

# A new adaptive exponential smoothing method for non-stationary time series with level shifts

Mohammad Ali Saniee Monfared · Razieh Ghandali ·  
Maryam Esmaeili

Received: 14 August 2013 / Accepted: 30 June 2014 / Published online: 30 July 2014  
© The Author(s) 2014. This article is published with open access at Springerlink.com

**Abstract** Simple exponential smoothing (SES) methods are the most commonly used methods in forecasting and time series analysis. However, they are generally insensitive to non-stationary structural events such as level shifts, ramp shifts, and spikes or impulses. Similar to that of outliers in stationary time series, these non-stationary events will lead to increased level of errors in the forecasting process. This paper generalizes the SES method into a new adaptive method called revised simple exponential smoothing (RSES), as an alternative method to recognize non-stationary level shifts in the time series. We show that the new method improves the accuracy of the forecasting process. This is done by controlling the number of observations and the smoothing parameter in an adaptive approach, and in accordance with the laws of statistical control limits and the Bayes rule of conditioning. We use a numerical example to show how the new RSES method outperforms its traditional counterpart, SES.

**Keywords** Time series analysis · Adaptive exponential smoothing · Level shifts · Statistical control limits

## Introduction

Daily sales of dairy products, annual harvest of a farm, daily oil prices, and monthly international air passengers are some examples of time series which show stochastic processes or sequences of random variables. Time series is an important tool in Operations Research which is used to forecast and analyze the future events in planning, control and management of projects, personnel, finance, production, operations and services (Sekar 2010).

To model stationary and non-stationary time series, different methods such as ARIMA, innovations state space, regression, standard exponential smoothing, and Holt's trend methods have been developed (Box et al. 2008; Fried 2007; Hyndman et al. 2002; Chen and Tiao 1990; Montgomery and Johnson 1997). Still, simple forecasting methods such as Simple Exponential Smoothing method (SES), originated in the seminal work of Brown (1959), is the most commonly used in practice for its simplicity, computational efficiency, reasonable accuracy, ease of adjusting the forecast responsiveness, and the fact that is competitive with respect to other more complicated forecasting methods (Montgomery and Johnson 1997; Makridakis et al. 1998; Croux et al. 2010; Xu and Perron 2013).

Despite the competitive advantages of the SES method against its more complex counterparts, we note that by using the SES method the outliers and structural breaks cannot be detected effectively. In fact, SES is generally less sensitive to rare events such as level shifts, ramp shifts, spikes, impulses, and the change of underlying probability function (Hu and Sun 2001; Hu et al. 2011; Koehler et al. 2012). These events are expected to be caused by institutional, legislative, technical, natural, or economic changes, e.g., the oil embargo crisis in 1973. The results, however, increase the error level (Gardner 2006). To cope with these

---

M. A. S. Monfared (✉) · M. Esmaeili  
School of Engineering, Alzahra University, 1993891176 Tehran,  
Iran  
e-mail: mas\_monfared@yahoo.com

M. Esmaeili  
e-mail: esmaeili\_m@alzahra.ac.ir

R. Ghandali  
Industrial Engineering Department, Yazd University of  
Technology, Yazd, Iran  
e-mail: rzyghandali@gmail.com

events more effectively, the researchers have developed different adaptive exponential smoothing methods in which the smoothing parameter is changed automatically in a controlled manner as the characteristics of the time series change (Trigg and Leach 1967; Whybark 1973; Gardner 2006). Those adaptive methods modify the constant parameter methods through constructing a convex combination of the current observations and the previous observations with weights depending on the last time point of a level shift. The value of the smoothing parameter in the existing adaptive methods depends on the magnitude of the most recent forecasting error (Tsay 1988; Yashchin 1997; Taylor 2004; Schelter et al. 2006; Fried and Gather 2007; Fried 2007; Koehler et al. 2012). For instance, in the Taylor's STES (2004)<sup>1</sup> method, the smoothing parameter is modeled as a continuous function of a transition variable, where the choice of the transition variable plays a vital role in the success of the method.

A new approach was developed by Monfared and Steiner (2000) to deal with level shifts in time series. The idea behind the new approach was to turn the infinite-horizon SES method into a finite-horizon one. The new method was implemented on an automated scheduling and control manufacturing system when the random order stream faces level shifts (Monfared and Yang 2005). It was shown that the new finite-horizon SES method could outperform the traditional one in dealing with non-stationary level shifts. In fact, the finite version of SES generalizes the traditional infinite one in a sense that it can easily be extended to the infinite one in the absence of level shifts. Monfared and Yang (2004) further developed the finite version of SES method by performing a sensitivity analysis on the smoothing parameters. It was shown that the number of observations which enters the forecasting process influences the accuracy of predictions. However, no mechanism was proposed to calculate the size of the observations which should be taken into the forecasting process in an adaptive manner.

Considering the fact that the weighted average of previous observations is the base of forecast in any SES method, it would be better to take more post-shift observations when a level shift occurs. In this paper, we introduce a new concept of shift arising probability into the context of our finite-horizon SES method to produce the Revised Simple Exponential Smoothing (RSES). In RSES, we estimate the exact number of observations using the shift arising probability, control charts, and Bayes rule of conditioning. We assign a probability to a level shift occurrence with the confidence level equal to  $1-\alpha$  in which  $\alpha$  is the Type I error, i.e., the error that is seeing an out-of-control value is confused with an actual level shift. The RSES method is different from the work of Taylor (2004)

in a sense that we do not incorporate transitional variables and we do not need any judgmental information.

