**Original Research Article**

**On the Effects of Missing Values on Estimates of Trend Parameters and Seasonal Indices in Descriptive Time Series Analysis**

**ABSTRACT**

In literature, when missing values are observed in a series, emphasis has always been on obtaining the estimates of the missing values while little or no attention has been given to assessing the effects of the missing values on parameter estimates. This study is aimed at assessing the effect of missing values on the estimates of trend parameters and seasonal indices, when trending curve is linear and decomposition method is additive. The position and number of missing values in the series are also been considered in this study. Estimates from trend parameters and seasonal indices were obtained using descriptive time series methods. The method of Decomposition Without Missing Value (DWMV) was adopted to obtain the estimates of the missing values while the values of summary statistics (MSE, RMSE and MAE) were used to assess the effect of missing values on the parameters. Parameter Recovery Index (PRI) was proposed to measure the percentage recovery rate of the parameters used in simulation. The results of the analysis show that the parameter estimates appear not to be affected by one missing value especially when they are located at different position. However, the estimates are affected when two or more missing values are located at consecutive positions. Furthermore, the effect of the missing value on the estimates appears not to be noticeable for at most ten missing values but appears more pronounced for more than ten missing observations located at separated positions. Specifically, the study discovered that trend parameters, error mean and error variance used in simulation are recovered better than seasonal indices. The study recommends that when there are less or equal to ten missing values in a series, DWMV should be utilized to obtain estimates of trend parameters and seasonal indices.

**Keywords:** Trend Parameters, Seasonal Indices, Consecutive, Separated and Accuracy Measures.

**1. Introduction**

In time series analysis,data is required to be complete, made sequentially in time and equally spaced. In other words, it makes no provision for missing observations. However, in real life missing values are unavoidable for some reasons. Most time series data such as financial time series does not require the data to have missing observations over a long period of time. This is because the statistical properties of the series are preserved by its sequence using such a complete data. Schmitts et al., (2015) confirmed that missing data does not only affect properties of statistical estimations, it also introduces some element of ambiguity in the data analysis and result in misleading inferences and conclusions. Similarly, the correlation structure of a dataset may not be captured if the decision is to replace missing values with zeros, Kerkri et al., (2015). Thus, the need to address the problem of missing values before proceeding with the analysis (Sohae 2015). Some conventional methods used in handling missing values include ignoring or deleting the missing observation, Arslan (2013). This is in contrast to one of the assumptions of the Box-Jenkins method which entails that the series be equally spaced over time and that there are no lost values in the series Yaffee and McGee (1999). Each record in time series is unique; hence, ignoring missing observation may result to having a series that is unusable for many purposes with truncated time series plot Tusell (2005). When analyst encounter missing observations in time series data, one of the remedial measures is to replace the missing observation by its estimate (David 2006 and Howell 2007). Imputation of missing value is an enormous field of study, where a lot of research has already been conducted. Popular techniques include the works of Rubin (1987), Dempster, Laird, Rubin (1977) and Vacek and Ashikaga (1980) on Multiple Imputation, Expectation-Maximization and Nearest Neighbor respectively. However, different methods of missing values imputation have been observed to give inconsistent estimates. This led to the comparison of estimates using different summary statistics by Iwueze, Nwogu, Nlebedim, Nwosu and Chinyem(2018). Adejumo, Onifade and Albert (2021) used similar summary statistics to compare different imputation methods on estimates of missing values. The study observed that the Kalman filter algorithm (KAL) produced thebest missing value estimates. However, the KAL involves complex computational algorithm, a major limitation to the use of the method. In other to generalize on the best method of estimating missing values, Afkanpour, Hosseinzadeh and Tabesh (2024) carried out a review on various imputation methods in the healthcare field. The study classified the missing value imputation methods into; conventional statistical, machine learning, deep learning, and applied hybrid imputation methods. The study found that greater percentage of researcher use the conventional statistical methods followed by machine learning methods of missing values imputation. From the foregoing, most studies have concentrated efforts on obtaining the estimates of missing values without considering the effects of the missing values on the estimates of the parameters of the models. This study is therefore aimed at assessing the effect of missing values on the estimates of trend parameters and seasonal indices in descriptive time series, when trending curve is linear and decomposition method is additive.

**2.** **Materials and method**

The traditional method of time series (also known as the descriptive time series analysis) is adopted for the decomposition the observed served series. The models are commonly used for the decomposition are:

The Additive model,

 (2.1)

the Multiplicative model,

 (2.2)

and the Mixed Model.

