# Nonparametric Kernel Density Estimation with Automatic Data-driven Bandwidth Selection Method: An Application to National Panel Survey Data Wave 4 in Tanzania

**Peter Aron Kanyelele**[1]

## Abstract

*Bandwidth selection is of great importance when statisticians want to estimate the functional forms of a data set by using nonparametric method. Bandwidth can be selected by several methods, but least squares cross-validation method may provide optimal bandwidth. This study used nonparametric kernel density estimation approach to estimate the function form of consumption of the household and computing an optimal bandwidth parameter, a bias-variance trade –off by using the least squares cross-validation data –driven automatic selection method. The study used National Panel Survey data wave 4 (2014/2015) in Tanzania with 3,352 households based on stratified, multi-stage cluster sample design to estimate the function form of the households' food shares. The results revealed that, optimal bandwidth for Gaussian and Epanechnikov kernel functions;-* $h_{optimal\ LSCV} = 0.087$, *factor* $= 0.935$ *for male/ female headed households,* $h_{optimal\ LSCV} = 0.609$, *scale factor* $= 0.708$ *for urban/rural households,* $h_{optimal\ LSCV} = 0.454$, *factor* $= 0.668$ *for households' sizes and* $h_{optimal\ LSCV} = 0.976$, *factor* $= 0.604$ *for adult equivalent households. MISE and ISE used as kernel functions criteria, depicted low value resulting to consistency and efficiency of* $h_{LSCV}$. *The use of nonparametric kernel density estimation method and Gaussian kernel functions, or Epanechnikov kernel function is recommended because it has fewer assumptions. Therefore, the findings are of greater value for mathematicians and statisticians.*

**Keywords:** *Non-parametric Kernel Density, LSCV-automatic data-driven bandwidth selection method, Gaussian Kernel function, Epanechnikov Kernel function, MISE, ISE.*

## 1.0 Introduction

Non-parametric density estimation (also known as non-parametric smoothing) is a statistical technique that does not require prior assumptions about the functional form of the model being estimated. Instead, the method allows the data to determine the structure of the model. Non-parametric density estimation is useful for assessing multimodality, skewness, and other distributional characteristics (Silverman, 1986). It is also applied in Bayesian posterior summarization, classification, and discriminant analysis (Simonoff, 1996). Additionally, it has proven valuable in Monte Carlo computational methods, such as bootstrap smoothing and particle filtering (Doucet *et al.,* 2001). The appeal of non-

1 Assistant Lecturer, Tanzania Public Service College- Dar es Salaam Campus peter.kanyelele@tpsc.go.tz

parametric methods lies in their flexibility, as they are free from the parametric constraints typically imposed on data-generating processes.

Over the past few decades, non-parametric techniques have gained significant attention from statisticians and mathematicians (Hardle *et al.,* 2005). However, despite the extensive literature on the subject, several challenges remain in the implementation and performance of kernel density estimators. First, the most commonly used data-driven bandwidth selection techniques, including the plug-in method (Sheather *et al.,* 1991; Jones *et al.,* 1996), are influenced by the normal reference rule. While plug-in estimators perform well under approximate normality, reliance on the normal reference rule contradicts the original motivation for using non-parametric methods. Second, the widely used Gaussian kernel density estimator lacks local adaptability, making it highly sensitive to outliers, prone to producing spurious bumps, and susceptible to bias—often flattening peaks and valleys in the density estimation (Marron *et al.,* 1992). Third, most kernel estimators suffer from boundary bias, particularly when data are nonnegative, as traditional kernels do not account for domain-specific knowledge (Park *et al.,* 2003). These issues have been mitigated to some extent through the development of more advanced kernel density estimation techniques. The first kernel density estimation method, introduced by Rosenblatt (1956), aimed to relax the parametric assumptions of discriminant analysis.

Two key developments have contributed to the widespread adoption of non-parametric methods in economics, econometrics, statistics, and mathematics. First, advances in computing power have made these methods feasible for practical applications. Without efficient computational resources and optimized algorithms, non-parametric estimation would be impractical. Second, the availability of statistical software packages, such as the "np" package in R (Hayfield & Racine, 2008), has further facilitated their use. The combination of powerful computing and accessible software has significantly popularized non-parametric methods in econometrics and economics.

In this study, we focus on non-parametric kernel density estimation methods that are robust to non-normal distributions, contaminated data, outliers, and leverage points. The primary objective is to estimate the density function of food shares among households in Tanzania using kernel density estimation with an automatic data-driven bandwidth selection method. Specifically, the study aims to estimate the density function of food shares among Tanzanian households and to examine the influence of gender, location, and household size on food shares in Tanzania.

The performance of the kernel density estimator (KDE) will be evaluated using the following criteria: Mean Square Error (MSE), Weighted Mean Absolute Error (MAE), Standard Errors (SE), Coefficients of Determination (R-squared), Mean Absolute Percentage Error (MAPE), Integrated Square Errors (ISE), Mean Integrated Square Errors (MISE), and Integrated Mean Square Errors (IMSE).

