IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-6, June 2018 E-ISSN: 2347-2693

Fitting the Best Model for ACEs Using Interrupted Time Series Data

A. Srinivasulu[1], B. Sarojamma[2*]

[1] Department of Statistics, S.V.University, Tirupati-517502, A.P, India
[2] Department of Statistics, S.V.University, Tirupati-517502, A.P, India

*Corresponding Author: saroja14397@rediffmail.com, Tel.: +91 8639652810

Available online at: www.ijcseonline.org

Accepted: 12/Jun/2018, Published: 30/Jun/2018

Abstract- The interrupted time series data plays a very important role in the evaluation of public health interventions and also to improve hospital antibiotic prescribing. We take the data from a study on the effects of the Italian smoking ban in public places on hospital admissions for Acute Coronary Events (ACEs). In January 2005, Italy introduced regulations to ban smoking in all indoor public places, which the aim of limiting adverse health affects of second hand smoke. The data used here are ACEs in the Sicily region between 2002 and 2006 among those aged 0-69 years. In this paper, we use three models for interrupted time series data. Root Mean Square Error (RMSE) measure is used for selecting the best model. Three models are empirically tested using interrupted time series smoke ban data in the Sicily region, Italy.

Keywords- ARIMA models, Before Intervention, After Intervention, Adaptive smoothing model, R^2, RMSE.

I. INTRODUCTION

In an interrupted time series (ITS) study, A time series of a particular outcome of interest is used to establish an underlying trend which is ‘interrupted’ by an intervention at a known point. In time, the counter factual scenario provides a comparison for the evaluation of the impact of the intervention by examining any change occurring in the post intervention period. ITS requires a clear differentiation of the pre-intervention period and the post intervention period. In some cases it may be difficult to define when the intervention began and to differentiate the effects of different components. The outcomes of ITS may take various forms such as counts, continuous data (or) binary variables. In our paper, we took the data from a study on the effects of the Italian smoking ban in public places on hospital admissions for Acute Coronary Events (ACEs). In January 2005, Italy introduced regulations to ban smoking in all indoor public places, which the aim of limiting adverse health affects of second hand smoke. The data used here are ACEs in the Sicily region between 2002 and 2006 among those aged 0-69 years.

The above paragraph explains the introduction of the research paper. Section II gives the brief history of review of literature and section III explains the methodology of three methods, one is quadratic regression method, second one is Interrupted time series method using ARIMA and the third one is Adaptive smoothing method. Using R^2 criteria, we will choose the best model. Section IV gives the empirical investigation given fo the above models. Section IV explains the summary and conclusions of the paper.

II. RELATED WORK

Interrupted time series (ITS) analysis is a valuable study design for evaluating the effectiveness of population level health interventions that have been implemented at a clearly defined point in time. The ITS has been used for the evaluation of a wide range of public health interventions including new vaccines, cycle helmet legislation, changes to paracetamol packaging, traffic speed zones, to reduce inappropriate use of key antibiotics, wearing helmets when riding bikes to reduce the death rates in bike accidents and precaution against nosocomial infections as well as in the evaluation of health impacts of un planned events such as the global financial crisis.

Traditional epidemiological study designs such as cohort and case-control studies can provide important evidence about disease etiology. Randomized controlled trails have been considered. The gold standard design for evaluating the effectiveness of an intervention, but Randomized controlled trails(RCT) are not always possible, in particular for health policies and programmes targeted at the population level. The interrupted time series is increasingly being used for the evaluation of public health interventions, it is particularly suited to interventions introduced at a population level over a clearly defined time period and that targets population level health outcomes [1].

Not only in the areas of health departments but also many departments, for example on 8 November 2016, the government of India announced the demonetization of all 500 and 1000 Rupees bank notes of Mahatma Gandhi series. Due to this the economic position of India was fell down drastically. The BSE Sensex and NIFTY stock indices fell over 6 percent on the day after the announcement. In the days following the demonetization, the country faced severe cash shortages with severe detrimental effects. So there is a situation of distinguish the effect of intervention from secular change, that is change that would have happened even in the absence of intervention. Estimating the intervention effect is done by comparing the trend in the outcome after intervention to the existing trend in the pre-intervention period.

