Statistics and Applications
Volume 12, Nos. 1&2, 2014 (New Series), pp. 61-70

Application of Time Series Intervention Modelling for Modelling and Forecasting Cotton Yield

Mrinmoy Ray[1], Ramasubramanian V.[2], Amrender Kumar[3] and Anil Rai[1]
[1]*Indian Agricultural Statistics Research Institute, New Delhi*
[2]*Central Institute of Fisheries Education, Mumbai*
[3]*Indian Agricultural Research Institute, New Delhi*

Abstract

The present study deals with investigations of time series intervention modelling in agriculture. Such models are employed in situations where it may be known that certain exceptional external events called 'interventions' could affect the time series phenomenon under study. As a case study, yield of cotton for Gujarat, Maharashtra and all India have been considered with the intervention being introduction of Bt Cotton variety in year 2002.When cotton yields were forecast, the performance of Auto Regressive Integrated Moving Average (ARIMA) intervention models was found to be superior to conventional ARIMA models for all the three datasets considered.

Keywords: Time series intervention modelling, ARIMA, step intervention, impulse response function, MAPE

1 Introduction

The most widely used technique for modelling and forecasting crop yield time-series data is the Box Jenkins' Autoregressive integrated moving average (ARIMA) methodology. However when the patterns of the time-series under study are affected by some external event such as incorporation of new environmental regulations, strikes and special promotion campaigns, introduction of new variety, severe disease of plant etc. then the forecasting performance of ARIMA model may be affected. However, it can be improved by employing appropriate techniques such as ARIMA-Intervention model. According to types of intervention, there are three kinds of intervention viz. step, pulse/ point and ramp. Step Intervention occurs at particular period of time and exists in the subsequent time-periods. The effect of step intervention may remain constant over time or it may increase or decrease over time. In agriculture, such type of intervention occurs due to introduction of new variety, pesticide, new economic policy etc. Introduction of Bt-cotton in India in 2002 is an example of this type of intervention. Pulse Intervention occurs only at particular period of time but the effect of these type of intervention may exist for that particular time period only or it may exist in the subsequent time period. In agriculture, these types of intervention is said to occur in specific years with severe drought or flood or severe insect-pest incidence. Severe drought occurred in the year 2002 in India can be considered as an example of this type of intervention. Ramp Intervention occurs at particular period of time and exists in the subsequent time-periods with an increasing magnitude. The effect of ramp intervention will always increase over time. In agriculture, this type of intervention exists in the price rise of an agricultural commodity. In this study, investigations of time series intervention modelling in agriculture for yield of cotton in Gujarat, Maharashtra and all India have been considered with the intervention being introduction of Bt Cotton variety in year 2002. Intervention modelling was introduced by Box and Tiao (1975) to study and quantify the impact of air pollution controls on smog-producing oxidant levels in the Los Angeles area and of economic controls on the consumer price index in the United States. Bianchi et al. (1998) analyzed existing and improved methods for forecasting incoming calls to telemarketing centers for the purposes of planning and budgeting. They found that ARIMA models with intervention performed better for the time series studied. Girard (2000) used ARIMA model with intervention in order to analyse the epidemiological situation of whooping-cough in England and Wales for the period of 1940-1990. ARIMA modeling of this illness contains intervention variable as the introduction of widespread vaccination3 in 1957. Mcleod and Vingilis (2005) used power function in intervention analysis to determine the probability that a proposed intervention analysis application will detect a meaningful change. Lam et al. (2009) used a time series intervention ARIMA model to measure the intervention effects and the asymptotic change in the simulation results of the business process reengineering that is based on the activity model analysis. Ismail et al. (2009) studied monthly data of five star hotels' occupancy in Bali city in the aftermath of occurrence of bombing in October, 2002 and have shown that intervention model is more appropriate for forecasting when compared to the conventional ARIMA model. A good account on ARIMA intervention modeling can be found in Box et al. (1994).

2 Material and Methods

2.1 Data Description

Yearly data on cotton yield for Gujarat, Maharashtra and all India were collected from Department of Agriculture and Cooperation (Agricultural Statistics at a Glance 2012) followed by the website http://agricoop.nic.in/agristatistics.htm. Given a data set, the whole of the data set was divided into observations belonging to pre and post intervention periods. The observations in the pre-intervention period were further be divided into two parts viz., data set for model fitting and for validation. Moreover, in the post-intervention period, the first few observations have been utilized for identification of the order of the impact and slope parameters (to be defined subsequently) of the intervention component in the model.