This paper is organized as follows. In Sect. 2, the RSES method is presented and its basic features are compared with its standard counterpart, SES. In Sect. 3, discussions are presented to determine the parameters of the RSES. In Sect. 4, an algorithm for RSES is presented which follows by an illustrative numerical example. Section 5 concludes the paper.

## The RSES

In this Section, the new method called RSES is defined and compared with the traditional version of SES. SES is defined as:

$$Y(t)^{\text{SES}} = \sum_{j=1}^{\infty} \alpha(1-\alpha)^{j-1} X(t-j) = \alpha X(t-1) + \alpha(1-\alpha)X(t-2) + \alpha(1-\alpha)^2 X(t-3) + \dots$$

$$0 < \alpha \leq 1 \quad (1)$$

In which,  $Y(t)^{\text{SES}}$  is the forecast value for the time period  $t$ ,  $X(t-j)$  is the past observation for the time period  $t-j$ , and  $\alpha$  is the smoothing parameter or the discounting parameter. It is important to note that, in SES method, the number of past observations which is taken into account is infinite.

Now, let us define the revised simple exponential smoothing (RSES) method as:

$$Y(t)^{\text{RSES}} = \frac{1}{\sum_{j=1}^{\tau} \rho^{j-1}} \sum_{j=1}^{\tau} \rho^{j-1} X(t-j) = \frac{1}{\sum_{j=1}^{\tau} \rho^{j-1}} \{X(t-1) + \rho X(t-2) + \rho^2 X(t-3) + \dots + \rho^{\tau-1} X(t-\tau+1)\}$$

$$0 \leq \rho \leq 1 \quad (2)$$

where,  $Y(t)^{\text{RSES}}$  is the forecast value for the time period  $t$ ,  $X(t-j)$  is the past observation for the time period  $t-j$ ,  $\tau$  is the maximum number of past observations entering the forecasting process which is now a variable, and  $\rho$  is the smoothing parameter or the discounting parameter. Note that,  $\rho$  in the RSES is the same as  $1-\alpha$  in the SES. This is while in the SES  $j$  changes from 1 to infinity, i.e.,  $\tau \rightarrow \infty$ . In both methods, however, coefficients of values change exponentially. This implies that by increasing the power of  $1-\alpha$  in the SES or  $\rho$  in the RSES, the influence of previous values is decreased in both methods. To compare the RSES against the SES, some mathematical properties should be considered as follows.

**Proposition 1**  $Y(t)^{\text{RSES}} = \hat{a}$ .

*Proof* To obtain the forecast value, sum of squared error (SSE) between the forecast value ( $\hat{a}$ ) and the actual value should be minimized such that:

<sup>1</sup> Smooth transition exponential smoothing.

$$\begin{aligned} \min : \text{SSE} &= \sum_{j=1}^{\tau} \rho^{j-1} [X(t-j) - \hat{a}]^2 \\ \frac{d\text{SSE}}{d\hat{a}} &= -2 \sum_{j=1}^{\tau} \rho^{j-1} [X(t-j) - \hat{a}] = 0 \\ &\Rightarrow \sum_{j=1}^{\tau} \rho^{j-1} [X(t-j) - \hat{a}] = 0 \\ &\Rightarrow \hat{a} = \frac{1}{\sum_{j=1}^{\tau} \rho^{j-1}} \sum_{j=1}^{\tau} \rho^{j-1} X(t-j) \end{aligned}$$

Therefore:  $Y(t)^{\text{RSES}} = \hat{a}$ .

**Proposition 2** SES and RSES methods are equal if the number of observations goes to infinity.

*Proof* It suffices to show that Eqs. 1 and 2 are equal if  $\tau \rightarrow \infty$ . Considering the geometric sequence we will have:

$$\begin{aligned} Y(t)^{\text{RSES}} &= \frac{1}{\sum_{j=1}^{\infty} \rho^{j-1}} \sum_{j=1}^{\infty} \rho^{j-1} X(t-j) = (1 - \rho) \\ &\times \sum_{j=1}^{\infty} \rho^{j-1} X(t-j) = (1 - \rho)X(t-1) \\ &+ (1 - \rho)\rho X(t-2) + (1 - \rho)\rho^2 X(t-3) \\ &+ \dots + (1 - \rho)\rho^j X(t-j-1) \end{aligned}$$

By considering  $\rho = 1 - \alpha$  then the last equation is the same as Eq. 1. Now, we show that by limiting the number of past observations from  $\infty$  in the SES to  $\tau$  in the RSES, we enable the RSES to detect sudden level shifts in non-stationary time series.

### Treating level shifts using RSES

In the RSES method, the number of previous observations ( $\tau$  in Eq. 1) which enters the forecasting process influences the accuracy of our predictions. This can be associated with the fact that the weighted average of the previous values is the basis for our forecast. If a level shift is happened for certain, i.e., with probability one, then, only post-shift values could be used to catch the level shift. But when a level shift occurs with certain probability, we would prefer to take more of post-shift values and less of pre-shift values into our estimation. In fact, higher weights are assigned to the post-shift observations and lower weights are given to the pre-shift observations. As we will show, this procedure makes the predictions more accurate.

