Forecasting Infections by Tuberculosis in the Nile River State "Sudan"

Dr. Abdalla Ahmed[1], Alkhalifa Abdalla[2]

Department of Statistics-Faculty of Science-Tabuk University (KSA)

Abstract: *The aim of this paper is to find a time series model to forecast the number of infections by Tuberculosis in the River Nile state "Sudan". The source of data is the national program to fight tuberculosis in the ministry of health and population of the Nile river state "Sudan". Box and Jenkins models were used to find the model. Analysis done by Minitabpackage. The paper concluded that, the Cumulative Incidence rate of Tuberculosis for the period studied is 0.15 and the appropriate model to estimate infections of Tuberculosis in the Nike river stat is moving average model "MA" and the forecast of infections by Tuberculosisfor the two coming years ranging from 74 to 140 case.*

Keywords: tuberculosis,, incidence rate, moving average, autoregressive, moving average-autoregressive.

1. Introduction

No doubt, thatTuberculosis is one of the major epidemics in Sudan, has negative economic effects on the country, problem arises from that in the River Nile state "Sudan" there is no significant declining in the number of infection, in 2007 there was457 infection and in 2013, there was 426infection. Studyaims to find out a time series model based on quarterly data for the infection in the state for the period 2007-2013.

The importance of this study is that it enables to develop requirements for the expected number of infected people, in addition to know the effectiveness of techniques used to reduce the infection by comparing the predicted number with the actual number. For descriptive and statistical analysis of dataBox _ Jenkins modelswere used.

1.1 Theoretical framework of the study

Definition of time series
Time series is a sequence of numerical data points in successive order, usually occurs in uniform intervals, or a time series is simply a sequence of numbers collected at regular intervals over a time of period.

Box Jenkins forecasting methodology
There are four steps before starting using the box-Jenkins models**:**
1) Checking if the series is stationary or not and transform it to stationary series if it is not.
2) Determination of the orders for auto regression and moving average models, using autocorrelation function ACF and partial autocorrelation function PACF.
3) Estimation of the model parameters and make sure they are statistically significant.
4) Test the validity of the model and check whether it's appropriateand reliable for predicting.

Models of box-Jenkins
Auto- regression model (AR):
The model take the form:
$$(1) \quad Z_t = \delta + \Phi_1 Z_{t-1} + ... + \Phi_p Z_{t-p}$$
Where:
$Z_t$represents the value of the seriesat time period t

$_\delta$is constant
Φ_i are Parameters of auto- regression

2. Moving average Model (MA)

The model takes the following form:
$$Z_t = a_t + \theta_1 a_{t-1} - ... - \theta_p a_{t-p}$$
Where:
$.a_t$A series of white noise which it random sequence with zero mean and a constant variance. . θ_i are Parameters of moving averages

3. Auto regression and moving average model (ARMA)

It is a mixed model of auto-regression and moving averages models and moving average, and takes the following form:
$$Z_t = \delta + \Phi_1 Z_{t-1} + ... + \Phi_p Z_{t-p} + a_t + \theta_1 a_{t-1} - ... - \theta_p a_{t-p}$$

3.1 Stationarity

Stationarity" is a fundamental property underlying almost all time series statistical models. To convert time series to stationary one of the following methods used:
1. Differencing continues until stationarity achieved. In practice, it is almost never necessary to go beyond second difference, because real data generally involve only first or second level non-stationarity.
2. Taking log
3. Taking square root

3.2 Model selection

Table 1: Auto correlation function "ACF" and PartialAuto correlation function "PACF" for different models

Model	ACF	PACF
MA(q)	Cuts off after ρk	Dies down
AR(p)	Dies down	Cuts off afterϕ_k
ARMA(p,q)	Dies down	Dies down

Whereρ_k is the partial correlation at lag k, andΦ_k is auto partial correlation at lag k.

3.3 Parameter Estimation

- The method of least squares can be used. However, for models involving an MA component, there is no simple formula that can be applied to obtain the estimates.
- Another method that is frequently used is

3.4 Diagnostic Checking

The model that finally chosen is the one considered best based on a set of diagnostic checking criteria. These criteria include
(1) t-tests for coefficient significance
(2) Residual analysis
(3) Model selection criteria

3.5 Forecasting:

- The h-period ahead forecast based on an ARMA(p,q) model is given by

$$\hat{y}_{t+h} = \hat{\delta} + \hat{\phi}_1 y_{t+h-1} + ...\hat{\phi}_p y_{t+h-p} + e_{t+h} - \hat{\theta}_1 e_{t+h-1} - .....\hat{\theta}_q e_{t+h-q}$$

Where elements on the r.h.s. of the equation may be replaced by their estimates when the actual values are not available.

3.6 Data description:

Males at risk are higher than females at risk although the dispersion of females is less. The differences between. Total numbers of people at risk at the period are very small, except in 2008.

