Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
`http://siba-ese.unisalento.it/index.php/ejasa/index`
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v12n3p674

Forecasting an explosive time series

By Chandra, Prabakaran

Published: 20 November 2019

Electronic Journal of Applied Statistical Analysis
Vol. 12, Issue 03, November 2019, 674-35
DOI: 10.1285/i20705948v12n3p674

Forecasting an explosive time series

K. Suresh Chandra*[a] and S. Prabakaran[b]

[a]*Former Professor, Department of Statistics, University of Madras*
[b]*Assistant Commissioner of Police, Puzhal range, Greater Chennai Police, Tamilnadu*

Published: 20 November 2019

Forecasting is an important exercise in Time series analysis. For a stationary time series, there are theoretically strong forecasting methods which can provide most accurate forecasts for the future (Karlin and Taylor, 1975). For most non stationary time series Box Jenkins methodology is a useful forecasting technique. Essentially, the Box Jenkins methodology assumes that any non stationarity time series can be conveniently modeled as an Autoregressive Intregrated Moving Averages (ARIMA) model with sufficient number of "unit roots" in the linear stochastic difference equation generating the time series. The non stationarity in such time series is then removed by successively differencing of the series until one obtains a stationary series, for which optimal forecasts can be computed. The forecasts for the original series are then computed by 'inverting' the difference operators that were used (Makridakis et al., 1998) on the forecasts computed for the stationary series. The main objective of this study is to demonstrate that the Box Jenkins methodology is not useful, especially in large time series, when the non stationarity in the time series is due to 'explosive' roots. An alternative method is proposed in such a situation and its performance is assessed both on a simulated as well as on a real life data.

keywords: Stochastic difference equation, Unit and explosive roots, ARIMA model,Rate of convergence in probability, Auxiliary processes.

1 Introduction

The unidirectional time dependence in time series paves way for forecasting the future, given the current and past. The stochastic properties of a stationary time series have

*Corresponding author: vijsur2012@gmail.com

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

provided several theorems (Karlin and Taylor, 1975, Box and Jenkins, 1976, Dickey and Fuller, 1979) that help computing efficient forecasts (like minimum mean square error forecasts). For a non stationary time series the forecasting methodology due to Box and Jenkins (BJ) (Box and Jenkins, 1976, Makridakis et al., 1998) has become a handy tool in the hands of researchers. In this paper, we demonstrate that this methodology will not be useful if the non stationarity in time series is due to the 'explosive roots' associated with the linear stochastic difference equation model that generates the time series, especially in long time series. An alternative method for forecasting is proposed which can be more accurate than the BJ method. The necessary theoretical background for the new method has been presented in the next section. In Section 3, we highlight some interesting features of the proposed estimates of the explosive roots. In Section 4, we review the tools for evaluation of forecasts. A limitation of the Box Jenkins forecasting methodology is highlighted in Section 5. Section 6 is devoted to the description of the new forecasting method. Sections 7 and 8 discuss the relative performance of the new forecasting method, in comparison to the BJ method through simulated as well as real data sets. The last section (Section 9) is reserved for discussion.

2 Theoretical background

Let the time series $X = \{X(t); t = 0, \pm 1, \pm 2, \ldots\}$ be generated by a linear stochastic difference equation given by

$$\begin{aligned} X(t) &= \alpha_0 + \alpha_1 X(t-1) + \cdots + \alpha_m X(t-m) + (\epsilon_t + \beta_1 \epsilon_{t-1} + \cdots + \beta_q \epsilon_{t-q}) \\ &= \alpha_0 + \alpha_1 X(t-1) + \cdots + \alpha_m X(t-m) + \eta_t, \text{ say} \end{aligned} \quad (1)$$

where m and q are non negative integers, and $\{\epsilon_t; t = 0, \pm 1, \pm 2, \ldots\}$ is a family of i.i.d random variables with

$$E(\epsilon_t) = 0 \text{ and } V(\epsilon_t) = \sigma^2, \text{ for all } t. \quad (2)$$

In this context, a given set of time series data $X(t); t = 1, 2, \ldots, n$ is considered as a partial realization of the time series X .

Strict and weak stationarity assumptions are well known in the literature. A simple way to define a non stationary time series is that it is neither strictly or weakly stationary.

Generally stationarity is defined for a doubly infinite time series. But, in most situations, one may have to consider a time series when time epochs takes values only on the positive side of origin. However, if one is interested in developing asymptotic theories, like convergence, rate of convergence etc, then the series can be set on the right side of origin. Such a series turns out to be asymptotically stationary in the sense that after large values of t the time series behaves like a stationary time series.

To be precise a time series X is said to be asymptotically stationary (in the wide sense) if (i) $E(X(t)) \to m$ (a constant) as $t \to \infty$ and (ii) $Cov\{X(t), X(t+h)\} \to c(h)$ (a function of h only) as $t \to \infty$.

In many situations, while deriving the asymptotic results, the aggravated algebra involved can be easily handled on assuming, without loss of generality, that $X(t) = \epsilon_t = 0$ for $t < 0$.

