



# Article Comparative Study on Exponentially Weighted Moving Average Approaches for the Self-Starting Forecasting

# Jaehong Yu<sup>1</sup>, Seoung Bum Kim<sup>2</sup>, Jinli Bai<sup>2</sup> and Sung Won Han<sup>2,\*</sup>

- <sup>1</sup> Department of Industrial and Management Engineering, Incheon National University, Incheon 22012, Korea; jhyu@inu.ac.kr
- <sup>2</sup> School of Industrial Management Engineering, Korea University, Seoul 04620, Korea; sbkim1@korea.ac.kr (S.B.K.); baijl0527@korea.ac.kr (J.B.)
- \* Correspondence: swhan@korea.ac.kr; Tel.: +82-2-3290-3384

Received: 2 September 2020; Accepted: 18 October 2020; Published: 20 October 2020



**Abstract:** Recently, a number of data analysists have suffered from an insufficiency of historical observations in many real situations. To address the insufficiency of historical observations, self-starting forecasting process can be used. A self-starting forecasting process continuously updates the base models as new observations are newly recorded, and it helps to cope with inaccurate prediction caused by the insufficiency of historical observations. This study compared the properties of several exponentially weighted moving average methods as base models for the self-starting forecasting techniques because of their superior performance as well as computational efficiency. In this study, we compared the performance of a self-starting forecasting process using different existing exponentially weighted moving average methods under various simulation scenarios and real case datasets. Through this study, we can provide the guideline for determining which exponentially weighted moving average methods under various simulation scenarios and real case datasets. Through this study, we can provide the self-starting forecasting process.

**Keywords:** comparative study; exponentially weighed moving average; non-stationary time series; self-starting forecasting

## 1. Introduction

Time series is a sequence of observations, measured typically at successive points in time spaced at uniform intervals. This exists in many situations, including the daily closing value of the stock market, manufacturing process, health status of patients, and economic indicators [1]. Using these time series data, forecasting future events has been of considerable interest in various fields, including control charts [2–5], health care surveillance [6–8], inventory controls [9], stock market prediction [10–12], pandemic occurrence prediction [13], and electricity demand forecasting [14]. In such industrial areas, the accurate forecasting of future time series values helps establish more effective decision or management policy. To achieve satisfactory forecasting performance, a number of time series forecasting methods have been proposed, including exponentially weighted moving average (EWMA; [15–18]), autoregressive integrated moving average (ARIMA; [19]), generalized autoregressive conditionally heteroskedastic (GARCH; [20]), and vector autoregression (VAR; [21,22]) methods. In recent years, neural network-based approaches [23,24] have been widely applied for the time series forecasting tasks [25–27]. Among these time series forecasting methods, in this study, we focus on the EWMA model.

The EWMA model assigns different weights for individual observations collected over time, and more recent observations are weighted more heavily than remote observations, and hence,

2 of 18

dynamics of time series can be effectively reflected into the forecasting models. Furthermore, such an EWMA model has a desirable property with a recursive form, which leads to computational efficiency. Finally, the EWMA models do not need any distributional assumption and prior knowledge on the time series [1,28]. Because of these advantages, a number of EWMA models have been proposed, including single exponential smoothing (single ES; [15,16]) model, double exponential smoothing (double ES; [17]) model, triple exponential smoothing (triple ES; [18]) model, and two-stage exponentially weighted moving average (two-stage EWMA; [29,30]) model. We presented a more detailed description on these existing EWMA models in Section 2.1.

To ensure the satisfactory forecasting performance of the EWMA models, historical time series observations should be sufficiently accumulated. However, lots of data analysists have suffered from the insufficiency of historical observations in many real situations [31]. For example, the demand prediction of newly launched products is not an easy task because there are no sales records of these products. In addition, the prognosis of a manufacturing system status is also difficult immediately after the preventive maintenance because meaningful process data are not sufficiently accumulated after these operations. Accordingly, the insufficiency of historical observations has been regarded as one of the challenging problems for time series data analysis [28].

To address the insufficiency of initial time series data, self-starting approaches [32–34] can be incorporated into the EWMA model. In the self-starting-based EWMA modeling process, an initial EMWA model is constructed only using small initial time series observations, and the EWMA model is continuously updated as a new observation is sequentially recorded. By doing so, forecasting the performance of EWMA models can be improved over time because historical records are continuously accumulated in the forecasting process. Besides, the self-starting forecasting process can effectively accommodate dynamically changing time-varying patterns in that the forecasting models are updated by adding the most recently recorded observation to the initial time series data. In Section 2.2, we also presented more details on the self-starting forecasting process.

The main purpose of this study was to examine the characteristics of EWMA models as base-forecasting models for the self-starting forecasting process. Although various properties on existing EMWA models have been thoroughly examined in a number of previous studies, their characteristics as a base model for self-starting forecasting have not been systematically analyzed. Hence, this study has a remarkable contribution in that this is the first attempt to comprehensively investigate the properties of the EWMA models as base models of the self-starting forecasting process.

The remainder of this paper is organized as follows. In Section 2, we describe both the characteristics of existing EWMA models and self-forecasting process. In Section 3, we present a simulation study to compare the methods under various scenarios with only small initial time series data. In Section 4, we present the comparative results from real case data, and Section 5 presents concluding remarks.