By considering the impact of large scale interventions (for example, population-based health interventions, media campaigns, and dissemination of professional guidelines) or public policy changes (for example, new laws or taxes), the researchers are faced with an effective sample size of N=1, where the treated group may be local community, state or even larger unit. It is also fairly common in these situations that the only data available are reported at an aggregate level. In multiple observations on an outcome variable of interest in the pre-intervention and post intervention periods can be obtained; an Interrupted Time Series Analysis (ITSA) offers a quasi-experimental research design with a potentially high degree of internal validity. ITSA has been used in many areas of study such as assessing the effects of community interventions, public policy regulatory actions and health technology assessment, etc. ITSA has also been proposed as a more flexible and rapid design to be considered in health research before defaulting to the traditional two-arm randomized controlled trial [2].

Randomized Controlled Trails (RCTs) are considered the ideal approach for assessing the effectiveness if interventions. But not all interventions can be assessed with an RCT, where as for many intervention trails can be prohibitively expensive. ITS analysis is a useful Quasi-experimental design with which to evaluate the longitudinal effects of interventions, through regression modeling [3].

In an interrupted time series study, a series of observations on the same outcome before and after the introduction of an intervention are used to test immediate and gradual effects of the intervention. A major strength of this design is its ability to distinguish the effect of the intervention from secular change. Estimating the intervention effect is done by comparing the trend in the outcome after intervention to the existing trend in the pre-intervention period and it is achieved through modifications to the standard regression analysis. In a basic segmented regression analysis, the time period is divided into pre and post intervention segments, and separate intercepts and slopes are estimated in each segment. Statistical tests of changes intercepts and slopes pre- to-post intervention are carried out [4].

ITS is to evaluate an intervention to reduce inappropriate use of key antibiotics with interrupted time series analysis. The intervention is a policy for appropriate use of alert antibiotics implemented through concurrent, patient-specific feedback is clinical pharmacists. Statistical significance and effect size were calculated by segmented regression analysis of interrupted time series of drug use and cost for 2 years before and after the intervention started. The segmented regression analysis of interrupted time series data explains us to assess in statistical terms, how much an intervention changed an outcome of interest, immediately and over time. When a separate control group is not available, the level and trend of pre-intervention segment serve as control for the post-intervention segment in single group time series. For estimating seasonal auto correlation, the auto regression model needs to evaluate correlations between error terms separated by multiples of 12 months. In the Three outcomes of ITS analysis, first, change in level immediately after the intervention, second, difference between pre-intervention and post-intervention slopes and third, the estimation of monthly average intervention effect after the intervention [5].

A segmented regression analysis is a powerful statistical method for estimates intervention effects in interrupted time series studies. Segmented regression model is fitted and constant of the model is estimated using least square regression line to each segment of the independent variable. The segmented regression equation is linear relationship between time and thus assume a linear relationship between time and the outcome within each segment. They proposed the below regression model to estimate the level and trend in mean number of prescriptions per patient before the three drug cap and the changes in level and trend following the camp in New Hampshire [6].

III. METHODOLOGY

In this research paper we are fitted three models for interrupted data:

Model I:

$Y_t = \beta_0 + \beta_1 * t + \beta_2$ (equation estimated value before intervention) + β_3 (equation estimated value after intervention)

The data is divided into two parts, one is before smoke ban and another one is after smoke ban, the interrupted point is smoke ban day.

Y_t is time series value at time point t.

β_0 is constant.

β_1 is constant concerned to time

β_2 is constant with equation before intervention

β_3 is constant with equation after intervention.

Equation before intervention and after intervention are chosen among straight line, power curve , exponential curve, parabola, polynomial of order 3, polynomial of order 4, Polynomial of order 5 and polynomial of order 6. Using R^2 criteria, by estimating observed values for the best model to before intervention and after intervention and substituting in the above equation, we get

$$Y_t = \beta_0 + \beta_1 * time + \beta_2 \hat{BI} + \beta_3 \hat{AI}$$

For estimating the constants β_0, β_1, β_2, and β_3, we are using minimum least square estimation method.

Model II: **Interrupted time series model using ARIMA**

The model fitted is as follows

$Y_t = \beta_0 + \beta_1 * t + \beta_2$ (equation estimated value before intervention using ARIMA) + β_3 (equation estimated value after intervention using ARIMA).

Here Y_t is time series value at time point t.

β_0, β_1, β_2, and β_3 are constant.

t is time.