The time path of X depends on the roots of the characteristic polynomial $P(z)$ given by

$$P(z) = z^m - \alpha_1 z^{(m-1)} - \cdots - \alpha_m \tag{3}$$

associated with (1). Let us assume that out of the m roots of $P(z) = 0$, k are numerically larger than 1, d are numerically equal to 1 and p are numerically less than 1, where k, d, and p are non negative integers such that $m = k + d + p$. For the sake simplicity let us assume that the roots that are numerically greater than unity are real positive and distinct. The roots that are numerically larger than one are called the explosive roots of $P(z) = 0$. The roots that are numerically smaller than one are called the non explosive roots of $P(z) = 0$. This assumption (the roots being real and distinct) can always be relaxed for establishing asymptotic properties relating to the time series (Venkataraman, 1968) Let the m roots $\rho_1, \rho_2, \ldots, \rho_m$ of $P(z) = 0$ have the placement

$$\begin{aligned} \rho_1 > \rho_2 > \cdots > \rho_k > 1 \\ \rho_{k+1} = \rho_{k+2} = \cdots = \rho_{k+d} = 1 \\ |\rho_i| < 1,\ i = k + d + 1, \ldots, m. \end{aligned} \tag{4}$$

Under this set up, the time series X is said to be *purely explosive* if $k > 0, d = 0$ and $q = 0$, *explosive* if $k \geq 1$, *partially explosive* if $k \geq 1$, $d = 0$ and $p \geq 1$ and *non explosive* if $k = d = 0$.

It can be noted that the autoregressive (AR), the moving average (MA), the autoregressive with moving average errors (ARMA), the Autoregressive Integrated Moving Average (ARIMA) models are particular cases of the model (1).

Note: The form of the characteristic polynomial $P(z)$ in (3) is different from the one given in standard text books (Kendall and Ord, 1990) on time series analysis. There should not be any confusion if the reciprocal of the roots are used for classification of explosive and non explosive roots.

The least squares estimation of the coefficients of (1) and their properties have been extensively studied by Venkataraman (1968, 1973, 1974), under both explosive and partially explosive assumptions. All results on least squares estimation relating to (1) use the properties of the following auxiliary processes defined by Venkataraman (1968) for the explosive case. With reference to the time series $X = \{X(t); t = 0, \pm 1, \pm 2, \ldots\}$ they are recursively defined by

$$\begin{aligned} X_1(t) &= X(t) \\ X_2(t+1) &= X_1(t+1) - \rho_1 X_1(t) \\ &\ \ \vdots \\ X_{k+1}(t+k) &= X_k(t+k) - \rho_k X_k(t+k-1) \end{aligned} \tag{5}$$

Note: The auxiliary processes are defined only for the explosive time series. It is not difficult to conceptualize such processes for imaginary roots. The process $\{X_{k+1}(t+k)\}$ becomes an ARIMA model if $d > 0$ else an ARMA model.

The following theorem summarizes an important property of these processes.

Theorem 1. *Let the time series X be generated by the model (1). Then under the underlying assumptions and notation in (2) to (5) and for each $r = 1, 2, \ldots, k$.*

$$\rho_r^{-t} X_r(t) \xrightarrow{p} V_r \text{ as } t \to \infty$$

where V_r is a random variable.

Proof. Let us assume, for the time being, that $d = 0$. For each r, $(1 \le r \le k)$, the assumptions in (4) lead to the factorization of $P(z)$ as

$$P(z) = P_1(z)P_2(z) \tag{6}$$

where $P_1(z)$ is a polynomial of degree $(r-1)$ with roots $\rho_1, \rho_2, \ldots, \rho_{r-1}$ and $P_2(z)$ is a polynomial of order $(m-r+1)$ with roots $\rho_r, \rho_{r+1}, \ldots, \rho_m$. When $r = 1$, we set $P_1(z) = 1$. Let

$$\begin{aligned} P_1(z) &= z^{r-1} + \gamma_1 z^{r-2} + \cdots + \gamma_{r-1} \\ P_2(z) &= z^{m-r+1} + \beta_1 z^{m-r} + \cdots + \beta_{m-r+1} \end{aligned} \tag{7}$$

This facilitates the identification of the stochastic difference equation generating the auxiliary process $X_r(t)$ as

$$X_r(t+m) = \alpha_0 + \beta_1 X_r(t+m-1) + \beta_2 X_r(t+m-2) + \cdots + \beta_{m-r+1} X_r(t+r-1) + \eta_{t+m}$$

η_t being defined in (1). Since the above equation holds for all $t \ge 1$,one can rewrite the above equation, as given below, that will hold for all $t \ge r$ or say for all $t \ge m$. Statements hence forth hold for all r and $t \ge m$.

$$X_r(t+m-r+1) = \alpha_0 + \beta_1 X_r(t+m-r) + \beta_2 X_r(t+m-r-1) + \cdots + \beta_{m-r+1} X_r(t) + \eta_{t+m-r+1} \tag{8}$$

Let us, for algebraic simplicity, assume that $X(t) = \epsilon_t = 0$ for $t < 0$. Solving the above difference equation, we have

$$X_r(t) = \alpha_0 \sum_{j=0}^{t-1} \lambda(j) + \sum_{j=0}^{t-1} \lambda(j) \eta_{t-r} \tag{9}$$

It is easy to check that, $\lambda(0) = 1$ and for $j > 1$

$$\lambda(j) - \beta_1 \lambda(j-1) - \beta_2 \lambda(j-2) - \cdots - \beta_{m-r+1} \lambda(j-(m-r+1)) = 0 \tag{10}$$

Solving (10) we get $\lambda(j) = \sum_{i=1}^{m-r+1} c_i \rho_{r+i-1}^j$ where $c's$ are well defined constants. In other words for each j, $\lambda(j)$ is a linear combination of j^{th} powers of $\rho_r, \rho_{r+1}, \ldots, \rho_m$.