#### 2. Background

#### 2.1. Exponentially Weighted Moving Average Model

In 1956, a single ES model was first proposed [15], which was designed to forecast demands in inventory control systems. In the single EWMA model, it is assumed that time series data,  $Y_1, Y_2, ..., Y_t$ , have no trend or seasonal pattern, and the level factor at time t,  $\hat{Y}_t$ , is given by

$$\hat{Y}_t = \alpha Y_t + (1 - \alpha) \hat{Y}_{t-1},\tag{1}$$

where  $\hat{Y}_1 = Y_1$  and  $\alpha$  is a smoothing parameter between 0 and 1. Equation (1) can be rewritten as follows:

$$\hat{Y}_{t} = \sum_{i=0}^{t} \alpha (1-\alpha)^{i} Y_{t-i}$$
(2)

Thus, the single ES model can be considered a weighted average of the current and all past observations. Because the multiplier of the past observation is  $(1-\alpha)$ , which is smaller than one for any  $\alpha$  values, it gives more weight to the current observation than past observations. Finally, the forecast for *s* times ahead from the time *t* can be determined as follows:

$$Y_{t+s}^* = \hat{Y}_t. \tag{3}$$

As shown in Equation (3), single ES forecasts the future observations as a constant value. Hence, this method might not yield satisfactory prediction performance in many real situations because the time series data are dynamically changed over time. To accommodate a time-varying pattern with a linear trend, a double ES model was proposed [17]. The double ES model assumes that the time series,  $Y_1, Y_2, \ldots, Y_t$ , exhibit a linear trend with no seasonal pattern. In this model, the level factor and trend factor at time *t*, denoted as  $\hat{Y}_t$ , and  $\Delta \hat{Y}_t$ , respectively, can be calculated by following equations:

$$\hat{Y}_t = \alpha Y_t + (1 - \alpha) \left( \hat{Y}_{t-1} + \Delta \hat{Y}_{t-1} \right), \tag{4}$$

$$\Delta \hat{Y}_t = \beta \left( \hat{Y}_t - \hat{Y}_{t-1} \right) + (1 - \beta) \Delta \hat{Y}_{t-1}, \tag{5}$$

where  $\hat{Y}_1 = Y_1$ , and  $\Delta \hat{Y}_2 = Y_2 - Y_1$ , and  $0 < \alpha < 1$ ,  $0 < \beta < 1$ . The parameter  $\alpha$  is a smoothing parameter for the level factor, and the parameter  $\beta$  is for a trend factor. The forecast for *s* times further from time *t* is given by

$$Y_{t+s}^* = \hat{Y}_t + s \times \Delta \hat{Y}_t. \tag{6}$$

This double ES model has successfully worked for forecasting the time series that has a constant trend, however, it does not properly work in a time series having a time-varying pattern of varying trends or seasonal variations.

To consider the seasonal variations, Winters [18] proposed the triple ES model to forecast the time series with both a linear trend and seasonality. Unlike to all the aforementioned methods, the triple ES method additionally estimates the parameters of a period term and seasonal factors with the historical data. This m is composed of three factors: the level factor, trend factor, and seasonality factor at time *t*, denoted as  $\hat{Y}_t$ ,  $\Delta \hat{Y}_t$ , and  $SN_t$ , respectively. Each factor is expressed as

$$\hat{Y}_{t} = \begin{cases} \alpha Y_{t} + (1 - \alpha)(\hat{Y}_{t-1} + \Delta \hat{Y}_{t}) \ if \ t < L\\ \alpha (Y_{t} - SN_{t-L}) + (1 - \alpha)\hat{Y}_{t-1} \ if \ t \ge L \end{cases}$$
(7)

$$\Delta \hat{Y}_t = \beta \left( \hat{Y}_t - \hat{Y}_{t-1} \right) + (1 - \beta) \Delta \hat{Y}_{t-1}, \tag{8}$$

$$SN_t = \begin{cases} Y_t - \hat{Y}_t \text{ if } t < L\\ \gamma (Y_t - \hat{Y}_t) + (1 - \gamma)SN_{t-L} \text{ if } t \ge L \end{cases}$$
(9)

where  $\hat{Y}_1 = Y_1$ , and  $\Delta \hat{Y}_2 = Y_2 - Y_1$ , and  $0 < \alpha < 1$ ,  $0 < \beta < 1$ ,  $0 < \gamma < 1$ . The parameter  $\alpha$  is a smoothing parameter for the level factor, the parameter  $\beta$  is for a linear trend factor, the parameter  $\gamma$  is for seasonality factor, and the parameter *L* represents the period. The forecast for *s* times further from the current time *t* can be computed as follows:

$$Y_{t+s}^* = \begin{cases} \hat{Y}_t + s \times \Delta \hat{Y}_t & \text{if } t < L \\ \hat{Y}_t + SN_{t+s-L} + s \times \Delta \hat{Y}_t & \text{if } t \ge L \end{cases}$$
(10)

If a time series has a seasonal pattern with increasing variation, the multiplicative triple ES method is more appropriate than the additive model described above. The calculation procedure of the multiplicative method is almost the same as the additive model except for some minor modifications. For multiplicative ES method,  $Y_t - SN_{t-L}$  in Equation (7) and  $Y_t - \hat{Y}_t$  in Equation (9) can be replaced with  $Y_t/SN_{t-L}$  and  $Y_t/\hat{Y}_t$ , respectively.

However, it is not easy to estimate such parameters accurately when the historical data are not sufficient or if prior knowledge on the time series is not available. Thus, the triple ES model might not be an appropriate base model for the self-starting forecasting process under insufficient initial historical observations. In addition, these traditional EWMA models are based on the assumption that the future observations will have a similar pattern to past observations. However, in various fields, such as signal processing, the time-varying patterns of future observations are unknown and far from the historical pattern [29,30]. Consequently, these traditional models often suffer from the bias of forecasting results from true future observations because of this assumption.