In recent years, the literature on non-parametric density estimation methods has grown rapidly, offering solutions to problems associated with parametric regression models. Unlike parametric models, non-parametric density estimation techniques do not require researchers to assume a specific functional form between the dependent and explanatory variables. Instead, the data determine the functional form, eliminating arbitrary model constraints. However, a major challenge in non-parametric density estimation is selecting

an optimal bandwidth that balances bias and variance while minimizing the integrated squared errors (ISE) or mean integrated squared errors (MISE).

Various bandwidth selection methods have been proposed, including "quick and dirty" approaches such as the rule of thumb (Deheuvels, 1977), maximal smoothing (Terrell, 1990), biased cross-validation (Scott & Terrell, 1985), smoothed cross-validation (Hall et al., 1992), Sheather and Jones' (1991) factorized smoothed cross-validation, one-sided cross-validation, modified cross-validation, and the bootstrap bandwidth method (Taylor, 1989). However, selecting an optimal bandwidth remains a critical challenge in obtaining accurate density estimates (Harpole *et al.,* 2014).

This paper employs a non-parametric kernel density estimation method combined with the least squares cross-validation data-driven bandwidth selection technique to analyze the **NPS4 (2014/2015)** dataset. The study aims to estimate the functional form of household food consumption, an area that has not been extensively explored. Household food consumption is analyzed in relation to factors such as male- vs. female-headed households, urban vs. rural households, household size, and adult-equivalent household measures. Both the Gaussian and Epanechnikov kernel functions are used to estimate the density functions of food shares.

## 2.0 Literature Review

The idea of bandwidth (smoothing parameter) methods has been discussed by many scholars; including Chen (2018) pointed out, that the problem of smoothing parameter techniques depends on the situation in which the data are serially dependent on time series and proposed localized bandwidth estimators. Gramacki (2018) discovered a new theorem deriving the asymptotic theory for linear combinations of smoothing parameters obtained from different areas. Heidenreich *et al.,* (2013) reviewed different methods of smoothing parameters (bandwidths) techniques and compared the methods by simulation. They found out that simple plug-in and cross-validation methods produce bandwidths with quite unstable performance. Jones *et al.* (1996) proposed a plug-in-method, this method depicted undesirable weakness, it is not a fully automatic method, because one needs to choose an initial value of bandwidth (h) to estimate the function. (Silverman, 1986), proposed rule- of-thumb. This method is inappropriate as it over smooth data. Extreme over-smoothing leads to an unimodal estimate which completely obscures the true density nature of the underlying distribution. Due to the lack of stability of these methods, different bandwidth techniques were introduced, among them is classical cross-validation, a fully automatic data-driven method of selecting the smoothing parameter (Rudemo, 1982; Stone, 1984; Bowman, 1984). Modified rule-of-thumb is probably the most popular and biased cross- validation method to select the smoothing parameter (Scott & Terrell, 1987; Park *et al*., 2003; Sheather & Jones, 1991; Hall *et al.,* 1992). Also, the bootstrap methods of Taylor (1989) as well as all its modifications by Cai *et al.* (2009).

Based on some literature reviews as mentioned above, it has been observed that there is a vast and rapid increase in the studies that are using nonparametric density estimation methods in econometric regression methods and semi-parametric regression methods for data analysis; such as Lin and Carrol (2000), Wilcox (2004), Ullah *et al.* (2005), Li and

Racine (2004), Henderson *et al*. (2006), but which an appropriate approaches, an optimal smoothing parameter and density estimation function most suitable to analyses and estimate the households consumption function form using least squares cross-validation method, is still an open question. To answer such a question, we adopted nonparametric kernel density estimation, LSCV data-driven automatic method, this method is based on the principles of selecting a bandwidth that minimizes the integrated squared error of the resulting estimate, and it provides an optimal bandwidth. Gaussian kernel function or Epanechnikov kernel function to estimate consumption functional form of food shares of households in Tanzania using National Panel Survey wave four (NPS4).

## 3.0 Methodology

This study focused on two nonparametric kernel density functions: The Gaussian Kernel and the Epanechnikov Kernel, with least squares cross-validation (LSCV) used as an automatic, data-driven method for bandwidth selection. The performance of these kernel density estimation functions was evaluated using mean integrated square error (MISE) and integrated square error (ISE), as well as coefficients of determination ($R^2$). Additionally, the statistical significance levels of explanatory variables were determined using bootstrapping, following the methods proposed by Racine (1997).