To detect a level shift, the control charts technique from the statistical quality control theory is employed. In such control charts, observation of an out-of-control limit event might be a sign of a level shift. At the same time, it might

be a sign of an outlier or a spike. To cope with this uncertainty, we assign a confidence level equal to  $1 - \alpha$  to our perceived level shift event in which  $\alpha$  is the Type I error, i.e., the error that seeing an out-of-control value such as a spike or an outlier confuses with a level shift. Also, we introduce the shift arising probability as a new measure of uncertainty and launch the shift change process according to this uncertainty measure as will be considered in the following.

### The shift arising probability

In the RSES, total weights (i.e., equal to 1) are assigned to a limited number of past observations ( $\tau$ ) to minimize the impact of pre-shift values or to maximize the impact of post-shift values. Also, the portion of the weights which will be assigned to the post-shift values is equal to the shift arising probability. If  $c$  denotes the shift arising probability, then  $100c\%$  of the total weight is assigned to the post-shift values and  $100(1-c)\%$  to the pre-shift values. The probability of a level shift is obtained using the result of Theorem below.

**Theorem** When an out-of-control observation has been detected, the probability of a level shift becomes:

$$p(B|A') = \frac{(1 - \beta)p(B)}{(1 - \beta)p(B) + \alpha(1 - p(B))} \tag{3}$$

In which,  $B$ : is the shift arising event from  $\mu_0$  to  $\mu_1$ .  $A'$ : is the event of seeing an observation out of the control limits of  $\mu_0$ ,  $\alpha$ : is the Type I error, i.e., a level shift has not happened but it is conceived as it happened (e.g., a spike or an outlier).  $\beta$ : is the Type II error, i.e., an actual level shift has happened but has not been detected, mistakenly.

*Proof* First, consider  $LCL_{\mu_0}$  and  $UCL_{\mu_0}$  as the control limits of  $\mu_0$  and  $A$  as the event of seeing an observation inside the control limits of  $\mu_0$  (complementing  $A'$ ), then  $\beta$  can be probabilistically defined as:

$$\beta = p(A|B) = p(\text{seeing\_an\_observation\_inside\_}UCL_{\mu_0} \text{ and } LCL_{\mu_0} | \text{level\_shift\_happened})$$

Hence,

$$\begin{aligned} 1 - \beta &= p(A'|B) \\ &= p(\text{seeing\_an\_observation\_outside\_}UCL_{\mu_0} \text{ and } LCL_{\mu_0} | \text{level\_shift\_happened}) \end{aligned}$$

Note that  $1 - \beta$  is not the shift arising probability,  $p(B|A')$ . Therefore, care must be taken to ensure that  $p(B|A')$  is calculated rather than  $p(A'|B)$ . To find out  $p(B|A')$ , the Bayes law of conditioning is used as follows:

$$p(x|y) = \frac{p(x \cap y)}{p(y)} = \frac{p(y|x)p(x)}{p(y)} = \frac{p(y|x)p(x)}{p(y|x)p(x) + p(y|x')p(x')}$$

Therefore:

$$p(B|A') = \frac{p(A'|B)p(B)}{p(A'|B)p(B) + p(A'|B')p(B')}$$

where,  $p(B)$  is the probability of having a shift change in the mean value of the process,  $p(B') = 1 - p(B)$  in which  $B'$  complements  $B$ , i.e., not facing a shift level event from  $\mu_0$  to  $\mu_1$ ,  $p(A'|B) = 1 - \beta$  as seen before. Moreover,  $p(A'|B')$  is evidently equal to  $\alpha$ , i.e.,  $p(A'|B') = \alpha$ , i.e., Type I error or the probability of being out of the control limits of  $\mu_0$  when a level shift has not been happened (e.g., a spike or an outlier took place). The proof is completed now such that:

$$p(B|A') = \frac{p(A'|B)p(B)}{p(A'|B)p(B) + p(A'|B')p(B')} = \frac{(1 - \beta)p(B)}{(1 - \beta)p(B) + \alpha(1 - p(B))} \tag{4}$$

Please note that the probability  $\beta$  can be estimated using OC curves as will be considered in the next sub section.

### Estimating $\beta$ using two-sided OC curves

The following three steps should be followed to estimate  $\beta$  using an OC curve:

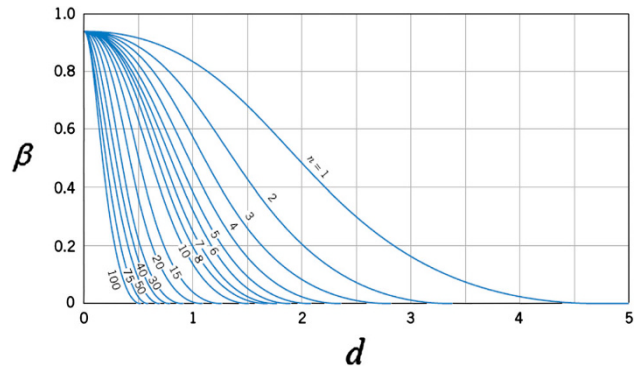
*Step 1* Set  $\alpha$  (i.e., the Type I error), and  $n$  (i.e., number of out-of-control limits observed),