2.2 ARIMA Model Fitting

An ARIMA model is given by: $\phi(B)(1-B)^d y_t = \theta(B)\varepsilon_t$

where

$\phi(B) = 1-\phi_1 B-\phi_2 B^2 - \cdots - \phi_p B^p$ (Autoregressive parameter)

$\theta(B) = 1-\theta_1 B-\theta_2 B^2 - \cdots - \theta_p B^q$ (Moving average parameter)

ε_t =white noise or error term

d= differencing term

B=Backshift operator i.e. $B^a Y_t = Y_{t-a}$

ARIMA methodology is carried out in three stages, viz. Identification, estimation and diagnostic checking. Parameters of ARIMA model are tentatively selected at the identification stage and at the estimation stage parameters are estimated using iterative least square techniques. The adequacy of the selected model is then tested at the diagnostic checking stage. If the model is found to be inadequate, the three stages are repeated until satisfactory ARIMA model is selected for the time-series under consideration.

2.2.1 Identification

Identification of d is necessary to make the non- stationary time series to stationary. A formal statistical test for the existence of stationarity, known as the test of the unit-root hypothesis or Augmented Dickey Fuller test was utilized to test the stationarity. A good account on Augmented Dickey Fuller test can be found in Makridakis et al. (1998).The null hypothesis is that the time series is not stationary and the alternative hypothesis is that the time series is stationary.

2.2.2 Estimation of Parameters

At the estimation stage, parameters are estimated for the ARIMA model tentatively chosen at the identification stage. Estimation of parameters for ARIMA model is generally done through iterative least squares method. Akaike information criterion (AIC) and Bayesian information criterion (BIC) values for ARIMA model are computed by:

$$AIC = T\log(\sigma^2) + 2(p+q+1) \tag{1}$$

and

$$BIC = T\log(\sigma^2) + (p+q+1)\log T \tag{2}$$

where T denotes the number of observations used for estimation of parameters and σ^2 denotes the Mean square error.

2.2.3 Diagnostic-checking

At this stage, testing is done to see if the estimated model is statistically adequate i.e. whether the error terms are white noise which means error terms are uncorrelated with zero mean and constant variance. For this purpose, Ljung-Box test is applied to the original series or to the residuals after fitting a model. A good account on Ljung-Box test can be found in Box et al. (1994).The null hypothesis is that the series is white noise, and the alternative hypothesis is that one or more autocorrelations up to lag m are not zero. The test statistics is given by:

$$Q^* = T(T+2)\sum_{k=1}^{m}\frac{r_k^2}{T-k} \tag{3}$$

where T is the number of observations used to estimate the model and m is the maximum number of lag. The statistics Q^* approximately follows a chi-squared distribution with (T-k) degrees of freedom, where k is the number of parameters estimated in the ARIMA model and r_k is the autocorrelation function of residual at lag k. If it is not satisfactory, we return to the identification stage to tentatively select another model.

2.3 Intervention Model Fitting

An intervention model is given by:

$$Y_t = \frac{\omega(B)}{\delta(B)} B^b I_t + \frac{\theta(B)}{\phi(B)}\varepsilon_t$$

where

Y_t= dependent variable

I_t =Indicator variable coded according to the type of intervention. The intervention type of step function starts from a given time till the last time period. Mathematically, the intervention type of step function is written as:

$$I_t = \begin{cases} 0 & t \neq T \\ 1 & t \geq T \end{cases} \tag{4}$$

with T is time of intervention when it first occurred.

$\delta(B) = 1 + \delta_1 B + ... + \delta_r B^r$ (Slope parameter)

$\omega(B) = \omega_0 + \omega_1 B + ... + \omega_s B^s$ (Impact parameter)

$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$ (Autoregressive parameter)

$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_p B^q$ (Moving average parameter)

ε_t =white noise or error term

b= delay parameter

B=Backshift operator i.e. $B^a Y_t = Y_{t-a}$

As with ARIMA model, fitting the intervention model fitting consists of the usual three stages i.e. identification, estimation, diagnostic checking. While the estimation process and diagnostic checking are similar to ARIMA modeling the identification procedure is somewhat different which is discussed subsequently.

