

# **Effect of Missing Observations on Buys-Ballot Estimates of Time Series Components**

**ABSTRACT** This study discusses the effects of missing observations on Buys-Ballot estimate when trend-cycle component of time series is linear. The method adopted in this study is Decomposing Without the Missing Value (DWMV) which is used to estimate missing observations in time series decomposition when data are arranged in a Buys-Ballot table. The model structure used is multiplicative. Results show that the trend parameters with and without missing observations have insignificant effect while there are significant differences in the seasonal indices only at the season points where missing observations occurred in the Buys-Ballot table.

**Keywords:** Time Series Decomposition, Missing Observation, Trend Parameter, Seasonal Effect, Multiplicative Model, Buys-Ballot Table.

## **1.1 Introduction**

The Buys-Ballot method is primarily used for the decomposition of a relatively short term period such that the trend and cyclical components are joint combined [1] Buys-Ballot decomposition models are.

Additive Model

$$X_t = M_t + S_t + e_t \quad (1)$$

Multiplicative Model:

$$X_t = M_t \times S_t \times e_t \quad (2)$$

and Mixed Model

$$X_t = M_t \times S_t + e_t. \quad (3)$$

It is always assumed that the seasonal effect, when it exists, has period  $s$ , that is, it repeats after  $s$  time periods.

$$S_{t+s} = S_t, \text{ for all } t \quad (4)$$

For Equation (1), it is convenient to make the further assumption that the sum of the seasonal components over a complete period is zero, ie ,

$$\sum_{j=1}^s S_{t+j} = 0. \quad (5)$$

Similarly, for Equations (2) and (3), the convenient variant assumption is that the sum of the seasonal components over a complete period is  $s$ .

$$\sum_{j=1}^s S_{t+j} = s. \quad (6)$$

In this study, one hundred and twenty (120) birth rate were considered from life spring specialist hospital Ikenegbu Owerri, Imo State, Nigeria from January 2009 to December 2018 in which five (5) births were not accounted for. The observed data was transformed and the trend parameters and seasonal indices estimated. The process was repeated with the estimated missing observations replaced. The estimated parameters and seasonal indices with missing values and without missing values were compared. This study is limited to time series decomposition with only linear trend and seasonal components combined in the multiplicative form, when more than one observations is missing in the Buys-Ballot table. The emphasis is to estimate the missing observations in descriptive time series with linear trend and seasonality using Decomposing without the Missing Value. [2] Proposed estimation of missing observations using different alternatives in stationary term series for autoregressive moving average models. According to him, missing observations occur commonly in descriptive time series and in some cases it is important to estimate them. [3] Provided the definition and computation of marginal likelihood of an ARIMA model with

missing data. They used the univariate version of the modified Kalman filter introduced by [4]. They started how to predict and interpolate missing values and obtain the mean squared error of the estimate. The method obtained for the estimation of models for discrete time series in the presence of missing observations are those of [5].

## 1.2 Estimation Methods for Replacing Missing Observations

The methods of estimating missing observations in time series decomposition include Mean Imputation, Series Mean, Linear Interpolation, Linear trend at point, Row Mean Imputation, Column Mean Imputation and Decomposing without the Missing Values.

[6] Observed that Decomposing without the Missing Value gave the best result when compared with the other six methods. In this study, we intend to use Decomposing without Missing Value (DWMV) to estimate missing observations

## 1.3 Decomposing Without the Missing Value (DWMV)

This method decomposes the remaining data series without the missing value to obtain the trend at point  $(i-1)S + j$  to be

$$M(i-1)S + j = \hat{a} + \hat{b}(i-1)S + j \quad (7)$$

Hence, the estimate at that point becomes

$$\hat{X}_{(i-1)S+j} = \hat{M}(i-1)S + j \times \hat{S}_j \quad (8)$$

## 1.4 Buys-Ballot Procedure for Time Series Decomposition.

According [7], if a time series contains seasonal affects with period  $s$  (length of the periodic interval), we expect observations separated by multiples of  $s$  to be similar:  $X_t$  should be similar to  $X_t \pm is$ ,  $i = 1, 2, 3, \dots, m$ . To analyze the data, it is helpful to arrange the series in a two – dimensional table (Table 1), according to the period and season, including the totals and/or averages. Such two – way tables that display within period pattern, that are similar from period to period are known as Buys – Ballot tables, [8] credits these arrangements to Buys-Ballot [9].