*Step 2* Set  $d = |\mu_1 - \mu_0|/\sigma$  (i.e., normalized level shift difference, in which,

- $\mu_0$  is the previous level mean,
- $\mu_1$  is the new level mean, which is revealed gradually as new data are generated and are averaged. For example, for the event of the first out-of-control limit data point, i.e., for  $n = 1$ ,  $\mu_1 = x(t)$  and for  $n = 2$ , we have  $\mu_1 = 1/2[x(t) + x(t + 1)]$  and henceforth

$\sigma$  is the standard deviation of the time series (past data). The standard deviation is assumed to be constant which is estimated from the smooth regions of time series. However, please note that the level shift or the change of mean value does not change the standard deviation as  $\sigma = (\sum (x - \bar{x})/n)^{0.5}$  and  $(x - \bar{x})$  remains constant both before the shift and after the shift. Notice, however, that the standard deviation is changing in the transient period where the stochastic process is transiting from  $\mu_0$  to  $\mu_1$

*Step 3* Set the value for  $p(B)$ . This can be estimated using the past experiences of the stochastic process under



**Fig. 1** OC curve to find  $\beta$  for given  $d, n$  and  $\alpha = 0.05$  (Montgomery and Runger 2011)

scrutiny. For instance, it is estimated by  $p(B) = n(A)/n(S)$  where  $n(A)$  denotes the number of weeks witnessing a level shift, and  $n(S)$  denotes total number of weeks that the process has been observed. In case of no past data,  $p(B)$  can be set equal to 0.5 that means we are indifferent toward having or not having a level shift. We will come out of this indifference when data goes outside the control limits and  $n$  counted as 1,2,...etc.

*Step 4* Use a two-sided normal operation characteristic curve (OC curve) as in Fig. 1 to estimate  $\beta$  which is the Type II error.

For example, from Fig. 1 when  $n = 1, d = 1$  then the error  $\beta$  is equal to 0.85. At the same  $d$  value, if  $n$  increases from 1 to 2 the error  $\beta$  will become 0.70 which shows an improvement. Also for  $n = 10$ , the error  $\beta$  is equal to 0.15. Note that in the OC curve shown in Fig. 1, as the difference between the new mean value and the previous mean value ( $\mu_1 - \mu_0$ ) decreases, the error  $\beta$  decreases too. For example, for  $d = 1$  and  $n = 100$ , the error  $\beta$  approaches zero (see Montgomery and Runger 2011).

### Determination of weighting coefficients

A level shift is perceived better as more observations are seen out-of-control limits, i.e., as  $n$  is increasing. However, even a single out-of-control limit observation can signal a true level shift with a certainty factor of  $p(B|A')$ . In other words, with probability  $1 - p(B|A')$ , such a single observation could be associated with a spike or an outlier and not a true level shift. Thus, our certainty factor improves as more out-of-control limit data are seen (i.e. as  $n$  increases), causing the probability  $p(B|A')$  to increase.

That is why a new concept of shift arising probability is introduced into our RSES method. Since  $0 < p(B|A') < 1$ , there are some observations from pre-shift event and some from post-shift event that should be

taken into the forecasting process. Those from post-shift event help to estimate  $\mu_1$  and those from pre-shift event hinders this process. The hindering side can only be eliminated if there is a certainty about a level shift, i.e., when  $p(B|A') = 1$ , which is in fact impossible realizing the nature of stochastic process inherited in a time series. However, as an extreme case assume that  $p(B|A') = 1$  then  $\tau$  would have been defined only by the post-shift observations. For example, for a single observation which is out-of-control limit with certainty, i.e., when  $p(B|A') = 1$  then  $\tau = 1$  and whole weight of  $\rho = 1$  is assigned to this new observation, i.e.,  $Y(t)^{RSES} = X(t)$ . On the other extreme, for  $p(B|A') = 0$  that means no out-of-control limit observations is seen for long run so that  $\tau = \infty$  and the RSES method in this case would act as SES method. In general, the shift arising probability is updated on the first and subsequent occurrences of out-of-control data; hence helps to reduce the forecast errors. Now to find the weight coefficients of  $\rho$ , the following definitions are considered.

**Definition 1** Equation 2 can also be represented in the following form:

$$Y(t)^{RSES} = \frac{1}{\sum_{j=1}^{\tau} \rho^{j-1}} \sum_{j=1}^{\tau} \rho^{j-1} X(t-j) = \frac{1}{\sum_{j=1}^{\tau} \rho^{j-1}} \{X(t-1) + \rho X(t-2) + \rho^2 X(t-3) + \dots + \rho^{\tau-1} X(t-\tau+1)\} = \rho_1 X(t-1) + \rho_2 X(t-2) + \rho_3 X(t-3) + \dots + \rho_{\tau} X(t-\tau+1)$$

in which,

$$\rho_1 = \frac{1}{\sum_{j=1}^{\tau} \rho^{j-1}}, \rho_2 = \rho \frac{1}{\sum_{j=1}^{\tau} \rho^{j-1}} = \rho \cdot \rho_1, \rho_3 = \rho^2 \cdot \rho_1 \text{ and hence } \rho_{\tau} = \rho^{\tau-1} \cdot \rho_1. \tag{5}$$

Thus,  $\rho_1, \rho_2, \rho_3,$  and  $\rho_{\tau}$  represent the weight coefficients of the RSES method. The value of  $\rho_1$  sets equal to the shift arising probability  $p(B|A')$  as considered in the following.

