RESEARCH ARTICLE OPEN ACCESS

Box –Jenkins Models For Forecasting The Daily Degrees Of Temperature In Sulaimani City.

Samira Muhammad Salh[1], Salahaddin A.Ahmed[2]
(1) University of Sulaimani- College of Administration and Economic / Statistic Department Sulaimani / Iraq.
(2) University of Sulaimani- Faculty of Science and Science Education / Physics Department Sulaimani / Iraq.

Abstract:
The Auto-regressive model in the time series is regarded one of the statistical articles which is more used because it gives us a simple method to limit the relation between variables time series. More over it is one of Box –Jenkins models to limit the time series in the forecasting the value of phenomenon in the future so that study aims for the practical analysis studying for the auto-regressive models in the time series, through one of Box –Jenkins models for forecasting the daily degrees of temperature in Sulaimani city for the year (2012-Sept.2013) and then for building a sample in the way of special data in the degrees of temperature and its using in the calculating the future forecasting . the style which is used is the descriptive and analyzing by the help of data that is dealt with statistically and which is collected from the official resources To reach his mentioned aim , the discussion of the following items has been done by the theoretical part which includes the idea of time series and its quality and the autocorrelation and Box –Jenkins and then the practical part which includes the statistical analysis for the data and the discussion of the theoretical part, so they reached to a lot of conclusions as it had come in the practical study for building autoregressive models of time series as the mode was very suitable is the auto-regressive model and model moving average by the degree (1,1,1).
Keywords: Box –Jenkins models, Time series, Auto-regressive model, Moving Average models, ARMA models, Sulaimani city.

I. Introduction

That there is a lot of studies and research economic and administrative, which focused on expectations of future because of its significant impact in this field well as study the time series for many phenomena and know the nature of the changes that will come out and what will happen with the change in the future and for a number of years and light of what happened to her in the past as researchers presented several studies to build models for time series and all amosmeh, and the mean time series study analyzed to its factors influencing, public Kalatjah, and seasonal changes and episodic and others. And is the regression model of self-time series and one of the statistical tools most widely used because it gives us an easy way to determine the relationship between the variables of the time series. Can express this relationship in the form of an equation as that one models Box _ Jenkins for time series analysis and forecasting values that appear in the future and his practical applications in the fields of economic and administrative and weather forecasters, for weather forecasting and measuring amounts of rain and temperatures that had significant effects in the areas of agricultural, industrial and marine navigation and others. Also for these models is particularly important in planning and forecasting future.

II. Objective of this research

The research aims to provide an analytical study applied to the regression model of autocorrelations-time series by the model Box – Jenkins for forecasting the daily temperature in the city of Sulaimani in (2012) and then for building a model of data on temperature and to be used in the calculation of future forecasts.

III. Research Methodology

This is applied research, which is trying to build an appropriate (suitable) model for the purpose of forecasting temperatures in the city of Sulaimani in 2012 and to be used for forecasting.

IV. The Theatrical part

The Concept of Types of Time Series

A time series is a collection of observations of well-defined data items obtained through repeated measurements over time. There are two types of time series called stationary and non-stationary time series. Time series is a set of observations $\{x_t\}$, each one being recorded at a specific time t.

A time series model for the observed data $\{x_t\}$ is a specification of the joint distributions (or possibly only the means and covariance) of a sequence of random variables $\{X_t\}$ of which $\{x_t\}$ is postulated to be a realization.

In reality, we can only observe the time series at a finite number of times, and in that case the underlying sequence of random variables $(X_1,X_2, . . . ,X_n)$ is just an n-dimensional random variable (or random vector). Often, however, it is convenient to allow the number of observations to be infinite. In that case $\{X_t, t = 1, 2, . . .\}$ is called a stochastic process. In order to specify its statistical properties we then need to `consider all n-dimensional distributions:

$P[X_1 \le x_1, . . . ,X_n \le x_n]$ for all $n = 1, 2, . . . ,$

A process $\{X_t, t \in Z\}$ is said to be an independent and identically distribution (IID) noise with mean (0) and variance (σ^2), written $\{X_t\}$ » IID(0, σ^2),

if the random variables X_t are independent and identically distributed with

$E(X_t) = 0$ and $Var(X_t) = \sigma^2$.

Then the mean is $\mu = E(w_t)$,and the variance:

$$\text{var}(w_t) = E(w_t - \mu)^2$$
$$\alpha_k = \text{cov}(w_t, w_{t+k}) = E(w_t - \mu)(w_{t+k} - \mu) \qquad (1)$$

If we assume that we have the following time series ($w_1, w_2, w_3, ..., w_n$) which represent the values of the time series (w_t) and ($(c_k, \sigma_w^2, \bar{w})$) each estimates variables $(\alpha_k, \text{var}(w_t), \mu)$ respectively. Then

$$\left.\begin{aligned} \bar{w} &= \frac{\sum_{t=1}^{N} w_t}{N} \\ \sigma_w^2 &= \frac{\sum_{t=1}^{N} (w_t - \bar{w})^2}{N-1} \\ C_k &= \frac{\sum_{t=1}^{N} (w_t - \bar{w})(w_{t+k} - \bar{w})}{N-1} \qquad , k = 1,2,3,... \end{aligned}\right\} ......(2)$$

We can distinguish between two types of time series through the values of correlation coefficients between the observations.

V. Models for time Series

Let us begin this section with the following wonderful quotation:

"Experience with real-world data, however, soon convinces one that both stationary and Gaussianity are fairy tales invented for the amusement of undergraduates." (Thomson 1994).

Bearing this in mind, stationary models form the basis for a huge proportion of time series analysis methods. As it is true for a great deal of mathematics, we can begin with very simple building blocks and then building structures of increasing complexity. In time series analysis, the basic building block is the purely random process.