To address this drawback of the traditional EWMA models, Ryu and Han proposed two-stage EWMA, which mimics the double ES model [29,30]. The two-stage EWMA can be formulated with three factors: level factor, adjustment factor, and drift factor. In the first step, the level factor at time t,  $\hat{Y}_t$ , can be computed as follows:

$$\hat{Y}_t = \alpha Y_t + (1 - \alpha) \hat{Y}_{t-1},\tag{11}$$

where  $\hat{Y}_1 = Y_1$ . Then, the adjustment factor at time t,  $d\hat{Y}_t$ , can be estimated as follows:

$$d\hat{Y}_t = \beta dY_t + (1 - \beta)d\hat{Y}_{t-1} \tag{12}$$

where  $dY_t$  is a first-order difference between time t and t-1, which can be defined as follows:

$$dY_t = Y_t - Y_{t-1}.$$
 (13)

In addition,  $d\hat{Y}_2 = Y_2 - Y_1$ . Finally, we estimate the drift factors of time t by

$$\Delta \hat{Y}_t = r dY_t + (1 - r) \Delta \hat{Y}_{t-1},\tag{14}$$

where  $\Delta \hat{Y}_2 = Y_2 - Y_1$ . The final forecast for *s* further times from the current time *t* can be computed by the following equation:

$$\Upsilon_{t+s}^* = \hat{Y}_t + d\hat{Y}_t + s \times \Delta \hat{Y}_t.$$
<sup>(15)</sup>

The drift factor plays similar role to the trend factor in the double ES model. However, it should be noted that the drift factor is defined as a difference between two successive observations, and hence, the drift factor can more accurately reflect the dynamically changing patterns than the trend factors, which is the difference between the estimated level factors of two successive time points. Moreover, this method involves the adjusted factors in order to alleviate the effects from sudden and unexpected changing patterns. Hence, the two-stage EWMA method can outperform the double ES method.

#### 2.2. Self-Starting Forecasting Process with EWMA Model

To ensure the satisfactory performance of these EWMA models, sufficient historical time series observations should be prepared beforehand. However, in many real situations, only a small number of time series observations are available. One of possible solutions to this problem is a self-starting approach, which both forecasting and model updating are simultaneously done as a new time series observation is added to the training instances. With that approach, the base EWMA model is built without requiring large preliminary initial time series observations. In addition, in the self-starting approach, the models can effectively accommodate the time-varying patterns of time series in that the base EWMA models are continuously updated as time series observations are newly recorded. It should be noted that other forecasting methods, including ARIMA, GARCH, and VAR, cannot be used for the self-starting forecasting process. These methods estimate a number of parameters, and in order to estimate them, more time series observations than the number of parameters are required [35,36]. Accordingly, other forecasting methods are not available for the self-starting forecasting process. Conversely, as described in Section 2.1, only a small number of time series observations (two) are

required to estimate the initial factors of EWMA models; therefore, we only consider the EWMA models as a base model for the self-starting forecasting process.

Because the self-starting forecasting process was proposed to address the situation where there was no or only very small initial training observations available, it starts using only the minimum observations required for the base model as initial training data. Thus, we also constructed the initial base EWMA model using only the first two observations. Then, with the base EWMA model, we forecasted *s*-step ahead future observations (i.e., forecasting window is denoted as *s*). The EWMA models were continuously updated when a new observation was added, and *s*-step ahead forecasting was also continuously conducted through the updated EWMA model. Figure 1 exemplifies the self-starting forecasting process, which was considered in this study, when the forecasting window, *s*, is set as 4, and the total length of the time series observation is 120.



**Figure 1.** Graphical illustration of self-starting forecasting process when the forecasting window (s) and total time series length are 4 and 120, respectively.

As shown in Figure 1, the initial EWMA model  $f(y_1, y_2)$  is constructed by only two initial observations  $(y_1 \text{ and } y_2)$ , and the  $f(y_1, y_2)$  generates the four-step ahead forecasting value,  $y_6^*$ . Then, when the observation is in the third time point,  $y_3$ , is newly recorded, the initial forecasting model is updated to  $f(y_1, y_2, y_3)$ . These forecasting and model updates are repeated until 116th observation is recorded. This continuous model update process was repeated until (120 s) time series observations were accumulated.

#### 3. Simulation Study

#### 3.1. Simulation Setup

We conducted simulation studies under nine time series patterns to evaluate the performance of four EWMA methods: (1) single ES; (2) double ES; (3) triple ES; and (4) two-stage EWMA methods. The simulation was conducted under the following nine patterns:

- Pattern 1:  $Y_t = N(0, \sigma_t^2)$ , t = 1, 2, ..., 120;
- Pattern 2:  $Y_t = t/3 + N(0, \sigma_t^2)$ , t = 1, 2, ..., 120;
- Pattern 3:  $Y_t = \exp(t/30) + N(0, \sigma_t^2)$ , t = 1, 2, ..., 120;
- Pattern 4:  $Y_t = \max(0, t 60) + N(0, \sigma_t^2)$ , t = 1, 2, ..., 120;
- Pattern 5:  $Y_t = \min(t, (120 t)) + N(0, \sigma_t^2), t = 1, 2, ..., 120;$
- Pattern 6:  $Y_t = \begin{cases} N(0, \sigma_t^2) \ 1 \le t \le 30 \\ 10 + N(0, \sigma_t^2) \ 31 \le t \le 60 \\ 20 + N(0, \sigma_t^2) \ 61 \le t \le 90 \\ 30 + N(0, \sigma_t^2) \ 91 \le t \le 120 \end{cases}$ ;
- Pattern 7:  $Y_t = 10\sin(t/10) + N(0, \sigma_t^2)$ , t = 1, 2, ..., 120;
- Pattern 8:  $Y_t = t + 10\sin(t/10) + N(0, \sigma_t^2)$ , t = 1, 2, ..., 120;

• Pattern 9: 
$$Y_t = \begin{cases} 10\sin(t/5) + N(0,\sigma_t^2) \ 1 \le t \le 20\\ 10\sin(t/15) + N(0,\sigma_t^2) \ 21 \le t \le 80\\ 10\sin(t/10) + N(0,\sigma_t^2) \ 81 \le t \le 120 \end{cases}$$
.