**Definition 2** The weight coefficient for the most recent observation, i.e.,  $\rho_1$  sets equal to the value of shift arising probability, i.e.,  $p(B|A')$ . Or,

$$\left\{ \rho_1 = \frac{\rho^{(1-1)}}{\sum_{j=1}^{\tau} \rho^{(j-1)}} = \frac{1}{\sum_{j=1}^{\tau} \rho^{(j-1)}} \right\} = \left\{ p(B|A') = \frac{(1-\beta)p(B)}{(1-\beta)p(B) + \alpha(1-p(B))} \right\} \tag{6}$$

In which, both  $\rho$  and  $\tau$  are unknown variables and need to be estimated simultaneously. Please note that in (6) both  $\alpha$  and  $p(B)$  are fixed parameters, and  $\beta$  is

estimated by using a two-sided normal operating characteristic curve (OC curve) as in Fig. 1. Now, for example, suppose that  $\alpha = 0.05, p(B) = 0.4,$  and  $\beta = 0$  (or  $1-\beta = 1$ ), then the weight coefficient for the first new forecast becomes

$$\rho_1 = \frac{(1-\beta)p(B)}{(1-\beta)p(B) + \alpha(1-p(B))} = \frac{1 * 0.4}{0.4 + 0.05(1-0.4)} = 0.93$$

This means that 93 % of prediction is based on the most recent data and only 7 % of the past data are used. Therefore, if an out-of-control value is observed which might be associated with a level shift, 0.93 of importance is assigned to the most recent data. From the OC curve in Fig. 1, it is conceivable that  $\beta$  can hardly get a value farther than 0.95, then the weight coefficient for the first new value is:

$$\rho_1 = \frac{(1-\beta)p(B)}{(1-\beta)p(B) + \alpha(1-p(B))} = \frac{0.05 * 0.4}{0.05 * 0.4 + 0.05 * 0.6} = 0.4$$

Now, let's find out  $\rho$  and  $\tau$  values for  $\rho_1 = 0.4$  such that:

$$\rho_1 = 0.4 = \frac{1}{2.5} = \frac{\rho^{(0)} = 1}{\sum_{j=1}^{\tau} \rho^{(j-1)}} \Rightarrow \sum_{j=1}^{\tau} \rho^{(j-1)} = 2.5 \Rightarrow 1 + \rho + \rho^2 + \rho^3 + \dots = 2.5$$

By assuming  $\tau = \infty$  we will have,

$$\Rightarrow 1 + \rho + \rho^2 + \rho^3 + \dots = 2.5 = \frac{1}{1-\rho} \Rightarrow \rho = \frac{3}{5} = 0.6$$

Therefore,  $\rho_2 = \rho \rho_1 = 0.6 * 0.4 = 0.24, \rho_3 = \rho^2 \rho_1 = 0.6^2 * 0.4 = 0.144,$  etc. Now, suppose that a second out-of-control event has occurred (i.e.,  $n = 2$ ). In this case, the sum of the two recent coefficients is,

$$\left\{ \rho_1 + \rho_2 = \frac{\rho^{(1-1)} = 1}{\sum_{j=1}^{\tau} \rho^{(j-1)}} + \frac{\rho^{(2-1)}}{\sum_{j=1}^{\tau} \rho^{(j-1)}} = \frac{1 + \rho}{\sum_{j=1}^{\tau} \rho^{(j-1)}} \right\} = \left\{ p(B|A') = \frac{(1-\beta)p(B)}{(1-\beta)p(B) + \alpha(1-p(B))} \right\}$$

Again, for  $\alpha = 0.05, p(B) = 0.4,$  and  $\beta = 0.95,$  the shift arising probability will become:

$$p(B|A') = \frac{(1-0.95)0.4}{(1-0.95)0.4 + 0.05(1-0.4)} = 0.4 \text{ so that this certainty factor is assigned to the current two observations which happened to be out-of-control limits. Therefore, weight coefficients for these two observations are: } 0.4 = \rho_1 + \rho_2. \text{ Now, } \rho \text{ values are found as:}$$



$$\begin{aligned}
 0.4 &= \frac{1 + \rho}{\sum_{j=1}^{\tau} \rho^{(j-1)}} \Rightarrow 0.4 * \sum_{j=1}^{\tau} \rho^{(j-1)} = 1 + \rho \\
 &\Rightarrow 0.4 * (1 + \rho + \rho^2 + \rho^3 + \dots) = 1 + \rho \\
 &\Rightarrow 0.4 * \left(\frac{1}{1 - \rho}\right) = 1 + \rho \\
 &\Rightarrow 0.4 = 1 - \rho^2 \\
 &\Rightarrow \rho^2 = 0.6 \\
 &\Rightarrow \rho = 0.7746.
 \end{aligned}$$