On noting that ρ_r is the largest explosive root of $P_2(z) = 0$, we can rewrite (9) as

$$X_r(t) = \alpha_0 \rho_r^t \sum_{j=0}^{t-1} \rho_r^{-t} \lambda(j) + \rho_r^t \sum_{j=0}^{t-1} \rho_r^{-t} \lambda(j) \eta_{t-r}. \quad (11)$$

Setting $u = t - j$, we have

$$\begin{aligned} \rho_r^{-t} X_r(t) &= \alpha_0 \sum_{j=0}^{t-1} \left(\frac{\lambda(j)}{\rho_r^t} \right) + \sum_{j=0}^{\infty} \left(\frac{\lambda(j)}{\rho_r^t} \right) \eta_{t-j} - \sum_{j=t+1}^{\infty} \left(\frac{\lambda(j)}{\rho_r^t} \right) \eta_{t-j} \\ &= \alpha_0 \sum_{j=0}^{t-1} \left(\frac{\lambda(j)}{\rho_r^t} \right) + \sum_{u=1}^{\infty} \left(\frac{\lambda(t-u)}{\rho_r^t} \right) \eta_u - \sum_{u=t+1}^{\infty} \left(\frac{\lambda(t-u)}{\rho_r^t} \right) \eta_u. \quad (12) \end{aligned}$$

On recalling the assumptions on ϵ_t, the placement of roots in (4) and the structure of $\lambda(.)$, it can be checked that, as $t \to \infty$ (i) the first term on the left hand side (LHS) of (12) has a finite limit, (ii) the second term converges in mean square to a random variable and (iii) the third term converges in mean square to 0. Hence the theorem holds on appealing to standard convergence theorems (Bhat, 1999, Venkataraman, 1968, Venkataraman, 1974). In fact the structure of V_r can be checked to be

$$V_r = \sum_{j=0}^{\infty} \left(\frac{\lambda(j)}{\rho_r^t} \right) + \sum_{j=0}^{\infty} \left(\frac{\lambda(j)}{\rho_r^t} \right) \eta_{t-j}. \quad (13)$$

□

Note: The theorem holds even when $d > 0$, on a slight modification in $\lambda(j)$ and on using the fact that $\sum_{j=0}^{\infty} \rho_r^{-t} t^a$ is convergent for $a > 0$ when $\rho_r > 1$.

Suresh Chandra et al. (1994, 1999) and Suresh Chandra and Janhavi (2008) have discussed the estimation of the explosive roots of (1). With reference to the partial realization of the time series $X(1), X(2), \ldots, X(n)$, let $\hat{\phi}_1, \hat{\phi}_2, \ldots \hat{\phi}_k$ be proposed as estimates of the explosive roots $\rho_1, \rho_2, \ldots, \rho_k$, recursively computed as follows:

$$\begin{aligned} \hat{\phi}_1 &= \frac{\sum_{t=1}^{n-1} X(t+1)X(t)}{\sum_{t=1}^{n-1} X(t)^2} \\ \hat{\phi}_i &= \frac{\sum_{t=1}^{n-i+1} X_i^*(t+i)X_i^*(t+i-1)}{\sum_{t=1}^{n-i} [X_i^*(t+i-1)]^2}, i = 2, 3, \ldots, k, \quad (14) \end{aligned}$$

where

$$X_i^*(t+i+1) = X_{i-1}^*(t+i+1) - \hat{\phi}_{i-1} X_{i-1}^*(t+i), i = 2, 3, \ldots, k. \quad (15)$$

Note: $\{X_i^*(t)\, t = 1, 2, \ldots, n-i+1\}, i = 2, 3, \ldots, k$ are essentially the estimated auxiliary processes associated with the given time series X and $\hat{\phi}_2, \hat{\phi}_3, \cdots \hat{\phi}_k$ are the second raw moments of these estimated auxiliary processes.

The following theorem, established by Suresh Chandra et al. (1999), which plays a crucial role in the sequel, is stated without proof.

Theorem 2. *Let the time series X be generated by the model (1). Then under the basic assumptions and notation in (2) to (5) and (14), each of*

$$\{(a(i))^n(\hat{\phi}_i - \rho_i)\}, i = 1, 2, \ldots, k$$

is bounded in probability, where

$$a(i) = min\left\{\frac{\rho_i}{\rho_{i+1}}, \frac{\rho_{i-1}}{\rho_i}\right\}, i = 1, 2, \ldots, k$$

on setting $\rho_0 = \infty$ *and* $\rho_{k+1} = 1$.

There are two points to be noted at this stage. The quantities $a(i)$ are numerically larger than unity. The quantities $(a(i))^n$ are the rates of convergence in probability of $\hat{\phi}_i$ to ρ_i respectively.

3 Some interesting features of the estimates of the explosive roots

3.1 Exponential rate of convergence

The rates of convergence of the estimates of the roots are exponential in nature. The exponential rates portray fast convergence of the estimates making the estimates very precise even at moderately large sample sizes.