The length of each time series pattern is 120, and the noises generated from the normal distribution whose mean is zero and standard deviation is  $\sigma_t^2$  were added to the baseline. In this study, we investigated the performance of EWMA methods on both homoscedasticity and the heteroscedasticity of the noises. Hence, the standard deviation of noise is  $\sigma_t$ , which can be defined as follows:

- Small noise (homoscedastic):  $\sigma_t = 0.5$ ,  $t = 1, 2, \dots, 120$ ;
- Medium noise (homoscedastic):  $\sigma_t = 1, t = 1, 2, \dots, 120;$
- Large noise (homoscedastic):  $\sigma_t = 2$ , t = 1, 2, ..., 120;
- Increasing noise (heteroscedastic):  $\sigma_t = 0.5 + t/80$ , t = 1, 2, ..., 120;
- Deceasing noise (heteroscedastic):  $\sigma_t = 2 t/80$ , t = 1, 2, ..., 120.

Figure 2 shows the baseline of the time series in each pattern.

Pattern 1 and 2 represent the simplest time series patterns, which are the stationary and linear trend, respectively. On the other hand, Pattern 3 is a time series with an exponential trend, one of the most well known nonlinearly time-varying pattern, and Pattern 4 represents a time series that changes from stationary to non-stationary. More specifically, the series has a stationary pattern from time 1 to 60, followed by a constantly increasing trend from time 61 to 120. In addition, in Pattern 5, the direction of the trend is changed from an increasing pattern to decreasing pattern. That is to say, the trend of the simulation data constantly increased from time 1 to 60, and then, it constantly decreased from time 61 to 120. The simulated data in Pattern 6 consist of four stationary patterns with step-like increments with a period size of 30. Pattern 7 considers time series that exhibit constant seasonal time series whose period size is 20, and Pattern 8 shows time series that exhibit seasonal variation (period size is 20) with a linearly increasing trend. Finally, Pattern 9 considers time series with seasonal variation with changing periods. In this scenario, the periods are 10, 20, and 30, respectively.

In each simulation study, we applied the self-starting EWMA model to each simulation case. That is, we built the initial EWMA models using only the first two observations as a training time series data,

and forecasted *s*-step ahead future observations using these initial models. Then, the EWMA models were continuously updated as a successive observation was added to the training time series data. Finally, the updated EWMA model forecasted *s*-step ahead future observations. This continuous model update and forecasting process was repeated until (120-s) time series observations were accumulated.



Figure 2. Baselines of nine time series patterns.

In this simulation study, we applied the forecasting window, *s*, from one to six. In each *s* value, we repeated this experiment 100 times and reported the average root mean square error (RMSE) values of *s*-step ahead forecasting obtained from these 100 repetitions. To use the EWMA methods, appropriate smoothing parameters should be determined. In this simulation, we changed the smoothing parameters between 0.1 and 0.9 with a step size of 0.1, and we reported the lowest average RMSE values among total smoothing parameter sets. The value of the smoothing parameter, 0.1, is sufficient to reduce the risk of involving the noises of latest time series observation, and also, the value of smoothing parameter, 0.9, is a sufficiently large value to consider the impact of the most current time series observations. For this reason, in previous works, the range of smoothing parameter values have been set between 0.1 and 0.9 [28,37]. Since simulation data for pattern from 1 to 6 do not have seasonality, we conducted the experiments without the triple ES model in these patterns. Finally, the self-starting forecasting process uses only two initial data, and the period and seasonal factors of the triple ES method cannot

be accurately estimated using only this small number of time series data. Hence, in this study, we used the ground truth of these factors to determine them.

#### 3.2. Simulation Results

From Figures 3–5, x-axis is the prediction windows and y-axis is average root mean square error (RMSE) values obtained from 100 repeated experiments. In Figure 3, the RMSE curves in Pattern 1 are exhibited. In this time series pattern, the single ES yields a smaller RMSE than the other two methods because the single ES is designed for stationary patterns. In addition, the trend factors in the double ES methods tend to be overestimated, and these inappropriately estimated trend factors lead to poor forecasting results. On the other hand, compared to the double ES method, the two-stage EWMA method produced better forecasting performance than the double ES because the adjustment factor alleviates the effect of the overestimated trend factors.



**Figure 3.** Comparison results between a single exponential smoothing (ES), double ES, and two-stage EWMA in Pattern 1.

Figure 4 shows that the comparative results from Pattern 2 to Pattern 5 are nonstationary time series having monotonically changing parts but no seasonality pattern. When the noises are not large, the double ES model shows comparable performance with other two methods because these patterns are composed of monotonically changing parts. In addition, the single ES model produced the worst results because it cannot consider the nonstationary time series. On the other hand, the two-stage EWMA model also produced the comparable performance with double ES in that the two-stage EWMA model also basically designed the forecasting of the monotonically changing time series. In particular, the two-stage EWMA method outperformed the other two methods in Pattern 3 because this model can properly deal with a nonlinearly increasing pattern by reducing the bias based on its adjustment term. On the other hand, when the large-sized noises are added, the double ES model performs the worst because of its inaccurately estimated trend factors. That is to say, the number of initial observations is too small to properly estimate the initial trend factors in the noisy time series, and these poor trend factor estimation results sequentially influence the successive model update process of the double ES model. Conversely, the two-stage EWMA model produces better forecasting performance because the adjustment factors can lessen the risk of the inaccurate estimation of trend factors due to the noises and small number of initial time series data. In addition, as mentioned earlier, the drift factor of the two-stage EWMA model helps to accommodate the dynamics of time series in that they are computed from the difference between the two successive time series observations—not from the difference

between the estimated level factors, as in the double ES model. These results imply that the two-stage EWMA model is robust to the noises because the adjustment factor can successfully compensate the inaccurately estimated trend factors.