ARIMA (p,d,q) with different values of p, d and q are fitted for before intervention data and after intervention data. For choosing the best model among several models we use R^2 criteria. The fitted model using MLE model is

$$Y_t = \beta_0 + \beta_1 * t + \hat{\beta}_2 (ARIMAforBI) + \hat{\beta}_3 (ARIMAforAI)$$

Model III: By using Adaptive smoothing model

The single exponential smoothing forecasting model requires the specifications of an α value and it has been shown that the mean absolute percentage error (MAPE) and Mean Square Error (MSE) measures depends on this choice. Adaptive Response Rate Single Exponential Smoothing (ARRSES) may have an advantage over Single Exponential Smoothing (SES), it allows the value of α to be modified in a controlled manner, as changes in the pattern of data occur. This characteristic seems attractive when hundreds (or) thousands of items require forecasting.

The basic equation for forecasting with the method of ARRSES is similar to an equation

$$F_{t+1} = \alpha_t Y_t + (1 - \alpha_t) F_t$$

$$\alpha_{t+1} = |A_t / M_t|$$

$$A_t = \beta e_t + (1 - \beta) A_{t-1}$$

$$M_t = \beta |e_t| + (1 - \beta) M_{t-1}$$

$$e_t = Y_t - F_t$$

Here β is a parameter between 0 and 1 and mod (||) denotes absolute values.

In equation A_t denotes a smoothed estimate of forecast error and is calculated as a weighted average of A_{t-1} and the last forecasting error $€_t$.

M_t denotes a smoothed estimate of the absolute forecast error; being calculated as a weighted average of M_{t-1} and the last absolute forecasting error $|€_t|$.

A_t and M_t gives single exponential smoothing estimates themselves.

$\alpha_{t+1} = |A_t/M_t|$ indicates that the value of α_t to be used for forecasting period(t+2) is defined as an absolute value of the ratio of A_t and M_t. Instead of α_{t+}, we could have used α_t in the above equation. we prefer α_{t+1} because ARRSES is often too responsive to changes, thus using α_{t+1} we introduce a small log of one period, which allows the system to "settle" a little and forecast in a more conservative manner.

R^2 CRITERIA:

In statistics, the coefficient of determination denoted by R^2 is the proportion of the variance in the dependent variable that is predictable from the independent variables(s). It is a statistic used in the context of statistical models whose main purpose is either the prediction of future outcomes or the testing of hypothesis, on the basis of other related information. It provides a measure of how well observed outcomes are replicated by the model, based on the proportion of total variation of outcomes explained by the model.

A data set has n values y_1, y_2,……, y_n each associated with a predicted value f_1,f_2,....f_n. Define the residual as $e_i = y_i - f_i$

If $\bar{y}$ is the mean of the observed data then

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^{n} y_i$$

The variability of the data set can be measured using three sums of squares formulas:

- the total sum of squares (proportional to the variance of the data):

$$SS_{tot} = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

- The regression sum of squares, also called the explained sum of squares:

$$SS_{reg} = \sum_{i=1}^{n} (f_i - \bar{y})^2$$

- The sum of squares of residuals, also called the residual sum of squares

$$SS_{res} = \sum_{i=1}^{n} (y_i - f_i)^2 = \sum_i e_i^{\,2}$$

The general definition of the coefficient of determination (R^2) is

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

Relation to unexplained variance:

In a general form, R^2 can be seen to be related to the fraction of variance unexplained (FVU), since the second term compares the unexplained variance (variance of the model's errors) with the total variance (of the data):

$$R^2 = 1 - FVU$$

Explained variance:

Suppose R^2 =0.49, this implies that 49% of the variability of the dependent variable has been accounted for, and the remaining 51% of the variability is still unaccounted for. In some cases the total sum of squares equals the sum of the two other sum of squares.

$$SS_{res} + SS_{reg} = SS_{tot}$$

By using this, R^2 is equivalent to

$$R^2 = \frac{SS_{res}}{SS_{tot}} = \frac{SS_{res}/n}{SS_{tot}/n}$$

Where n is the number of observations on the variables. In this, R^2 is expressed as the ratio of the explained variance to the total variance.

As squared correlation coefficient:

In linear least squares regression with an estimated intercept term R^2 equals the square of the Pearson correlation coefficient between the observed y and modeled (predicted) f data values of the dependent variable. In linear least squares regression with an intercept term and a single explanatory that is also equal to the squared Pearson correlation coefficient of the dependent variable y and the explanatory variable x. R^2 value can be calculated as the square of the correlation coefficient between the original y and modeled f data values.