As it is analytically difficult to derive the exact sampling distribution of $\hat{\phi}_i$. An extensive simulation study (Prabakaran, 2015) was carried out for the case $m = 1, p = 0$ and $q = 0$ so that $k = 1$ varying ρ in the interval $[\,1 < \rho < 1.2\,]$. The data size for the simulation study were set at n= 50, 100 and 250 (matching the values set in Dickey and Fuller, 1979). The standard deviation of the error term were set at 1, 10 and 50. These simulations are simple and easy to carry out in R environment and the details are omitted for the sake of brevity. Important features of the outcome of the simulation study are that the sampling distribution of $\hat{\phi}_i$ are (i) unimodal (ii) free from the variance of the error term (as expected even from a theoretical point of view), (iii) degenerate (even at sample size=100) and shift to the right asρ increases and (iv) the effect of the constant term on the estimate diminishes exponentially as sample size increases.

3.2 Unit root tests and explosive roots

Unit root tests, especially the Dickey-Fuller test or the Augmented Dickey-Fuller (ADF) test (Dickey and Fuller, 1979), are constructed to test the null hypothesis H : a root of the stochastic difference equation generating the given time series is equal to one against the alternative K : that root is less than one.

For any unit root test, the left tail is used as the critical region to test $H : \rho = 1$ under the alternative $K : \rho < 1$ (as in the case of Dickey Fuller tests). The degeneracy and the shifts to right (as ρ increases from unity) of the sampling distributions of $\hat{\phi}_i$ revealed in the simulation study, suggest that, that (i) any level α test for H against K is also a level α test for testing $H_1 : \rho \geq 1$ against K and (ii) the power of the level α test for H against K is larger than any level α test for testing $H_1 : \rho \geq 1$ against K. This information can lead us to the empirically justifiable conclusion that the distribution of $\hat{\phi}_i$ when $\rho = 1$ is least favourable (vide Lehmann, 1986, p.105).

Hence it is possible to conclude that, even in moderately large samples, the acceptance of the null hypothesis H in unit root tests can indicate presence of explosive roots and the rejection of the hypothesis will certainly reject the presence of explosive roots. This finding is of practical relevance in the determination of a stopping rule in the recursive estimation of $\hat{\phi}_i$ and hence the value of k.

4 Evaluation of the forecasts

There are several criteria for assessing the performance of a forecasting model based on the difference between the observed value and the forecast value. Since we normally do not have the observed value for a future time, the evaluation should be based on the differences observed when both observation and forecast are available. For this reason, a commonly used procedure is to divide the period of observation into two mutually exclusive intervals: a training set and a test set. Usually the test set is identified very near to the forecast horizon and the training set is a period just before the test set. The observations in the training set are used to develop a forecast model, in terms of estimation of the coefficients in the assumed model and the training set for the evaluation of the forecasts. Let e_t denote the difference between observed value and the forecast value at time t in the test set. Some of the commonly used absolute measures (Makridakis et al., 1998) for the evaluation of a forecasting method are listed below.

$$\text{Mean Error} = \text{ME} = \{(1/r)\sum_{t=1}^{r} e_t\}$$
$$\text{Mean Absolute Error} = \text{MAE}= \{(1/r)\sum_{t=1}^{r} |\, e_t\,|\}$$
$$\text{Mean Square Error} = \text{MSE}= \left\{(1/r)\sum_{t=1}^{r} e_t^2\right\}$$

where r is the number of observations in the test set.

5 The inadequacy of the Box Jenkins forecasting methodology for explosive models

ARIMA models are the most frequently used by researchers for analyzing a non stationary time series data.Typically (with reference to the underlying stochastic difference equation) an ARIMA model has three specification parameters: $p =$ the number of the non explosive roots which are numerically less than one, $d =$ the number of unit roots and $q =$ the length of the moving average error.

The Box-Jenkins (BJ) methodology provides a systematic procedure to identify the values of the specification parameters, p, d and q. A simple heuristic proof to the fact that differencing does not eliminate the explosive root follows from Theorem 1. For large t, $X(t) \sim \rho_1^t$ which implies that $(X(t+1) - X(t)) \sim \rho_1^t$, thereby both series have the same explosive path. Alternatively, from (1), a simple algebra will reveal that the characteristic polynomial of the differenced series will be the same as that of the original series unless there is a unit root, in which case the differencing would remove that root keeping others unchanged. This fact has motivated the search for an alternative method of forecasting for explosive models.

However, when the time series is not long enough and explosive root is nearer to unity, the exponential path of the time series can be approximated by a polynomial trend by virtue of the Stone-Weierstrass Theorem and hence the differencing may remove the effect of the explosive root. But, if the explosive root is away from unity, the differencing can not be useful.

6 The proposed forecasting procedure

A feature of an explosive time series is that it has an exponential growth so that the values increase continuously. There are several examples for explosive time series. For instance, the population size, some economic indicators like the Gross Domestic Product, not to speak of examples in biology and nuclear science.

The auxiliary processes (5) are useful in eliminating explosive roots. The possibility of consistently estimating the explosive roots and their properties as revealed in the discussions in the Section 2 are exploited to describe the proposed forecasting procedure.

We retain the basic philosophy of Box-Jenkins methodology of identifying and removing the non stationary part of the given time series, not by differencing the series , but by using the estimated auxiliary processes to eliminate the explosive roots. Thus, the algorithm for forecasting is described as follows.

1. At the first instance, we compute $\hat{\phi}_1$. If $\hat{\phi}_1 \leq 1$, we proceed along Box-Jenkins for identifying an ARIMA model and use it for forecasting. If $\hat{\phi}_1 > 1$ we move to stage 2.

