BENIN JOURNAL OF STATISTICS
ISSN 2682-5767
Vol. 4, pp. 8– 17 (2021)

In honour of Prof. Sunday Martins Ogbonmwan at retirement

Density Estimation using a Modified Multivariate Cluster Sampling Kernel Approach

E. M. Ogbeide[1*] and J. E. Osemwenkhae[2]

[1]*Department of Mathematics and Statistics, Ambrose Alli University, Ekpoma, Edo State, Nigeria*; [2]*Department of Statistics, University of Benin, Nigeria*

(*Received: 17 September 2020; accepted: 13 January 2021*)

Abstract. This work is based on the use of the modified multivariate cluster sampling kernel density estimation (MMCKDE). The method is demonstrated using the Nigerian crime rate data reported to the Police as contained in the publication of the National Bureau of Statistics in 2009. The method is data-driven. The quality of the estimates from the MMCKDE method showed some significant improvements over the fixed H smoothing and the multivariate cluster sampling kernel density estimation methods in terms of the variance and the asymptotic mean squared error.

Keywords: density function, kernel density estimation, multivariate cluster sampling.

Published by: Department of Statistics, University of Benin, Nigeria

1. Introduction

Data density estimation provides estimates of the probability function from which a set of data is drawn. Density is better estimated from the data. In density estimation, the true density is unknown. One of the popular approaches is the multivariate kernel density estimation. It is the nonparametric estimation approach which requires a kernel function and a bandwidth (window size or smoothing parameter H). Early research (Little and Rubin, 2002; Wu et al., 2007) showed that a data set with missing observations has a density curve with points of discontinuities that may be corrected when the missing data are accounted for in the original data set. This can be done via good imputation methods with very low mean squared errors (Little and Rubin, 2002). The multivariate kernel density estimator herein is a direct extension of the univariate

*Corresponding author. Email: ogbeideoutreach@gmail.com

estimator.

Let $X_1, X_2, \cdots, X_n$ denote a d-variate random sample having a density f. We shall use the notation $X_i = (X_{i1}, X_{i2}, \cdots, X_{in})^T$ and $x \in \mathrm{R}^d$ as a generic vector with $x = (x_1, x_2, \cdots, x_d)^T$. The kernel estimator of the d-variate random sample $X_1, X_2, \cdots, X_n$ drawn from f evaluated at x is given by

$$\hat{f}(X, H) = \frac{1}{n} \sum_{i=1}^{n} K_H(x - X_i), \tag{1}$$

where n is the sample size, H is a symmetric positive definite $d \times d$ matrix called the matrix of window widths or smoothing parameters or the bandwidth matrix and $K_H(x) = |H|^{-\frac{1}{2}} K\left(H^{-\frac{1}{2}} x\right)$, $|\bullet|$ stands for the determinant of H and K_H is the d-variate kernel indexed by H satisfying $\int K_H(x)dx = 1$. The integral is over R^d unless stated otherwise. The matrix H specifies the 'width' of the kernel around each sample point X_i. For the Gaussian kernel, which has gained popularity in the literature, $K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right)$ (Bowman and Azzalini, 1997; Kathovnik and Shmulevich, 2002; Yang et al. 2020).

In this work, the modified multivariate cluster sampling kernel density estimation is adopted. The method is a data-driven approach that requires only the knowledge of the use of pilot plots and the bandwidth sizes from the data set. This is relevant to lowering the asymptotic mean integrated squared error (AMISE) and ensuring faster rates of convergence. The study is aimed at fitting density to multivariate data sets. This is illustrated using the Nigerian crime rate data as reported by the National Bureau of Statistics from 2002–2006. The data set contains some missing observations. The missing observations are imputed using existing methods in the literature – the available data case (AV) and mode-related expectation maximization (MEAM) (see Ogbeide, 2018).

