	2.52574	0.451151	-0.729053
ARIMA(1.1.1)x(1.0.0)4	0.675011	0.508919	2.52946	0.454678	-0.72012
$\Delta RIMA(1 \ 1 \ 1)_{V}(0 \ 0 \ 1)_{A}$	0 675055	0 509129	2 5306	0/15175/	-0 719988
ANIMA(1,1,1)A(0,0,1)4	0.075055	0.303123	2.5500	0.431734	0.715500
$ADIMAA(1   1   1)_{y}(2   0   0)_{A}$	0 669696	0 50741	2 52254	0 421705	0 71607
ARIIVIA(1,1,1)X(2,0,0)4	0.008080	0.50/41	2.52254	0.451/95	-0./109/

It is shown in table (3) that the SARIMA (1,1,1)X(2,0,0)4 model produced the value of each RMSE, MAE, MAPE and AIC criteria with the smallest values. This means that the SARIMA (1,1,1)X(2,0,0)4 model is the best among all the other models, which is the most suitable model that can be obtained for the monthly weight of imported equipment in Sulaymaniyah international airport.

# 4.5 Parameters Estimation:

Since we concluded in the previous section that the SARIMA (1,1,1)X(2,0,0)4 model is the best model with the smallest value of RMSE, MAE, MAPE and AIC criteria, the parameters had been estimated using the method of maximum likelihood estimation as it is the best and most appropriate method of estimation. The results of the parameters estimation of the model are shown in table (4) below.

Estimate	Stnd. Error	t	P-value
0.45049	0.118144	3.81308	0.000258
0.93299	0.043249	21.5721	0.000000
0.03376	0.019229	1.75570	0.077732
0.19799	0.015614	12.6807	0.000396
	Estimate 0.45049 0.93299 0.03376 0.19799	EstimateStnd. Error0.450490.1181440.932990.0432490.033760.0192290.197990.015614	EstimateStnd. Errort0.450490.1181443.813080.932990.04324921.57210.033760.0192291.755700.197990.01561412.6807

Table 4: Parameter Estimates of SARIMA (1,1,1)X(2,0,0)4 Model Estimate model coefficients



It is shown in table (4) that the p-value for the parameters AR(1), MA(1), and SAR(2) coefficients are less than  $\alpha = 0.05$ . This indicates that these coefficients are significantly different from zero, however the p-value of SAR(1) parameter is less than  $\alpha = 0.10$ , means that this coefficient is significantly different from zero at level of significant (%10). As it is shown for this model, the RMSE, MAPE, MAE and AIC criteria are the smallest values among the other models. Thus, the final model is SARIMA (1,1,1)x(2,0,0)4.

### 4.6 Forecasting for weight of imported equipment:

After getting the final model SARIMA (1,1,1)X(2,0,0)4 of the data of the monthly weight of imported equipment in Sulaymaniyah international airport that has been expressed above, the researcher used it for forecasting future quantities weight of imported equipment. We forecasted quantities of monthly weight of imported equipment in Sulaymaniyah international airport in 2017-2018 for 12 months. The forecasting of time series for monthly weight of imported equipment in Sulaymaniyah international airport have been plotted as in figure (7).



Figure 7: Plot of the data and the forecasts with 95% confidence interval are represented

Figure 7 shows the result that the behavior of forecasted values is the same as original series of weight of imported equipment in Sulaymaniyah international airport. The results of the forecasted values in table (5) for the year 2017-2018 are all between the upper and lower boundaries of the 95% confidence intervals. This confirms that the forecasting is very efficient.

Table 5 shows that the quantities of monthly weight of imported equipment in Sulaymaniyah international airport in 2017 – 2018 for 12 months have been forecasted. It is also shown from these results that the forecasted values are all between the upper and lower boundaries of the 95% confidence intervals. This supports that the forecasting is efficient.