Table (1) displays the within periods relationship, which represent the correlation among observations in the row  $(\dots, X_{t-2}, X_{t-1}, X_t, X_{t+1}, X_{t+2}, \dots)$  and between – periods relationship which represent the correlation among observations in the column,  $(\dots, X_{t-2s}, X_{t-s}, X_{t+s}, X_{t+2s}, \dots)$ . The within periods relationship represent the non-seasonal components of a study series while the between periods represent the seasonal component of a study series. Therefore, the estimates of non-seasonal components are derivable from the row means while the estimates of the seasonal components are derivable from the column means.

For seasonal data with length of period, interval,  $s$ , the Buys-Ballot naturally partitions the observed data into  $m$ -row for easy application.

**Table 1: Buys - Ballots Table for Seasonal time series**

Rows/ Period (i)	Columns (season) j								
	1	2	...	$j$	...	$s$	$T_{i.}$	$\bar{X}_{i.}$	$\hat{\sigma}_{i.}$
1	$X_1$	$X_2$	...	$X_j$	...	$X_s$	$T_{1.}$	$\bar{X}_{1.}$	$\hat{\sigma}_{1.}$
2	$X_{s+1}$	$X_{s+2}$	...	$X_{s+j}$	...	$X_{2s}$	$T_{2.}$	$\bar{X}_{2.}$	$\hat{\sigma}_{2.}$
3	$X_{2s+1}$	$X_{2s+2}$	...	$X_{2s+j}$	...	$X_{3s}$	$T_{3.}$	$\bar{X}_{3.}$	$\hat{\sigma}_{3.}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$i$	$X_{(i-1)s+1}$	$X_{(i-1)s+2}$	...	$X_{(i-1)s+j}$	...	$X_{(i-1)s+s}$	$T_{i.}$	$\bar{X}_{i.}$	$\hat{\sigma}_{i.}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$m$	$X_{(m-1)s+1}$	$X_{(m-1)s+2}$	...	$X_{(m-1)s+j}$	...	$X_{ms}$	$T_{m.}$	$\bar{X}_{m.}$	$\hat{\sigma}_{m.}$
$T_{.j}$	$T_{.1}$	$T_{.2}$	...	$T_{.j}$	...	$T_{.s}$	$T_{..}$		
$\bar{X}_{.j}$	$\bar{X}_{.1}$	$\bar{X}_{.2}$	...	$\bar{X}_{.j}$	...	$\bar{X}_{.s}$		$\bar{X}_{..}$	
$\hat{\sigma}_{.j}$	$\hat{\sigma}_{.1}$	$\hat{\sigma}_{.2}$	...	$\hat{\sigma}_{.j}$	...	$\hat{\sigma}_{.s}$			$\hat{\sigma}_{..}$

In this arrangement each time period  $t$  is represented in terms of the period  $i$  (e.g. year) and season  $j$  (e.g. month of the year), as  $t = (i-1)s + j$ . Thus, the period (row), season (column) and overall totals, means and variances are defined as

$$T_{i.} = \sum_{j=1}^s X_{(i-1)s+j},$$

$$\bar{X}_{i.} = \frac{T_{i.}}{s},$$

$$\hat{\sigma}_{i.}^2 = \frac{1}{s-1} \sum_{j=1}^s (X_{ij} - \bar{X}_{i.})^2$$

$$T_{.j} = \sum_{i=1}^m X_{(i-1)s+j},$$

$$\bar{X}_{.j} = \frac{T_{.j}}{m},$$

$$\hat{\sigma}_{.j}^2 = \frac{1}{m-1} \sum_{i=1}^m (X_{ij} - \bar{X}_{.j})^2$$

$$T_{..} = \sum_{i=1}^m \sum_{j=1}^s X_{(i-1)s+j},$$

$$\bar{X}_{..} = \frac{T_{..}}{n}, n = ms,$$

$$\hat{\sigma}_x^2 = \frac{1}{n-1} \sum_{i=1}^m \sum_{j=1}^s (X_{ij} - \bar{X}_{..})^2$$