There are a number of methods of estimating bandwidths in the multivariate kernel density. Some of these methods use a fixed window width, while a few others are adaptive, that is they use varied window widths in the course of density estimation (Ogbeide et al., 2012; Wang et al., 2018; Tang et al., 2020). For a review of the available variable methods, see Silverman (1986), Doung and Hazelton (2005), Zhang and Chan (2011) and Jornet et al. (2020). More very data sensitive techniques have been developed in the literature (Wu and Tsai, 2004; Wu et al., 2007). These include the cluster and the average cluster methods. The bandwidth controls the smoothness of the fitted density curve (Cortes and Sanz, 2020). The true density is unknown. The variable window sizes of the multivariate cluster kernel density estimation (MCKDE) and the intersection of confidence interval (ICI) approaches for estimating densities by Wu et al. (2007) and Kathovnik and Shmulevich (2002), respectively, can be considered to be improvements with a view of developing methods that could be adaptive to the MKDE. In most cases, the above methods could lead to underfitting, an indication that the methods are often less optimal (Bowman and Azzalini, 1997; Dicu and Stanga, 2012; Ogbeide et al., 2016).

When we consider the studies on variable window sizes on the average cluster approach and the intersection of confidence interval (ICI) method as applied to MKDE, one is tasked with how sensitive these methods are, and the errors committed? What are the effects when we extend them to multivariate kernel density? These questions are usually considered when developing new methods. For instance, when constructing methods that are more adaptive (Ogbeide et al., 2017). The approach of MMCKDE as presented in Ogbeide et al. (2016) reduced the points of discontinuities in the graphical densities of the data sets. The present study is concerned with the methods of achieving adaptive multivariate kernel density estimation.

2. Materials and Method

It is well known that it is better to estimate the optimal MISE element-wise (Wand and Jones, 1995; Horova et al., 2008; Ogbeide and Osemwenkhae, 2019; Semeyutin and O'Neill, 2020). The ideal optimal bandwidth selector that is point-wise adaptive is given by

$$H_{AMISE} = agr \min_{h \in H} AMISE(H), \tag{2}$$

where the $agr(\bullet)$ is the point-wise element selection scheme.

This study adopts point-wise adaptive bandwidth procedures in estimating densities. Let H be equivalent to the selection of optimal h_{ij} from $\{H_1, H_2, \cdots, H_n\}$. The modified multivariate cluster sampling kernel density estimation (MMCKDE) improves on the cluster sampling kernel density estimate by adjusting the bandwidths using some idea from the kernel nearest neighbour estimation of the density of a multivariate data set. Its smoothing parameter is an $n \times d$ dimensional matrix obtained from the relevant number of clusters in an information matrix. The Euclidean distance is used to construct the bandwidths.

The modified multivariate cluster sampling kernel density estimation (MMCKDE) algorithm is stated below. Steps 5, 6 and 7 are the main improvements over other existing methods.

Step 1 Start with n clusters, each containing a single observation and an $n \times n$ symmetric matrix of distances $D = \{d_{ij}\}$ from the observations.

Step 2 Search the distance matrix for the nearest pair of clusters. Let the distance between the "nearest" clusters S and T be $d_{ST} = \{n_S n_T\} \sum_{i=1}^{n_S} \sum_{j=1}^{n_T} d_{ij}$ in the case of observation i in the cluster S and observation j in the cluster T, and n_S and n_T are the number of observations in cluster S and cluster

T, respectively. See Gray (1997) for length and distance details on the definition of d_{ij}.

Step 3 Merge (combine) cluster S and T. Label the newly formed cluster (ST). Update the entries in the distance matrix by
(a) deleting the rows and columns corresponding to clusters S and T and
(b) adding a row and a column giving the distances between cluster (ST) and the remaining clusters.

Step 4 Repeat steps 2 and 3 a total of $n - 1$ times so that all observations will be in a single cluster at termination of the algorithm. Record the identity clusters that are merger and the distance levels at which the mergers take place.

