

Citation: L'Hadj, M., Selmane, S., "Time Series Analysis of Tuberculosis in Medea Province in Algeria".
Journal of Engineering Technology and Applied Sciences 4 (2) 2019 : 85-94.

TIME SERIES ANALYSIS OF TUBERCULOSIS IN MEDEA PROVINCE IN ALGERIA**

Mohamed L'Hadj^a , Schehrazad Selmane^{b*} 

^a *Health and Hospital Reform Services, Ministry of Health, Population and Hospital Reform.*
Algiers, Algeria
lhadjmoh@yahoo.fr

^b *Faculty of Mathematics, University of Sciences and Technology Houari Boumediene.*
Algiers, Algeria
*cselmane@usthb.dz (*corresponding author)*

Abstract

Despite Algeria has been able to join the group of countries with moderate tuberculosis (TB) prevalence since the 1980s, the disease remains one of the major public health issues in the country. Over the past decade, the annual incidence rate has hovered around 55 per 100 000 people. The incidence rate remains, however, very high in some provinces. The aim of this study was to describe the temporal patterns of TB in Médéa province which records the highest incidence rate in the country. In this retrospective study, the monthly pulmonary TB (PTB) and extrapulmonary TB (EPTB) data from 2008 to 2017, extracted from the national surveillance system, were analyzed and seasonal fluctuation was examined. The Box-Jenkins approach to fit seasonal autoregressive integrated moving average (SARIMA) model to the monthly PTB and EPTB notification data from 2008 to 2016 was performed. The models were used to predict the monthly cases of PTB and EPTB for the year 2017. The models were found to be adequate. Our findings indicate that SARIMA models are useful tools for monitoring and for predicting trends of TB incidence in Médea province.

Keywords: Algeria, biostatistics, Box-Jenkins, seasonal autoregressive integrated moving average, times series analysis, tuberculosis

1. Introduction

Tuberculosis (TB) is an infectious disease caused by a bacterium called Mycobacterium Tuberculosis (MTB). This bacterium was discovered by Robert Koch in 1882. The bacteria spread through the air when a person with active TB of the lungs coughs, sneezes, or talks. The probability of TB transmission per contact is low when compared of that of airborne

**Part of this study was presented in ICFAS2019 as short abstract.

transmitted diseases such influenza, but nonetheless the infectious period of TB is longer. The risk of infection is substantial only for individuals who have frequent and close contacts with a person with active TB. The bacteria usually attack the lungs, causing pulmonary TB (PTB). But, these bacteria can also attack any other part of the body, causing extrapulmonary TB (EPTB). TB can be cured in almost all cases, if treated; its treatment takes six to nine months and sometimes longer. Immunization with the Bacillus Calmette-Guérin (BCG) vaccine is used in many parts of the world. This vaccine is 70-80% effective against the most severe forms of TB, such as TB meningitis, however, it is less effective in preventing the form of TB that affects the lungs [1-3].

TB is the top infectious killer in the world and it is one of the top 10 causes of death worldwide. In 2017, 1.6 million people died from TB and 10 million people fell ill with TB. Yet, 87% of new TB cases occurred in the 30 high TB burden countries. Two-thirds of the new TB cases are reported in eight countries, namely, India, China, Indonesia, the Philippines, Pakistan, Nigeria, Bangladesh and South Africa. An estimated 54 million lives were saved through TB diagnosis and treatment between 2000 and 2017. Globally, TB incidence is falling at about 2% per year. Multidrug-resistant TB (MDR-TB) remains a public health issue. Only one in four people had access to MDR-TB treatment. Ending the TB epidemic by 2030 is among the health targets of the Sustainable Development Goals. However, funding gaps for research and also for the implementation of Directly Observed Treatment Short course (DOTS) in some affected regions may be an obstacle to achieving this target [1].

Tuberculosis is one of the main public health issues in Algeria. Since 1962, the disease has become a health priority and benefits from a national control programme which is based on recommendations of the World Health Organization (WHO) and the International Union against Tuberculosis and Lung Disease. Algeria has been able to join the group of countries with moderate prevalence since the 1980s. The annual incidence rate has hovered around 55 per 100000 people. The incidence rate remains, however, very high in some provinces, exceeding 100 per 100000 people [4].