UNDER PEER REVIEW

## 2.1 Estimation of Trend Parameters and Seasonal Indices

To choose the appropriate trend of the entire series, the plot of the transformed periodic averages is considered. The expression of the linear trend:

$$\bar{X}_i = a + bi \quad (9)$$

[10] provided the estimation of the trend parameters and seasonal indices as

$$\hat{a} = \bar{X}_{..} - \frac{\hat{b}}{2}(n+1) \quad (10)$$

$$\hat{b} = \frac{b'}{s} \quad (11)$$

$$\hat{S}_j = \frac{\bar{X}_{.j}}{\bar{X}_{..} + \frac{\hat{b}}{2}(2j-s-1)} \quad (12)$$

Estimation of missing observations in the transformed and original time series. To estimate the missing observations in the transformed data, the estimated parameters of trend and seasonal effect for multiplicative model is given as

$$\hat{X}_{ij} = \hat{a} + \hat{b}[(i-1)s + j] \times \hat{S}_j \quad (13)$$

The exponent of the estimated missing observation in the transformed series gives the estimate of the missing observations in the original data

$$\text{Original } X_{ij} = e^{\hat{a} + \hat{b}[(i-1)s + j]} \hat{S}_j \quad (14)$$

### 3.1 Real Life Example

The real life example is based on monthly data on number of birth rate collected from life Spring Specialist Hospital Ikenegbu Owerri, Imo State, Nigeria for a period of 2009 to 2018 given in Appendix A while the time plots of original and transformed series with missing observations are given in Figures 1 and 2.

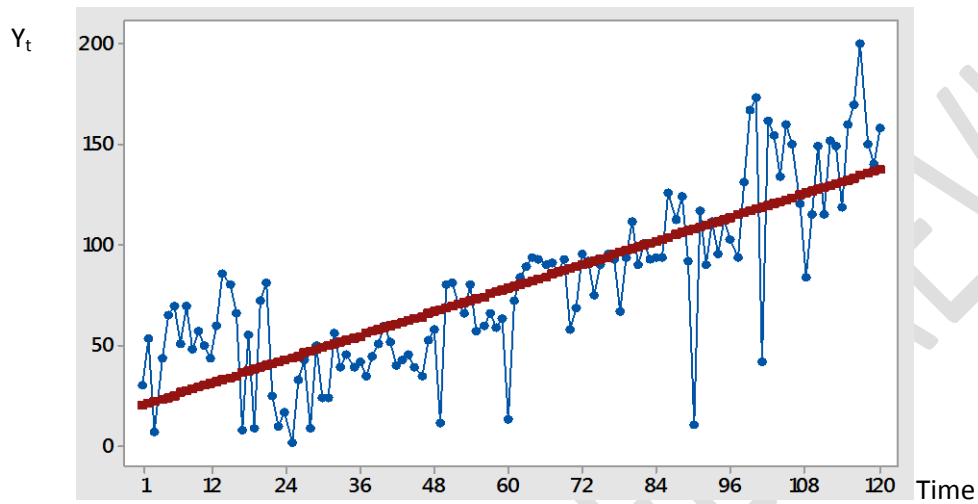


Figure 1: Time plot of the actual series on number of Birth rate with missing observation (2009-2018)