Then,

$$\begin{aligned}
 \rho_1 + \rho_2 &= \rho_1 + \rho \cdot \rho_1 = \rho_1(1 + \rho) \\
 \Rightarrow \rho_1 &= (\rho_1 + \rho_2)/(1 + \rho) = 0.4/(1 + 0.7746) = 0.2254 \\
 \Rightarrow \rho_2 &= \rho \cdot \rho_1 = 0.7746(0.2254) = 0.17459 \\
 \Rightarrow \rho_3 &= \rho^2 \cdot \rho_1 = 0.7746^2(0.2254) = 0.13524 \\
 \Rightarrow \rho_4 &= \rho^3 \cdot \rho_1 = 0.7746^3(0.2254) = 0.10475 \\
 \Rightarrow \rho_5 &= \rho^4 \cdot \rho_1 = 0.7746^4(0.2254) = 0.08114 \\
 \Rightarrow \rho_6 &= \rho^5 \cdot \rho_1 = 0.7746^5(0.2254) = 0.06285 \\
 \Rightarrow \rho_7 &= \rho^6 \cdot \rho_1 = 0.7746^6(0.2254) = 0.04868
 \end{aligned}$$

Hence, based on the fact that  $p(BIA') = 0.4$ , 40 % of total weights are assigned to the two recent observations (i.e.  $\rho_1$  and  $\rho_2$ ) and 60 % of total weights are assigned to the other past observations of which  $\rho_3$  to  $\rho_7$  are shown above which contains about 43 % of the total weight.

It should be noted that when an observation happened to be out-of-control limit, the RSES method is triggered to replace the SES method. This transformation is in fact undertaken by limiting the infinite number of data used in SES (or  $\tau = \infty$ ) to some finite number of  $\tau < \infty$ . However, the estimation of  $\tau$  comes indirectly from the estimation of  $\rho_j$  as shown in the above example.

It is trivial to show that for  $\beta = 0$ ,  $p(B) = 1$ , or  $\alpha = 0$ , the contribution of the most recent data will be 100 %. In other words, the most recent observation is the most valuable data for the forecasting process of the next period. The current distribution of weight is based on the primary value of  $p(B)$ , i.e., initial shift change probability as estimated from the past performance of the non-stationary time series.

When a level shift takes place, any consideration of pre-shift data is harmful since it introduces bias into the estimated  $\mu_1$ . However, this is the case when  $p(B) < 1$  or  $\beta > 0$  both the after and pre-shifts data are used. By considering  $n = 0$ ,  $\tau$  would be infinite and the RSES method reduces to SES method. As  $n$  increases more out-of-control limits data are seen then, the total weights which is assigned to the recent data becomes larger approaching 100 %. Now, the level shift from  $\mu_0$  to  $\mu_1$  becomes certain.

In the next Section, the RSES algorithm is presented which follows with a numerical example to illustrate the working logic of the new adaptive method, the RSES.

### The algorithm of the RSES

In this Section, the algorithm for the RSES is presented as follows:

*Step 1* Set values for  $\alpha$  (the Type I error),  $p(B)$  i.e., the prior shift arising probability, and  $n = 0$ , i.e., initializing the variable for counting the out-of-control data.

*Step 2* Calculate  $\sigma$  and  $\mu_0$  for smooth regions of time series by canceling level shifts.

*Step 3* Determine the control limits to detect the level shift as:

$$\begin{aligned}
 LCL_{\mu_0} &\leq X(t) \leq UCL_{\mu_0} \text{ or} \\
 \mu_0 - z_{\alpha/2}\sigma &\leq X(t) \leq \mu_0 + z_{\alpha/2}\sigma
 \end{aligned}$$

*Step 4* Check for an out-of-control event, i.e.,  $LCL_{\mu_0} \geq X(t) \geq UCL_{\mu_0}$  then set  $n = n + 1$  and apply the RSES.

*Step 5* Set  $\mu_1 = \sum_n X(t)/n$ ,  $d = |\mu_1 - \mu_0|/\sigma$ , and estimate  $\beta$  from a two-sided OC curve.

*Step 6* Use  $\alpha$ ,  $p(B)$  and  $\beta$  to determine the posterior probability  $p(BIA')$  as in Eq. 4.

*Step 7* Calculate the weighting coefficients  $\rho_j$  for  $j = 1, 2, \dots, \tau$  using Definition 2 and forecast the new event using Eq. 5.

*Step 8* Switch back to the SES method and set  $\mu_0 = \mu_1$  if two consecutive values for  $\mu_1$  is less than  $\varepsilon$ , where  $\varepsilon$  denotes the error tolerance (e.g.,  $\varepsilon = 0.001$ ). Or if  $p(BIA')$  approaches unity (e.g.,  $p(BIA') \geq 0.9$ ), then the level shift certainly happened. Otherwise, check for the next observation and go to Step 4.