Loosely speaking a stationary process is one whose statistical properties do not change over time. More formally, a strictly stationary stochastic process is one where given $t_1, . . . , t_\ell$ the joint statistical distribution of $X_{t1} , . . . ,X_{t\ell}$ is the same as the joint statistical distribution of $X_{t1+\tau} , . . . ,X_{t\ell+\tau}$ for all ℓ and τ . This is an extremely strong definition: it means that all moments of all degrees : expectations, variances, third order and higher of the process, any where are the same. It also means that the joint distribution of (X_t,X_s) is the same as (X_{t+r},X_{s+r}) and hence cannot depend on s or t but only on the distance between s and t, i.e. $s - t$.

Since the definition of strict stationary is generally too strict for everyday life, a weaker definition of second order or weak stationary is usually used. Weak stationary means that the mean and the variance of a stochastic process do not depend on t (that is they are constant) and the auto-covariance between X_t and $X_{t+\tau}$ only can depend on the lag τ (τ is an integer, the quantities also need to be finite). Hence for stationary processes $\{X_t\}$ the definition of auto-covariance is:

$$\gamma(\tau) = \mathrm{cov}(X_t, X_{t+\tau}),$$

for integers τ . It is vital to remember that, for the real world, the auto-covariance of a stationary process is a model, albeit a useful one. Many actual processes are not stationary as we will see in the next section. Having said this, much fun can be given with stationary stochastic processes!

One also routinely comes across the autocorrelation of a process which is merely a normalized version of the auto-covariance to values between −1 and 1 and commonly uses the Greek letter ρ as its notation:

$$\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)}$$

for integers τ and where $\gamma(0) = \mathrm{cov}(X_t, X_t) = \mathrm{var}(X_t)$.

VI. Time Series Data

A time series is a set of statistics, usually collected at regular intervals. Time series data occur naturally in many application areas.

- Economics e.g., monthly data for unemployment, hospital admissions, etc.
- Finance e.g., daily exchange rate, a share price, etc.
- Environmental e.g., daily rainfall, air quality readings.
- Medicine e.g., ECG brain wave activity every 2^{-8} sec.

The methods of time series analysis pre-date those for general stochastic processes and Markov Chains. The aims of time series analysis are to describe and summarize time series data, to fit low-dimensional models, and to make forecasts.

We write our real-valued series of observations as.. $,X_{-2},X_{-1},X_0,X_1,X_2,$ ….., a doubly infinite sequence of real-valued random variables indexed by Z.

VII. Trend, seasonality, cycles and residuals

One simple method of describing a series is that of classical decomposition. The notion is that the series can be decomposed into four elements:

1) Trend (Tt) — long term movements in the mean;
2) Seasonal effects (It) — cyclical fluctuations related to the calendar;
3) Cycles (Ct) — other cyclical fluctuations (such as a business cycles);
4) Residuals (Et) — other random or systematic fluctuations.

The idea is to create separate models for these four elements and then to combine them, either additively

$$X_t = T_t + I_t + C_t + E_t \qquad (3)$$

or multiplicatively

$$X_t = T_t \cdot I_t \cdot C_t \cdot E_t . \qquad (4)$$

Stationary processes

1. A sequence $\{X_t, t \in Z\}$ is strongly stationary or strictly stationary if

$$(X_{t1}, \ldots, X_{tk}) D= (X_{t1+h}, \ldots, X_{tk+h})$$

for all sets of time points t1, . . . , tk and integer h.

2. A sequence is weakly stationary, or second order stationary if

(a) $E(X_t) = \mu$, and

(b) $\mathrm{cov}(X_t, X_{t+k}) = \gamma_k$,

Where μ is constant and γ_k is independent of t.

3. The sequence $\{\gamma_k, k \in Z\}$ is called the auto-covariance function.

4. We also define $\rho_k = \dfrac{\gamma_k}{\gamma_0} = corr(X_t, X_{t+k})$

and call { $\rho_k, k \in Z$ } the auto-correlation function (ACF).

Models for time series: (MA, AR and ARMA models)
This section considers some basic probability models extensively used for modeling time series.

Moving Average models: (MA)
The moving average process of order q is denoted MA(q) and defined by

$$X_t = \sum_{s=0}^{q} \theta_s \varepsilon_{t-s}$$

where θ1, . . . , θq are fixed constants, θ0 = 1, and { ε_t } is a sequence of independent (or uncorrelated) random variables with mean 0 and variance σ2. It is clear from the definition that this is the second order stationary

$$\gamma_k = \begin{cases} 0 & |k| \le q \\ \sigma^2 \sum_{s=0}^{q-|k|} \theta_s \theta_{s+k} & |k| \le q \end{cases} \qquad (5)$$

We remark that two moving average processes can have the same auto-correlation function. Let

$$Xt = \varepsilon_t + \theta\varepsilon_{t-1} \quad and \quad Xt = \varepsilon_t + \left(\tfrac{1}{\theta}\right)\varepsilon_{t-1}$$

Both have

$$\rho_1 = \frac{\theta}{\left(1+\theta^2\right)} \quad , \rho_k = 0 \quad |k| \succ 1 \qquad (6)$$

However, the first gives

$$\varepsilon_t = X_t - \theta\varepsilon_{t-1} = X_t - \theta(X_{t-1} - \theta\varepsilon_{t-2}) = X_t - \theta X_{t-1} + \theta^2 X_{t-2} - \ldots \qquad (7)$$

This is only valid for |θ| < 1, so-called invertible process. No two invertible processes have the same autocorrelation function.