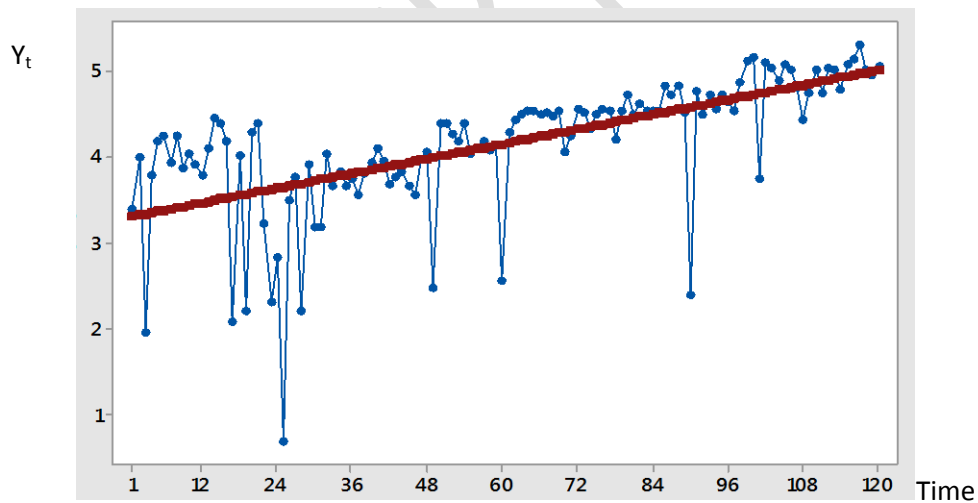


Figure 2: Time plot of the transformed series on number of Birth rate with missing observation (2009-2018)

### 3.2 Buys-Ballot Estimates of Trend and Seasonal Effect with Missing Observations

Hence, the linear trend of the row means in the Buys-Ballot table with missing observations is given as:

$$\bar{X}_i = 3.168 + 0.1820i \quad (15)$$

Using (10) and (11)

$$\hat{b} = \frac{0.1820}{12} = 0.0152$$

$$\hat{a} = 4.1687 - 0.0076(120 + 1) = 3.2491$$

Using (12)

$$\hat{S}_j = \frac{\bar{X}_{.j}}{4.1687 + 0.0076(2j - 12 - 1)}$$

**Table 2: Estimates of Trend and Seasonal Effect with missing observations**

$j$	$\bar{X}_j$	$S_j$
1	3.5682	0.8735
2	4.3590	1.0631
3	4.1990	1.0203
4	4.2630	1.0320
5	4.0640	0.9602
6	4.2307	1.0167
7	4.1050	0.9829
8	4.4190	1.0542
9	4.3492	1.0339
10	4.1980	0.9994
11	4.1220	1.0073
12	3.8537	0.9365
$\sum_{j=1}^s \hat{S}_j$		12.0000



### 3.3 Buys-Ballot Estimates of Trend and Seasonal Effect Without Missing Observations

The linear trend of row means in the Buys-Ballot table without missing observations is given as

$$\bar{X}_i = 3.2107 + 0.1719i$$

Using (10) and (11)

$$\hat{b} = \frac{0.1719}{12} = 0.0143$$

$$\hat{a} = 4.156 - 0.0072(120 + 1) = 3.2893$$

Using (12)

$$\hat{S}_j = \frac{\bar{X}_j}{4.1563 + 0.0072(2j - 12 - 1)}$$

**Table 3: Estimates of Trend and Seasonal Effect without missing observations**

$j$	$\bar{X}_j$	$\hat{S}_j$
1	3.686	0.9041
2	4.3591	1.0654
3	4.199	1.0227
4	4.263	1.0346
5	4.064	0.9829
6	4.047	0.9754
7	4.105	0.9859
8	4.419	1.0577
9	4.391	1.0474
10	4.198	0.9979
11	4.122	0.9768
12	4.022	0.9492
$\sum_{j=1}^s \hat{S}_j$		12.0000