Figure 4. Cont.



**Figure 4.** Comparison results between the single ES, double ES, and the two-stage EWMA in (**a**) Pattern 2, (**b**) Pattern 3, (**c**) Pattern 4 and (**d**) Pattern 5.

In the current study, we also investigated the properties of three EWMA models when the size of noise varies over time (heteroscedastic settings). When the initial noises are small, trend factors in the double ES are accurately estimated in early phase, and they are also appropriately updated, not affected by the noise in the early update process. Although the noise is increased over time, the double ES model tends to still accurately forecast because these accurately estimated trend factors in the early phase can alleviate the effects of noises. Conversely, the two-stage EWMA model showed smaller RMSE values when the time series started with large-sized noise because the adjustment factor can successfully compensate the inaccurately estimated trend factors, as the same as the experimental results in homoscedastic settings. The simulation results in the heteroscedastic settings demonstrate that the initial condition of time series observations might significantly affect the performance of a self-starting EWMA model when the time series patterns a have monotonically increasing or decreasing part.

Figure 5 shows the comparative results from Pattern 6 to Pattern 9 that have a seasonality pattern. Please note that the initial seasonal factors in the triple ES model should be carefully estimated for more accurate forecasting. However, in our problem setting, the number of initial historical data is too small to accurately estimate the initial seasonal factors. Moreover, when the noises become large, the seasonal factors also tend to be inaccurately estimated. For these reasons, the triple ES model showed the worst forecasting performance in these scenarios, although this model is proposed for forecasting the seasonal time series. These results imply that the triple ES model is not effective for seasonality-patterned data when the initial time series is insufficiently prepared and the time series data contain large sized noises. Finally, when the periods of the seasonality are changing (Pattern 9), the triple ES model performs the worst because this model is based on the assumption that the period is constant. The double ES model also yielded unsatisfactory forecasting results because this model was designed for only considering the linear trend factors. In addition, as mentioned earlier, this model is quite sensitive to the noises, and hence, when large-sized noises are added to the time series, forecasting accuracies become worse. Conversely, the two-stage EMWA model tends to produce the best forecasting results among all EWMA models because the adjustment factor helps adjust the bias of the estimation results of the level factors for the patterns whose changing rates between successive time series observations is not constant. However, the two-stage EMWA model cannot accurately forecast when the forecasting windows become large in the seasonality-patterned data because this method is basically proposed for forecasting the monotonically changing patterns. Thus, these results confirmed that the two-stage EWMA can also be effectively applied to short-term forecasting for the seasonality time series data.





Figure 5. Cont.



**Figure 5.** Comparison results between the single ES, double ES, triple ES, and two-stage EWMA in (**a**) Pattern 6, (**b**) Pattern 7, (**c**) Pattern 8 and (**d**) Pattern 9.

In addition, the experimental results in heteroscedastic settings are similar to those in homoscedastic settings. That is to say because the seasonal factors cannot be accurately estimated under

insufficient time series observations, the triple ES model cannot produce a satisfactory forecasting performance regardless of the varying patterns of noises. Besides, single ES and double ES models are proposed for the stationary and monotonically changing time series patterns, and thus, they cannot yield superior performance in seasonal time series patterns for both cases that the noises are homoscedastic and heteroscedastic.

## 4. Case Study

In this section, we investigated the performance of EWMA methods using 18 real time series datasets. Table 1 and Figure 6 show the characteristics and time series patterns of these real data, respectively. These real data are available in public data repositories (Federal Reserve Bank of St. Louis (https://research.stlouisfed.org/); https://datamarket.com/data/list/?q=provider%3Atsdl; https://datahub.io/core/global-temp#data).

Case	Duration of Time Series	Time Scale
Case 1: Unemployment rate in USA	2000.01-2019.01	Monthly
Case 2: USD–EUR exchange rate	2000.01-2019.02	Monthly
Case 3: USD-KRW exchange rate	2000.01-2019.02	Monthly
Case 4: WTI crude oil price	2000.01-2019.01	Monthly
Case 5: Brent crude oil price	2000.01-2019.01	Monthly
Case 6: S&P 500 index	2009.01-2019.01	Monthly
Case 7: NASDAQ-100 index	2000.01-2019.02	Monthly
Case 8: Australian expenditure on financial services	1969.09-1994.03	Quarterly
Case 9: Chemical process temperature readings	NA	Minute
Case 10: Changes in the Earth's rotation day length	1821–1970	Yearly
Case 11: Total building and construction activities in Australia	1973.09-1995.03	Quarterly
Case 12: Money supply in USA	1890–1974	Yearly
Case 13: Birth per 10,000 of 23 year old people	1961–2019	Yearly
Case 14: Numbers on unemployment benefit in USA	1984.01-2019.01	Monthly

Tal	ole 1.	Summary	of real	cases.
-----	--------	---------	---------	--------

As shown in Figure 6, most time series data in real cases are nonstationary patterns, and they also have large-sized noises. In real situations, prior knowledge on periods and seasonality factors for the triple ES is not available. However, we already demonstrated that the triple ES is not effective for our problem settings in the simulation study. Hence, we only considered the single ES, double ES, and two-stage EWMA models in the real time series cases.