Probably the next simplest model is that constructed by simple linear combinations of lagged elements of a purely random process, { ε_t } with $\mathrm{E}(\varepsilon_t) = 0$.
A moving average process {X_t} of order q is defined by :

$$X_t = \beta_0\varepsilon_t + \beta_1\varepsilon_{t-1} + \ldots + \beta_q\varepsilon_{t-q} = \sum_{i-0}^{q} \beta_i \varepsilon_{t-i} \qquad (8)$$

and the shorthand notation is MA(q). Usually with a newly defined process
it is of interest to discover its statistical properties. For an MA(q) process the mean is simple to find (since the expectation of a sum is the sum of the expectations):

$$\mathrm{E}(X_t) = \mathrm{E}\left(\sum_{i=0}^{q} \beta_i \varepsilon_{t-i}\right) = \sum_{i=0}^{q} \beta_i \mathrm{E}\left(\varepsilon_{t-i}\right) = 0 \qquad (9)$$

Because $\mathrm{E}(\varepsilon_t) = 0$ for any r.
A similar argument can be applied for the variance calculation:

$$\mathrm{var}(X_t) = \mathrm{var}\left(\sum_{i=0}^{q} \beta_i \varepsilon_{t-i}\right) = \sum_{i=0}^{q} \beta_i^2 \mathrm{E}\left(\varepsilon_{t-i}\right) = \sigma^2 \sum_{i=0}^{q} \beta_i^2 \qquad (10)$$

$$\sin ce\ \mathrm{var}(\varepsilon_r) = \sigma^2 \quad for\ all\ t.$$

The auto-covariance is slightly trickier to work out.

$$\left.\begin{aligned}\gamma(t) &= \operatorname{cov}(X_t, X_{t-\tau}) \\ &= \operatorname{cov}\Big(\sum_{i=0}^{q}\beta_i\varepsilon_{t-i}, \sum_{j=0}^{q}\beta_j\varepsilon_{t-\tau-j}\Big) \\ &= \sum_{i=0}^{q}\sum_{j=0}^{q}\beta_i\beta_j \operatorname{cov}(\varepsilon_{t-i}, \varepsilon_{t-\tau-j}) \\ &= \sigma^2\sum_{i=0}^{q}\sum_{j=0}^{q}\beta_i\beta_j\delta_{j,i+\tau}\end{aligned}\right\} \quad (11)$$

where $\delta_{u,\upsilon}$ is the Kronecker delta which is 1 for $u=\upsilon$ and zero otherwise

(This arises because of the independence of the ε values. Thus since $\delta_{j,i+\tau}$ is involved only terms in the j sum where $j=i+\tau$ survive).

Hence continuing the summation gives

$$\gamma(\tau) = \sum_{i=0}^{q-\tau}\beta_i\beta_{i+\tau}$$

In other words, the β_j becomes $\beta j+\tau$ and also the index of summation ranges only up to $q-\tau$ since the largest $\beta_{i+\tau}$ occurs for $i=q-\tau$.

The formula for the auto-covariance of an MA(q) process is fascinating:

It is effectively the convolution of { β_i } with itself (an "auto convolution").

One of the most important features of an MA(q) auto covariance is that it is zero for $\tau \succ q$. The reason for its importance is that when one is confronted with an actual time series $x_1, \ldots, x_n$., one can compute the sample auto covariance given by:

$$c(\tau) = \sum_{i=1}^{n-\tau}\left(x_i - \bar{x}\right),\left(x_{i+\tau} - \bar{x}\right) \quad (12)$$

for τ = 0, . . . , n − 1. The sample autocorrelation can be computed as

r(τ) = c(τ)/c(0). If, when one computes the sample auto covariance, it cuts off at a certain lag q, i.e. it is effectively zero for lags of q + 1 or higher, then one can postulate the MA(q) model in (11.5) as the underlying probability model. There are other checks and tests that one can make but comparison of the sample auto covariance with reference values, such as the model auto covariance given in (11.12), is the first major step in the model identification.

Also, at this point one should question what one means by "effectively zero". The sample auto covariance is an empirical statistic calculated from the random sample at hand. If more data in the time series were collected, or another sample stretch used then the sample auto covariance would be different (although for long samples and stationary series the probability of a large difference should be very small). Hence, sample auto covariance and autocorrelations are necessarily random quantities and hence "is effectively zero" translates into a statistical hypothesis test on whether the true autocorrelation is zero or not. Finally, whilst we are on the topic of sample auto covariance notice that at the extremes of the range of τ :

$$c(0) = \sum_{i=0}^{n}\left(x_i - \bar{x}\right)^2$$

Te sample variance will be:

$$c(n-1) = x_1 x_n$$

The lesson here is that c(0), which is an estimate of γ(0), is based on n pieces of information, where a c(n − 1), an estimate of γ(n − 1), is only based on 1 piece of information. Hence, it is easy to see that estimates of sample auto covariance for higher lags are more unreliable for those of smaller lags.

VIII. Autoregressive processes: (AR)

The autoregressive process of order p is denoted AR(p), and defined

$$X_t = \sum_{r=1}^{p} \phi_r X_{t-r} + \varepsilon_t$$

where $\phi_1, \phi_2, ..., \phi_r$ are fixed constants and { ε_t } is a sequence of independent (or uncorrelated) random variables with mean 0 and variance σ^2 .

The AR(1) process is defined by

$$X_t = \phi X_{t-1} + \varepsilon_t$$

To find its auto covariance function we make successive substitutions, to get

$$X_t = \varepsilon_t + \phi_1(\varepsilon_{t-1} + \phi 1(\varepsilon_{t-2} + ...)) = \varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_1^2 \varepsilon_{t-2} + ... \quad (13)$$

The fact that {Xt} is the second stationary follows from the observation that E(Xt) = 0 and that the auto covariance function can be calculated as follows:

$$\left.\begin{aligned} \gamma_0 &= \mathrm{E}(\varepsilon_t + \phi_1\varepsilon_{t-1} + \phi_1^2\varepsilon_{t-2} + ...)^2 = (1 + \phi_1^2 + \phi_1^4 + ...)\sigma^2 \\ &= \frac{\sigma^2}{1-\phi_1^2} \\ \gamma_k &= \mathrm{E}(\sum_{r=0}^{\infty} \phi_1^2 \varepsilon_{t-r} \sum_{s=0}^{\infty} \phi_1^s \varepsilon_{t-k-s}) = \frac{\sigma^2 \phi_1^k}{1-\phi_1^2} \end{aligned}\right\} \quad (14)$$