The intervention model consists of three parameters ω, δ and b where ω is known as impact parameter which implies change (either positive or negative) due to intervention and δ is known as slope parameter which has different meanings in case of different types of intervention.. In case of step intervention, if δ is near to zero, the effect of the intervention remains constant over time and if δ is near to one, the effect of intervention increases over time. The delay parameter b usually takes value 0, 1 or 2; b=0 implies that the effect of intervention has occurred at the time of intervention itself, b=1 implies, the effect of intervention is felt after a delay of one period and so on.

The order of b can be determined by examining the data visually and the form of the model is ascertained by comparing computed impulse response functions with theoretical impulse response functions. The impulse response function is obtained by plotting the residual which is the absolute difference between the actual values of the post-intervention observations with the forecasted value obtained by ARIMA model which fitted on the basis of pre-intervention data.

2.4 Forecasting Performance

Forecasting performance of the model has been judged by computing Mean Absolute Percent Error (MAPE). The model with less MAPE is preferred for forecasting purposes. The MAPE is computed as

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|Y_t - \hat{Y}_t\right| / Y_t \times 100 \tag{5}$$

where n is the total number of forecast value. Y_t is the actual value at period t and $\hat{Y}_t$ is the corresponding forecast value. Less the MAPE better the forecast. Statistical analysis of ARIMA with Intervention has been done using Statistical Analysis Systems (SAS), USA, Version 9.2, Module SAS-ETS, available at Indian Agricultural Statistics Research Institute, New Delhi.

3 Results and Discussion

All India

Data on cotton yield during 1961-1991 is used for pre-intervention ARIMA model fitting and 1992-2001 is used for model validation. Data during 2002-2006 i.e. 5 post-intervention observations have been used to know the intervention component form. Generally two or three postintervention is sufficient to identify the final intervention model but as in case of All India by examining the data visually it is observed that b=2 which is also observed in impulse response function more data points are required to know the model form. ARIMA (1, 1, 1) is selected for preintervention data as its MAPE is least. Now using ARIMA (1, 1, 1) model forecasting is done to first 5 post intervention observations for computing the impulse response function because here the delay period is 2 and there is no significant difference between the year 2004 and 2005. The impulse response function is given below in Fig-1.

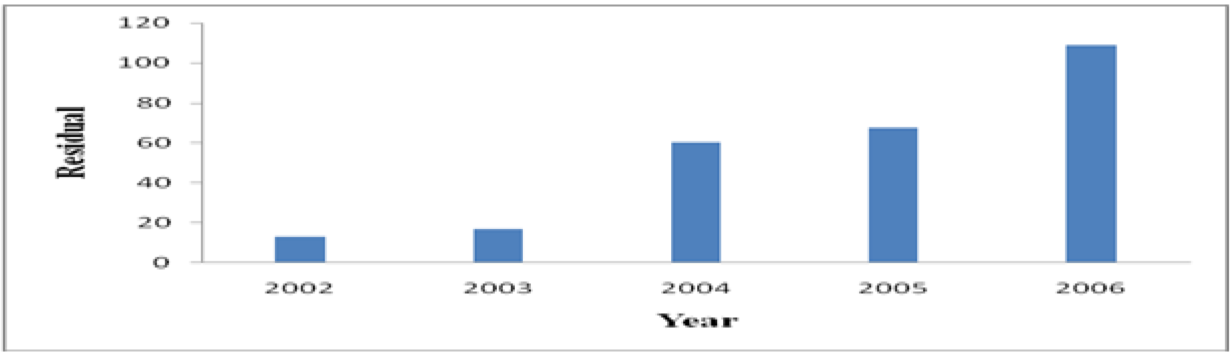

Fig-1 Impulse Response Function (All-India)

From impulse response function, it can be inferred that b=2 i.e. though the introduction of Bt-cotton has occurred in 2002, its effect was felt only in 2004.It can be also inferred that there are two intervention components ω and δ involved in the model.

Parameter estimates of ARIMA-Intervention are given in Table 1.