In the case study, we applied the same experimental setting as the simulation study. At first, we constructed the initial EWMA models using only the first two observations as training time series data and forecasted *s*-step ahead future observations using these initial models. In the next step, we updated the EWMA models by adding the successive new observation to the training time series. Finally, the updated model forecasted the *s*-step ahead future observations. This continuous model update process was repeated until the (*L*-s) time series observations were accumulated where *L* was the length of the total time series data. In addition, as the same in the simulation study, we considered the smoothing parameter values between 0.1 and 0.9 with a step size of 0.1, and we reported the lowest average RMSE values among the total smoothing parameter sets. Finally, the RMSE values are computed by varying the forecasting window (*s*) from one to six.

Table 2 shows the comparison results between three EWMA models.



Figure 6. Time series data of 14 real cases.

Case	EMWA Model	<i>s</i> = 1	<i>s</i> = 2	<i>s</i> = 3	<i>s</i> = 4	<i>s</i> = 5	<i>s</i> = 6	Average
Case 1	Single ES	0.17	0.26	0.35	0.44	0.53	0.62	0.40
	Double ES	0.15	0.22	0.28	0.35	0.42	0.51	0.32
	Two-stage EWMA	0.15	0.21	0.28	0.33	0.39	0.47	0.31
	Single ES	0.03	0.05	0.06	0.07	0.08	0.09	0.06
Case 2	Double ES	0.03	0.05	0.07	0.08	0.09	0.11	0.07
	Two-stage EWMA	0.03	0.05	0.06	0.07	0.08	0.09	0.06
Case 3	Single ES	28.46	45.93	57.41	66.86	76.29	85.51	60.08
	Double ES	28.91	48.23	61.96	74.09	87.02	100.51	66.79
	Two-stage EWMA	25.48	45.35	56.37	65.63	73.43	82.76	58.17
	Single ES	5.74	9.35	12.37	14.78	16.78	18.39	12.90
Case 4	Double ES	5.67	9.90	13.58	16.82	19.77	22.38	14.69
	Two-stage EWMA	5.08	8.34	<u>11.46</u>	13.77	15.59	17.54	<u>11.96</u>
	Single ES	5.92	9.58	12.50	14.81	16.75	18.42	13.00
Case 5	Double ES	5.99	10.10	13.63	16.66	19.43	22.01	14.64
	Two-stage EWMA	5.33	8.95	11.86	14.41	16.09	17.86	12.42
	Single ES	53.49	80.72	99.46	115.55	129.62	144.23	103.85
Case 6	Double ES	54.49	85.14	106.53	124.03	140.21	156.16	111.09
	Two-stage EWMA	49.47	76.97	93.08	<u>105.16</u>	114.09	122.17	<u>93.49</u>
	Single ES	151.05	230.85	289.31	340.43	390.81	442.15	307.43
Case 7	Double ES	158.75	265.42	342.90	407.23	477.73	552.12	367.36
	Two-stage EWMA	139.70	222.72	275.33	321.19	361.35	402.55	287.14
Case 8	Single ES	40.40	67.80	96.53	125.10	153.53	181.59	110.83
	Double ES	31.51	44.53	61.65	80.38	102.38	121.51	73.66
	Two-stage EWMA	29.64	41.23	56.46	72.36	91.03	107.80	66.42
	Single ES	0.25	0.46	0.66	0.85	1.03	1.19	0.74
Case 9	Double ES	0.14	0.31	0.49	0.68	0.87	1.07	0.59
	Two-stage EWMA	<u>0.14</u>	0.28	0.44	0.60	0.76	0.92	0.52
Case 10	Single ES	29.57	53.40	74.62	92.72	108.53	122.14	80.16
	Double ES	19.87	44.42	74.80	102.48	124.28	144.99	85.14
	Two-stage EWMA	17.46	36.79	<u>60.13</u>	<u>81.02</u>	98.35	<u>111.90</u>	<u>67.61</u>
Case 11	Single ES	599.66	713.71	805.95	853.11	1050.50	1049.90	845.47
	Double ES	621.67	778.90	925.22	1047.40	1371.00	1546.80	1048.50
	Two-stage EWMA	608.68	721.59	795.26	777.03	998.28	1048.70	824.92
Case 12	Single ES	15.05	26.93	37.78	47.55	55.69	64.03	41.17
	Double ES	5.70	10.40	16.23	22.91	25.54	31.73	18.75
	Two-stage EWMA	5.45	9.57	14.72	21.13	24.10	30.23	17.53
Case 13	Single ES	12.65	19.88	25.26	30.32	35.89	41.92	27.65
	Double ES	12.05	18.28	22.11	26.27	32.51	40.55	25.30
	Two-stage EWMA	11.69	18.52	22.98	26.14	28.04	32.01	23.23
	Single ES	14404	24111	30938	36193	41164	46366	32196.00
Case 14	Double ES	13835	23541	30264	35100	39573	44480	31132.17
	Two-stage EWMA	11485	22070	29333	33969	38237	42563	29609.50

Table 2. RMSE values of EWMA models in 14 real cases varying forecasting windows.

In Table 2, the lowest RMSE values in each case and prediction window are highlighted as bold and underline for more effectively showing the best EWMA model. As shown in Table 2, the experimental results in most cases are the almost same as those in simulation studies. That is to say, the double ES model tends to show the worst forecasting results because the trend factors in double ES are sensitive to large-sized noises. Moreover, single ES model also showed unsatisfactory performance in that this model cannot accommodate the nonstationary time series encountered in many real situations. On the other hand, as confirmed in a simulation study, the two-stage EWMA model shows the best

performance in most cases because the adjustment factor of the two-stage EWMA model helps to lessen the effect of noises and the bias of forecast caused by insufficient time series observations. Besides, the two-stage EWMA model yields much smaller RMSE values than the double ES model because the drift factor better models the changing patterns of time series than the trend factor in the double ES model. In other words, the drift factor is calculated as the first-order difference of two successive observations, and it is more effective to accommodate the inherent dynamics of the time series than the trend factor which is defined as the difference of two successive estimated level factors.