There is an easier way to obtain these results. That is to multiply equation (13) by Xt−k and taking the expected value, to give

$$\mathrm{E}(X_t X_{t-k}) = \mathrm{E}(\phi_1 X_{t-1} X_{t-k}) + \mathrm{E}(\varepsilon_t X_{t-k})$$

$$thus \quad \gamma_k = \phi_1 \gamma_{k-1} \ , k = 1, 2, ...$$

Similarly, squaring (13) and taking the expected value gives

$$\left.\begin{aligned} &\mathrm{E}(X_t^{\ 2}) = \phi_1 \mathrm{E}(X_{t-1}^{\ 2}) + 2\phi_1 \mathrm{E}(X_{t-1}\varepsilon_t) + \mathrm{E}(\varepsilon_t^2) = \phi_1^2 \mathrm{E}(X_{t-1}^{\ 2}) + 0 + \sigma^2 \\ &and \quad so \quad \gamma_0 = \frac{\sigma^2}{(1-\phi_1^2)} \end{aligned}\right\} \quad (15)$$

More generally, the AR(p) process is defined as

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + ... + \phi_p X_{t-p} + \varepsilon_t$$

Again, the autocorrelation function can be found by multiplying (1.3) by Xt−k, taking the expected value and dividing by γ0, thus producing the Yule-Walker equations

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + ... + \phi_p \rho_{k-p} \ \ , k = 1, 2, ...$$

These are linear recurrence relations, with general solution of the form

$$\rho_k = C_1 \omega_1^{|k|} + ... + C_p \omega_p^{|k|}$$

where $\omega_1, \omega_2, ..., \omega_p$ are the roots of

$$\omega^p - \phi_1 \omega^{p-1} - \phi_2 \omega^{p-2} - .... - \phi_p = 0 \quad (16)$$

and C1, . . . ,Cp are determined by $\rho_0 = 1$ and the equations for k = 1, . . . , p − 1. It is natural to require $\gamma_k \to 0 \ \ as \ \ k \to \infty$, in which case the roots must lie inside the unit circle, that is, |ωi| < 1. Thus there is a restriction on the values of φ1, . . . , φp that can be chosen.

IX. ARMA Models

Both AR and MA models express different kinds of stochastic dependence. AR processes encapsulate a Markov-like quality where the future depends on the past, whereas MA processes combine elements of randomness from the past using a moving window. An obvious step is to combine both types of behaviors into an ARMA(p, q) model which is obtained by a simple concatenation. The process autoregressive moving average process, ARMA(p, q), is defined by:

$$X_t - \sum_{r=1}^{p} \phi_r X_{t-r} = \sum_{s=0}^{q} \theta_s \varepsilon_{t-s} \qquad (17)$$

Where again { ε_t } is white noise. This process is stationary for appropriate φ, θ.

Also if the original process {Yt} is not stationary, we can look at the first order difference process

$$X_t = \nabla Y_t = Y_t - Y_{t-1}$$

or the second order differences

$$X_t = \nabla^2 Y_t = \nabla(\nabla Y)_t = Y_t - 2Y_{t-1} + Y_{t-2} \qquad (18)$$

and so on. If we ever find that the differential process is a stationary process we can look for a ARMA model of that.

The process {Yt} is said to be an autoregressive integrated moving average process, ARIMA(p, d, q), if $X_t = \nabla^d Y_t$ is an ARMA(p, q) process. AR, MA, ARMA and ARIMA processes can be used to model many time series. A key tool in identifying a model is an estimate of the auto covariance function. We can estimate the mean by

$$\bar{X} = \frac{\sum_{t=1}^{T} X_t}{T}$$, the auto covariance by

$$c_k = \hat{\gamma}_k = \frac{\sum_{t=k+1}^{T} \left(X_t - \bar{X}\right)\left(X_{t-k} - \bar{X}\right)}{T} \quad . \qquad (19)$$

and the autocorrelation by $r_k = \hat{\rho}_k = \dfrac{\hat{\gamma}_k}{\hat{\gamma}_0}$

X. Application part

In order to build an appropriate model for time series during the search and through the model of box-Jenkins to predict the daily temperatures in the city of Sulaimani, data has been collected from the recording data of the meteorological directorate of Sulaimani, for the one year period (2012), as illustrated in the table (1) in appendix. A statistical analysis of data was tested to see how they need to convert and complete stability, in contrast, then to obtain a homogeneous and appropriate data. According to that, mean and range of the data has been taken, and then the time series is divided into 12 sub-groups, including 12 observations for each, and then calculating the mean and the range of each sub-group, as shown in the table (1). This is also followed by a figure (1) of the mean against range, which can observe that the values are scattered in a random manner suggesting the extent and independence of the average and thus a homogeneous time series values and no need for a conversion process. In order to know the stationary of the time series, we found the values of auto-correlation coefficients, as it is shown in (table 2) in the Appendix that time series is non-stationary ,so the first difference is taken for the purpose to converting stationary time series, and to see the significance of the correlation coefficients in several periods of time, as well as to see the series containing seasonal effects which is by finding correlation coefficients and partial correlation coefficients . In order to diagnose appropriate model through the existing values in the tables (2 and 3) respectively, all the models are tested, the result is that the appropriate model is ARIMA (1,1,1) and the following form $\nabla W_t = \phi W_{t-1} - \theta a_{t-1} + a_t$.

Table (1) shows the mean and the range of daily temperatures (2012-Sept. 2013) for Sulaimani by sub-groups, each of with size 12, Mean series =20.68, and Variance=104.91.