In the above table all the parameters except μ are significant at 5% level of significance and

Table 1: Parameter estimates of the ARIMA-Intervention model (All-India)

MODEL Step:44/(1) + ARIMA(1,1,1)				
Model Parameter	**Estimate**	**Std.error**	**T**	**Prob>|t|**
Gen. mean (μ)	2.59	1.54	1.68	0.0998
AR (1) coefficient (ϕ_1)	0.71	0.22	3.22	0.0024
MA (1) coefficient (θ_1)	0.11	0.31	2.37	0.0138
Impact (ω)	104.05	19.96	3.41	0.0014
Slope (δ)	0.18	0.09	7.52	<.0001
Model variance($\hat{\sigma}^2_\omega$)	789.58	.	.	.

μ is significant at 10% level of significance.

The final ARIMA-Intervention model obtained is given by (deviated from mean)-

$$Y_t' = \frac{104.05}{1+0.18B} B^2 I_t + \frac{1-0.11B}{1-0.71B}\varepsilon_t \qquad \text{where, } \; Y_t' = Y_t - Y_{t-1}$$

The model can be written in the form-

$$Y_t' = (0.71-0.18)Y_{t-1}' + (0.18\times 0.71)Y_{t-2}' + 104.05 I_{t-2} - (0.71\times 104.05) I_{t-3} - (0.18\times 0.11)\varepsilon_{t-2} + (0.18-0.11)\varepsilon_{t-1} + \varepsilon_t$$

Now for the remaining 3 observations forecasting is done one with ARIMA Model other with ARIMA-Intervention model and MAPE is computed. The result is given in Table 2.

Table 2

		MODEL	
		ARIMA-INTERVENTION	**ARIMA**
YEAR	**ACTUAL**	**FORECAST**	**FORECAST**
2007	467	439.67	413.12
2008	403	451.48	418.56
2009	403	463.78	423.43
2010	499	487.16	428.23
2011	491	502.84	433.09
	MAPE	7.54	9.28

From table-2, it can be inferred that the accuracy of forecasting of ARIMA-intervention model is more than ARIMA model.

Gujarat

Data during 1961-1991 is used for pre-intervention ARIMA model developing and 1992-2001 is used for model validation. Data during 2002-2004 i.e. 3 post-intervention observations have been used to know the intervention component form. By examining the data visually, it is observed that b=1 which is also observed in impulse response function. Model ARIMA (2, 0, 0) the model has least BIC as well as least MAPE. Now using ARIMA (2, 0, 0) model, forecasting is done to first 3 post-intervention observations for computing the impulse response function which is given below in Fig-2.

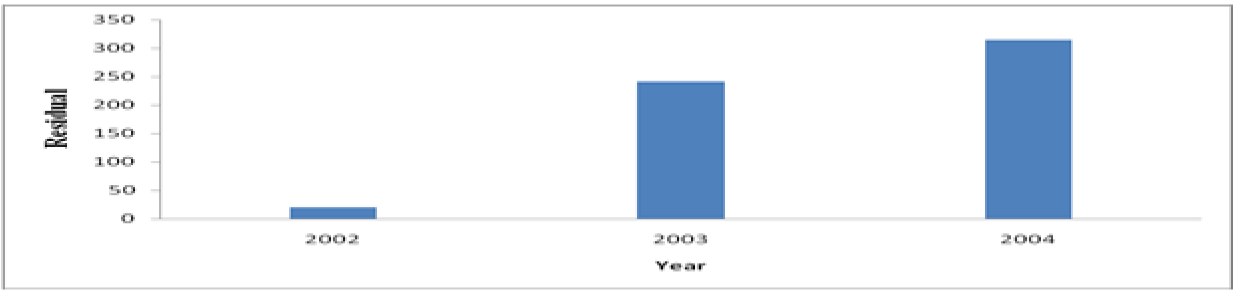

Fig-2 Impulse Response Function (Gujarat)

From impulse response function, it can be inferred that b=1 i.e. though the introduction of Bt-cotton has occurred in 2002, its effect only felt in 2003. It can be also inferred that there are two intervention components ω and δ involved in the model.

Parameters estimate of the ARIMA-Intervention are given in Table 3.

Table 3: Parameters estimates of the ARIMA-Intervention model

Model Parameter	Estimate	Std.error	t	Prob>|t|
General mean (μ)	172.75	7.82	22.09	< .0001
AR(1) coefficient(ϕ_1)	0.49	0.17	2.79	0.0093
AR(2) coefficient(ϕ_2)	-0.36	0.18	-1.98	0.0571
Impact(ω)	89.05	23.59	2.66	0.0138
Slope(δ)	0.38	0.56	7.52	0.0354
Model variance($\hat{\sigma}^2_\omega$)	807	.	.	.