This study aimed to describe the temporal patterns of TB in Médéa province that records a high incidence rate in Algeria. In this retrospective study, the monthly PTB and EPTB data from 2008 to 2017, extracted from the national surveillance system, were analysed and seasonal fluctuation was examined. The Box-Jenkins approach to fit seasonal autoregressive integrated moving average (SARIMA) model to the monthly PTB and EPTB notification data from 2008 to 2016 was performed. The models were used to predict the monthly cases of PTB and EPTB for the year 2017.

Section 2 of this paper, describes the materials and methodology. Section 3 investigates the results of the data analysis and their discussion. Section 4 provides some concluding remarks.

2. Materials and methods

2.1 The study Region: Médéa province

Médéa province is located in the heart of the Tellian Atlas, in northern Algeria. The province covers a land size of 8776 square kilometres. As 2017, the total resident population of the province reached 834809. Its climate is warm and temperate with an average annual temperature of 14.4 °C and an average annual rainfall of 736 mm [5].

Table 1. Monthly notified cases of pulmonary TB (P) and extrapulmonary TB (EP).

Year	2008		2009		2010		2011		2012		2013		2014		2015		2016		2017	
Month	P	EP	P	EP	P	EP	P	EP	P	EP	P	EP	P	EP	P	EP	P	EP	P	EP
Jan	7	31	5	29	11	26	10	48	12	56	15	57	10	38	7	96	14	74	12	67
Feb	14	46	10	42	9	33	11	50	12	40	12	71	5	46	6	70	13	72	10	102
Mar	15	37	22	53	26	34	15	58	9	56	10	55	17	50	13	81	9	88	14	110
Apr	22	44	25	40	17	55	14	82	12	90	16	55	15	67	21	91	11	75	12	119
May	15	46	23	49	9	52	23	93	13	59	17	46	14	66	22	81	10	75	12	106
Jun	15	35	25	46	20	36	12	67	18	44	11	57	9	51	12	88	7	71	8	79
Jul	22	34	15	40	16	33	11	70	21	49	19	71	16	39	14	55	8	33	15	79
Aug	22	32	23	23	24	30	18	57	10	52	14	28	18	46	15	54	15	56	4	73
Sep	18	26	24	20	23	24	11	48	11	52	22	53	12	49	15	50	10	41	7	55
Oct	14	29	14	31	9	35	19	63	13	70	19	34	16	57	7	56	10	68	16	83
Nov	4	39	14	25	8	33	20	46	7	44	14	23	11	82	11	61	7	39	9	82
Dec	9	40	19	27	11	40	13	39	17	52	12	29	9	62	3	53	2	50	7	51

2.2 Data

The data sets of notified cases of PTB and EPTB used comprise 120 months from January 2008 to December 2017 and are displayed in Table 1. They were provided by the Ministry of Health, Population and Hospital Reform. This study was approved by the National Council on Ethics of Health Sciences, Algeria, and the Committee of National Tuberculosis Control Programme.

2.3 Box-Jenkins method

Time series modelling allows to study the collected past observations and then to develop an appropriate model which describes the inherent structure of the series of observations. The built model is thereafter used to make forecasts. Thus, time series forecasting allows predicting the future by understanding the past. One of the most popular and frequently used stochastic time series models is the Autoregressive Integrated Moving Average (ARIMA) model. If the series is observed to be seasonal of period S , Box and Jenkins had proposed a quite successful variation of ARIMA model, namely, the Seasonal ARIMA (SARIMA). The Box-Jenkins methodology is used in this study [6-7].

In adequate modelling, time series should be stationary. Often, in practice, a time series is not stationary, that is, has a non-stationary variance, or has a non-stationary mean, or has periodic or seasonal components. To stabilise the variance, various transformations can be applied to each observation X_t ($t = 1, \dots, n$) such as the logarithm, the square root, the reciprocal. To stabilise the mean, differencing to an appropriate order could render a non-stationary series to a stationary one. A $(p, d, q) \times (P, D, Q)_S$ SARIMA model is defined by the following difference equation:

$$A(L)\Phi(L^S)\nabla^d \nabla_S^D X_t = B(L)\Theta(L^S)\varepsilon_t \quad (1)$$

where the backward shift operator L is defined by $L^k X_t = X_{t-k}$ and where ∇^d is the d^{th} difference of X_t , ∇_S^D is the D^{th} seasonal difference operator with $\nabla_S = 1 - L^S$, and $\{\varepsilon_t\}$ is a white noise process. The polynomials $A(L)$, $B(L)$, $\Phi(L)$, and $\Theta(L)$ are respectively the autoregressive (AR), the moving average (MA), the seasonal autoregressive and the seasonal moving average polynomials. All roots of the polynomials $A(L)$, $B(L)$, $\Phi(L)$, and $\Theta(L)$ must be outside the unit circle so that the model is stationary and invertible.