2. We eliminate the largest explosive root,ρ_1, by constructing the data $X_2^*(t); t = 2, 3, \ldots, n$ relating to the second order auxiliary process . Using this data, we compute $\hat{\phi}_2$ which is an estimate of the largest root of the process $\{X_2^*(t); t = 2, 3, \ldots, \infty\}$ which is, incidentally, the second largest root of the original process X. If $\hat{\phi}_2 \leq 1$, we proceed along Box-Jenkins methodology for identifying an ARIMA model and use it for forecasting for the series $\{X_2(t)\}$. If $\hat{\phi}_2 > 1$ we move to the next stage, which is a repetition of this stage, but relating auxiliary process of third order $\{X_3^*(t); t = 3, 4, ..., \infty\}$.

3. We continue the process till we reach a stage $\hat{\phi}_r \leq 1$. At this stage k will be equal to $(r-1)$ and, and we would have eliminated all explosive roots, so that the resulting estimated auxiliary processes $\{X_{k+1}^*(t); t = k+1, k+2, ..., \infty\}$, is modeled and forecasted as an ARIMA model, using the Box-Jenkins methodology.

4. The forecasts for the given explosive time series is obtained from ARIMA forecasts for the series $X^*_{k+1}(t); t = k+1, k+2, \ldots, n$, by inverting the operators used in the definition of estimated auxiliary processes used up to this stage.

In order to elaborate what exactly we mean by inverting the operators used for auxiliary process we note from the defining equations in (5) for the auxiliary processes that, for a specified k

$$X_{k+1}(t+k) = X(t+k) - a_1 X(t+k-1) + a_2 X(t+k-2) - \cdots - (-1)^k a_k X(t) \quad (16)$$

where a_r is the sum of products of $\rho_1, \rho_2, \ldots \rho_k$ taken r at a time, for $r = 1, 2, \cdots, k$. For forecasting the values for the original process (X), we use the estimated version of (16) given by

$$X^*_{k+1}(t+k) = X(t+k) - a^*_1 X(t+k-1) + a^*_2 X(t+k-2) - \cdots - (-1)^k a^*_k X(t) \quad (17)$$

where a^*_r is the sum of products of $\hat{\phi}_1, \hat{\phi}_2, \cdots \hat{\phi}_k$, defined in (13) taken r at a time, for $r = 1, 2, \ldots, k$ and $X^*_{k+1}(t+k+1)$ is defined in (14). Let $f_{k+1}(t)$ denote the BJ forecast at t. Let $f(n+h)$ denote the (required) forecast at for a horizon $(h > 1)$ for the original process X. This can be computed using (16) recursively by

$$f(n+h) = f_{k+1}(n+h) + a^*_1 X(n+h-1) - a^*_2 X(n+h-2) - \cdots + (-1)^{k+1} a^*_k X(n+h-k) \quad (18)$$

If the previous values of $X(t)$, used in (17) at any time point $(n+h)$, is not known, then it has to be replaced by its forecast. The recursive nature of computation suggests that for any h, the procedure should start from $t = 1$.

7 Illustration based on simulated data

In this section we give an illustration based on a simulated data that exposes the better performance of the proposed method when compared to BJ method. We simulated a time series data of 115 observations generated by the model (1) setting $k = 1, d = 1$ and $p = 1$ and $q = 0$. In the study the three roots associated with the generated data were 1.1, 1 and 0.5, with intercept $\alpha_0 = 100$. The error terms were assumed to be iid normal random variables with mean 0 and variance 1. Simulated data is appended to this paper as Appendix-I. The the data clearly indicates the explosive nature of the time series.

The first 100 observations were used as the training set to develop the model. The observations 101-110 (test set) were used for evaluating the forecast. The last five observations (forecasting horizon) were forecasted on assuming that the observations were not available. With reference to the simulated data, we will first illustrate the Box-Jenkins forecasting procedure. To obtain stationarity the series required six successive differencing (based on ADF tests) and at the end of the sixth differencing the best model suited for the six times differenced series was obtained as an ARMA (1,1) with mean

0. The estimates of the parameters were ar1=0.5111, ma1=-0.802and $\sigma^2 = 337.9$. The goodness of fit indicators were Log-likelihood= -411.52, AIC=829.03, AICc=829.29 and BIC=836.69.

It may be noted that, although the simulated series had only two roots larger than or equal to unity, the Box-Jenkins Methodology required six successive differencing to arrive at a stationary process. The increased number of differencings required for the stationarity in the presence of explosive root has also been noted, as a feature when Box Jenkins method is used for explosive time series, by Suresh Chandra et al. (1994).

Table 1: Forecasts by BJ and the proposed methods and their errors

Test set	O	BJ	New	BJ error	New Method error
101	30452460	30452385.20	30452450.51	74.80	9.49
102	33497880	33497997.99	33497885.07	-117.99	-5.07
103	36847840	36847699.20	36847845.84	140.80	-5.84
104	40532790	40532864.70	40532790.08	-74.70	-0.08
105	44586240	44586162.40	44586240.30	77.60	-0.3
106	49045040	49045061.23	49045037.22	-21.23	2.78
107	53949710	53949660.63	53949716.95	49.37	-6.95
108	59344850	59344770.32	59344835.53	79.68	14.47
109	65279510	65279550.16	65279519.80	-40.16	-9.8
110	71807630	71807560.08	71807632.04	69.92	-2.04

Note: O : the simulated data, BJ : BJ forecasts and New : forecasts by the new method