Step 5 Let b_{i^*j} be the average distance level of X_i in the dendrogram. Suppose n_i denotes the total number of times that a cluster containing X_i is merged into a larger cluster (that is, total number of mergers that involve X_i), and $\ell_1, \cdots, \ell_{ni}$ are the distance levels at which the n_i mergers take place. Then

$$b_{i^*j} = n_i^{-1} \sum_{i=1}^{n_i} \ell_i.$$

Step 6 Apply inter-quartile range to total number of mergers that involve X_i at the boundary values. That is, take the range of the b_i of the clusters in the nearest neighbour situation, using formed clusters of

$$\{X_i\} = C_{i0} \subset C_{i1} \subset \cdots C_{in_i} = \{X_1, \cdots, X_n\}.$$

Then

$$d_{ST}(opt)_i(H) = \{n_S n_T\} \sum_{i=1}^{n_S} \sum_{j=1}^{n_T} d_{ij}, \tag{3}$$

The above scheme clusters are formed in the nearest nested sequence of clusters. Eventually, b_i will be small if n_i is large (that is, a large number of mergers involve X_i).

Step 7 For the boundary value problem,

$$H = d_{ST}(opt)_i(H) = \frac{d_{ST}(H_i)}{2} \tag{4}$$

and

$$d_{ST}(opt)(H_{i+1}) \leq d_{ST}(opt)(H_i). \tag{5}$$

This result is consistent with the intuitive idea of kernel estimator having to find a compromise between estimating two distinct values of f on either side of discontinuity. Since the location of the boundary of $\hat{f}(x; H)$ is usually known, we adopted this to achieve better performance in its vicinity. So we have H in equation (3) as the bandwidths.

Given a data observation X_{ij}. We have bandwidth factors b_{i^*j} according to the number of clusters form (starting from Step 3 in the proposed algorithm) from the element wise groups from the information data rows. From the bandwidth factors, b_{i^*j}, and the global smoothing parameter, h_i^*, we have

$$H = h_i^* b_{i^*j} \tag{6}$$

where $i = 1, 2, \cdots, n$, $i^* = 1, 2, \cdots, n_i$, $j = 1, 2, \cdots, d$ and H is a finite set of optimal bandwidths $\{H_1, \cdots, H_n\}$ with each $H_i = h_{ij}$. We choose h_i^* via each of the information data rows' h_{MSE}. That is, using the MSE approach to get each h_i^*. Then b_{i^*j} will be small if as n_{i^*} is large (that is, a large number of mergers involving X_i). Thus, the range of the b_{i^*j} of the clusters in the nearest neighbour situation is taken using the formed clusters of

$$\{X_i\} = C_{i0} \subset C_{i1} \subset \cdots C_{in_i} = \{X_1, \cdots, X_n\}.$$

From the data set, the above scheme clusters are formed from the nearest nested sequence of clusters information data rows' elements. This procedure gives a full bandwidth matrix of vary smoothing parameters for possible values of the data sizes for i rows and j columns, $i \leq j$ and $i > j$.

To correct the problem of discontinuities at some points in the cluster sampling approach to MKDE, points of discontinuity in the estimation are identified using the cluster sampling approach as a pilot guide. In this case, the location of the boundary of $\hat{f}(X; H)$ is usually known and this is utilized to achieve better performance in its vicinity. Suppose, we have for S number of row clusters and T number of column clusters, we have

$$H_i = d_{ST} = \sum_{i=1}^{n_S} \sum_{j=1}^{n_T} d_{ij}. \tag{7}$$

To obtain the density estimates, the bandwidth sizes are substituted into equation (1).

3. Results and Discussion

The data for this study are drawn from the Nigeria National Bureau of Statistics Annual Abstract of Statistics published in the year 2009 on reported crime rate to the police from 2002 – 2006 (source: www.nbs.gov.ng). The bandwidth selections, density, error and convergence rate for the multivariate cluster sampling kernel density estimation (MCKDE) and the modified multivariate cluster sampling kernel density estimation (MMCKDE) methods are presented in Tables 1 – 4. Table 1 contains the calculated bandwidth selection errors and convergence rate for the crime rate data.