# 5. Conclusions

In this study, we examined the properties of the EWMA models as a base forecasting model of a self-starting forecasting process. To this end, we conducted situation studies under the situations that only small initial time series data (number of initial observations is two). In the simulation study, we generated a number of scenarios that reflected a variety of practical situations. These simulation scenarios consider different time series patterns and sizes of noises. In addition to the simulation study, we also applied the self-starting forecasting using the EMWA models on the various real case data. Through comparative experiments in such a number of cases, we can summarize the characteristics of EMWA models as base models of self-starting the forecasting process as follows:

- Single ES performs best only in the stationary time series and yields unsatisfactory results in nonstationary patterns. Thus, this model is not proper for the base model for the self-starting forecasting process in that, in many real situations, there is no assurance that the time series data is changed with stationary patterns.
- Double ES shows comparable or better performances than other EMWA models when the time series observations are monotonically increased or decreased. However, this model is vulnerable to the noises when there are no sufficient time series observations to compensate the effect of noise. In other words, the trend factor in the double ES model cannot be accurately estimated when large-sized noises are added to the insufficient time series observations, and these poorly estimated trend factors sequentially influence the successive model update process of the double ES model. Therefore, the double ES model might not be proper for a base model of the self-starting forecasting process in that the time series data in many real situations often contain large-sized noises.
- The seasonal factor in the triple ES model should be carefully estimated for the sake of more accurate forecasting. However, the seasonal factors are poorly estimated when the initial time series observations are not sufficient, and thus, the triple ES shows the worst performance although this model is designed for handling seasonality patterns. In addition, prior knowledge on the true period is not available when time series observations are not sufficiently accumulated. For these reasons, the tripe ES model is not appropriate to be used for a base model of the self-starting forecasting process.
- Conversely, the two-stage EWMA model tends to yield comparable or better performance than
  other EWMA models in all cases. In particular, this model outperforms other EWMA models as a
  base model for the self-starting forecasting process in the complex time series (i.e., non-stationary
  and noisy time series) because of the drift factor and adjustment factor. That is to say, the drift
  factor calculated as the first-order difference of two successive observations helps to accommodate
  the dynamics of the time series, and the adjustment factor helps to lessen the intrinsic bias caused
  by the noises and insufficient initial time series data. Finally, these appropriately estimated factors
  also lead to desirable EWMA model updates in the self-starting process.

Based on these various comparative results, we suggest using the two-stage EMWA model as a base model of the self-starting forecasting process. For the further study, we will carefully investigate the effects of the smoothing parameters on the self-starting forecasting performance. In addition, we will also study the systematic way to select the optimal smoothing parameters of the EWMA model for the self-starting forecasting process. Finally, for the sake of exploiting these advantages of

the two-stage EWMA as a base model for the self-starting approach, we will attempt to apply it to self-starting approach-based control chart techniques.

**Author Contributions:** J.Y. is responsible for the whole part of the paper (conceptualization, methodology, formal analysis and writing—original draft preparation), S.B.K. is responsible for review and editing, J.B. is responsible for validation and formal analysis, and S.W.H. is responsible for formal analysis, review and editing, and supervision. All authors have read and agreed to the published version of the manuscript.

**Funding:** The fourth author (S.W.H.) of this research was supported by the Korea Institute for Advancement of Technology (KIAT) grant funded by the Korea Government (MOTIE) (P0008691, The Competency Development Program for Industry Specialist).

Conflicts of Interest: The authors declare no conflict of interest.