3.7 Incidence rate of Tuberculosis:

Table 2: Incidence rate of Tuberculosis

year	Population at risk	cases of Tuberculosis	Cumulative cases of Tuberculosis	Incidence rate		Cumulative Incidence rate	
				For the year	From the start	For the year	From the start
2007	285	111	111	0.48	0.48	0.39	0.39
2008	174	119	230	0.51	0.79	0.68	0.41
2009	293	30	260	0.11	0.95	0.10	0.30
2010	263	1	261	0.01	0.95	0.00	0.23
2011	264	3	264	0.01	0.97	0.01	0.17
2012	261	27	291	0.10	1.02	0.10	0.17
2013	288	3	294	0.01	1.02	0.01	0.15
Total for period	294	294	294		1.02		0.15

3.8 Application

C1: represent original series
C4: represents series after taking the first difference
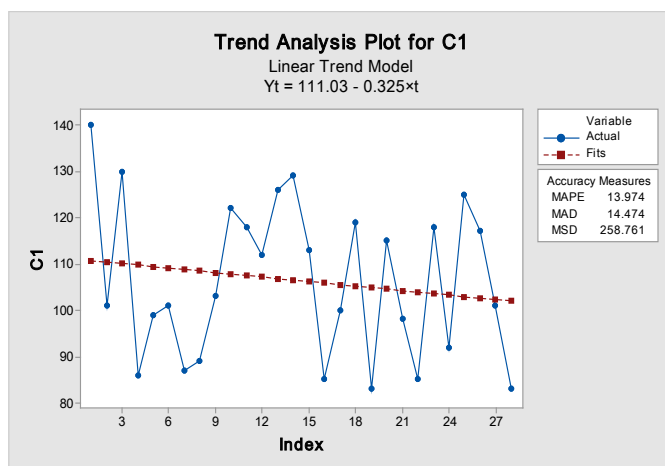**Checking stationarity**:

Figure 1: trend analysis plot for c1

Figure 1 shows that the series affected by general trend, so it's not stationary. By taking the first difference and graph data as in figure 2 below:
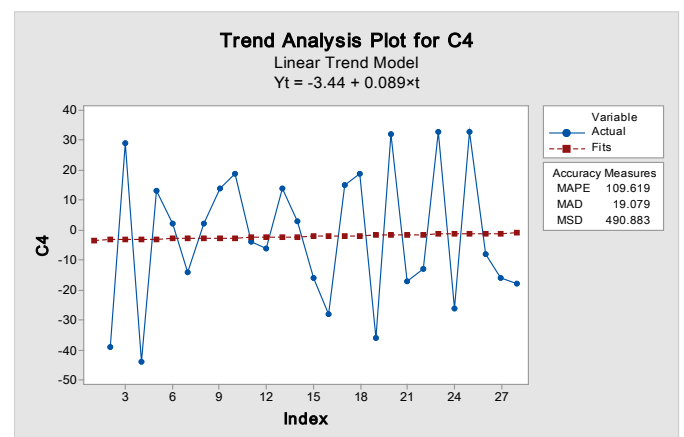

Figure 2: trend analysis plot for c4

Figure 2: indicate that the series is stationary.

Diagnostic stage:
To determine the appropriate model autocorrelation and partial autocorrelation drawn as in figure 3 and figure 4.
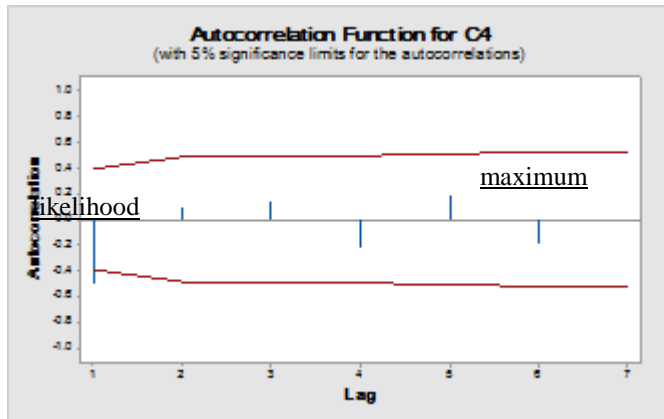
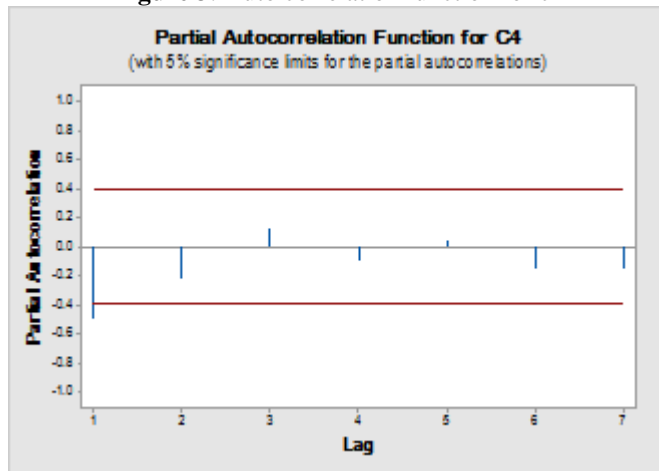**Figure 3**: Auto correlation function for c4