N	***range***	***mean***	***N***	***range***	***mean***	***N***	***range***	***mean***
1	3.4	7.792	20	3.6	33.613	40	5.45	**19.462**
2	6.06	4.604	21	4.2	33.241	41	7.05	**17.312**
3	6.55	5.483	22	3.4	31.585	42	10.25	**23.025**
4	3.75	7.579	23	4.6	29.837	43	8.05	**21.420**

5	9.5	6.795	24	3.55	27.808	44	8.25	**25.425**
6	9.7	5.862	25	5.4	25.895	45	4.2	**27.541**
7	12.6	11.058	26	9.75	21.854	46	3	**30.850**
8	9.35	14.429	27	2.35	19.145	47	3.6	**32.208**
9	6.65	19.183	28	7.1	15.408	48	4.9	**34.104**
10	7.2	20.458	29	3.15	12.800	49	5.15	**33.083**
11	7.3	23.962	30	4.35	10.450	50	4.86	**32.637**
12	3.4	7.792	31	6.65	8.660	51	5.65	**33.241**
13	7.5	25.263	32	4.25	8.660	52	7	**31.716**
14	7.35	26.095	33	6.25	4.004	53	5.45	**19.462**
15	7.35	31.337	34	4.05	10.404			
16	5.55	31.383	35	8.95	9.217			
17	2.2	30.662	36	3.95	10.420			
18	3.85	34.504	37	7.65	11.943			
19	**3.75**	**35.195**	**38**	**11.45**	**14.729**			

In this model, the parameters that have been estimated make minimum mean square error (3.221) and $\phi = -0.193, \theta = 0.249$, parameters and the model is as follows $W_t = W_{t-1} = -0.193W_{t-1} + 0.249a_{t-1} + a_t$. Through using this model, we can find the estimate and the residual values for January of 2013 .as shown in (table 3) in appendix. The model was also used to forecast efficiently, as a result the auto-correlation coefficient for residual were found as shown in table (4) in appendix. The same process was applied on other 15 days of September 2013.As illustrated in the figure (2).

Table (2) shows auto correlation coefficient after taking the first difference.

K	*rk*	*k*	*rk*	*k*	*rk*	*k*	*rk*
1	0.048	41	0.068	81	0.018	121	**-0.021**
2	-0.195	42	-0.054	82	-0.025	122	**-0.015**
3	-0.079	43	-0.021	83	0.065	123	**-0.001**
4	-0.140	44	0.004	84	0.016	124	**-0.058**
5	-0.089	45	0.021	85	0.004	125	**0.041**
6	-0.008	46	0.024	86	0.085	126	**-0.003**
7	-0.039	47	0.069	87	-0.049	127	**-0.041**
8	-0.118	48	0.004	88	-0.039	128	**0.088**
9	-0.070	49	0.043	89	0.007	129	**0.008**
10	-0.136	50	0.015	90	-0.028	130	**-0.017**
11	-0.001	51	0.083	91	-0.073	131	**-0.041**
12	0.032	52	-0.028	92	-0.001	132	**0.002**
13	-0.027	53	-0.006	93	-0.020	133	**0.024**
14	-0.046	54	-0.003	94	-0.028	134	**-0.038**
15	0.048	55	-0.013	95	-0.033	135	**0.006**
16	-0.018	56	0.042	96	0.009	136	**0.069**
17	0.120	57	0.031	97	-0.015	137	**0.048**
18	-0.083	58	-0.004	98	0.004	138	**-0.041**
19	0.006	59	0.070	99	-0.009	139	**-0.009**
20	-0.044	60	-0.007	100	0.011	140	**-0.029**
21	0.010	61	0.004	101	-0.016	141	**0.000**
22	0.078	62	-0.057	102	0.008	142	**-0.036**
23	-0.033	63	-0.044	103	-0.002	143	**-0.023**
24	0.031	64	-0.006	104	-0.002	144	**0.019**
25	0.023	65	-0.035	105	0.053	145	**-0.058**
26	0.106	66	0.006	106	-0.020	146	**-0.057**
27	0.049	67	0.009	107	-0.075	147	**-0.006**
28	0.063	68	-0.099	108	0.020	148	**0.021**
29	0.001	69	-0.035	109	0.036	149	**-0.101**
30	0.030	70	-0.090	110	-0.009	150	**-0.004**
31	-0.044	71	-0.016	111	-0.043	151	**-0.005**

32	-0.051	72	-0.001	112	-0.011	152	**-0.020**
33	0.126	73	-0.003	113	-0.044		
34	0.157	74	-0.070	114	0.024		
35	0.051	75	-0.004	115	0.060		
36	0.134	76	0.010	116	-0.052		
37	0.044	77	-0.028	117	-0.061		
38	-0.017	78	0.002	118	-0.017		
39	0.001	79	-0.020	119	0.023		
40	**0.021**	**80**	**-0.042**	**120**	**0.015**		

Table (3) Shows partial auto correlation coefficient after taking the first difference.

k	***rk***	***k***	***Rk***	***k***	***rk***	***k***	***rk***
1	0.048	41	0.057	81	0.037	121	**-0.032**
2	-0.193	42	-0.084	82	0.019	122	**-0.022**
3	-0.096	43	-0.059	83	0.044	123	**-0.006**
4	-0.105	44	-0.029	84	0.054	124	**-0.060**
5	-0.058	45	0.055	85	-0.020	125	**0.010**
6	0.046	46	0.065	86	0.009	126	**0.066**
7	0.017	47	0.065	87	-0.072	127	**-0.002**
8	-0.089	48	0.021	88	-0.076	128	**-0.001**
9	-0.048	49	-0.001	89	-0.021	129	**0.027**
10	-0.077	50	0.011	90	0.018	130	**-0.014**
11	0.045	51	0.064	91	-0.065	131	**-0.018**
12	0.116	52	-0.035	92	0.037	132	**0.043**
13	0.019	53	-0.054	93	0.059	133	**0.005**
14	-0.046	54	-0.044	94	-0.037	134	**-0.041**
15	0.042	55	-0.051	95	0.006	135	**-0.027**
16	-0.010	56	0.058	96	0.028	136	**0.014**
17	0.091	57	0.033	97	0.013	137	**0.027**
18	-0.048	58	0.014	98	0.034	138	**-0.073**
19	-0.047	59	0.039	99	0.000	139	**-0.037**
20	-0.049	60	0.010	100	0.012	140	**-0.007**
21	-0.009	61	-0.013	101	-0.010	141	**0.001**
22	0.082	62	-0.006	102	-0.041	142	**0.013**
23	-0.006	63	-0.007	103	-0.023	143	**0.023**
24	0.017	64	-0.026	104	-0.018	144	**0.011**
25	-0.008	65	-0.057	105	0.062	145	**-0.028**
26	0.071	66	0.005	106	0.003	146	**-0.050**
27	0.022	67	0.059	107	-0.038	147	**-0.029**
28	0.035	68	-0.039	108	0.018	148	**0.076**
29	-0.022	69	-0.018	109	0.050	149	**-0.003**
30	-0.042	70	-0.017	110	-0.004	150	**-0.024**
31	-0.088	71	0.029	111	-0.066	151	**0.007**
32	-0.066	72	0.039	112	0.018	152	**-0.003**
33	0.127	73	0.077	113	-0.057		
34	0.183	74	-0.022	114	0.000		
35	-0.015	75	-0.005	115	0.059		
36	0.008	76	0.010	116	-0.040		
37	-0.020	77	-0.043	117	-0.029		
38	-0.109	78	0.006	118	0.021		
39	-0.007	79	0.003	119	0.025		
40	**0.033**	**80**	**-0.027**	**120**	**0.044**		