Fitting of a SARIMA model begins with the search of the orders d , p , q , D , P , Q and S . The seasonality is often apparent from the time plot of the time series. Moreover, the seasonality of period S is suggestive of a significant spike at the lag S in the correlogram with neighbouring spikes non-significant. A positive spike at lag S indicates the involvement of a seasonal AR component of lag 1, that is, $P = 1$ whereas a negative spike at lag S indicates the involvement of a seasonal MA component of lag 1, that is, $Q = 1$. The orders p and q are lags of the autocorrelation functions (ACF) and partial autocorrelation functions (PACF) respectively. Once orders are set, the model equation (1) is then estimated. To test the model for goodness-of-fit, the residuals should be analysed. The residuals should be uncorrelated with zero mean and follow a Gaussian distribution; moreover, the autocorrelations of the residuals should not be significantly different from zero [7-8].

Model equation estimation and analyses were performed using EViews 9, a statistical software offering access to powerful statistical, forecasting, and modelling tools [9].

3. Results and discussion

Of the 8128 TB cases notified between 2008 and 2017 in Médéa province, 1632 (20.1%) were patients affected by PTB and 6496 (79.9%) were patients affected by EPTB. The yearly evolution of notified cases of PTB and EPTB along with the yearly incidence rate per 100000 people from 2008 to 2017, are displayed in Figure 1. The highest annual number of notified PTB (respectively EPTB) cases occurred in 2009 (respectively 2017) with 219 (respectively 1006) cases and the lowest one occurred in 2016 (respectively 2009) with 116 (respectively 425) cases. TB is notified throughout the year with peaks in the spring months. Indeed, 29% (respectively 28%, 24%, 19%) of PTB cases occurred in the spring (respectively summer, autumn, winter) and 31% (respectively 24%, 23%, 22%) of EPTB cases occurred in the spring (respectively winter, summer, autumn).

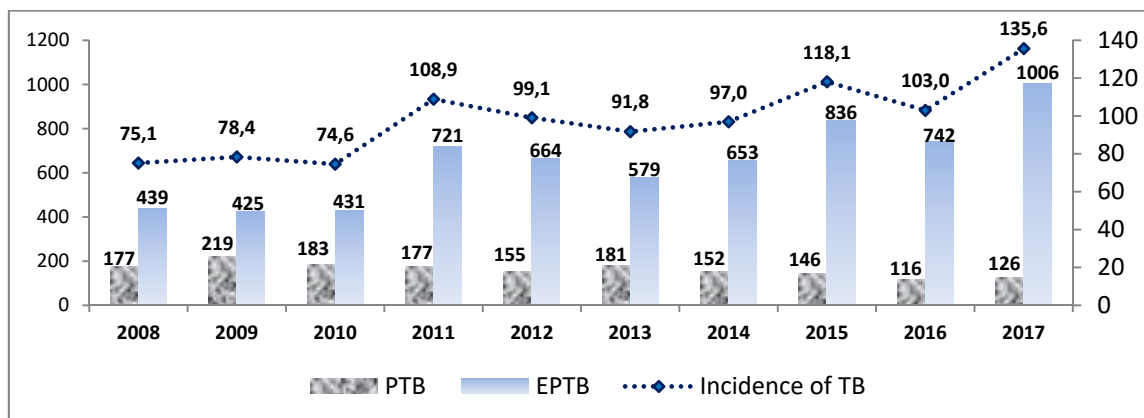


Figure 1. Annual evolution of cases of PTB and EPTB, and the incidence rate per 100000 people in Médéa Province 2008-2017.