We now proceed to illustrate the proposed forecasting procedure for the same data. The ADF test for the data (as stated above) revealed non stationarity. To begin, the value of $\hat{\phi}_1$ turned out to be 1.100013244, which is very close to the value of $\phi_1 = 1.1$ that we had set for simulation. One should recall that the closeness of the estimate is due to the exponential rate of convergence in probability. Next step is to construct the second order auxiliary series $X_2^*(t); t = 2, 3, \ldots, n$ using $\hat{\phi}_1$. We then proceed to compute $\hat{\phi}_2$, the estimate of the second largest root using (13). This estimate turned out to be 0.99682, which is very close to the second value of we had set for simulation. Since this value was less than unity, the BJ method was adopted to forecast the series $X_2^*(t+1)$. While doing so, it was observed that the series required two successive differencing to obtain stationarity. The differenced series was modeled as an ARMA(1,1) model with a drift. The estimates of the parameters were intercept = -0.4678, ar1=-0.9801, ma1=-0.6956 and $\sigma^2 = 6.007$. The goodness of fit indicators were Log-likelihood=-225.4, AIC=458.8, AICc=459.23 and BIC=469.1. The forecasts for the original series are obtained by inverting procedure described earlier.

Table 1 summarizes the forecasts of both procedures for the test period, and their errors.

Flowing out from Table 1 is Table 2 which depicts the superiority of the new method over the BJ method in terms of ME, MAE and MSE of forecasts.

Table 2: Evaluation of forecasts relating to simulated data

Criterion	BJ method	New method
ME	23.8084	-4.3350
MAE	74.6254	9.6814
MSE	6668.2809	153.5585

The forecast for the forecasting horizon 111-115 are presented in Table 3.

Table 3: Evaluation of forecasts relating to simulated data

horizon	Actual value	BJ forecast	New forecast
111	78988560	78988480.01	78988642.59
112	86887590	86887100.37	86887787.66
113	95576510	95574741.56	95576885.81
114	105134300	105129303.02	105134940.7
115	115647900	115635777.74	115648857.49

It can be seen that the proposed forecasting procedure performs better than Box-Jenkins forecasting procedure even in the forecasting horizon. As expected, the forecasting errors in the period of horizon are numerically larger than in the test period and increases as the horizon extends, with the new procedure having relatively smaller numerical increase cautioning the reliability of forecasts in large horizons.

8 Illustration based on National crime data

The real life data used for demonstrating the utility of the proposed forecasting procedure relates to the annual number of cases registered under Indian Penal Code,in India from 1955 to 2012. The source of the data is from the published records of National Crime Records Bureau, New Delhi. The data is appended as Appendix-2. The data clearly indicates the explosive nature of the time series.

The period of observations is 59 years of which the first 53 observations (1955-2007) were used as the training set, next 5 observations (2008-2012) as the test set and the year 2013 was identified for forecasting horizon.

First, the the Box-Jenkins forecasting procedure was applied to the data. The ADF test for the given data supported the existence of unit root. Hence the data was differenced once and the ADF test was applied to the differenced series. The ADF test rejected the unit root hypothesis ($p = 0.01$) and therefore we conclude that the first differenced series was stationary. The model for the once differenced series was ARMA(0,0) with a non zero mean of 27969.942 and $\sigma^2 = 2.52.(10^9)$. The goodness of fit measures turned out to be Log-likelihood= -734.56, AIC=1471.12, AICc=1471.19 and BIC=1473.21. Using this ARMA (0,0) model with drift, the Box-Jenkins forecast for the original series were computed.

In parallel, the new proposed forecasting method was used for forecasting the future. The estimate $\hat{\phi}_1$ of the largest root ρ_1, which turned out to be 1.0198. Using this estimate, the second auxiliary series $X_2^*(t)$; $t = 2, 3, ..., n$ was constructed. The estimate of $\hat{\phi}_2$ was -0.079 which is numerically less than 1, leading to the conclusion that the second auxiliary series was stationary. The model fitted for the auxiliary series was ARMA (0,0) with mean zero and $\sigma^2 = 2.66.(10^9)$. The goodness of fit measures turned out to be Log-likelihood= -638.05, AIC=1278.09, AICc=1278.17 and BIC=1280.04. The following table summarizes the details of forecasting and errors involved in the two methods.

Table 4: Forecasts by BJ and the proposed methods and their errors for the crime data

Test set	O	BJ	New	BJ error	New Method error
2008	2093379	2017642.94	2024612.46	75736.06	68766.54
2009	2121345	2121348.94	2130139.58	-3.94	-8794.58
2010	2224831	2149314.94	2158596.68	75516.06	66234.32
2011	2325575	2252800.94	2263899.93	72774.06	61675.07
2012	2387188	2353544.94	2366413.04	33643.06	20774.96

Note: O : the simulated data, BJ : BJ forecasts and New : forecasts by the new method

Flowing out from Table 4 is the superiority of the new method in terms of ME, MAE and MSE of forecasts as summarized below:

The forecast for the year 2013 (one step horizon) in the Box-Jenkins methodology was 2415157.942 and that in the proposed procedure was 2429107.99, the latter being closer to 2647722, the actual value for the year 2013.

This application also exposes the superiority of the proposed procedure over the Box-Jenkins methodology even when the explosive root is close to unity.