Figure 4: Partial Auto correlation function for c4

From figure 3 : ACF cuts after one lag and from figure 4 PACF dies down, indicating that the proposed model is MA

$$Z_t = a_t + \theta_1 a_{t-1} - ... - \theta_p a_{t-p}$$

Output:
Final Estimates of Parameters

Type Coef SECoef T P
MA 1 1.0742 0.1446 7.43 0.000
Constant -0.5884 0.2032 -2.90 0.008
Mean -0.5884 0.2032

Number of observations: 27
Residuals: SS = 6050.29 (backforecasts excluded)
 MS = 242.01 DF = 25

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag 12 24 36 48
Chi-Square 9.7 21.9 * *
DF 10 22 * *
P-Value 0.465 0.463 * *

And the above output support the model, since p-value is 0.00 sowing the significant of the model, also
The p-values for the modified Box-Pierce all are well above .05, indicating "non-significance." This is a desirable result.

Testing model validity:

Figure 5 and **figure 6** below showing that all ACF and PACF for residuals are within the intervals, which means the validity of the model.
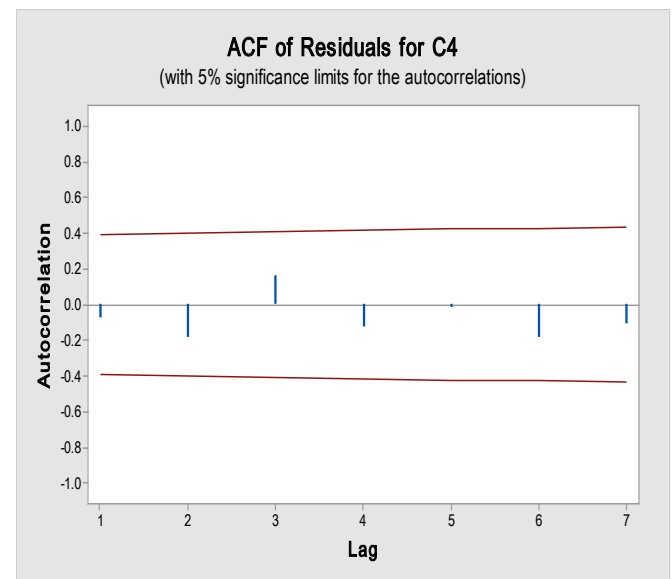

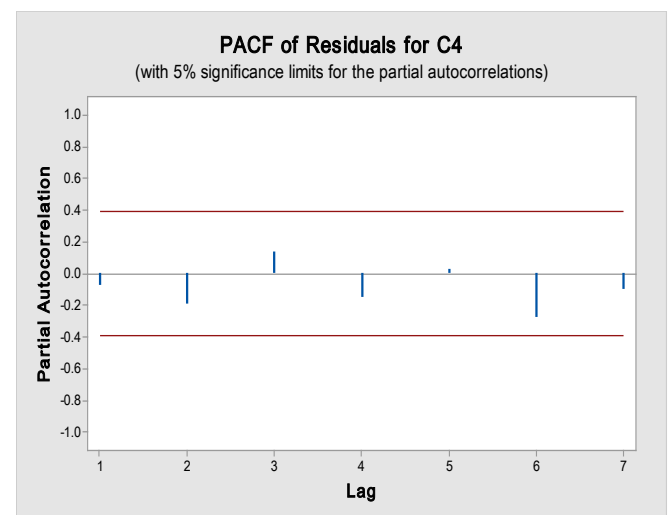**Figure 5:** ACF of residuals for c4

Figure 6: PACF of residuals for c4

4. Results

1. The Cumulative Incidence rate of Tuberculosis for the period studied is 0.15.
2. The appropriate model to estimate infections ofTuberculosis in the Nike river stat is MA.
3. The forecast of infections by Tuberculosisfor the two coming years ranging from 74 to 140 case.

5. Recommendation

1. According to previous results, Tuberculosis in the Nile river state is under control, but more attention must be given to reduce the incidence of the disease.
2. MA model should be use to compare predicted number with actual number to evaluate the efforts used for reduction of the disease.

Paper ID: SUB15545
1463

References

[1] Box, G.E.P and Jenkins, G.M., "Time Series Analysis Forecasting andcontrol", Holden day, London. (1976)

[2] Box, G.E.P. & Pierce, D.A., "Distribution of the Residual

[3] Autocorrelation in Autoregressive-integrated moving Average Time SeriesModels", JASA, VOL.65, P. (1520-1526). (1970),

[4] Douglas,C.M&Contreas, J.GNote On Forecasting With Adaptive

[5] Filtering, O.P.Q,Vol.24,No.4,P(87-90). .,(1976),"

[6] Kaiser, R. and Maravall, A. "Notes on Time Series Analysis ARIMA Modelsand Signal Extraction", Benco de EsponaservicioEstudios(2001),

[7] NunoCrato, "Some Results on the Spectral Analysis of stationary TimeSeries" Portugal Mathematic Vol. 53. Fasc. 2-1996.