Period	Forecast	Lower Limit 95.0%	Upper Limit 95.0%
Nov – 2017	899005000	236445000	3418170000
Dec – 2017	1131240000	251454000	5089240000
Jan – 2018	1379040000	290854000	6538540000
Feb <b>–</b> 2018	1182200000	243605000	5737150000
Mar – 2018	1653020000	333661000	8189400000
Apr – 2018	1351040000	269062000	6784020000
May – 2018	1306920000	257503000	6633130000
Jun – 2018	961071000	187560000	4924600000
Jul – 2018	1153740000	212541000	6262820000
Aug – 2018	1199730000	214966000	6695700000
Sep – 2018	1246540000	219386000	7082810000
Oct – 2018	1196720000	207650000	6896880000

Table 5: Forecast future value with the lower and upper 95% confidence interval

# 5. Conclusion and Recommendations

### 5.1 Conclusion

From the results, the following conclusions can be summarized:

- 1. The statistical tests show that the time series of the monthly weight of imported equipment in Sulaymaniyah international airport is stable. In addition to have the seasonal changes it repeats itself every 4 months.
- 2. The best and most efficient model is SARIMA (1,1,1)x(2,0,0)4 among the possible models which was chosen using the balancing standards (the smallest value of each : AIC, RMSE, MAPE and MAE criteria).
- 3. Parameters Estimate of SARIMA (1,1,1)x(2,0,0)4 Model are significant thus, the SARIMA (1,1,1)x(2,0,0)4 is efficient.
- 4. According to SARIMA (1,1,1)x(2,0,0)4, the monthly weight of imported equipment in Sulaymaniyah international airport for the year 2017-2018 for 12 months have been forecasted. The forecasted values are showed harmony with its counterparts in the original series values. Moreover, the forecasted values are all between the upper and lower boundaries of the 95% confidence intervals. Thus, it provided a future image of the reality of monthly weight of imported equipment.

# 5.2 Recommendations:

Through the results that have been reached, we recommend the following:



- 1. Adopting the results of this research that there is a real problem facing Sulaymaniyah international airport according to weight of imported equipment through the upcoming years which is the decrease in the monthly weight of imported equipment in Sulaymaniyah international airport. Furthermore, it helps the officials and the decision makers in finding solutions and quick alternatives to face this problem and putting the future plans of the monthly weight of imported equipment in Sulaymaniyah international airport to stop aggravating the problem.
- 2. Generalizing this study to similar studies on the cities and districts level and the other cities level and comparing between them.
- 3. Weight of imported equipment at each airport is a vital element of the economic development and progress, thus it should be available plan to use this fortune from monthly weight of imported equipment.