 (2.3)

where for time ,  is the trend,  is the seasonal effect,  is cyclical and is irregular or error term. According to Chatfield (2004) the cyclical component is superimposed into the trend when a short period of time is involved, when this happens, the trend-cycle component is given by . This adjustment reduces equations (2.1) – (2.3) to;

The Additive,

 (2.4)

the Multiplicative model,

 (2.5)

and the Mixed Model.

 (2.6)

This study is limited to the model given in equation (2.4). It is assumed from the model Equations that the seasonal effect, when it exists, has periods. That is, it repeats after time periods.

, for all  (2.7)

Furthermore, for Equation (1.4), it is assumed that the sum of the seasonal components over

a complete period is zero.

 (2.8)

It is also assumed that for Equation (2.4), the irregular component is normal with mean zero and constant variance, that is, .

In traditional method of decomposition, the trend component in the observed series is first identified and isolated either by subtraction for Equation (2.4). The de-trended series is obtained as for Equation (2.4). From the detrended series, the seasonal component is estimated and isolated by estimating the average of the de-trended series at each season. The de-trended, de-seasonalized series is obtained as for Equation (2.4). This gives the irregular component which may or may not be random is left.

**3. Assessing performance of trend parameter and seasonal indices**

The effect of missing values on the model parameters will be assessed using the accuracy measures. The accuracy measures; Mean square error (MSE), Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) are based on deviation of the parameter estimates from the parameters. When denotes the parameters used in simulation while  denotes the corresponding estimates in the presence of the missing values. That is,

 (3.1)

The accuracy measures are defined as:

 (3.2)

 (3.3)

 (3.4)

; t is the number of parameters

Also proposed is Parameter Recovery Index (PRI). The index is based on the error in estimate denoted by given in (3.1). It measures the percentage recovery rate of the parameters used in simulation. The PRI is given in (3.5).

,  (k categories). (3.5)

The estimates of the trend parameters and seasonal indices for linear trend-cycle components a3nd additive model are;

Linear trend-cycle components

 (3.6)

Additive Model

 (3.7)

where,  and . The cyclical component is superimposed into the trend when a short period of time is involved, when this happens, the trend-cycle component is given by ( Chatfield,2004).

 (3.8)

where, , and are the trend parameters, is the seasonal indices and is the error.

The estimates of missing values are obtained using decomposing without the missing value (DWMV) proposed by Iwueze et al., (2018).

**Results and discussion**

The empirical examples in this Section consist of 100 series of 120 observations each simulated from . The trend-cycle component used are:,, and ,  (the error component) assumed, , is the period while is the season. The seasonal indices, are shown in Table 1. The simulations were done with Minitab Statistical Software, Version 20. For want of space, only six series, one each for none, one, two, five, ten and fifteen, missing values are shown, while the corresponding graphs are shown in Figures 1 through 4.6.

Table 1.: The seasonal indices for the simulated series

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|  | -1.5 | 2.5 | 3.5 | -4.5 | -1.5 | 2.5 | 3.5 | -4.5 | -1.5 | 2.5 | 3.5 | -4.5 |