Table 1: Evaluated estimates of the crime rate data with missing observations

Offence	Case	2002	2003	2004	2005	2006
False Pretence/ Cheating	ADC	7913	9508	–	9580	6395
	MEAM	7913	9508	9544	9580	6395
Unlawful Possession	ADC	3790	4142	5358	8772	8666
Receiving stolen property	ADC	1161	1289	2733	3892	7308
Arson	ADC	2005	1499	1289	1268	1010
Perjury	ADC	17	50	4	3	5
Gambling	ADC	199	148	–	631	473
	MEAM	199	148	390	631	473
Breach of trust	ADC	7055	7298	–	7967	5945
	MEAM	7055	7298	7633	7967	5945
Escape from custody	ADC	220	272	–	99	132
	MEAM	220	272	186	99	132
Local acts	ADC	2885	5171	–	3072	2610
	MEAM	2885	5171	4122	3072	2610
Others	ADC	3262	3322	–	891	914
	MEAM	3262	3322	2107	891	914

Note: ADC – Available Data Case; MEAM – Mode-related Expectation Adaptive Maximization.

The results favour the use of the MEAM approach for missing data imputation as the MEAM has relative lower error propagation and relative faster convergence rate (Ogbeide, 2018).

Table 2: Estimated bandwidths for the crime rate data from 2002–2006

Offence	Case bandwidth	2002	2003	2004	2005	2006	Var
False pretence/ Cheating	Fixed H	25.8500	25.8500	25.8500	25.8500	25.8500	–
	MCKDE	31.900	0.6500	0.7200	60.1000	28.9000	459.1561
	MMCKDE	31.900	16.2750	8.4975	30.4100	28.9000	116.7849
Unlawful possession	Fixed H	27.3920	27.3920	27.3920	27.3920	27.3920	–
	MCKDE	7.0400	24.3200	68.2800	2.1200	324300	559.0200
	MMCKDE	7.0400	15.6900	45.3000	35.2000	17.2750	179.3006
Receiving Stolen property	Fixed H	33.7380	33.7380	33.7380	33.7380	33.7380	–
	MCKDE	2.5600	28.8800	23.1800	68.3200	45.7500	489.4700
	MMCKDE	2.5600	28.8800	23.1800	45.7700	52.5100	310.8864
Arson	Fixed H	4.5380	4.5380	4.5380	4.5380	4.5380	–
	MCKDE	10.1200	3.6700	0.4200	4.8900	2.7900	10.3800
	MMCKDE	5.0600	3.6700	2.0420	4.8900	2.7900	1.3664
Perjury	Fixed H	0.3340	0.3340	0.3340	0.3340	0.3340	–
	MCKDE	0.6600	0.9200	0.0200	0.0400	0.0300	0.14
	MMCKDE	0.6600	0.5200	0.2700	0.0400	0.0300	0.0343
Gambling	Fixed H	2.2622	2.2622	2.2622	2.2622	2.2622	–
	MCKDE	1.02	3.356	4.8485	3.1600	3.0209	1.18
	MMCKDE	1.0200	3.3560	4.3410	3.1600	3.2700	1.0207
Breach of Peace	Fixed H	15.476	15.476	15.476	15.476	15.476	–
	MCKDE	4.8600	2.9890	6.6900	21.2300	23.5650	164.1300
	MMCKDE	4.8600	3.9245	6.6900	21.2300	23.5650	72.6648
Escape from custody	Fixed H	1.2615	1.2615	1.2615	1.2615	1.2615	–
	MCKDE	1.0400	1.5120	1.5010	0.6600	1.1475	0.10
	MMCKDE	1.0400	1.2760	1.5010	1.0805	0.9035	0.0432
Local acts	Fixed H	22.5310	22.5310	22.5310	22.5310	22.5310	–
	MCKDE	45.7200	20.9900	21.3900	9.2400	15.3150	134.0300
	MMCKDE	27.8800	20.9900	21.3900	13.1200	14.2175	28.5666
Others	Fixed H	12.5260	12.5260	12.5260	12.5260	12.5260	–
	MCKDE	3.0110	24.3100	24.3100	9.7600	12.3500	78.3682
	MMCKDE	1.5055	12.1550	24.3100	4.8800	10.9841	62.2800

For the MEAM imputation using the fixed H, the MCKDE and the MMCKDE, it can be inferred that the MMCKDE bandwidths selection approach has a lower variance compared to the MCKDE approach.