In the above table all the parameters (except ϕ_2) are significant at 5% level of significance and ϕ_2 is significant at 6% level of significance.

The final ARIMA-Intervention model obtained is given by (deviated from mean) is-

$$Y_t = \frac{89.05}{1+0.38B} B^1 I_t + \frac{1}{1-0.49B+0.36B^2} \varepsilon_t$$

The model can be linearly represented as

$$Y_t = (0.49-0.38)Y_{t-1} + (0.49\times0.38-0.36)Y_{t-2} - 0.36\times0.38Y_{t-3} + 89.05I_{t-1} - 89.05\times0.49I_{t-2} + 0.36\times89.05I_{t-3} + 0.38\varepsilon_{t-1} + \varepsilon_t$$

Now for the remaining 4 observations forecasting is done one with ARIMA Model and other with ARIMA-Intervention model and MAPE is computed. The result is given in Table 4.

Table 4

		MODEL	
		ARIMA-INTERVENTION	**ARIMA**
YEAR	**ACTUAL**	**FORECAST**	**FORECAST**
2005	604	609.39	437.05
2006	625	632.98	443.88
2007	581	691.32	451.08
2008	507	732.12	459.90
2009	672	757.98	464.67
2010	689	765.56	471.02
2011	653	773.45	479.72
	MAPE	15.41	25.33

From table-4 it can be inferred that the accuracy of forecasting of ARIMA-intervention model is better than ARIMA model.

Maharashtra

Data of 1961-1991 has been used for pre-intervention ARIMA model development and 1992-2001 has been used for model validation. Data during 2002-2004 i.e. 3 post-intervention observations have been used to know form of the intervention component. Model ARIMA (2, 1, 0) is selected for preintervention data as its MAPE is least. Now using ARIMA (2, 1, 0) model forecasting has been done only for the first 3 post intervention observations for computing the impulse response function which is given below in Fig-3.

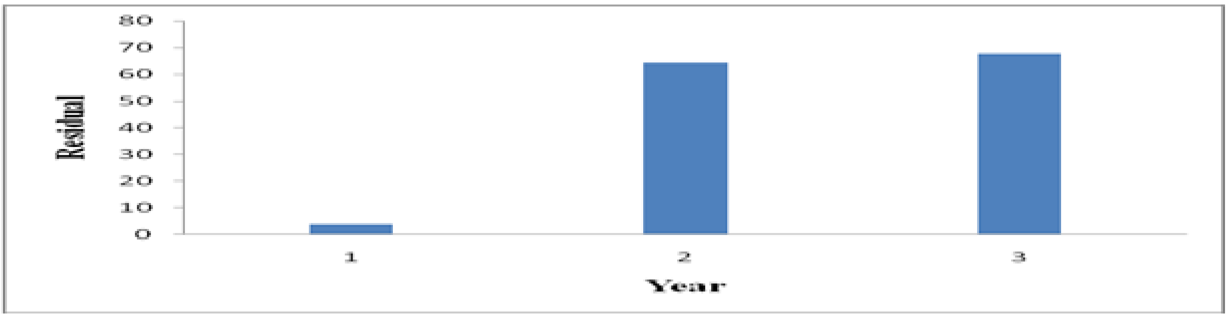

Fig-3 Impulse response function (Maharashtra)

From impulse response function, it can be inferred that b=1 i.e. though the introduction of Bt-cotton has occurred in 2002, its effect has only been felt in 2003. It can be also inferred, that there are two intervention parameters ω and δ involved in the model.

Parameters estimate of the ARIMA-Intervention is given in Table 5.

Table 5: Parameters estimates of the ARIMA-Intervention model

Model Parameter	Estimate	Std.error	T	Prob>|T|
General mean(μ)	1.71	1.9588	1.8731	0.0879
AR (1) coefficient(ϕ_1)	-0.34	0.1318	-4.8985	<.0001
AR(2) coefficient (ϕ_2)	-0.56	0.1340	-4.2041	0.0001
Impact(ω)	61.48	28.3128	2.1715	0.0360
Slope(δ)	0.49	0.4817	2.1430	0.0411
Model variance($\hat{\sigma}_{\omega}^2$)	753.09	.	.	.

In the above table, all the parameters except μ is significant 5% level of significance and μ is significant from 10% level of significance.