*denotes Seasonal indices for Additive model*

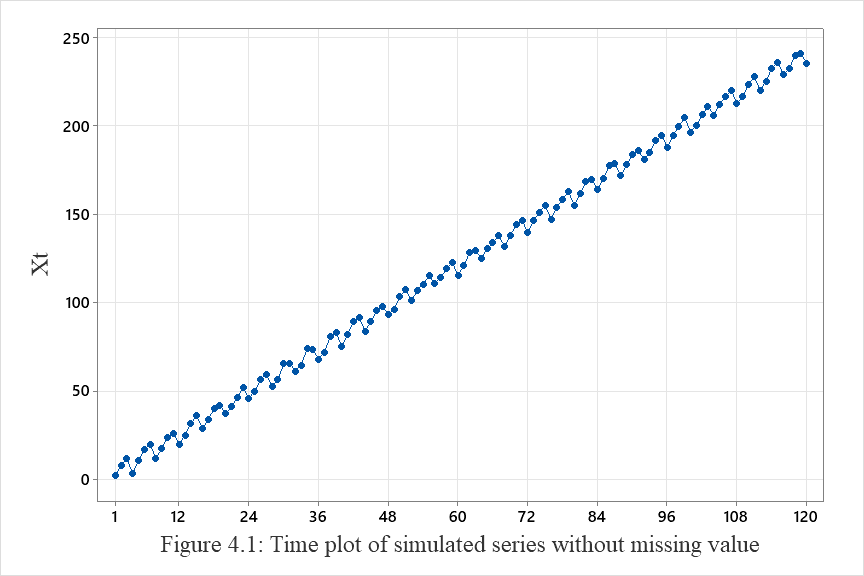
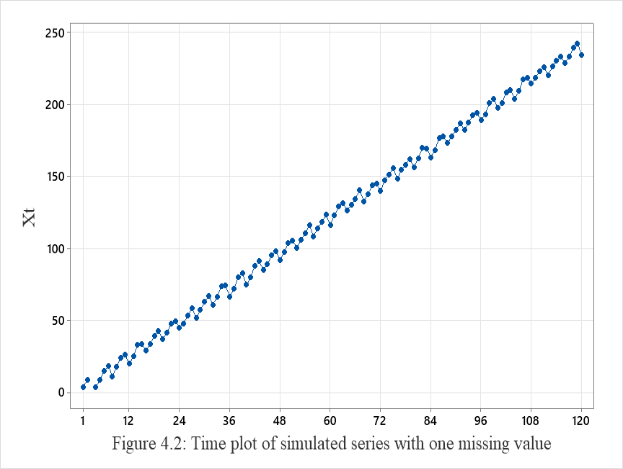


Figure 1: Time plot of simulated series without missing value

Figure 2: Time plot of simulated series with missing value

Figure 1 shows, the plot is the normal plot of a series without missing value moving in an upward direction in a linear form, Figure 2 is the time plot of series with a missing value at position 3.

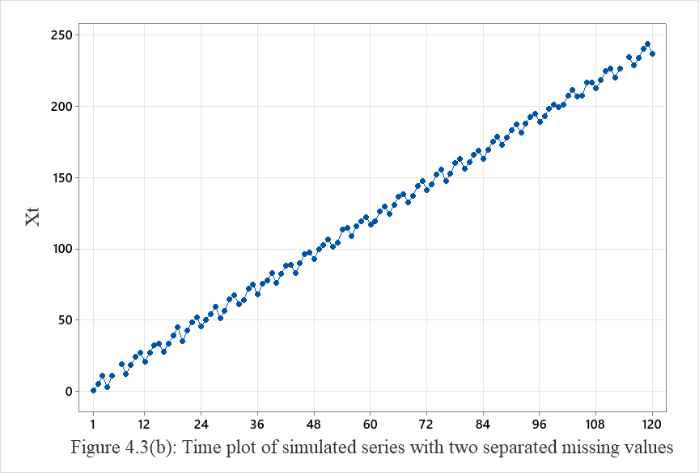
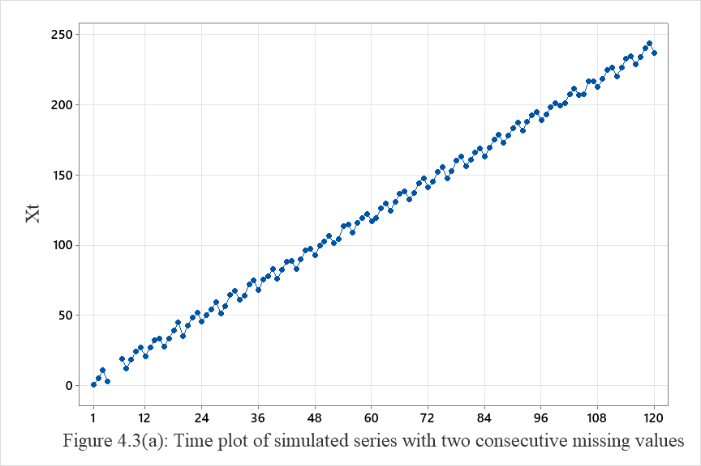


Figure 3(a): Time plot of simulated series with two consecutive missing values

Figure 3(b): Time plot of simulated series with two separated missing values

Figures 3(a) and (b) are the time plots of series with two consecutive and separated missing values.