Table 3 contains the density estimates of the various bandwidths for the crime rate from 2002–2006. It can be seen from Table 3 that the density sum is closer to unity for the MMCKDE compared to the MCKDE.

Table 3: Density estimates of the various bandwidths for the crime rate from 2002–2006

Offence	Case: Densities	2002	2003	2004	2005	2006	Density sum
False pretence/ Cheating	Fixed H	0.1842	0.2214	0.2231	0.2231	0.1411	0.9929
	MCKDE	0.1901	0.2123	0.2337	0.2231	0.1465	1.0057
	MMCKDE	0.1842	0.2189	0.2231	0.2231	0.1489	0.9982
Unlawful possession	Fixed H	0.1321	0.1352	0.1765	0.2991	0.1566	0.8995
	MCKDE	0.1233	0.1331	0.1944	0.2577	0.282	0.9705
	MMCKDE	0.1321	0.1352	0.1765	0.2991	0.211	0.9939
Receiving stolen property	Fixed H	0.0708	0.0787	0.1668	0.2399	0.3941	0.9503
	MCKDE	0.0811	0.0787	0.1668	0.2376	0.4023	0.9665
	MMCKDE	0.0811	0.0768	0.1881	0.2376	0.4147	0.9983
Arson	Fixed H	0.2514	0.1576	0.1322	0.1311	0.3195	0.9918
	MCKDE	0.2514	0.1713	0.1314	0.1231	0.3171	0.9943
	MMCKDE	0.2514	0.1713	0.1322	0.1244	0.3192	0.9985
Perjury	Fixed H	0.2151	0.6101	0.0503	0.0366	0.0633	0.9754
	MCKDE	0.2198	0.6088	0.0506	0.0379	0.0633	0.9804
	MMCKDE	0.2198	0.6198	0.0531	0.0399	0.0633	0.9959
Gambling	Fixed H	0.1081	0.0799	0.2001	0.3392	0.257	0.9843
	MCKDE	0.1083	0.0795	0.2116	0.3388	0.257	0.9952
	MMCKDE	0.1083	0.0801	0.2116	0.3428	0.257	0.9998
Breach of Peace	Fixed H	0.1965	0.2033	0.2102	0.2219	0.1656	0.9975
	MCKDE	0.1968	0.2001	0.2102	0.2219	0.1656	0.9946
	MMCKDE	0.1968	0.2033	0.2122	0.2219	0.1656	0.9998
Escape from custody	Fixed H	0.2422	0.2887	0.2041	0.1089	0.1453	0.9892
	MCKDE	0.2476	0.2887	0.2041	0.1089	0.1453	0.9946
	MMCKDE	0.2476	0.2939	0.2041	0.1089	0.1453	0.9998
Local acts	Fixed H	0.1615	0.2833	0.2289	0.1701	0.1461	0.9899
	MCKDE	0.1615	0.2833	0.2298	0.1701	0.1461	0.9908
	MMCKDE	0.1615	0.2882	0.2308	0.172	0.1461	0.9986
Others	Fixed H	0.3108	0.2998	0.2007	0.0823	0.0871	0.9807
	MCKDE	0.3108	0.2998	0.2007	0.0849	0.0871	0.9833
	MMCKDE	0.3108	0.3123	0.2007	0.0849	0.0871	0.9958

The AMISE* and the convergence rates of methods are given in Table 4. From Table 4, it can be concluded that there the relative errors, h^* (which is the error in relation to the fixed optimal bandwidth value) and the AMISE* of the proposed method (i.e., MMCKDE) are smaller. More so, the proposed method has faster convergence rates compared to its classical version. Simply put, the MMCKDE has lower error propagation and faster convergence rates when used to estimate the data vis-á-vis the fixed optimal H and MCKDE approaches.