Table 2. Descriptive statistics of the monthly EPTB and PTB data from 2008 to 2017

Statistics	EPTB	PTB
Mean	54.13	13.6
Median	51.5	13
Maximum	119	26
Minimum	20	2
Std. Dev.	20.66	5.33
Jarque-Bera	11.19	3.19
Probability	0.003709	0.202207

The mean \pm SD of the monthly PTB cases and monthly EPTB cases is 13.6 ± 5.3 (95% CI 12.7 to 14.5) and 54.1 ± 20.6 (95% CI 50.4 to 57.8) respectively. The median (IQR) of the monthly PTB and EPTB is 13 (17 – 10 = 7) and 51.5 (67.75 - 39 = 28.75) respectively, and the mode is 12 and 46 respectively. The PTB data are normally distributed (p-value 0.202207) while the EPTB data are not normally distributed (p-value 0.003709) see Table 2.

The data were partitioned into two periods: the data from January 2008 to December 2016 have been used for model identification, and the data from January 2017 to December 2017 have been used for model validation and for forecasting. The time series plots (see Figure 2) show an increasing trend of EPTB and a decreasing trend of PTB; a similar trend has been observed in other countries [10-13]. Both series are not stationary. A 12-month differencing of the EPTB original data and logarithm transformation were required to obtain a stationary time series, which we have noted by D12LNEPTB, and a 12-month differencing of the PTB original data were required to obtain a stationary time series, which we have noted by D12PTB. Both series D12LNEPTB and D12PTB exhibit an overall horizontal trend and no observable regular seasonality (see Figure 3).

The autocorrelation functions (ACF) and partial autocorrelation functions (PACF) have allowed us to identify the appropriate SARIMA form to model the obtained stationary series D12LNEPTB and D12PTB.

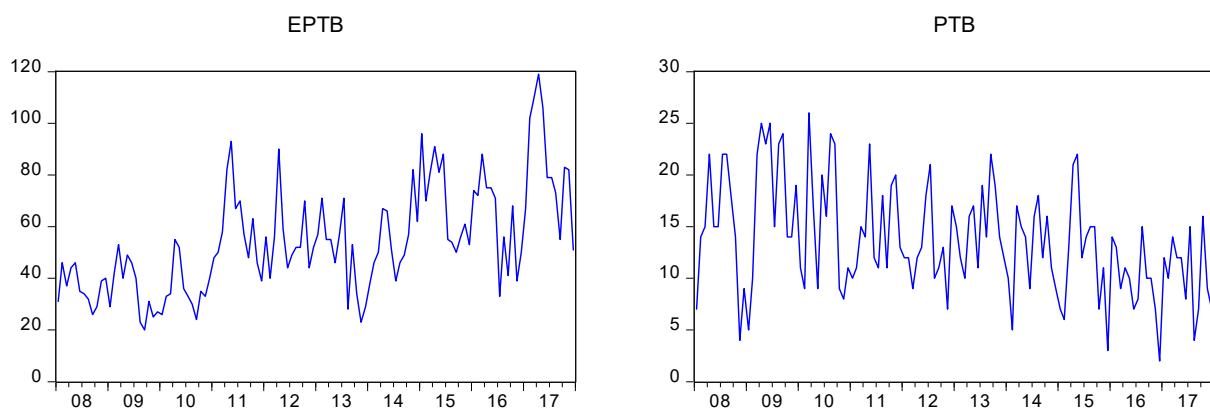


Figure 2. Monthly notified cases of EPTB and PTB between 2008 and 2017 in Médea province.

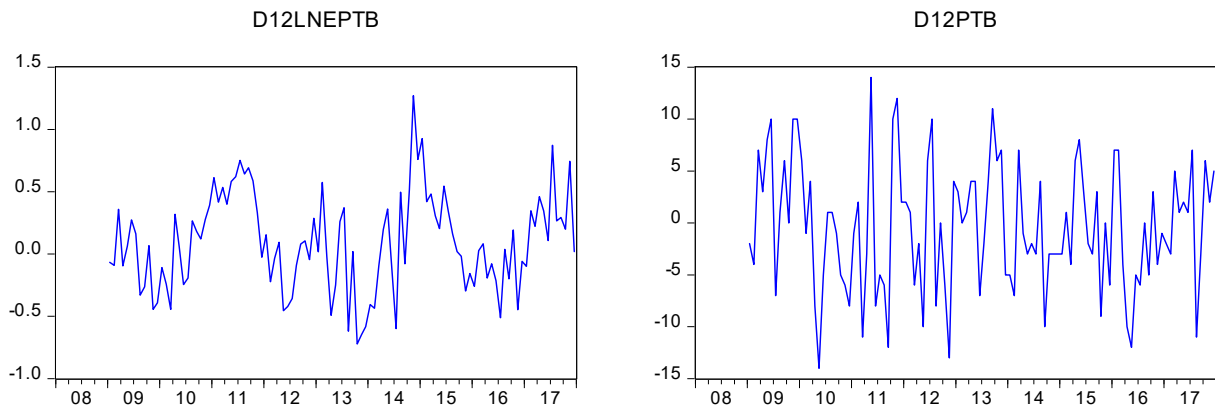


Figure 3. The obtained stationary time series D12LNEPTB and D12PTB.