### 3.4 Estimation of Missing Observations in Transformed Data

Estimation of transformed and original missing observations

Using (13)

$$\text{Transformed } \hat{X}_{10,1} = 3.2491 + 0.0076[(10-1)12+1]0.8735 = 3.9727$$

$$\text{Transformed } \hat{X}_{8,6} = 3.2491 + 0.0076[(8-1)12+6]1.0167 = 3.9445$$

$$\text{Transformed } \hat{X}_{2,9} = 3.2491 + 0.0076[(2-1)12+9]1.0339 = 3.4141$$

$$\text{Transformed } \hat{X}_{8,9} = 3.2491 + 0.0076[(8-1)12+9]1.0339 = 3.9799$$

$$\text{Transformed } \hat{X}_{7,12} = 3.2491 + 0.0076[(7-1)12+12]0.9365 = 3.8470$$

Using (14)

$$\text{Original } \hat{X}_{10,1} = e^{3.9727} = 53.1384 \square 53$$

$$\text{Original } \hat{X}_{8,6} = e^{3.9445} = 51.6505 \square 52$$

$$\text{Original } \hat{X}_{2,9} = e^{3.4141} = 30.3816 \square 30$$

$$\text{Original } \hat{X}_{8,9} = e^{3.9799} = 53.5117 \square 53$$

$$\text{Original } \hat{X}_{7,12} = e^{3.8470} = 4.68523 \square 47$$

**Table 4: Buys-Ballot Estimates of Parameters of Trend and Seasonal Effect With and Without Missing Observations**

Parameter	With missing values	Without missing values	Difference
$\hat{a}$	3.2491	3.2893	0.040
$\hat{b}$	0.0076	0.0072	0.0004
$\hat{S}_1$	0.8735	0.9041	0.031
$\hat{S}_2$	1.0631	1.0654	0.002
$\hat{S}_3$	1.0203	1.0227	0.002
$\hat{S}_4$	1.0320	1.0346	0.003
$\hat{S}_5$	0.9802	0.9829	0.003
$\hat{S}_6$	1.0167	0.9754	0.041
$\hat{S}_7$	0.9829	0.9859	0.003
$\hat{S}_8$	1.0542	1.0577	0.004
$\hat{S}_9$	1.0073	1.0474	0.014
$\hat{S}_{10}$	0.9994	0.9979	0.002
$\hat{S}_{11}$	1.0073	0.9768	0.0305
$\hat{S}_{12}$	0.9365	0.9492	0.013

The difference in the trend parameters and seasonal effects with and without missing observations are contained in Table 4. From table 4, it is clear that, the trend parameters with and without missing observations have an insignificant difference. Approximately, they are the same. This is indication that missing observations have insignificant effect on trend parameters. For seasonal effects, there are significant differences with and without missing observations. Significantly, the difference occurred at  $j = 1, 6, 9, 9, 12$ . These are the points in the seasons of the Buys-Ballot table that had missing observations. The original estimates for the unobserved number of Birth rate are fifty three (53) in January 2018, fifty two (52) in June 2016, thirty (30) in September 2010, fifty three(53) in September 2016 and forty seven (47) in December 2015.

### **Concluding Remark**

This paper has discussed the effects of missing observations on Buys-Ballot estimate of time series components for a linear trending curve. Estimates of missing observations in descriptive time series with linear trend and seasonality when more than one observations are missing in the Buys-Ballot table are discussed. The model structure adopted is multiplicative. Results show that missing observations have insignificant effect on trend parameters but significant in seasonal indices.

### **REFERENCES**

- [1] Chatfield, C. (2004). *The analysis of time Series: An introduction*. Chapman and Hall,/CRC Press, Boca Raton.
- [2] Ferreiro, O. (1987). Methodologies for the Estimation of Missing Observations in Time Series. *Statistical and Probability Letters*, 5(1): 65-69.
- [3] Kohn, R. and Ansely, C.F. (1986). Estimation, Prediction and Interpolation of ARIMA Models with Missing Data. *Journal of the American Statistical Association* 81(395): 751-761.
- [4] Ansely, C.F. and Kohn, R. (1985). Estimation, Filtering and Smoothing in State Space Models with Incompletely Specified Initial conditions. *Annals of Statistics* 13: 1286-1316.
- [5] Robinson, P.M. and Dunsmuir, W. (1981). Estimation of Time Series Models in the Presence of Missing Data. *Journal of the Royal Statistical Association* 76(375): 560-568.
- [6] Nwosu, U.I. (2015). Comparison of Methods of Estimating Missing Values in Descriptive Time Series. MSc. Thesis, Department of Statistics, Federal University of Technology, Owerri.
- [7] Iwueze, I. S. & Nwogu E.C. (2004). *Buys-Ballot estimates for time series decomposition*, Global Journal of Mathematics, 3(2), 83-98.
- [8] Wold, H. (1938). A Study in the Analysis of Stationary Time Series. *Almqvist and Wiksell*, Sweden.