Final ARIMA-Intervention model obtained is given from by -

$$Y_t' = \frac{61.48}{1+0.49B} B^1 I_t + \frac{1}{1+0.34B+0.56B^2}\varepsilon_t \qquad \text{where} \quad Y_t' = Y_t - Y_{t-1}$$

Model is linearly written as-

$$Y_t' = -(0.49+0.34)Y_{t-1}' - (0.49\times 0.34+0.56)Y_{t-2}' - (0.49\times 0.56)Y_{t-3}' + 61.48I_{t-1} + 0.34\times 61.48I_{t-2} + 0.56\times 61.48I_{t-3} + 0.49\varepsilon_{t-1} + \varepsilon_t$$

Now for the remaining 4 observations, forecasting is done with ARIMA Model and the other with ARIMA-Intervention model and MAPE is computed. The result is given in Table 6.

Table 6

		MODEL	
		ARIMA-INTERVENTION	ARIMA
YEAR	ACTUAL	FORECAST	FORECAST
2005	230	263.45	205.13
2006	253	277.73	215.54
2007	373	308.35	220.08
2008	395	335.60	229.50
2009	367	373.18	235.68
2010	297	412.89	245.76
2011	313	445.12	257.78
	MAPE	19.94	25.59

From table-6, it can be inferred that the accuracy of forecasting of ARIMA-intervention model is better than ARIMA model.

4 Conclusions

Intervention analysis or event study is used to assess the impact of a special event on the time series of interest. Alternatively, intervention analysis may be undertaken to adjust for any unusual values in the series Y_t that might have resulted as a consequence of the intervention event. This will ensure that the results of the time series analysis of the series, such as the structure of the fitted model, estimates of model parameters, and forecasts of future values, are not seriously distorted by the influence of these unusual values. The present study deals with investigations of time series intervention modeling in the domain of agriculture. As a case study, Cotton yield of India at all-India level and for two major states viz., Gujarat and Maharashtra have been considered with the step intervention being introduction of Bt-Cotton variety in 2002.It has been found that in all the three locations there is a significant change in cotton yield due to intervention of Bt-cotton as indicated by the impact parameter. When cotton yields were forecasted, the performance of ARIMA intervention models was found to be superior to the conventional ARIMA models for all the three locations because MAPE was always less in case of ARIMA-Intervention model than ARIMA model. In addition, there appeared to be a small delay in the effect of intervention in all cases at varied magnitudes and differential slopes over time. Thus it can be concluded that time series intervention modelling can be usefully employed for forecasting cotton yield.

References

Bianchi, L., Jarrett, J., Hanumara, R.C. (1998). Improving forecasting for telemarketing centres by ARIMA modeling with intervention. *International Journal of Forecasting,* **14,** 497-504.

Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (2009), *Time Series Analysis: Forecasting and Control (3rd ed.),* San Francisco: Holden-Day.

Box, G.E.P., and Tiao, G.C. (1975).Intervention Analysis with Application to Economic and Environment Problems. *Journal of the American Statistical Associations*, **70,** 70-79.

Brakel, J. V. D., Roels, J. (2010). Intervention analysis with state-space models to estimate discontinuities due to a survey redesign. *The Annals of Applied Statistics*, **4,** 1105-1138.

Girard, D. Z.(2000).Intervention times series analysis of pertussis vaccination in England and Wales. *Health Policy*, **54,** 13-25.

Ismail, Z., Suhartono., Yahaya, A., and Efendi, R.(2009). Intervention Model for Analyzing the Impact of Terrorism to Tourism Industry. *Journal of Mathematics and statistics*, **4,** 322-329.

Lam, C.Y., Ip, W.H., and Lau, C.W.(2009). A business process activity model and performance measurement using a time series ARIMA intervention analysis. *Expert Systems with Applications,* **36,** 925-932.

Makridakis, S., Wheelwright, S.C., and Hyndman, R. J.(1998).*Forecasting: Methods and Applications (3rd ed.),* Chichester: Wiley.

Mcleod, A. I., and Vingilis, E. R.(2005). Power computations for intervention analysis. *Technometrics,* **47,**174-181

Author for correspondence

Mrinmoy Ray
Indian Agricultural Statistics Research Institute,
Library Avenue
New Delhi

Received: 10 March, 2014 *Revised: 2 April, 2014* *Accepted: 10 November, 2014*