To select the best-fit model, we applied some criteria such as the smallest Akaike information criterion (AIC), Standard Error Regression (SEE), the highest adjusted R^2 , the stationary and invertibility condition, and the white noise condition for residuals. After several trials, $(1,0,1) \times (0,1,1)_{12}$ has been selected as appropriate to the EPTB time series data and $(4,0,0) \times (0,1,1)_{12}$ has been selected as appropriate for the PTB time series data. Estimation outputs are displayed in Table 3 and Table 4.

All features of the best model were fulfilled. The models can, therefore, be used to estimate the forecasts. The simulated EPTB and PTB cases for the period 2008-2016 were closely approximated to the recorded data (Figure 4 and Figure 5) with strong correlation (Pearson product-moment correlation coefficient: $r = 0.80288$ and $r = 0.6933$). The models were then used for predicting the monthly EPTB and PTB cases from January 2017 to December 2017. The fitted data agree considerably with the actual data. All these conclusions attest to the adequacy of the models.

Climate factors were associated with TB disease in some studies [14]. We therefore examined the relationship between monthly PTB and EPTB cases and some climate variables. An extension of SARIMA modelling in an attempt to predict the PTB and EPTB cases using temperature, relative humidity, precipitation, sunshine hours and wind speed variables have been undertaken, however no best model was found.

4. Conclusion

This study revealed that SARIMA models are useful tools for monitoring and for predicting trends of TB incidence in Médea province, and may help public health authorities to get a better handle on the incidence trends of both PTB and EPTB cases in this endemic area. A $(1,0,1) \times (0,1,1)_{12}$ (respectively $(4,0,0) \times (0,1,1)_{12}$) SARIMA model was fitted to EPTB (respectively PTB) data in Médéa province and the models proved to be adequate.

The forecasts were solely based on tuberculosis notified cases. All other conditions that may affect TB incidence, such as risk factors, socio-economic determinants, climate and environmental factors, and control programme, which may vary over time, were not taken on board. Therefore, the outcomes should be read with care.

Table 3. Model estimation

Dependent Variable: D12LNEPTB
 Method: ARMA Maximum Likelihood (OPG - BHHH)
 Sample: 2009M01 2016M12
 Included observations: 96
 Convergence achieved after 36 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.072515	0.025605	2.832119	0.0057
AR(1)	0.855435	0.106824	8.007882	0.0000
MA(1)	-0.519694	0.150774	-3.446850	0.0009
SMA(12)	-0.873322	0.173650	-5.029200	0.0000
SIGMASQ	0.054395	0.010555	5.153499	0.0000
R-squared	0.636394	Mean dependent var		0.062235
Adjusted R-squared	0.620412	S.D. dependent var		0.388812
S.E. of regression	0.239550	Akaike info criterion		0.200744
Sum squared resid	5.221947	Schwarz criterion		0.334303
Log likelihood	-4.635697	Hannan-Quinn criter.		0.254731
F-statistic	39.81778	Durbin-Watson stat		2.088899
Prob(F-statistic)	0.000000			
Inverted AR Roots	.86			
Inverted MA Roots	.99	.86+.49i	.86-.49i	.52
	.49-.86i	.49+.86i	.00-.99i	-.00+.99i
	-.49-.86i	-.49+.86i	-.86+.49i	-.86-.49i
	-.99			