Table (4) shows the actual and the forecast data for 15 days of September 2013

time	*actual data*	*forecast data*	*error*
1	29.3	28.31	**0.99**
2	27.8	25.9	**1.9**
3	27.5	26.1	**1.4**
4	26.7	24.4	**2.3**
5	28	23.8	**4.2**
6	27.3	26.9	**0.4**
7	28.2	27.3	**0.9**
8	23.9	20.2	**3.7**
9	21.8	19.5	**2.3**
10	24.6	20.2	**4.4**
11	32.4	29.32	**3.08**
12	33.2	28.46	**4.74**
13	32	30.23	**1.77**
14	31.4	31.22	**0.18**
15	**33.1**	**31.29**	**1.81**

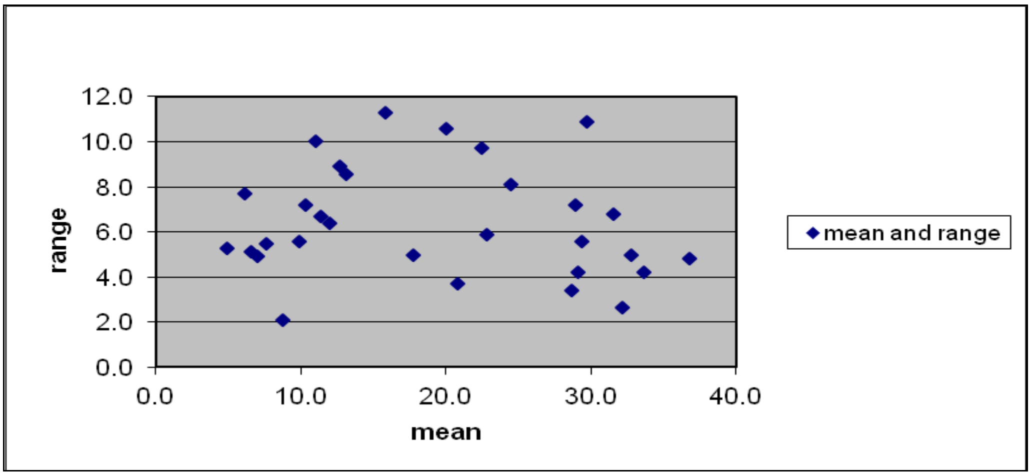

Figure (1) Mean and Range

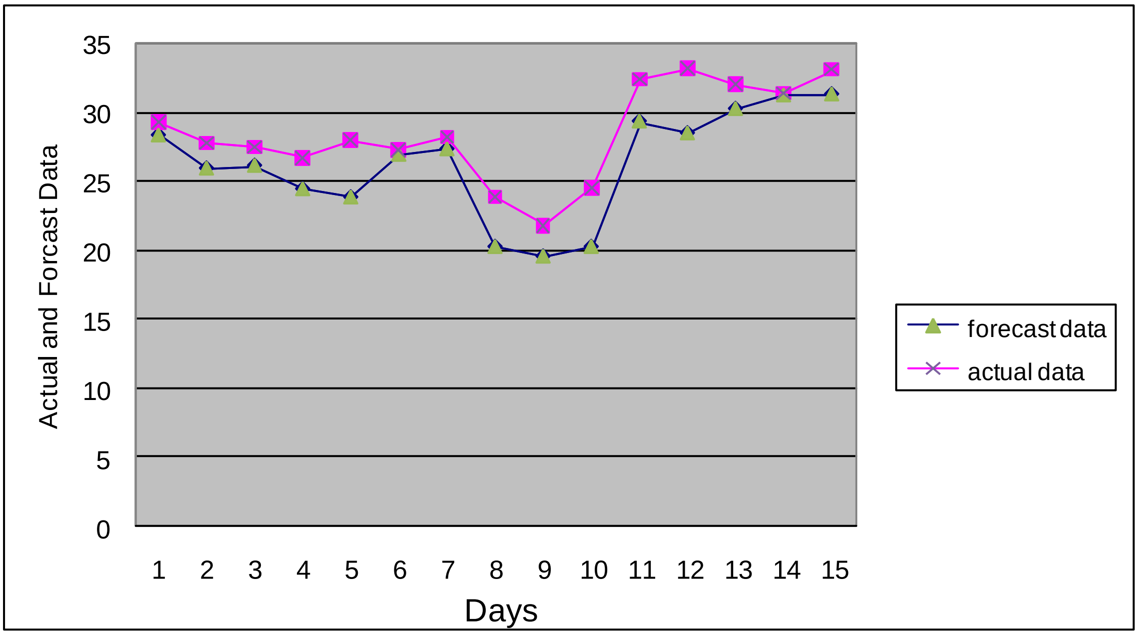

Figure (2) represent actual and forecast data for 15 days of September 2013

According to what is stated in this applied study, which is to build an auto-regression model for time series through the box-Jenkins models, the researcher has reached a number of conclusions and same recommended points.

XI. conclusions

1) Through the values of auto-correlation coefficients in (Table 2) in Appendix, it has been observed that daily temperature in the city of Sulaimani makes up the non-stationary time series.

2)According to the values that have been found from the auto-correlation coefficients and partial auto-correlation coefficients based on the table(2 and 3) respectively, in appendix, it has been observed that the appropriate model is the model ARIMA(1,1,1), which can be written in the following form: $\nabla W_t = \phi W_{t-1} - \theta a_{t-1} + a_t$.