#### References

- 1. Bowerman, B.L.; O'Connell, R.T.; Koehler, A.B. *Forecasting, Time Series, and Regression: An Applied Approach;* Thomson Brooks/Cole: Pacific Grove, CA, USA, 2005.
- 2. Hunter, J.S. The Exponentially Weighted Moving Average. J. Qual. Technol. 1986, 18, 203–210. [CrossRef]
- 3. Crowder, S.V.A. Simple Method for Studying Run Length Distributions of Exponentially Weighted Moving Average Control Charts. *Technometrics* **1987**, *29*, 401–407.
- 4. Crowder, S.V. Design of Exponentially Weighted Moving Average Schemes. J. Qual. Technol. 1989, 21, 155–162. [CrossRef]
- 5. Lucas, J.M.; Saccucci, M.S. Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements. *Technometrics* **1990**, *32*, 1–12. [CrossRef]
- 6. Friker, R.D.; Knitt, M.C.; Hu, C.X. Comparing Directionally Sensitive MCUSUM and MEWMA Procedures with Application to Biosurveillance. *Qual. Eng.* **2008**, *20*, 478–494. [CrossRef]
- 7. Joner, M.D.; Woodall, W.H.; Reynolds, M.R.; Fricker, R.D.A. One-sided MEWMA Chart for Health Surveillance. *Qual. Reliab. Eng. Int.* **2008**, *24*, 503–519. [CrossRef]
- 8. Han, S.W.; Tsui, K.-L.; Ariyajunyab, B.; Kim, S.B. A Comparison of CUSUM, EWMA, and Temporal Scan Statistics for Detection of Increases in Poisson Rates. *Qual. Reliab. Eng. Int.* **2010**, *26*, 279–289. [CrossRef]
- 9. Snyder, R.D.; Koehler, A.B.; Ord, J.K. Forecasting for Inventory Control with Exponential Smoothing. *Int. J. Forecast.* **2002**, *18*, 5–18. [CrossRef]
- De Faria, E.L.; Albuquerque, M.P.; Gonzalez, J.L.; Cavalcante, J.T.P.; Albuquerque, M.P. Predicting the Brazilian Stock Market through Neural Networks and Adaptive Exponential Smoothing Methods. *Expert Syst. Appl.* 2009, *36*, 12506–12509. [CrossRef]
- 11. Rundo, F.; Trenta, F.; di Stallo, A.L.; Battiato, S. Machine Learning for Quantitative Finance Applications: A Survey. *Appl. Sci.* **2019**, *9*, 5574. [CrossRef]
- 12. Jiang, H.; Fang, D.; Spicher, K.; Cheng, F.; Li, B. A New Period-Sequential Index Forecasting Algorithm for Time Series Data. *Appl. Sci.* **2019**, *9*, 4386. [CrossRef]
- 13. Shilbayeh, S.A.; Abonamah, A.; Masri, A.A. Partially versus Purely Data-Driven Approaches in SARS-CoV-2 Prediction. *Appl. Sci.* **2020**, *10*, 5696. [CrossRef]
- 14. Taylor, J.W. Short-term Electricity Demand Forecasting using Double Seasonal Exponential Smoothing. *J. Oper. Res. Soc.* 2003, *54*, 799–805. [CrossRef]
- 15. Brown, R.G. *Exponential Smoothing for Predicting Demand*; Arthur, D., Ed.; Little Inc.: Cambridge, MA, USA, 1956.
- 16. Brown, R.G. Statistical Forecasting for Inventory Control; McGraw-Hill: New York, NY, USA, 1959.
- Holt, C.C. Forecasting Seasonals and Trends by Exponentially Weighted Moving Averages. *Int. J. Forecast.* 2004, 20, 5–10. [CrossRef]
- 18. Winters, P.R. Forecasting Sales by Exponentially Weighted Moving Averages. *Manag. Sci.* **1960**, *6*, 324–342. [CrossRef]
- 19. Box, G.E.; Jenkins, G.M.; Reinsel, G.C.; Ljung, G.M. *Time Series Analysis: Forecasting and Control*; John Wiley & Sons: Hoboken, NJ, USA, 2015.
- 20. Akgiray, V. Conditional Heteroscedasticity in Time Series of Stock Returns: Evidence and Forecasts. *J. Bus.* **1989**, *62*, 55–80. [CrossRef]
- 21. Sims, C.A. Macroeconomics and Reality. Econom. J. Econom. Soc. 1980, 48, 1-48. [CrossRef]

- 22. Watson, M.W. Vector Autoregressions and Cointegration. Handb. Econom. 1994, 4, 2843–2915.
- 23. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. Neural Comput. 1997, 9, 1735–1780. [CrossRef]
- 24. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. Nature 2015, 521, 436-444. [CrossRef]
- 25. Yang, H.; Pan, Z.; Tao, Q. Robust and Adaptive Online Time Series Prediction with Long Short-Term Memory. *Comput. Intell. Neurosci.* 2017, 2017, 1–9. [CrossRef] [PubMed]
- Divina, F.; Torres Maldonado, J.F.; García-Torres, M.; Martínez-Álvarez, F.; Troncoso, A. Hybridizing Deep Learning and Neuroevolution: Application to the Spanish Short-Term Electric Energy Consumption Forecasting. *Appl. Sci.* 2020, *10*, 5487. [CrossRef]
- 27. Hao, Y.; Gao, Q. Predicting the Trend of Stock Market Index Using the Hybrid Neural Network Based on Multiple Time Scale Feature Learning. *Appl. Sci.* **2020**, *10*, 3961. [CrossRef]
- 28. Montgomery, D.C. Statistical Quality Control, 5th ed.; Wiley: New York, NY, USA, 2004.
- 29. Ryu, E.S.; Han, S.W. The Slice Group Based SVC Rate Adaptation Using Channel Prediction Model. *IEEE COMSOC MMTC E Lett.* **2011**, *6*, 39–41.
- 30. Ryu, E.S.; Han, S.W. Two-Stage EWMA-Based H.264 SVC Bandwidth Adaptation. *Electron. Lett.* **2012**, *48*, 127–1272. [CrossRef]
- 31. Thomassey, S.; Fiordaliso, A. A Hybrid Sales Forecasting System Based on Clustering and Decision Trees. *Decis. Support Syst.* 2006, 42, 408–421. [CrossRef]
- 32. Zou, C.; Zhou, C.; Wang, Z.; Tsung, F. A Self-starting Control Chart for Linear Profiles. J. Qual. Technol. 2007, 39, 364–375. [CrossRef]
- Hawkins, D.M. Self-starting CUSUM charts for Location and Scale. J. R. Stat. Soc. Ser. D 1987, 36, 299–316.
   [CrossRef]
- 34. Menzefricke, U. Control Charts for the Variance and Coefficient of Variation Based on Their Predictive Distribution. *Commun. Stat. Theory Methods* **2010**, *39*, 2930–2941. [CrossRef]
- 35. Linden, A.; Adams, J.L.; Roberts, N. Evaluating Disease Management Program Effectiveness: An Introduction to Time-Series Analysis. *Dis. Manag.* **2003**, *6*, 243–255. [CrossRef]
- 36. Hyndman, R.J.; Kostenko, A.V. Minimum Sample Size Requirements for Seasonal Forecasting Models. *Foresight* **2007**, *6*, 12–15.
- 37. Kang, J.H.; Yu, J.; Kim, S.B. Adaptive Nonparametric Control Chart for Time-varying and Multimodal Processes. *J. Process Control.* **2016**, *37*, 34–45. [CrossRef]

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).