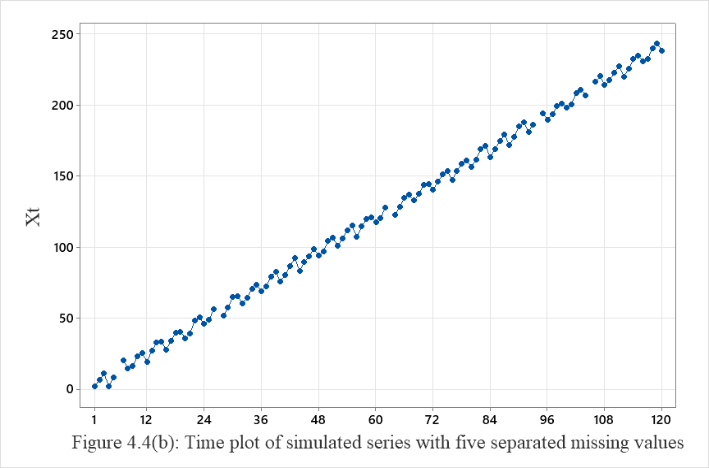
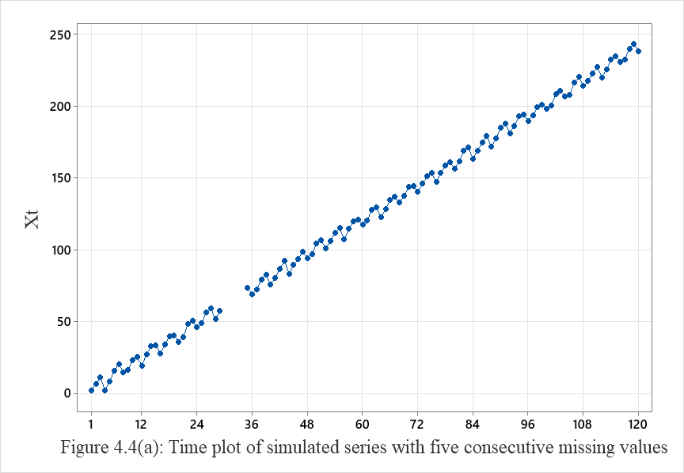


Figure 4(a): Time plot of simulated series with five consecutive missing values

missing value

Figure 4(b): Time plot of simulated series with five separated missing values

Similarly, Figures 4(a) and (b) show the time plots series with five consecutive and separated missing values. The time plots for series with ten consecutive and separated missing are shown in Figures 5(a) and (b). Figures 5-time plots for series with ten consecutive and separated missing values

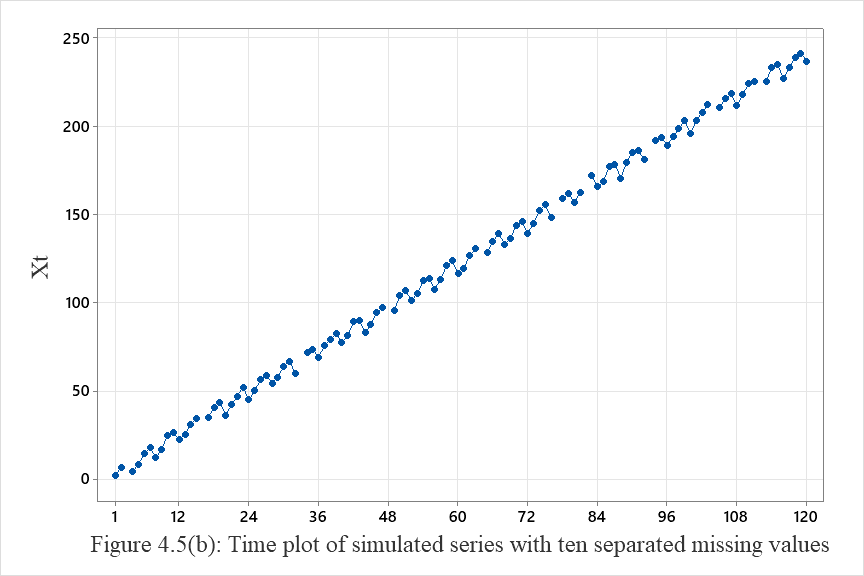
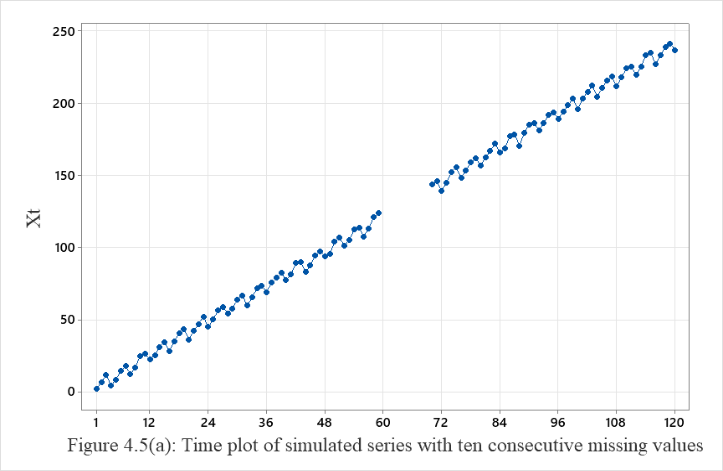