### Numerical example

To test the performance of the RSES method, we consider the sales of a hypothetical manufacturer as shown in Table 1. Data are generated from a normally distributed stochastic process  $X(t) \sim N(\mu_0 = 0, \sigma = 1)$ . A level shift is introduced at time period  $t = 0$  converting the  $X(t) \sim N(\mu_0 = 0, \sigma = 1)$  process into the  $X(t) \sim N(\mu_1 = 2, \sigma = 1)$  process. With  $\alpha = 0.05$ , SES method uses the following equation to forecast the results for time period  $t = 1$  as:

$$\begin{aligned}
 Y(1)^{SES} &= (0.05)X(0) + (0.05)(0.95)X(-1) + (0.05) \\
 &(0.95)^2X(-2) + (0.05)(0.95)^3X(-3) + \dots = 0.05(2.27) + \\
 &0.04(-0.21) + 0.04(-0.01 - 0.15 + 2.72 + 0.97) + 0.03
 \end{aligned}$$

**Table 1** Forecasting using RSES and SES

$t$	$X(t)$	$Y(t)^{RSES}$	$Y(t)^{SES}$
-8	0.69		
-7	1.51		
-6	-0.2		
-5	0.97		
-4	2.72		
-3	-0.15		
-2	-0.10		
-1	-0.21		
0 (level shift)	2.27		
1	2.06	1.99	0.56
2	2.28	1.91	0.88
3	4.26	2.03	1.17
4	4.12	2.94	1.78
5	2.78	3.06	2.25
6	2.01	2.87	2.36
7	4.36	2.65	2.29
8	2.2	2.89	2.7
Mean squared error	1.34	3.02	
MSE <sub>SES</sub> to MSE <sub>RSES</sub> ratio	2.25		

$(-0.2-0.2 + 1.51 + 0.69) = 0.56$ . Results for time periods  $t = 2$  to  $t = 8$  are reported in Table 1.

Now, if the RSES is adopted for time period  $t = 0$  we have  $x(0) = 2.27$ ,  $\sigma = 1$ ,  $n = 1$ , and  $d_1 = [x(t) - \mu_0]/\sigma = [2.27 - 0]/1 = 2.27$ ,  $P(B) = 0.4$  and  $\alpha = 0.05$ , then from Fig. 1 the value of  $\beta$  becomes 0.37 and the weight coefficient of the first observation will be:

$$\rho_1 = \frac{(1 - \beta)p(B)}{(1 - \beta)p(B) + \alpha(1 - p(B))} = \frac{0.63 * 0.4}{0.63 * 0.4 + 0.05 * 0.6} = 0.89$$

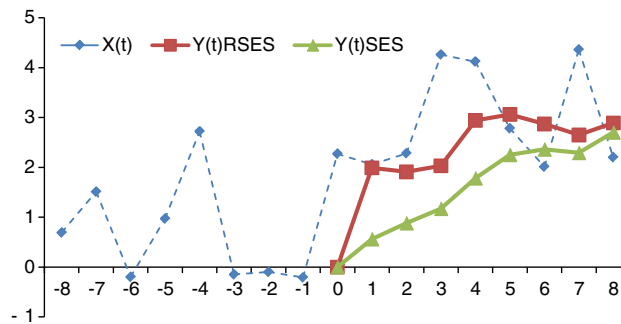
Now, to find the other weight coefficients  $\rho$  is estimated as:

$$\begin{aligned} \Rightarrow 0.89 &= \frac{\rho^{(0)} = 1}{\sum_{j=1}^{\tau_1} \rho^{(j-1)}} \Rightarrow \sum_{j=1}^{\tau_1} \rho^{(j-1)} = 1.123 \\ \Rightarrow 1 + \rho + \rho^2 + \rho^3 + \dots &= 1.123 \\ \Rightarrow \frac{1}{1 - \rho} &= 1.123 \Rightarrow \rho = \frac{0.123}{1.123} = 0.1095 \end{aligned}$$

Then, the weight coefficient for the second observation is  $\rho_2 = \rho\rho_1 = 0.1095(0.8905) = 0.0975$ . The third weight coefficient is  $\rho_3 = \rho^2\rho_1 = 0.1095^2(0.8905) = 0.01067$  and henceforth for the rest. Therefore, the forecast for  $t = 1$  will become:

$$\begin{aligned} Y(1)^{RSES} &= (0.89)X(0) + (0.097)X(-1) + (0.01)X(-2) \\ &+ (0.001)X(-3) = (0.89)(2.27) + (0.097)(-0.21) + (0.01) \\ &(-1) + (0.001)(-0.15) = 1.99. \end{aligned}$$

Using the similar method



**Fig. 2** Comparing results of SES and RSES for a level shift at  $t = 0$

as considered above, the results for time periods  $t = 2$  to  $t = 8$  are calculated and then shown in Table 1.

The results for both the RSES and the SES are also shown in Fig. 2 from which it is quite clear that the level shift at  $t = 0$  is successfully and quickly perceived by RSES method. With the SES, however, it takes a large number of time steps to detect the level shift change from  $\mu_0 = 0$  to  $\mu_1 = 2$ . The mean square error for both RSES and SES are estimated as:

$$\begin{aligned} MSE_{SES} &= \{(2.06-0.566)^2 + (2.28-0.88)^2 + (4.26- \\ &1.17)^2 + (4.12-1.78)^2 + (2.78-2.25)^2 + (2.01-2.36)^2 + \\ &(4.36-2.29)^2 + (2.2-2.7)^2\}/8 = 3.02, \end{aligned}$$

$$\begin{aligned} MSE_{RSES} &= \{(2.06-1.99)^2 + (2.28-1.91)^2 + (4.26- \\ &2.03)^2 + (4.12-2.94)^2 + (2.78-3.06)^2 + (2.01-2.87)^2 + \\ &(4.36-2.65)^2 + (2.2-2.89)^2\}/8 = 1.34. \end{aligned}$$

It is clear that error associated with the SES method is 2.25 times larger than the error associated with the RSES method.