INTERPRETATION:

R^2 is a statistic that will give some information about the goodness of fit of a model. In regression, the R^2 coefficient of determination is a statistical measure of how well the regression predictions approximate the real data points. An R^2 of 1 indicates that the regression predictions perfectly fit the data. Values of R^2 outside the range 0 to 1.In all instances where R^2 is used, the predictors are calculated by ordinary least-squares regression: that is, by minimizing SS_{res}. In this case R^2 increases as we increase the number of variables in the model.

ADJUSTED R^2 or ($\bar{R}^2$):

The use of an adjusted R^2 or ($\bar{R}^2$) is an attempt to take account of the phenomenon of the R^2 automatically and spuriously increasing when extra explanatory variables are added to the model. It is a modification due to Henri Theil of R^2 that adjusts for the number of explanatory terms in a model relative to the number of data points. The adjusted R^2 can be negative, and its value will always be less than or equal to that of R^2. Unlike R^2, the adjusted R^2 increases only when the increase in R^2 (due to the inclusion of a new explanatory variable) is more than one would expect to see by chance. If a set of explanatory variables with a predetermined hierarchy of importance are introduced into a regression one at a time, with the adjusted R^2 computed each time, the level at which adjusted R^2 reaches a maximum, and decreases afterward, would be the regression with the ideal combination of having the best fit without excess/unnecessary terms. The adjusted R^2 is defined as

$$\bar{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-p-1}$$

Where p is the total number of explanatory variables in the model, and n is the sample size.

The principle behind the adjusted R^2 statistic can be seen by rewriting the ordinary R^2 as

$$R^2 = 1 - \frac{VAR_{res}}{VAR_{tot}}$$

Where $VAR_{res} = SS_{res}/n$ and $VAR_{tot} = SS_{tot}/n$ are the sample variances of the estimated residuals and the dependent variable respectively, which can be seen as biased estimates of the population variances of the errors and of the dependent variable.

IV. RESULTS AND DISCUSSION

The three intervention models are as follows.

MODEL I:

$Y_t = \beta_0 + \beta_1 * t + \beta_2$ (equation estimated value before intervention) + β_3 (equation estimated value after intervention)

For choosing the best model among the three, before intervention and after intervention are as follows

(i) Straight line is $y = A + Bx$

(ii) Polynomial equation of order 2 is $y = A + Bx + Cx^2$

(iii) Polynomial equation of order 3 is $y = A + Bx + Cx^2 + Dx^3$

(iv) Polynomial equation of order 4 is $y = A + Bx + Cx^2 + Dx^3 + Ex^4$

(v) Polynomial equation of order 5 is $y = A + Bx + Cx^2 + Dx^3 + Ex^4 + Fx^5$

(vi) Polynomial equation of order 6 is $y = A + Bx + Cx^2 + Dx^3 + Ex^4 + Fx^5 + Gx^6$

(vii) Power curve is $y = Ax^B$

(viii) Exponential curve is $y = Ae^{Bx}$

(ix) Logarithmic equation is $y = A*\ln(x) + B$

For choosing the best model among different models using R^2 criteria, the below table explains about R^2 value for fitted models.

Table-1

MODEL	Before Intervention	
	R^2 value	***Equation***
Exponential curve	0.488	$Y = 728.9*e^{0.005x}$
Linear Equation	0.479	Y = 4.453*x + 728.4
Logarithmic Equation	0.462	Y =53.23*ln(x) + 669.3
Polynomial Equation of order 2	0.494	$Y = -0.085x^2+7.614x+708.4$
3	0.494	$Y=-0.000x^3-0.048x^2+7.060x+710.2$
4	0.501	$Y = -0.000x^4+0.050x^3-1.281x^2+17.52x+688.5$
5	0.504	$Y=0.00005x^5-0.005x^4+0.213x^3-.591x^2+30.43x+669.5$
6	0.584	$Y = 0.00003x^6-0.003x^5+0.134x^4-2.600x^3+23.41x^2-78.21x+791.6$
Power Curve	0.479	$Y = 675.9*x^{0.067}$

By using R^2, the best model for before intervention is 6th degree polynomial, i.e.