The estimated bandwidth selection errors and the convergence rates for the crime rate data with missing observations using various methods favour the use of the MMCKDE approach over the fixed H and MCKDE approaches. This is because MMCKDE's bandwidth errors are smaller and the scheme has a higher convergence rate. The MMCKDE has some improvements over the MCKDE approach as shown in Table 3 and Table 4. Apart from the fact that the MM-CKDE is data sensitive, it also provides full bandwidths matrix for the multivariate kernel density estimation. As in other improved methods, the MMCKDE

Table 4: Bandwidth selection errors and convergence rate for the estimated bandwidths

Issue	Approach	Mean	Convergence Rate	Variance	δ	h^*	AMISE*
False pretence/ Cheating	MCKDE	24.4540	1.0907	459.1561	21.4279	10.3869	2.1828×10^{-1}
	MMCKDE	17.4540	1.4255	116.7849	10.8067	5.2384	1.2624×10^{-1}
Unlawful possession	MCKDE	27.3920	1.0000	559.0200	23.6436	11.4609	2.3616×10^{-1}
	MMCKDE	24.3020	1.1991	179.3006	13.3903	6.4907	1.4985×10^{-1}
Receiving stolen Property	MCKDE	33.1460	1.1215	489.4700	22.1239	10.7242	2.2394×10^{-1}
	MMCKDE	30.5800	1.1625	310.8864	17.6319	8.5468	1.8676×10^{-1}
Arson	MCKDE	4.3780	1.0285	10.3800	3.2218	1.5617	4.7943×10^{-2}
	MMCKDE	3.4910	1.3502	1.3664	1.1689	0.5661	2.1304×10^{-2}
Perjury	MCKDE	0.3340	1.0000	0.1400	0.3741	0.1813	8.5632×10^{-3}
	MMCKDE	0.2380	1.5881	0.0343	0.1852	0.0897	4.8793×10^{-3}
Gambling	MCKDE	3.0290	0.5795	1.1800	1.0862	0.5265	2.0089×10^{-2}
	MMCKDE	3.0209	0.5829	1.0419	1.0207	0.4947	1.9114×10^{-2}
Breach of Peace	MCKDE	14.7749	1.1058	164.1300	12.8113	6.2101	1.4465×10^{-1}
	MMCKDE	12.0539	1.4273	72.6648	8.5243	4.1320	1.0441×10^{-1}
Escape from custody	MCKDE	1.1721	1.1206	0.1000	0.3162	0.1532	7.4854×10^{-3}
	MMCKDE	1.1602	1.1379	0.0432	0.2078	0.1007	5.5851×10^{-3}
Local acts	MCKDE	22.5310	1.0000	134.0300	11.5771	5.6118	1.3339×10^{-1}
	MMCKDE	18.6925	1.3152	28.3639	5.3447	2.5907	7.1876×10^{-2}
Others	MCKDE	14.7482	0.7516	78.3682	8.8525	4.2911	1.0762×10^{-1}
	MMCKDE	10.7482	1.2189	62.2800	7.8919	3.8254	9.8171×10^{-2}

scheme requires only simple but two additional steps when compared to the MCKDE approach. These additional procedures are in the choice and application of the smoothing parameters to multivariate density estimation. These steps introduce the adaptive density.

4. Conclusion

This work proposes the use of the MMCKDE approach for kernel density estimation. This approach is based on the data at hand. The study demonstrated the usefulness of the MMCKDE approach with the Nigerian crime rate data reported to the police. The quality of the obtained estimates of the MMCKDE approach showed some improvements over the fixed H and MCKDE approaches in terms of the variance, the asymptotic mean squared error and the rate of convergence. The fixed H approach is limited by both over fitting and under fitting as the case may be. The MMCKDE is able to circumvent points of discontinuities and displays adaptive density.

Acknowledgement

The authors are grateful to the anonymous reviewers for their constructive criticisms and suggestions which significantly improve the paper.

References

Bowman, A.W. and Azzalini, A. (1997). Applied Smoothing Techniques for Data Analysis. Clarendon Press, Oxford.

Cortes, J-C. L and Sanz, J. M. (2020). Improving Hernel methods for density estimation in random differential equations problems. Mathematical and Computational Application, **25**(2): 33.

Dicu, I. and Stanga, I.C. (2013). Exposure and triggering factors of roads (UN) safety and risk in AISI municipality (Romania). Analele Stintifice Ale 'Alexandru IOAN CUZA', **LIX**(1), 171-190.