**Conclusion**

In this paper, a new adaptive method called the RSES is developed to predict time series with non-stationary level shifts. The RSES method enjoys the traditional properties of simple exponential smoothing (SES) method except that only a limited number of observations are taken into considerations. In fact, the RSES is a finite version of the infinite-horizon SES method. The RSES method employs a shift arising probability to monitor the event of level shifts and to adopt the weight coefficients accordingly

To implement this new scheme, the timing of a level shift is detected using the control chart limits from the statistical quality control theory. A new concept of shift arising probability has been developed to measure the certainty factor associated with occurrence of a level shift and updated this measure using the Bayes rule of conditioning. The shift arising probability is used to determine the weight coefficients of the forecasting process. It is shown that the RSES method enables one to more

accurately predict time series with sudden level shifts. A numerical example is used to illustrate the working logic of the RSES method in comparison with its traditional counterpart (the SES method).

Still, there are some important topics of interests that need to be investigated. Included are whether the RSES is sensitive toward deviations from normality, could the RSES produce robust results, how the choice of prior probability  $P(B)$  will influence the results, what if level shift augments with variance changes, and how the RSES could deal with ramp shifts and long term cycles. Also, performing extensive experiments using real and simulated data is another path of further research.

**Acknowledgments** Authors are thankful to Mr. Mohamad Mahmoudi for his kind assistance in checking the English language of this manuscript.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

## References

- Box GEP, Jenkins GM, Reinsel GC (2008) Time series analysis: forecasting and control, 4th edn. Wiley, London
- Brown RG (1959) Statistical forecasting for inventory control. McGraw-Hill, NY
- Chen C, Tiao GC (1990) Random level-shift time series models, ARIMA approximations, and level-shift detection. *J Bus Econ Stat* 8(1):83–97
- Croux C, Gelper S, Mahieu K (2010) Robust exponential smoothing of multivariate time series. *Comp Stat Data Anal* 54(12):2999–3006
- Fried R (2007) On the robust detection of edges in time series filtering. *Comput Stat Data Anal* 52(2):1063–1074
- Fried R, Gather U (2007) On rank tests for shift detection in time series. *Comput Stat Data Anal* 52:221–233
- Gardner ES Jr (2006) Exponential smoothing: the state of the art, part II. *Int J Forecast* 22:637–666
- Hu S, Sun G (2001) Process monitoring technology and application. National Defence Industry Press, Beijing, pp 68–103
- Hu S, Xiaofeng W, Meinke K, Ouyang H (2011) Outlier to lerant fitting and online diagnosis of outliers in dynamic process sampling data series. *Artificial intelligence and computational intelligence*. Springer Press, Beerlin, pp 195–203
- Hyndman RJ, Koehler AB, Snyder RD, Grose S (2002) A state space framework for automatic forecasting using exponential smoothing methods. *Int J Forecast* 18:439–454
- Koehler AB, Snyder RD, Ord JK, Beaumont A (2012) A study of outliers in the exponential smoothing approach to forecasting. *Int J Forecast* 28(2):477–484
- Makridakis S, Wheelwright S, Hyndman R (1998) *Forecasting Methods and Applications*. Wiley, NY
- Monfared MAS, Steiner SJ (2000) Fuzzy adaptive scheduling and control systems. *Fuzzy Sets Syst* 115(2):231–246
- Monfared MAS, Yang JB (2004) Design of an intelligent manufacturing scheduling and control system using fuzzy logic. *J Intell Fuzzy Syst* 15(2):89–104
- Monfared MAS, Yang JB (2005) Multilevel intelligent scheduling and control system for an automated flow shop manufacturing environment. *Int J Prod Res* 43(1):147–168
- Montgomery DC, Johnson LA (1997) *Forecasting and time series analysis*. McGraw-Hill, New York
- Montgomery DC, Runger GC (2011) *Applied statistics and probability for engineers*, 5th edn. Wiley, London
- Schelter B, Winterhalder M, Timmer J (eds) (2006) *Handbook of time series analysis*. WILEY-VCH Verlag GmbH & Co KGaA, Weinheim
- Sekar P (2010) Application of time series models. *Indian J Sci Technol* 3(9):1032–1037
- Taylor JW (2004) Volatility forecasting with smooth transition exponential smoothing. *Int J Forecast* 20(2):273–286
- Trigg DW, Leach AG (1967) Exponential smoothing with an adaptive response rate. *Op Res Quart* 18:53–59
- Tsay RS (1988) Outliers, level shifts, and variance changes in time series. *J Forecast* 7:1–20
- Whybark DC (1973) Comparison of adaptive forecasting techniques. *Logist Transp Rev* 8:13–26
- Xu J, Perron P (2013) *Forecasting return volatility: level shifts with varying jump probability and mean reversion*, Boston University, March 1
- Yashchin E (1997) *Nonlin Anal Theory Meth Appl* 30(3997):1997