$$Y = 0.00003x^6-0.003x^5+0.134x^4-2.600x^3+23.41x^2-78.21x+791.6$$

Table-2

MODEL	After Intervention	
	R^2 value	***Equation***
Exponential curve	0.244	$Y = 795.1*e^{0.006x}$
Linear Equation	0.237	Y = 5.141*x + 795.8
Logarithmic Equation	0.188	Y =37.62*ln(x) + 773.1
Polynomial Equation of order 2	0.244	$Y = -0.156x^2+8.906x+780.1$
3	0.310	$Y = -0.078x^3-2.683x^2-18.94x+841.6$
4	0.403	$Y = 0.016x^4-0.871x^3+15.1x^2-88.56x+941.0$
5	0.567	$Y=0.003x^5-0.216x^4+4.156x^3-31.73x^2+86.50x+759.3$
6	0.605	$Y=0.000x^6+0.028x^5-0.881x^4+12.9x^3-87.35x^2+238.6x+635.2$
Power Curve	0.189	$Y = 775.3*x^{0.043}$

By using R^2, the best model for after intervention is 6^{th} degree polynomial, i.e.

$$Y=0.000x^6+0.028x^5-0.881x^4+12.9x^3-87.35x^2+238.6x+635.2$$

The fitted equation for model I is

Y_t=-70.5783+ (-0.06926)*T + 1.086385*($0.00003x^6-0.003x^5+0.134x^4-2.600x^3+23.41x^2-78.21x+791.6$) + 1.081243*($0.000x^6+0.028x^5-0.881x^4+12.9x^3-87.35x^2+238.6x+635.2$)

The RMSE for model I is 35.31417

Model II:

The different ARIMA models for before and after intervention and their respective R^2, stationary R^2, The equation for model II is as follows

$Y_t = \beta_0 + \beta_1 * t + \beta_2$ (ARIMA before best estimated) + β_3 (ARIMA after best estimated)

The fitted equation for model II is

Y_t = -59.6905 + 0.050109*T + 1.069112 (ARIMA (6, 0,6) model) + 1.061186(ARIMA(6,0,5) model)

The RMSE for model II is 34.1375

MODEL III:

Another model for intervention is Adaptive smoothing model.