Table 4. Model estimation

Dependent Variable: D12PTB
 Method: ARMA Maximum Likelihood (BFGS)
 Sample: 2009M01 2016M12
 Included observations: 96
 Convergence achieved after 12 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.713888	0.179257	-3.982482	0.0001
AR(1)	0.169820	0.094350	1.799906	0.0753
AR(2)	-0.078120	0.105752	-0.738712	0.4620
AR(3)	0.099089	0.113874	0.870168	0.3865
AR(4)	-0.256158	0.103854	-2.466510	0.0156
MA(12)	-0.833139	0.185935	-4.480804	0.0000
SIGMASQ	20.63767	4.328834	4.767490	0.0000
R-squared	0.470863	Mean dependent var	-0.635417	
Adjusted R-squared	0.435191	S.D. dependent var	6.277981	
S.E. of regression	4.718142	Akaike info criterion	6.157833	
Sum squared resid	1981.217	Schwarz criterion	6.344817	
Log likelihood	-288.5760	Hannan-Quinn criter.	6.233415	
F-statistic	13.19972	Durbin-Watson stat	1.964138	
Prob(F-statistic)	0.000000			
Inverted AR Roots	.53-.47i	.53+.47i	-.44-.56i	-.44+.56i
Inverted MA Roots	.98	.85+.49i	.85-.49i	.49+.85i
	.49-.85i	-.00-.98i	-.00+.98i	-.49-.85i
	-.49+.85i	-.85+.49i	-.85-.49i	-.98

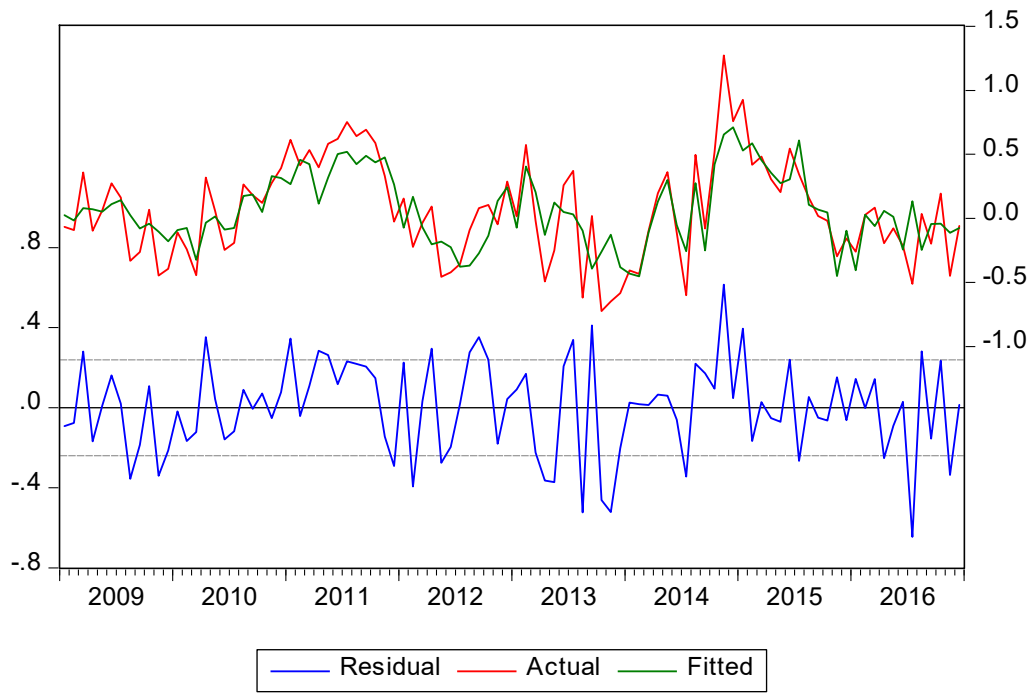


Figure 4. Residual Actual and Fitted (D12LNEPTB).