3)Through the Figure 2, we have found that the model mentioned in the previous paragraph has given good estimates close to the actual values.

Recommendations:
Through the reported findings, the researcher recommends:
1) The Meteorological Department and the relevant authorities can depend on different statistical models to forecast weather.
2) Different box-Jenkins models can be used to predict daily temperature in different regions of Kurdistan.

References

[1] Anderson, "Forecasting "new York, North Holland. (1980).
[2] 2)Box,G.E. and Jenkins, " Time series Analysis Forecasting and control "san Francisco holdenday ,1979.
[3] Box, G.E.P., and D.A. Pierce (1970). "Distribution of Residual Autocorrelations in Autoregressive-integrated Moving Average Time Series Models," Journal of the American Statistical Association, 65, 1509-1526.
[4] Granger, C.W.J. and M.J. Morris (1976). "Time Series Modeling and Interpretation," Journal of the Royal Statistical Society, Series A, 139, 246-257.
[5] Hamilton, J.D. (1994). Time Series Analysis. Princeton University Press, New Jersey.
[6] Harvey, A.C. (1993). Time Series Models, Second Edition. MIT Press, Massachusetts.
[7] Hayashi, F. (2000).Econometrics. Princeton University Press, New Jersey.
[8] Johnston, Jack & John DiNardo.1997.Econometric Methods. New York McGraw Hill,Inc.
[9] L jung, T. and G.E.P. Box (1979). "The Likelihood Function for a Stationary Autoregressive Moving Average Process," Biometrika,66, 265-270.
[10] Wei, William W.S. "Time series Analysis "2nd (2006), new York .united states of America. Holden Day, San Francisco.

Appendices

Table (1) Daily temperatures in the city of Sulaimani, (2012 Sept.2013).

6.3	7.6	13.2	25.3	28.3	36.5	34.4	24.8	15.8	10.4	8.2	14.6	13.5	14.0	29.3	33.1	34.0
8.3	7.6	12.7	25.5	28.4	37.0	33.1	26.9	14.2	10.3	7.8	14.4	17.2	15.8	28.0	34.3	33.2
8.0	8.1	13.1	25.2	31.3	34.9	33.3	25.9	13.5	8.5	9.4	10.6	10.3	10.4	28.8	35.2	31.3
7.4	8.6	13.2	23.5	32.4	35.8	30.8	24.3	12.3	9.2	8.1	8.3	10.6	9.5	30.2	34.7	
7.1	9.9	12.9	20.5	33.2	35.7	30.1	25.8	13.5	7.8	6.2	10.1	14.2	12.1	34.3	32.5	
7.0	6.8	15.0	22.2	32.9	35.2	31.3	26.8	13.6	9.5	8.8	12.3	15.3	13.8	33.1	33.3	
7.8	7.5	10.5	22.5	34.5	35.2	30.0	24.2	14.3		6.6	11.9	15.8	13.8	30.7	33.0	
7.2	8.2	11.4	22.0	35.8	36.1	29.0	23.7	15.2		6.9	10.9	14.6	12.7	29.9	35.3	
9.2	4.3	13.1	23.0	33.8	37.4	32.3	23.4	13.8		6.4	10.3	15.2	12.7	30.3	31.2	
8.5	3.4	11.4	23.9	29.8	35.6	33.4	24.7	13.5		1.0	10.3	18.9	14.6	29.8	30.1	
7.2	4.6	10.6	26.3	28.3	34.3	33.5	26.2	11.9		0.7	9.0	20.9	14.9	29.9	30.6	
9.7	3.2	13.7	27.8	28.9	33.6	31.3	23.8	12.2		1.2	10.0	21.4	15.7	31.2	32.5	
7.1	6.4	17.3	27.1	31.8	36.1	28.9	24.9	11.6		3.0	10.8	21.0	15.9	30.5	31.2	
5.5	5.6	18.3	28.3	32.0	34.1	31.2	23.9	11.1		4.0	10.9	19.6	15.3	31.8	31.2	
6.1	6.9	19.3	26.3	32.2	34.1	31.8	21.3	11.3		4.5	11.7	15.5	13.6	33.0	35.5	