Figure 5(a): Time plot of simulated series with ten consecutive missing values

missing value

Figure 5(b): Time plot of simulated series with ten separated missing values

value

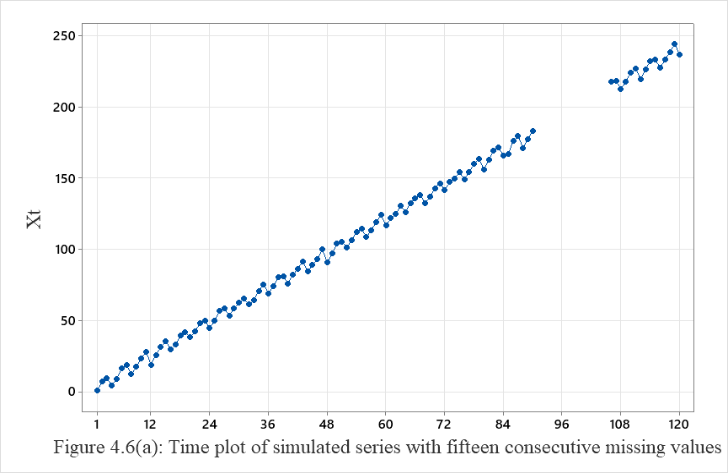


Figure 6(a): Time plot of simulated series with fifteen consecutive missing values

missing value

Figure 6(b): Time plot of simulated series with fifteen separated missing values

value

Furthermore,

Figures 6(a) and (b) are the time plots of series with fifteen consecutive and separated missing values.

Furthermore, Figures 6(a) and (b) are the time plots of series with fifteen consecutive and

separated missing values.

**4.2. Empirical Results: Estimation of trend parameters and seasonal indices**

Table 2 presents the estimate of trend parameters and seasonal indices for series with no missing value and series with one missing value at different positions in the series. Contained in Table 2 are the errors , that is, difference between actual value and estimated values and summary statistics for these series. As Table 2 show, the value of the slope recovered almost perfectly well. The value of the intercept  recovered but not as much as the slope, with absolute error ranging from 0.06 to 0.11. On the other hand, the seasonal indices recovered with greater absolute error. The poor recovery of the seasonal indices is the same for series without missing value and series with one missing value and this is true for one missing value at different position. From the PRI, it is clear that recovery is better with trend parameters, error mean and variance than seasonal indices. The estimate of trend parameters and seasonal indices for series with no missing value and series with two or more missing values at consecutive and separated positions. The errors, that is, difference between actual parameter values used in simulation and corresponding estimated values and summary statistics are presented in Table .3 for missing values at consecutive positions and Table 4 for missing values at separated positions. Similarly, the value of the slopewas recovered almost perfectly well for series in Table 3 and table 4 while the value of the intercept was recovered but not as much as the slope with absolute error ranging from 0.04 to 0.12 for Table 3 and 0.05 to 0.12 for Table.4. The seasonal indices were recovered with greater absolute error. The poor recovery of the seasonal indices is the same for all number of missing values in both consecutive and separated position. The value of PRI shows that recovery is better with trend parameters error mean and variance, than seasonal indices as seen in Table 3 and 4.