Table-3

FORECAST

Month	***ACE***	***forecast Ft***	***et***	***smoth err At***	***Absmoothed error Mt***	***alpha t***	***1-alpha t***	***ESS***
1	728	0	0	0	0	0	0	0
2	659	728	-69	-41.4	41.4	1	0	4761
3	791	659	132	62.64	95.76	0.65	0.35	17424
4	734	791	-57	-9.14	72.5	0.13	0.87	3249
5	757	783.59	-26.59	-19.61	44.95	0.44	0.56	707.028
6	726	757	-31	-26.44	36.58	0.72	0.28	961
7	760	734.68	25.32	4.62	29.82	0.15	0.85	641.102
8	740	760	-20	-10.15	23.93	0.42	0.58	400
9	720	751.6	-31.6	-23.02	28.53	0.81	0.19	998.56
10	814	720	94	47.19	67.81	0.7	0.3	8836
11	795	785.8	9.2	24.4	32.64	0.75	0.25	84.64
12	858	795	63	47.56	50.86	0.94	0.06	3969
13	887	854.22	32.78	38.69	40.01	0.97	0.03	1074.528
14	766	887	-121	-57.12	88.6	0.64	0.36	14641
15	851	809.56	41.44	2.02	60.3	0.03	0.97	1717.274
16	769	851	-82	-48.39	73.32	0.66	0.34	6724
17	781	796.88	-15.88	-28.88	38.86	0.74	0.26	252.174
18	756	781	-25	-26.55	30.54	0.87	0.13	625
19	766	759.25	6.75	-6.57	16.27	0.4	0.6	45.563
20	752	766	-14	-11.03	14.91	0.74	0.26	196
21	765	755.64	9.36	1.2	11.58	0.1	0.9	87.61
22	831	765	66	40.08	44.23	0.91	0.09	4356
23	879	825.06	53.94	48.4	50.06	0.97	0.03	2909.524
24	928	879	49	48.76	49.42	0.99	0.01	2401
25	914	927.51	-13.51	11.4	27.87	0.41	0.59	182.52
26	808	914	-106	-59.04	74.75	0.79	0.21	11236
27	937	830.26	106.74	40.43	93.94	0.43	0.57	11393.428
28	840	937	-97	-42.03	95.78	0.44	0.56	9409
29	916	894.32	21.68	-3.8	51.32	0.07	0.93	470.022
30	828	916	-88	-54.32	73.33	0.74	0.26	7744
31	845	850.88	-5.88	-25.26	32.86	0.77	0.23	34.574
32	818	845	-27	-26.3	29.34	0.9	0.1	729
33	860	820.7	39.3	13.06	35.32	0.37	0.63	1544.49
34	839	860	-21	-7.38	26.73	0.28	0.72	441
35	887	854.12	32.88	16.78	30.42	0.55	0.45	1081.094
36	886	887	-1	6.11	12.77	0.48	0.52	1
37	831	886.52	-55.52	-30.87	38.42	0.8	0.2	3082.47
38	796	831	-35	-33.35	36.37	0.92	0.08	1225
39	833	798.8	34.2	7.18	35.07	0.2	0.8	1169.64
40	820	833	-13	-4.93	21.83	0.23	0.77	169
41	877	830.01	46.99	26.22	36.93	0.71	0.29	2208.06
42	758	877	-119	-60.91	86.17	0.71	0.29	14161
43	767	792.51	-25.51	-39.67	49.77	0.8	0.2	650.76
44	738	767	-29	-33.27	37.31	0.89	0.11	841
45	781	741.19	39.81	10.58	38.81	0.27	0.73	1584.836
46	843	781	62	41.43	52.72	0.79	0.21	3844
47	850	829.98	20.02	28.58	33.1	0.86	0.14	400.8
48	908	850	58	46.23	48.04	0.96	0.04	3364
49	1021	905.68	115.32	87.68	88.41	0.99	0.01	13298.702
50	859	1021	-162	-62.13	132.56	0.47	0.53	26244
51	976	944.86	31.14	-6.17	71.71	0.09	0.91	969.7
52	888	976	-88	-55.27	81.48	0.68	0.32	7744
53	962	916.16	45.84	5.4	60.1	0.09	0.91	2101.306
54	838	962	-124	-72.24	98.44	0.73	0.27	15376
55	810	871.48	-61.48	-65.78	76.26	0.86	0.14	3779.79
56	876	810	66	13.29	70.1	0.19	0.81	4356
57	843	822.54	20.46	17.59	40.32	0.44	0.56	418.612
58	936	843	93	62.84	71.93	0.87	0.13	8649
59	912	923.91	-11.91	17.99	35.92	0.5	0.5	141.848
							TOTAL	237106.6
							MSE	4018.756
							RMSE	63.39165

The above equations belongs to Adaptive smoothing model

$$F_{t+1} = \alpha_t Y_t + (1-\alpha_t)F_t \quad (1)$$

$$\alpha_{t+1} = |A_t / M_t| \quad (2)$$

$$A_t = \beta e_t + (1-\beta)A_{t-1} \quad (3)$$

$$M_t = \beta|e_t| + (1-\beta)M_{t-1} \quad (4)$$

$$e_t = Y_t - F_t \quad (5)$$

The values, for example for e11, A11, M11, α12 and F11 are calculated as follows

e11 = 795-785.4484 = 9.55157 (using 5)

A11=0.6*9.55157+0.4*47.20868=24.61441 (using 3)

M11 = 0.6*|9.55157|+0.4*67.8032= 32.8522(using 4)

α12 = 24.61441/32.85224 = 0.749246(using 2)

The forecast value for the period 12 can be computed using equation 1 is

F11 = 0.69626*814+0.30374*720 = 785.4484.

The RMSE value for adaptive smoothing model is 63.39165.

V. CONCLUSIONS

Interrupted time series analysis is used month wise for smoke ban before intervention (2002-2004) and after intervention (2005-2006).

We are fitted three models for interrupted smoke ban data. The three models are

Model I

$Y_t = \beta_0 + \beta_1 * t + \beta_2$ (equation estimated value before intervention) + β_3 (equation estimated value after intervention)

The fitted equation is

Y_t = -70.5783+ (-0.06926)*t + 1.086385*(0.00003x^6-0.003x^5+0.134x^4-2.600x^3+23.41x^2- 78.21x+791.6) + 1.081243*(0.000x^6+0.028x^5-0.881x^4+12.9x^3-87.35x^2+238.6x+635.2)

Model II

The equation for model II is as follows

$Y_t = \beta_0 + \beta_1 * t + \beta_2$ (ARIMA before best estimated) + β_3 (ARIMA after best estimated)

The fitted equation for model II is

Y_t = -59.6905 + 0.050109*T + 1.069112 (ARIMA (6,0,6) model) + 1.061186(ARIMA(6,0,5) model)

Model III

Model III is constructed by using adaptive smoothing model.