Table 5: Evaluation of forecasts relating to crime data

Criterion	BJ method	New method
ME	257665.3	208656.31
MAE	257673.2	226245.47
MSE	17866544382	13428580029

9 Discussion

The forecasting procedure proposed in this paper for explosive models appears to be simple and novel. It is an improvement of the earlier works in Suresh Chandra et al. (1994); Suresh Chandra and Janhavi (2008) in terms of both theoretical and empirical support it has. This methodology works mainly due to the possibility of estimating the explosive roots with exponential rate of convergence and estimation of the auxiliary processes.

The assumption of real root can always be relaxed (Venkataraman, 1973) since the imaginary roots occur in pairs. Corresponding modification may not be conceptually difficult both in terms of estimating the roots and defining the auxiliary processes and estimating them.

The proposed method of forecasting can also be viewed as an extension of the Box Jenkins forecasting methodology. In fact, we settle down to BJ methodology once the series is free from explosive roots. This fact has also been clearly shown in the appended flow chart.

It is well known that both conventional and not so conventional procedures are adopted for forecasting the future values of the time series with the sole aim to get forecasts as close as possible to the realized values of the time series. In this context, it is perhaps necessary to evaluate a new forecasting methodology with all known and used forecasting procedures. But, we have chosen to compare the new methodology with Box-Jenkins methodology. This comparison was not only because of its popularity, but also because of the philosophical similarity among the two procedures.

Finally it is interesting to note that, in both applications, the new method provided a better fit for the stationary part of the model than the BJ method. Since the stationary part plays a crucial role in the forecasting methodology for non stationary time series this observation is relevant.

Acknowledgement

The authors wish to thank the referees for their suggestions to improve the quality of the paper.

References

Bhat, B. R. (1999). *Modern Probability Theory.* 3/e, New Age International (P) Ltd., Publishers, New Delhi.

Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis Forecasting and Control.* Holden Day, San Francisco.

Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74 , pages 427–431.

Karlin, S. and Taylor, H. M. (1975). *A First Course in Stochastic Processes.* 2/e, Academic Press, Inc. New York.

Kendall, M. and Ord, J. (1990). *Time Series.* Edward Arnold.

Lehmann, E. L. (1986). *Testing Statistical Hypothesis.* (2/e), John Wiley and Sons, New York.

Makridakis, S., Wheelwright, S. C., and Hyndman, R. J. (1998). *Forecasting Methods and Applications.* (3/e), John Wiley and Sons. Inc., New York.

Prabakaran, S. (2015). *Statistical analysis of heinous crimes in tamilnadu and related statistical inferences.* Ph.D thesis submitted to the University of Madras, Chennai. India.

Suresh Chandra, K., Bhat, B, R., and Srinivasan, T, M. (1994). A modified Box-Jenkins methodology for explosive time series. In *Proceedings of the Third Ramanujan Symposium on Stochastic Processes and their applications,Ramanujan Institute, University of Madras, Chennai,India*, pages 89–96.

Suresh Chandra, K. and Janhavi, J. (2008). Unit root tests for time series in the presence of an explosive rootunit root tests for time series in the presence of an explosive root. *Working Paper 27/2008, Madras School of Economics, Chennai ,India.*

Suresh Chandra, K., Manjunath, S. M., and Vaman, H. J. (1999). Least squares estimation for linear explosive models and its ramifications. *Statistical Methods CUSAT Kochi*, 1 , pages 54–64.

Venkataraman, K. N. (1968). Limit theorems on a linear stochastic difference equation with a constant term and their statistical applications. *Sankhya*, 10 , pages 51–74.

Venkataraman, K. N. (1973). Some convergence theorems on a second order linear explosive stochastic difference equation with a constant term. *Journal of the Indian Statistical Association*, 11, pages 47–69.

Venkataraman, K. N. (1974). Convergence theorems on the least squares estimators of a linear explosive model. *Annals of the Institute of Statistical Mathematics*, 26 (1), pages 61–72.