[9] Buys-Ballot, C.H.D., (1847). *Leo Claemert Periodiques de Temperature*, Kemintet Fills, Utrecht.

[10] Iwueze, I.S., Nwogu, E.C., Ohakwe, J. & Ajaraogu J.C. (2011). *Best Linear Unbiased Estimate using Buys-Ballot Procedure when trend-cycle Component is Linear. Journal of Applied Statistics Vol.2*

**Appendix A: Buys-Ballot table for the actual data on number of Birthrate with missing observation (2009-2018)**

Year	Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sept.	Oct.	Nov.	Dec.	$\bar{X}_i$	$\sigma_i^2$
2009	30	54	7	44	65	70	51	70	48	57	50	44	49.17	309.79
2010	60	86	80	66	8	55	9	72	-	25	10	17	47.42	972.81
2011	2	33	43	9	50	24	24	56	39	46	39	42	33.92	266.27
2012	35	45	51	60	52	40	43	46	39	35	53	58	46.42	71.36
2013	12	80	81	71	66	80	57	60	66	59	63	13	59.00	541.27
2014	72	84	89	94	93	90	91	87	93	58	69	96	84.67	142.24
2015	92	75	90	96	93	67	94	112	90	101	93	-	91.42	129.54
2016	94	126	113	124	92	-	117	90	-	96	113	103	99.25	923.84
2017	94	131	167	173	42	162	155	134	160	150	121	84	131.10	1582.40
2018	-	149	115	152	149	119	160	170	200	150	140	158	148.08	594.27
$\bar{X}_j$	60.6	86.3	83.6	88.9	71.0	71.8	80.1	89.7	92.8	77.7	75.1	70.9		
$\sigma_j^2$	1521.6	1474.2	1941.6	2493.7	1462.9	2007.1	2736.3	1504.9	2782.0	2011.6	1657.7	1992.8		

**Appendix B: Buys-Ballot table for the transformed data on number of Birthrate with missing observation (2009-2018)**

Year	Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sept.	Oct.	Nov.	Dec.	$\bar{X}_i$	$\sigma_i^2$
2009	3.401	3.989	1.946	3.784	4.174	4.249	3.932	4.249	3.871	4.043	3.912	3.784	3.778	0.388
2010	4.094	4.454	4.382	4.190	2.079	4.007	2.197	4.277	-	3.219	2.303	2.833	3.458	0.894
2011	0.693	3.497	3.761	2.197	3.912	3.178	3.178	4.025	3.664	3.829	3.664	3.738	3.278	0.902
2012	3.555	3.807	3.932	4.094	3.951	3.689	3.761	3.829	3.664	3.555	3.970	4.060	3.822	0.034
2013	2.485	4.382	4.394	4.263	4.190	4.382	4.043	4.094	4.190	4.078	4.143	2.565	3.934	0.448
2014	4.277	4.431	4.489	4.543	4.533	4.410	4.511	4.466	4.533	4.060	4.234	4.564	4.428	0.024
2015	4.522	4.317	4.410	4.564	4.533	4.205	4.543	4.719	4.410	4.615	4.533	-	4.505	0.017
2016	4.543	4.836	4.727	4.820	4.522	-	4.762	4.410	-	4.564	4.727	4.635	4.663	0.443
2017	4.543	4.875	5.118	5.153	3.738	5.088	5.043	4.898	5.075	5.011	4.796	4.431	4.814	0.167
2018	-	5.004	4.745	5.024	5.004	4.779	5.075	5.136	5.298	5.011	4.942	5.063	5.007	0.027
$\bar{X}_j$	3.568	4.359	4.199	4.263	4.064	4.231	4.105	4.419	4.349	4.198	4.122	3.964		
$\sigma_j^2$	1.589	0.235	0.779	0.709	0.627	0.623	0.807	0.167	0.311	0.355	0.585	0.652		