Table 2: Estimates of trend parameter and seasonal indices for series with no/one missing value at different Position in the data

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | One missing value at different position; (k = 8) | | | | | | | |  |  |
|  | NONE | 1 | 2 | 5 | 10 | 20 | 50 | 100 | 120 |  |  |
| Parameter | Error | Error | Error | Error | Error | Error | Error | Error | Error | Range | PRI |
|  | -0.09 | -0.06 | -0.10 | -0.07 | -0.10 | -0.08 | -0.11 | -0.10 | -0.08 | 0.06 - 0.11 | 91.22 |
|  | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0 | 100.00 |
|  | 0.54 | 0.54 | 0.54 | 0.53 | 0.57 | 0.55 | 0.58 | 0.46 | 0.54 | 0.46 - 0.58 | 46.11 |
|  | -0.18 | -0.18 | -0.18 | -0.19 | -0.15 | -0.20 | -0.37 | -0.20 | -0.18 | 0.15 - 0.37 | 79.67 |
|  | -0.66 | -0.66 | -0.66 | -0.68 | -0.63 | -0.68 | -0.62 | -0.74 | -0.66 | 0.62 - 0.76 | 33.44 |
|  | 0.56 | 0.56 | 0.56 | 0.55 | 0.59 | 0.54 | 0.60 | 0.61 | 0.56 | 0.54 - 0.61 | 43.00 |
|  | -0.48 | -0.48 | -0.48 | -0.49 | -0.45 | -0.69 | -0.44 | -0.56 | -0.48 | 0.44 - 0.69 | 49.44 |
|  | -0.05 | -0.05 | -0.05 | -0.06 | -0.02 | -0.06 | -0.01 | -0.13 | -0.05 | 0.01 - 0.13 | 94.67 |
|  | -0.11 | -0.11 | -0.11 | -0.12 | -0.08 | -0.13 | 0.30 | 0.18 | -0.11 | 0.08 - 0.3 | 86.11 |
|  | -0.32 | -0.32 | -0.32 | -0.34 | -0.30 | 0.02 | -0.28 | -0.03 | -0.32 | 0.02 - 0.34 | 75.00 |
|  | 0.08 | 0.08 | 0.08 | 0.07 | -0.11 | 0.06 | -0.10 | 0.00 | 0.08 | 0 - 0.11 | 92.67 |
|  | -0.09 | -0.09 | -0.09 | 0.04 | -0.19 | -0.11 | -0.46 | -0.15 | -0.09 | 0.04 - 0.46 | 85.44 |
|  | 0.13 | 0.13 | 0.13 | 0.11 | 0.15 | 0.11 | 0.17 | 0.04 | 0.13 | 0.04 - 0.17 | 87.78 |
|  | 0.59 | 0.59 | 0.59 | 0.58 | 0.62 | 0.57 | 0.63 | 0.51 | 0.59 | 0.51 - 0.63 | 41.44 |
|  | -0.07 | -0.07 | -0.07 | -0.07 | -0.07 | -0.07 | -0.08 | -0.07 | -0.08 | 0.07 - 0.08 | 92.78 |
|  | 0.05 | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 | 0.05 | 0.04 - 0.05 | 95.78 |
| **MSE** | **0.11** | **0.11** | **0.11** | **0.11** | **0.11** | **0.12** | **0.13** | **0.11** | **0.11** |  |  |
| **RMSE** | **0.33** | **0.33** | **0.33** | **0.33** | **0.33** | **0.34** | **0.36** | **0.33** | **0.33** |  |  |
| **MAE** | **0.24** | **0.24** | **0.24** | **0.23** | **0.24** | **0.23** | **0.28** | **0.23** | **0.24** |  |  |

Table 3: Estimates of trend parameter and seasonal indices when the number of missing values is consecutive