Duong, T. and Hazelton, M.L.(2005). Convergence rates for unconditional bandwidth matrix selector for multivariate kernel density estimation. J. Multivariate Anal., **93**: 417-433.

Gray, A. (1997). The intuitive idea of distance on surfaces in "Modern differential geometry of curves and surfaces with Mathematica" (2nd ed.). Boca Ration, FL, CRC press.

Horova, I., Kolacek, J., Zelinka, J. and Vopatova, K. (2008). Bandwidth choice for kernel density estimates. IASC, Yokohama, Japan.

Jornet, M., Calatayud, J., Le Moitre, O.P. and Cortes, J-C. (2020). Second order linear differential equations with analytic uncertainties, stochastic analysis via the computation of the probability density function. J. Computation, Apppli. Math., **374**, 112770.

Katkovnik, V. (1999). A new method for varying adaptive band width selection. IEEE Transaction on Signal process, **47**(9): 2567-2571.

Katkovnik, V. and Shmulevich, I. (2002). Kernel density estimation with adaptive varying window size. Pattern Recognition Letters, **23**(14): 1641-1076.

Little R.J.A. and Rubin, D.B. (2002). Statistical analysis with missing data (2nd ed.). Wiley & Sons Publisher, New Jersey, USA.

National Bureau of Statistics (2009). Annual Abstract of Statistics. Federal Republic of Nigeria, Abuja.

Ogbeide, E. M. (2018). A new iterative imputation method based on adaptive Expectation Maximization. Science Tech. Journal, **3**(1): 133-142.

Ogbeide, E.M., Osemwenkhae, J. E. and Oyegue F.O. (2011). A new adaptive kernel density estimation bandwidth approach. Journal of Engineering Science and Application, **7**(3): 14-25.

Ogbeide, E.M., Osemwenkhae, J. E. and Oyegue, F. O. (2016). On a modified multivariate cluster sampling kernel approach to multivariate density estimation. Journal of Nigerian Association of Mathematical Physics, **34**: 123-132.

Ogbeide, E.M., Osemwenkhae, J. E. and Oyegue F.O. (2017). A modified intersection of confidence intervals (MICIH) approach in the multivariate data density estimation. Journal of Nigeria Statistical Association, **29**: 47-57.

Ogbeide, E. M. and Osemwenkhae, E. J. (2019). Some improvements in nonparametric multivariate kernel density estimation. Athens Journal of Science, **6**(4): 287-308.

Semeyutin, A. and O'Neill, R. (2020). A brief survey on the choice of parameters for kernel density estimation for time series. North American journal of Economics and Finance, **50**, 101038.

Silverman, B.W. (1986). Density Estimation for Statistics and Data Analysis. Chapman and Hall, London.

Tang, K. Wan, X. and Liao, Q. (2020). Deep density estimation via invertible block-triangular mapping. Theor. Application Mech. Lett., **10**: 143-148.

Wand, M. P. and Jones, M. C. (1995). Kernel Smoothing. Chapman & Hall, London.

Wang, X. and Saghafi, A. (2018). Improved parameter estimation of time dependent kernel density estimation using Artificial Neural network. Journal of Finance and Data science, **4**(3): 172-182

Wu, T.J., Chen, C.F and Chen, H.Y. (2007). A variable bandwidths selectors in multivariate kernel density estimation. Stat. and Prob. Letters, **77**: 462-467.

Wu, T.J. and Tsai, M.H. (2004). Root n bandwidths selectors in multivariate kernel density estimation. Probab. Theory Related Fields, **129**: 537-558.

Yang, N., Huang, Y., Hou, D,. Liu, S., Ye, D., Dong, B. and Fan, Y. (2019). Adaptive nonparametric kernel density estimation approach for joint probability density function modelling of multiple wind farms. Enerigies, **12**, 1346.

Zhang, Z. G. and Chan, S.C. (2011). On kernel selection of multivariate local polynomial modelling and its application to image smoothing and Reconstruction. Journal of Signal Process System, **64**: 361-374.