**Appendix C: Buys-Ballot table for the actual data on number of Birthrate without missing observation (2009-2018)**

Year	Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sept.	Oct.	Nov.	Dec.	$\bar{X}_i$	$\sigma_i^2$
2009	30	54	7	44	65	70	51	70	48	57	50	44	49.17	309.79
2010	60	86	80	66	8	55	9	72	81	25	10	17	47.42	972.81
2011	2	33	43	9	50	24	24	56	39	46	39	42	33.92	266.27
2012	35	45	51	60	52	40	43	46	39	35	53	58	46.42	71.36
2013	12	80	81	71	66	80	57	60	66	59	63	13	59.00	541.27
2014	72	84	89	94	93	90	91	87	93	58	69	96	84.67	142.24
2015	92	75	90	96	93	67	94	112	90	101	93	94	91.42	129.54

2016	94	126	113	124	92	11	117	90	112	96	113	103	99.25	923.84
2017	94	131	167	173	42	162	155	134	160	150	121	84	131.10	1582.40
2018	115	149	115	152	149	119	160	170	200	150	140	158	148.08	594.27
$\bar{X}_{.j}$	60.6	86.3	83.6	88.9	71.0	71.8	80.1	89.7	92.8	77.7	75.1	70.9		
$\sigma_{.j}^2$	1521.6	1474.2	1941.6	2493.7	1462.9	2007.1	2736.3	1504.9	2782.0	2011.6	1657.7	1992.8		

**Appendix D: Buys-Ballot table for the transformed data on number of Birthrate without missing observation (2009-2018)**

Year	Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sept.	Oct.	Nov.	Dec.	$\bar{X}_i$	$\sigma_i^2$
2009	3.401	3.989	1.946	3.784	4.174	4.249	3.932	4.249	3.871	4.043	3.912	3.784	3.778	0.388
2010	4.094	4.454	4.382	4.190	2.079	4.007	2.197	4.277	4.394	3.219	2.303	2.833	3.458	0.894
2011	0.693	3.497	3.761	2.197	3.912	3.178	3.178	4.025	3.664	3.829	3.664	3.738	3.278	0.902
2012	3.555	3.807	3.932	4.094	3.951	3.689	3.761	3.829	3.664	3.555	3.970	4.060	3.822	0.034
2013	2.485	4.382	4.394	4.263	4.190	4.382	4.043	4.094	4.190	4.078	4.143	2.565	3.934	0.448
2014	4.277	4.431	4.489	4.543	4.533	4.410	4.511	4.466	4.533	4.060	4.234	4.564	4.428	0.024
2015	4.522	4.317	4.410	4.564	4.533	4.205	4.543	4.719	4.410	4.615	4.533	4.543	4.505	0.017
2016	4.543	4.836	4.727	4.820	4.522	2.398	4.762	4.410	4.719	4.564	4.727	4.635	4.663	0.443
2017	4.543	4.875	5.118	5.153	3.738	5.088	5.043	4.898	5.075	5.011	4.796	4.431	4.814	0.167
2018	4.745	5.004	4.745	5.024	5.004	4.779	5.075	5.136	5.298	5.011	4.942	5.063	5.007	0.027
$\bar{X}_{.j}$	3.568	4.359	4.199	4.263	4.064	4.231	4.105	4.419	4.349	4.198	4.122	3.964		
$\sigma_{.j}^2$	1.589	0.235	0.779	0.709	0.627	0.623	0.807	0.167	0.311	0.355	0.585	0.652		