Among the above three models, the best model for data is selected using error criteria is given in table-4

Table-4

MODEL	RMSE
Model I	35.31417
Model II	34.1375
Model III	63.39165

From the above table, the least RMSE is for model II (34.1375). So model II is the best model for the smoke ban data.

REFERENCES

[1] James Lopez Bernal, Steven Cummins, Antonio Gasparrini, "*Interrupted Time Series Regression for The Evaluation of Public Health Interventions: a tutorial*", International Journal of Epidemiology, 348-355, 2017.

[2] Ariel Linden, "*Conducting Interrupted Time Series Analysis for Single and Multiple Group Comparisions*", The Stata Journal 15, (2), 480-500, 2015.

[3] Evangelos Kontopantelis, Tim Doran, David A Springate, Lain Buchan, David Reeves,"*Regression Based Quasi-Experimental Approach When Randomisation is not an option: Interrupted Time Series Analysis*", The BMJ; 350:h2750, 2015.

[4] Monica Taljaard, Joanne E McKenzie, Craig R Ramsay, Jeremy M Grimshaw, " *The Use of Segmented Regression in Analysing Interrupted Time Series Studies: An Exmple in Pre-Hospital Ambulance care*", Implementation Science, 9:77, 2014.

[5] Faranak Ansari, Kiesteen Gray, Dilip Nathwani, Gabby Philips, Simon Ogston, Craig Ramsay, Peter Davely "*Outcomes of an Intervention to Improve Hospital Antibiotic Prescribing: Interrupted Time Series with Segmented Regression Analysis*", Journal of Antimicrobial Chemotherapy, 52, 842-848, 2003.

[6] A.K. Wagner, S,B. Soumerai, F.Zhang, D.Ross-Degnan "*Segmented Regression Analysis of Interrupted Time Series Studies in Medication Use Research*" Journal of Clinical Pharmacy and Therapeutics, 27, 299-309, 2002.

[7] Gene V Glass,"*Interrupted Time Series Quasi-Experiments"*, American Educational Research Association., 2nd edition 589-608, 1997.

[8] Donald P.Hartmann, John M. Gottman, Richard R. Jones, William Gardner, Alan E. Kazdin, Russell S. Vaught "*Interrupted Time Series Analysis and Its Application to Behavioral Data*",Journal of Applied Behavior Analysis,13, 543-559, 1980.

[9] http://ije.oxfordjournals.org/lookup/suppl/doi:10.1093/ije/dyw098/-/DC1.

Authors Profile

Mr.A.Srinivasulu. pursed Bachelor of Science from Sri Venkateswara University, Tirupai, A.P, India 2002 and Master of Science from Sri Venkateswara University, Tirupai, A.P, India year 2004. He is currently pursuing Ph.D in the Department of Statistics, Sri Venkateswara University, Tirupati, Andhra Pradesh, India. He has published two papers in National and International Journals. He has presented two papers in International conference and two in National Conference.

Dr.B.Sarojamma received Ph.D. degree in statistics in 2006 from Sri Venkateswara University, Tirupati, Andhra Pradesh, India. She taught the subjects Time series and Forecasting, Operations Research, Distribution Theory, Sampling Techniques and Biostatistics, etc. to the students of M.Sc. Statistics, Applied

Statistics, Statistics with Computer applications, M.Sc.Home science, M.Sc. Aquaculture, Biotechnology and Bio-informatics. She has given several invited lectures in national/International workshops/Conferences. She has published more than 64 research papers in internationall reputed journals and presented more than 83 papers at National/International Conferences/Simposia. She is the Life Member in manyProfessional Bodies like Indian Society for Probability and Statistics (ISPS), Andhra Pradesh Society for Mathematics Sciences Hyderabad, International Indian Statistics Association (IISA) ,Cochin,Indian Mathematical Society (IMS), Aurangabad and Indian Statistical Association(AISA), Pune. At present she is working as an Assistant Professor in the Deparment of Statistics, S.V.University, Tirupati.