6.4	8.1	19.9	21.5	31.3	35.3	29.2	18.9	11.9	4.8	12.0	17.3	14.6	30.8	35.0
6.7	7.6	19.0	22.2	31.7	35.6	30.7	20.0	12.5	3.4	13.3	20.1	16.7	31.1	32.9
5.6	8.5	18.9	24.8	31.6	34.9	29.5	19.7	13.2	5.8	12.2	22.9	17.5	33.7	32.6
6.0	10.6	20.2	24.5	33.3	32.0	29.7	16.5	10.5	7.2	12.8	22.4	17.6	32.0	34.5
2.9	12.7	21.3	25.9	32.3	33.0	31.5	19.3	10.5	9.9	14.0	17.8	15.9	31.8	32.0
-0.1	6.9	23.2	29.0	31.5	33.3	32.0	18.2	10.4	9.4	10.9	16.6	13.7	34.0	31.4
2.4	4.5	18.2	27.3	31.0	32.2	28.7	17.5	8.9	9.9	11.5	18.3	14.9	34.4	32.5
3.0	2.4	16.5	23.7	29.5	32.9	27.9	17.2	9.5	11.7	16.0	19.1	17.5	31.4	33.3
4.0	0.5	17.6	22.8	29.6	33.2	27.2	18.2	9.8	12.4	13.2	19.1	16.2	30.9	35.3
4.3	1.5	18.4	24.4	30.3	34.9	26.8	18.1	10.5	11.0	10.4	19.1	14.7	31.6	34.6
6.8	3.5	18.5	25.5	31.0	33.0	27.1	19.3	11.5	11.4	5.6	20.2	12.9	32.2	33.3
9.1	7.5	19.6	23.6	29.3	33.8	28.9	19.8	9.3	11.9	8.1	19.8	13.9	33.7	34.8
4.9	10.5	19.0	24.0	30.4	34.8	27.7	20.4	9.0	12.3	9.8	19.0	14.4	31.9	36.7
6.8	9.5	18.5	24.2	31.0	34.4	27.7	21.5	8.8	9.7	12.4	18.6	15.5	32.8	31.8
4.0	9.5	20.8	25.0	31.5	31.5	26.7	21.4	9.1	8.3	13.1	17.3	15.2	35.2	33.7
2.5	6.8	17.4	26.4	31.3	31.3	26.2	19.6	7.5	9.0	12.6	14.1	13.3	36.8	31.3
5.8	7.8	18.0	28.3	31.7	32.1	26.4	18.9	5.5	5.2	17.1	13.2	15.1	35.3	31.7
6.2	13.4	19.0	26.5	31.9	32.0	26.1	19.4	7.2	5.6	19.8	13.7	16.8	32.5	31.0
4.9	15.5	19.3	25.8	32.6	31.4	29.2	19.2	5.4	5.4	19.6	14.9	17.2	33.8	31.8
3.4	13.9	20.0	28.3	33.0	33.0	30.6	20.8	10.3	7.1	21.6	16.7	19.2	34.3	32.1
7.4	9.1	21.0	31.4	34.9	33.2	30.7	18.8	12.1	8.3	21.2	19.4	20.3	35.9	33.4
6.3	5.8	22.3	32.3	32.8	34.6	29.1	14.9	9.7	9.0	11.7	19.6	15.7	33.9	30.2
6.2	2.9	21.7	29.8	33.3	35.5	27.7	12.4	9.2	8.9	10.0	25.0	17.5	33.5	29.7
7.3	7.8	23.3	28.8	34.8	34.7	26.3	11.0	9.1	11.1	13.0	26.7	19.8	32.8	30.4
7.2	12.3	24.6	28.6	36.8	35.6	25.5	12.9	10.3	12.3	12.6	27.0	19.8	31.8	31.4

Table (2) Correlation coefficient.

LAG	ACF	LAG)	ACF	LAG	ACF
1	0.981283	52	0.495725	103	-0.20173
2	0.961147	53	0.479264	104	-0.21273
3	0.94704	54	0.464623	105	-0.22289
4	0.936336	55	0.45206	106	-0.23495
5	0.929103	56	0.44156	107	-0.24638
6	0.923845	57	0.428761	108	-0.25678
7	0.917171	58	0.415333	109	-0.26814
8	0.909232	59	0.402009	110	-0.28047
9	0.904855	60	0.387998	111	-0.29393
10	0.901917	61	0.372261	112	-0.30526
11	0.901277	62	0.355887	113	-0.31745
12	0.899624	63	0.339266	114	-0.32719
13	0.893691	64	0.322769	115	-0.33625
14	0.886351	65	0.307301	116	-0.34667
15	0.880945	66	0.293818	117	-0.35613
16	0.873425	67	0.281502	118	-0.36493
17	0.866575	68	0.267933	119	-0.37376
18	0.856783	69	0.255081	120	-0.3833
19	0.848407	70	0.243019	121	-0.39418
20	0.840697	71	0.231082	122	-0.40434
21	0.834215	72	0.21828	123	-0.41416
22	0.828689	73	0.205698	124	-0.42416
23	0.820826	74	0.191094	125	-0.43284
24	0.813255	75	0.176757	126	-0.4412
25	0.804752	76	0.161562	127	-0.45072
26	0.797356	77	0.145139	128	-0.46025

27	0.788137	78	0.129023	129	-0.46906
28	0.776863	79	0.113613	130	-0.47826
29	0.764777	80	0.09977	131	-0.48628
30	0.753486	81	0.087215	132	-0.49328
31	0.743212	82	0.073468	133	-0.50198
32	0.736282	83	0.058953	134	-0.51114
33	0.731866	84	0.043546	135	-0.5196
34	0.723121	85	0.026227	136	-0.52836
35	0.707894	86	0.010209	137	-0.53755
36	0.693375	87	-0.00737	138	-0.54712
37	0.678691	88	-0.02259	139	-0.55418
38	0.664554	89	-0.03525	140	-0.55971
39	0.654313	90	-0.04735	141	-0.56431
40	0.643905	91	-0.06035	142	-0.5692
41	0.632429	92	-0.07052	143	-0.57502
42	0.619249	93	-0.08065	144	-0.58198
43	0.609098	94	-0.09246	145	-0.5897
44	0.601153	95	-0.10283	146	-0.59635
45	0.594257	96	-0.11362	147	-0.60135
46	0.584746	97	-0.12545	148	-0.60487
47	0.573201	98	-0.1383	149	-0.61048
48	0.559375	99	-0.15172	150	-0.6165
49	0.54422	100	-0.16448	151	-0.62165
50	0.528953	101	-0.1768	152	-0.62667
51	0.513779	102	-0.1897		

Table (3) Shows actual data and residual for temp.(January 2012).

T	Wt	wt	E	t	Wt	Wt	E
2	6.25	7.047	-0.797	17	6.35	5.509	0.841
3	8.25	6.713	1.537	18	6.65	4.973	1.677
4	8	7.923	0.077	19	5.6	5.499	0.101
5	7.4	8.594	-1.194	20	6	5.957	0.043
6	7.1	8.012	-0.912	21	2.85	6.466	-3.616
7	7	8.114	-1.114	22	-0.1	3.547	-3.647
8	7.8	6.879	0.921	23	2.35	-0.073	2.423
9	7.15	7.105	0.045	24	3	0.135	2.865
10	9.2	6.968	2.232	25	3.95	1.042	2.908
11	8.5	9.195	-0.695	26	4.25	1.612	2.638
12	7.2	9.360	-2.160	27	6.75	11.123	-4.373
13	9.65	7.692	1.958	28	9.05	10.675	-1.625
14	7.1	9.454	-2.354	29	4.85	8.503	-3.653
15	5.45	7.890	-2.440	30	6.75	3.500	3.250
	6.05	5.649	0.401	31.00	4	12.352	-8.352