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Number of consecutive missing values; (k = 4) | | | | |  | |
| TWO | FIVE | TEN | FIFTEEN |  |  | |
| Parameter | Error | Error | Error | Error | Range | PRI |
|  | -0.07 | -0.04 | -0.12 | -0.11 | 0.04 - 0.12 | 91.50 |
|  | 0.00 | 0.00 | 0.00 | 0.00 | 0 | 100.00 |
|  | 0.51 | 0.51 | 0.84 | 0.57 | 0.51 - 0.84 | 39.25 |
|  | -0.21 | -0.12 | -0.19 | -0.35 | 0.12 - 0.35 | 78.25 |
|  | -0.69 | -0.75 | -0.72 | -0.65 | 0.65 - 0.75 | 29.75 |
|  | 0.53 | 0.44 | 0.63 | 0.16 | 0.16 - 0.63 | 56.00 |
|  | -0.51 | -0.71 | -0.46 | -0.67 | 0.46 - 0.71 | 41.25 |
|  | -0.08 | 0.28 | -0.21 | -0.67 | 0.08 - 0.67 | 69.00 |
|  | 0.23 | -0.33 | -0.37 | -0.45 | 0.23 - 0.45 | 65.50 |
|  | -0.35 | 0.13 | -0.52 | 0.68 | 0.13 - 0.68 | 58.00 |
|  | 0.05 | -0.07 | 0.12 | 0.09 | 0.05 - 0.12 | 91.75 |
|  | -0.12 | -0.06 | -0.15 | -0.01 | 0.01 - 0.15 | 91.50 |
|  | 0.09 | 0.09 | 0.53 | 0.62 | 0.09 - 0.62 | 66.75 |
|  | 0.56 | 0.59 | 0.52 | 0.70 | 0.52 - 0.7 | 40.75 |
|  | -0.07 | -0.05 | -0.11 | -0.05 | 0.05 - 0.11 | 93.00 |
|  | 0.03 | 0.09 | 0.04 | 0.05 | 0.03 - 0.09 | 94.75 |
| **MSE** | **0.11** | **0.12** | **0.17** | **0.20** |  |  | |
| **RMSE** | **0.33** | **0.35** | **0.42** | **0.45** |  |  | |
| **MAE** | **0.24** | **0.25** | **0.33** | **0.35** |  |  | |

Table 4: Estimates of trend parameter and seasonal indices when the number of missing values is separated

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Number of separated missing values; (k = 4) | | | | |  |
|  | TWO | FIVE | TEN | FIFTEEN |  |  |
| Parameter | Error | Error | Error | Error | Range | PRI |
|  | -0.05 | -0.09 | -0.06 | -0.12 | 0.05 - 0.12 | 92.00 |
|  | 0.00 | 0.00 | 0.00 | 0.00 | 0 | 100.00 |
|  | 0.54 | 0.16 | 0.22 | 0.40 | 0.16 - 0.54 | 67.00 |
|  | -0.18 | 0.10 | -0.30 | -0.16 | 0.1 - 0.3 | 81.50 |
|  | -0.66 | -0.48 | -0.53 | -1.00 | 0.48 - 1 | 33.25 |
|  | 0.56 | 0.39 | 0.26 | 0.09 | 0.09 - 0.56 | 67.50 |
|  | -0.48 | -0.49 | -0.10 | -0.10 | 0.1 - 0.46 | 70.75 |
|  | -0.05 | -0.06 | -0.32 | 0.18 | 0.05 - 0.32 | 84.75 |
|  | -0.11 | -0.28 | -0.12 | 0.72 | 0.11 - 0.72 | 69.25 |
|  | -0.32 | -0.03 | 0.47 | -0.01 | 0.01 - 0.47 | 79.25 |
|  | 0.08 | 0.17 | 0.08 | 0.05 | 0.05 - 0.17 | 90.50 |
|  | -0.09 | -0.25 | -0.08 | -0.71 | 0.08 - 0.71 | 71.75 |
|  | 0.13 | 0.01 | -0.05 | 0.11 | 0.01 - 0.13 | 92.50 |
|  | 0.59 | 0.77 | 0.49 | 0.44 | 0.44 - 0.77 | 42.75 |
|  | -0.08 | -0.05 | -0.05 | -0.11 | 0.05 - 0.11 | 92.75 |
|  | 0.04 | 0.10 | 0.02 | 0.03 | 0.02 - 0.1 | 95.25 |
| **MSE** | **0.11** | **0.09** | **0.07** | **0.15** |  |  |
| **RMSE** | **0.33** | **0.29** | **0.26** | **0.38** |  |  |
| **MAE** | **0.24** | **0.20** | **0.19** | **0.25** |  |  |

**5. Concussion and Recommendation**

Previous studies have shown that when an analyst encounters missing observations in time series data, the remedial measure is to replace the missing observation by its estimate David (2006). However, reviewed literatures have shown that different methods of handling missing values give inconsistent estimates. This major setback led to the comparisons by Iwueze et al., (2018) and Adejumo et al., (2021). Hence, the need to also investigate the effect of missing values on the estimates of parameter. This study is focused on assessing the effect of missing value based on the locations and number of missing values on the estimates of trend parameter and seasonal indices. The results of the summary statistics MSE, RMSE and MAE in Table 2 revealed that the parameter estimates appear not to be affected by one missing value, especially when they are located at different position. However, the consistent increment in the values of summary statistics in Table 3 suggest that estimates of trend parameters and seasonal indices are affected when two or more missing values are located at consecutive positions. Furthermore, when the number of missing values is separated as showed in Table 4, the effect on the estimates only appears to be noticeable for more than ten missing values. The results of the PRI showed that the trend parameters, error mean and standard deviation used in simulation recovered with high precision while the seasonal indices recovered but not as much as the trend parameters. This indicates that recovery is better with trend parameters than seasonal indices. The study recommends that when there are at most ten missing values in a series, DWMV should be utilized to obtain estimates of trend parameters and seasonal indices.