Flow chart of steps in the proposed forecasting procedure

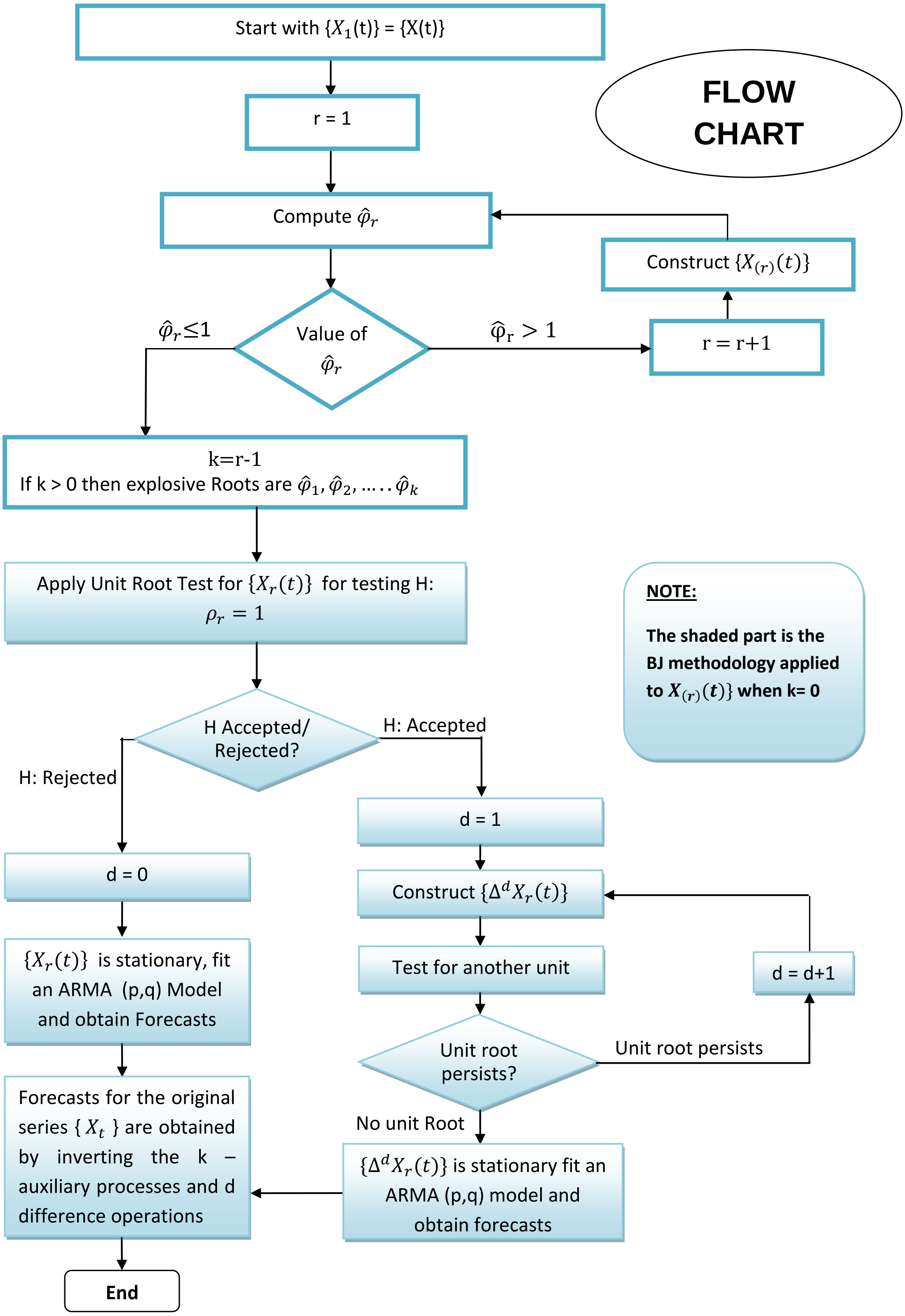

Figure 1: flowchart

APPENDIX-1

Table 6: Simulated time series data

t	x	t	x	t	x	t	x
1	260.42	31	36680.42	61	671086.5	91	11739670
2	461.6	32	40542.66	62	738375.8	92	12913810
3	695.16	33	44790.1	63	812395.5	93	14205370
4	959.7	34	49461.07	64	893817.7	94	15626080
5	1252.18	35	54597.38	65	983383	95	17188860
6	1574.33	36	60245.17	66	1081905	96	18907920
7	1929.35	37	66456.63	67	1190279	97	20798890
8	2319.63	38	73289.55	68	1309490	98	22878950
9	2749.29	39	80806.02	69	1440620	99	25167020
10	3221.41	40	89072.69	70	1584863	100	27683890
11	3741.95	41	98165.06	71	1743528	101	30452460
12	4315.51	42	108166.7	72	1918058	102	33497880
13	4946.22	43	119169.1	73	2110041	103	36847840
14	5640.37	44	131271.4	74	2321224	104	40532790
15	6404.91	45	144585.3	75	2553524	105	44586240
16	7247.39	46	159230.7	76	2809053	106	49045040
17	8174.88	47	175342	77	3090134	107	53949710
18	9195.73	48	193065	78	3399322	108	59344850
19	10318.15	49	212561.4	79	3739427	109	65279510
20	11552.04	50	234007.3	80	4113542	110	71807630
21	12909.37	51	257597.3	81	4525070	111	78988560
22	14402.75	52	283546.4	82	4977753	112	86887590
23	16046.3	53	312090.3	83	5475705	113	95576510
24	17853.75	54	343486.8	84	6023452	114	105134300
25	19842.17	55	378020.7	85	6625974	115	115647900
26	22028.31	56	416006	86	7288747	-	-
27	24432.18	57	457788.6	87	8017798	-	-
28	27073.84	58	503748.3	88	8819754	-	-
29	29977.91	59	554303.3	89	9701905	-	-
30	33169.93	60	609914.4	90	10672270	-	-

APPENDIX-2

Table 7: Number cases registered under Indian Penal Code

Year	Number of cases	Year	Number of cases
1955	535236	1985	1384731
1956	585217	1986	1405835
1957	581371	1987	1406992
1958	614184	1988	1440356
1959	620326	1989	1529844
1960	606367	1990	1604449
1961	625651	1991	1678375
1962	674466	1992	1689341
1963	658830	1993	1629936
1964	759013	1994	1635251
1965	751615	1995	1695696
1966	794733	1996	1709576
1967	881981	1997	1719820
1968	861962	1998	1778815
1969	845167	1999	1764629
1970	955422	2000	1771084
1971	952581	2001	1769308
1972	984773	2002	1780330
1973	1077181	2003	1716120
1974	1192277	2004	1832015
1975	1160520	2005	1822602
1976	1093897	2006	1878293
1977	1267004	2007	1989673
1978	1344968	2008	2093379
1979	1336168	2009	2121345
1980	1368529	2010	2224831
1981	1385757	2011	2325575
1982	1353904	2012	2387188
1983	1349866	2013	2647722