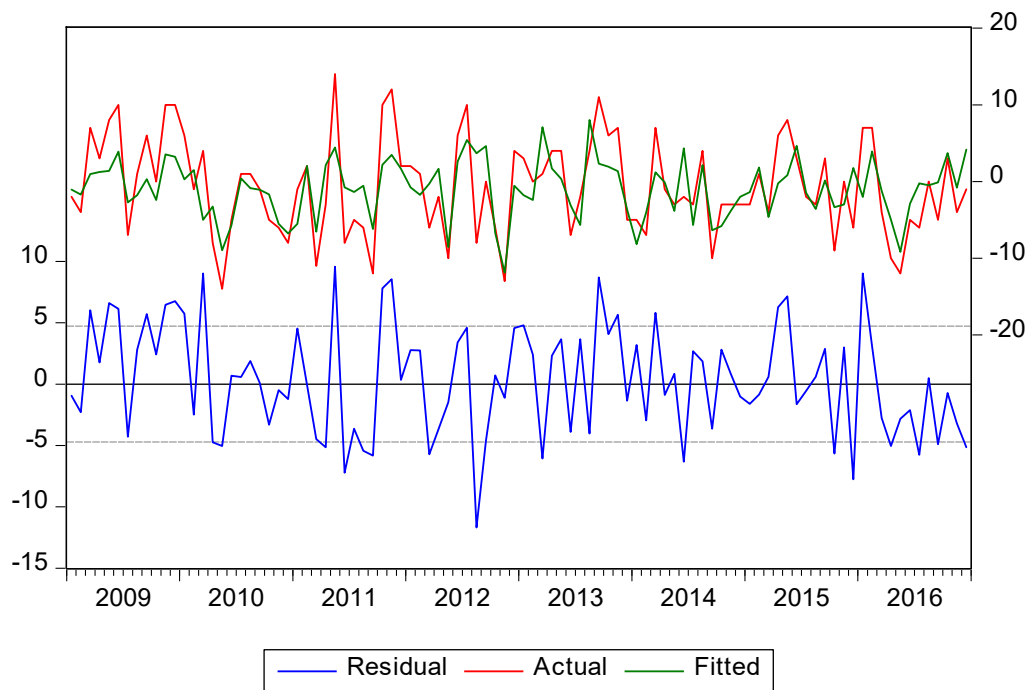


Figure 5. Residual Actual and Fitted (D12PTB).

References

- [1] World Health Organization. Global tuberculosis. 2017 Report. Geneva; 2017.
- [2] Manuel de la lutte antituberculeuse à l'usage des personnels médicaux. Programme National de Lutte Contre la Tuberculose-MSP, Direction de la prévention-INSP. INSP, édition 2007.
- [3] Selmane, S., "The impact of isolation of identified active Tuberculosis cases on the number of latently infected individuals", ICCSA 5 (2011) : 527-536.
- [4] Manuel de la lutte antituberculeuse à l'usage des personnels médicaux. Programme National de Lutte Contre la Tuberculose- MSP, Direction de la prévention-INSP. Edition & Publications-ANDS 2012.
- [5] Office national des statistiques. URL : <http://www.ons.dz>
- [6] Allard R., "Use of time-series analysis in infectious disease surveillance", Bull World Health Organ 76 (1998) : 327-333.
- [7] Box G.E.P., Jenkins G.M., "Time series analysis, forecasting and control", Holden-Day, San Francisco (1976).
- [8] Selmane, S., L'hadj, M., "Forecasting and prediction of scorpion sting cases in Biskra Province, Algeria, using a seasonal autoregressive integrated moving average model", Epidemiol Health 8 (2016) : 1-8.
- [9] IHS EViews. EViews 9. <https://www.eviews.com/EViews9/ev9main.html>.
- [10] Solovic, I., Jonsson, J., Korzeniewska- Kosela, M., Chiotan, D.I., Pace-Asciak, A., Slump, E., Rumetshofer, R., Abubakar, I., Kos, S., Svetina-Sorli, P., Haas, W., Bauer, T., Sandgren, A., van der Werf, M.J., "Challenges in diagnosing extrapulmonary tuberculosis in the European Union, 2011", Euro Surveill 18(12) (2013) : 1-9.
- [11] Kruijshaar M.E., Abubakar I., "Increase in extrapulmonary tuberculosis in England and Wales 1999-2006", Thorax 64(12) (2009) : 1090-5.
- [12] Ates Guler, S., Bozkus, F., Inci M.F., et al., "Evaluation of pulmonary and extrapulmonary Tuberculosis in immunocompetent adults: A retrospective case series analysis", Med Princ Pract 24 (2015) : 75-79.
- [13] Gomes T., Reis-Santos B., Bertolde A., et al., "Epidemiology of extrapulmonary tuberculosis in Brazil: a hierarchical model", BMC Infectious Diseases 14 (2014) : 9.
- [14] Fernanda Monteiro de Castro, F., Eder de Souza, M., Daniella Melo Arnaud Sampaio, P., Maria do Socorro Nantua, E., "Relationship between climatic factors and air quality with tuberculosis in the Federal District, Brazil, 2003–2012", The Brazilian Journal of Infectious Diseases 21(4) (2017) : 369-375.