**6. Competing interest**

The authors declare that no known competing exist on this work.

**7. Disclaimer (Artificial intelligence)**

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc.) and text-to-image generators have been used during the writing or editing of this manuscript.

**8. Reference**

1. Adejumo O.A., Onifade O.C., Albert S. (2021), Kalman Filter Algorithm versus Other Methods of Estimating Missing Values: Time Series Evidence. African Journal of Mathematics and Statistics Studies 4(2), 1-9. DOI: 10.52589/AJMSS-VFVNMQLX.
2. Afkanpour, M., Hosseinzadeh, E., and Tabesh, H. (2024). Identify the most appropriate imputation method for handling missing values in clinical structured datasets: a systematic review. BMC Med Res Methodol 24, 188. <https://doi.org/10.1186/s12874-024-02310-6>
3. Arslan, A., and Aydilek, 1. B. (2013). “A hybrid method for imputation of missing values using optimized fuzzy c-means with support vector egression and a genetic algorithm,” Inf. Sci. (Ny)., vol. 233, pp. 25 – 35, 2013. <https://doi.org/10.1016/j.ins.2013.01.021>
4. David, S.C.F. (2006) Methods for the Estimation of Missing Values in Time Series. Cowan University Press, Western Australia.  https://ro.ecu.edu.au/theses/63
5. Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977), `Maximum Likelihood from Incomplete Data via the EM Algorithm', Journal of the Royal Statistical Society. Series B (Methodological) 39(1), pp. (1-38). http://links.jstor.org/sici?sici=0035-9246%281977%2939%3A1%3C1%3AMLFIDV%3E2.0.CO%3B2-Z.
6. Howell, D.C. (2007) The Analysis of Missing Data. In: Outhwaite, W. and Turner, S., Eds., Handbook of Social Science Methodology, Sage, London. https://doi.org/10.4135/9781848607958.n11.
7. Iwueze, I.S., Nwogu, E.C., Nlebedim, V.U., Nwosu, U.I. and Chinyem, U.E. (2018). Comparison of Methods of Estimating Missing Values in Time Series. Open Journal of Statistics, 8, 390-399. <https://doi.org/10.4236/ojs.2018.82025>.
8. Kerkri, A, Zarrouk, Z. & Allal, J. (2015). A comparison of NIPALS algorithm with two other missing data treatment methods in a principal component analysis. http://paperssondages14.sfds.asso.fr/submission\_41.pdf
9. Rubin, D.B. (1987) Multiple Imputation for Nonresponse in Surveys. John Wiley & Sons Inc., New York. http://dx.doi.org/10.1002/9780470316696
10. Schmitt, P; Mandel, J; & Guedj, M. (2015). A comparison of six methods for missing data imputation. Journal of Biometrics & Biostatistics, 6(224), 1-6. http://dx.doi.org/10.472/2155-6180.1000224
11. Sohae, O. (2015). Multiple Imputation in missing values in time series data (Master's thesis). Duke University, North California. https://dukespace.lib.duke.edu/server/api/core/bitstreams/027b3bc4-6c6b-40e4-a8ff-bef39fc59785/content
12. Tusell, F.P. (2005). Multiple imputation of time series with an application to the construction of historical price indices. Lehendakari Aguirre, Bilbao: University of the Basque Country Go. https://addi.ehu.es/handle/10810/5663
13. Vacek, P. and Ashikaga, T. (1980). `An examination of the nearest neighbor rule for imputing missing values', ASA Proceedings of the Statistical Computing Section, pp. 326–331. https://hellanicus.lib.aegean.gr/bitstream/handle/11610/17202/Missing\_Data\_in\_Time\_Series\_and\_Imputation\_methods.pdf?sequence=1
14. Yaffee, R & McGee, M. (1999). Introduction to Time Series Analysis and Forecasting: With Applications of SAS and SPSS. Brooklyn, NY: Academic Press Incorporation. https://www.researchgate.net/publication/234789625