# 6. References

- 1. Adhistya Erna Permanasari, Indriana Hidayah and Isna Alfi Bustoni (2013), "SARIMA (Seasonal ARIMA) implementation on time series to forecast the number of Malaria incidence" Conference Paper October 2013 with 1,996 Reads, Conference: Conference: Information Technology and Electrical Engineering (ICITEE), 2013 International Conference.
- 2. Agyemang, Boakye. (2012), Autoregressive Integrated Moving Average (ARIMA) Intervention Analysis Model for the Major Crimes in Ghana. (The case of the Eastern Region), BSc, Department of Mathematics, Kwame Nkrumah University of Science and Technology, Ghana.
- 3. Akaike. A new look at the statistical model identi\_cation. IEEE Transactions on Automatic Control, 19(6):716{723, 1974.
- 4. Box, G. E. P., Jenkins, G. M., (1970), Time Series Analysis, Forecasting and Control, 3rd ed. Prentice Hall, Englewood Cliffs, New Jersey.
- 5. Biljana Petrevska and Goce Delcev (2016), "Predicting tourism demand by A.R.I.M.A. models", Economic Research-Ekonomska Istraživanja, ISSN: 1331-677X (Print) 1848-9664 (Online) Journal homepage: http://www.tandfonline.com/loi/rero20, 2017
- 6. Chatfield, C. (1996), The Analysis of Time Series, 5th ed., Chapman & Hall, New York.
- 7. Charles, Akinmutmi. (1998-2011), A seasonal Arima Modelling of Residential Coal Consumption Series in Nigeria, Department of Mathematics, Ahmadu Bello University, Nigeria.
- 8. Cochrane, J. H. (2005). *Time Series for Macroeconomics and Finance*. Chicago: Graduate School of Business, University of Chicago.
- 9. Cryer, J. D, Chan, K.(2008), Time Series Analysis, Second Edition, Springer Science Business Media, New York.
- 10. Elangbam Haridev Singh, 2013, "Forecasting Tourist Inflow in Bhutan using Seasonal ARIMA", International Journal of Science and Research (IJSR), India Online ISSN: 2319-7064, Volume 2 Issue 9, September 2013.
- 11. Daniel Eni and Fola J. Adeyeye (2015), "Seasonal ARIMA Modeling and Forecasting of Rainfall in Warri Town, Nigeria", Journal of Geoscience and Environment Protection, 2015, 3, 91-98 Published Online August 2015 in SciRes.
- 12. DTREG. (2010). *Time Series Analysis*. Retrieved March 2015, from DTREG: <u>https://www.dtreg.com/methodology.</u>
- 13. Ette H. Etuk (21 Aug 2014), "An additive seasonal Box-Jenkins model for Nigerian monthly savings deposit rates", Conference Paper on (18-Apr 2014).
- 14. Gyasi-agyei, Kwame. (june,2012), Analysis and modeling of prevalence of measles in the ashnti region of chana, Kwame Nkrumah University of Science and Technology, West Africa.



- 15. Hamilton, J. D.( (1994), Time series analysis, Princeton university press United State of America.
- Hudu, Muhammed. (2009), Interrupted Time Series Analysis of the Rate of Inflation in Ghana, MSc, Department of Mathematics, Kwame Nkrumah University of Science and Technology, West Africa.
- 17. Hurvich, C.M., and Tsai, C.L (1989). "Regression and Time Series Model Selection in Small Sample", Biometrika, 76, 297-307.
- 18. John A. D. aston, David F. Findley, Tucker S. Mcelroy, Kellie C. Wills and Donald E. K. Martin (2007), "New ARIMA Models for Seasonal Time Series and Their Application to Seasonal Adjustment and Forecasting", Conference Paper on (18-oct 2007).
- 19. Kibar, Mustafa alptekin. (2007), Building Cost Index Forcasting with Time Series Analysis, MSc, Civil Engineering Department, Middle East Technical University, Turkey.
- 20. Mynsbrugge, Jorrit V. (2010), "Bidding Strategies Using Price Based Unit Commitment in a Deregulated Power Market", K.U.Leuven, 62, 24-28.
- 21. Reinsel, G. C. (1994), Time Series Analysis, Forecasting and Control, 3<sup>rd</sup> ed. Prentice Hall, Englewood Clifs, New Jersey.
- 22. Sariaslan, Nazli. (2010), the Effect of Temporal Aggregation on Univariate Time Series Analysis, department of statics, Meddle East Technical University, Turkey.
- 23. Shumway, R. Stoffer, H. (2006), Time Series Analysis and Its Applications, Second Edition, Springer Science, Business Media, New York
- 24. Spyros, M., Wheelwright, S. C. and Hyndman, R. J. (1998), "Forecasting: Methods and Applications", Third Edition, John Wiley & Sons, Inc, U.S.A.
- 25. Wei. William W. S. (1989). "Time Series Analysis: Univariate and Multivariate Method". University of Tempie, department of statistic. Addison-Wesley publishing company, Inc.

# 7. Appendix:

Data used in the study:

Months	Pax. (Domestic. Int.)
January (2010)	3051
February	2595
March	2830
April	3766
Мау	4453
June	4490
July	6160
September	13336
October	9664
November (2017)	10763