Given a dataset of independent and identically distributed samples from an unknown univariate distribution with density function, the kernel density estimator provided an estimate of the probability distribution by averaging over a localized region determined by a bandwidth parameter. The kernel function must satisfy key properties, including normalization, symmetry, and finite second moments. The least squares cross-validation (LSCV) method, was used to minimize the integrated square error (ISE). The method optimizes the bandwidth selection process, ensuring that the kernel density estimator provides an accurate representation of the underlying distribution. Minimizing the LSCV function helped achieve the lowest mean integrated square error (MISE), leading to an optimal bandwidth choice. This study was conducted in Tanzania, chosen for its diverse analytical units (households). The National Panel Survey Wave 4 (NPS4, 2014-2015) explicitly defines three analytical strata: Dar es Salaam, urban and rural areas in Mainland Tanzania, and urban and rural areas in Zanzibar. Within each stratum, clusters were randomly selected as primary sampling units, with selection probability proportional to population size. In urban areas, clusters corresponded to census enumeration areas, while in rural areas, clusters were equivalent to villages.

The NPS4 (2014-2015) dataset was used, consisting of 3,352 households and 419 clusters. The survey employed a stratified, multi-stage cluster sampling design, drawing from the 2002 Population and Housing Census (PHC). The sample design maintained the three analytical strata, ensuring proportional representation across Tanzania.

This study relied on secondary data from NPS4 (2014-2015), a national longitudinal survey conducted by the National Bureau of Statistics (NBS) in collaboration with the Office of the Chief Government Statistician – Zanzibar. Data collection spanned 13 months, from October 2014 to November 2015. The NPS aims to track national and international development progress, analyze poverty dynamics, and evaluate policy impacts.

## 4.0 Data Analysis, Presentation, and Discussion of the Findings

This part presents the study's results with their respective discussions based on the objective of the study which was first, to estimate the density functional form of food shares of households in Tanzania, *second,* to find out the influence of gender, location, and households size of food consumption in Tanzania. This section presents data estimation and analysis, sample size, estimation type, kernel density estimation, and bandwidth selection. We estimated the total household consumption function form of food shares by using the nonparametric kernel density estimation method. Furthermore, we adopted Gaussian kernel function and Epanechnikov kernel function to achieve study objectives. We make the frequency used assumptions that the bandwidths (smoothing parameters) for the variables can differ between variables, but are constant over the domain of each explanatory variable. Then we used LSCV as a data-driven automatic bandwidths selection method which trade-off between bias and variance.

### 4.1 Presentation of the Results

All estimations, calculations, and presentation of results were conducted and written within the statistical software environment "R" and "np" packages. The households' consumption data contains (3,344 households' consumptions of food shares replications, and 8 variables)

**Table 1: Least Squares Cross-Validation data-driven bandwidths summary of food shares in Tanzania.**

| Variables | Gaussian kernel function | | Epanechnikov kernel function | |
|---|---|---|---|---|
| | Bandwidth | Scale factor | Bandwidth | Scale factor |
| Region | 0.763 | 0.534 | 0.763 | 0.534 |
| District | 0.988 | 0.571 | 0.988 | 0.571 |
| Ward | 0.445 | 0.045 | 0.442 | 0.044 |
| Household size | 0.454 | 0.668 | 0.454 | 0.668 |
| Male/Female headed | 0.087 | 0.935 | 0.087 | 0.935 |
| Urban or Rural | 0.609 | 0.708 | 0.609 | 0.708 |
| Adult Equivalent | 0.967 | 0.605 | 0.968 | 0.605 |
| Year | 0.133 | 0.405 | 0.132 | 0.404 |

**Source:** Field data (2024)

**Table 2: Criteria of nonparametric kernel density estimation function of food shares in Tanzania**

| Gaussian kernel Function | | Epanechnikov kernel Function |
|---|---|---|
| MSE | 0.003 | 0.002 |
| SE | 0.053 | 0.059 |
| R-squared | 0.888 | 0.854 |
| MAE | 0.024 | 0.024 |
| MAPE | 0.039 | 0.039 |
| ISE | 0.003 | 0.004 |
| MISE | 0.004 | 0.005 |

**Source:** Field data (2024)

**Table 3: Individual Significance Tests of variables of food shares in Tanzania**

| Gaussian kernel Function | | Epanechnikov kernel Function |
|---|---|---|
| **Variable** | **P-Value** | **P-Value** |
| Region | 0.000 | 0.000 |
| District | 0.000 | 0.000 |
| Ward | 0.000 | 0.000 |
| Household size | 0.852 | 0.852 |
| Male/Female headed | 0.311 | 0.311 |
| Urban or Rural | 0.000 | 0.000 |
| Adult Equivalent | 0.995 | 0.995 |
| Year | 0.000 | 0.000 |

**Source:** Field data (2024)

Table 3 shows significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1. The above results depict that the region, district, ward, household weights, education, urb an/rural, and year (time) variables are statistically significant (at 0 % significant level), fe male-headed households, household size, and adult equivalent variables are not statisticall y significant for both Gaussian kernel and Epanechnikov kernel function. According to th e bootstrap significance test proposed by Racine (1997), and Hart and Li (2006), Urban/ru ral and female-headed household variables have a greater effect on the dependent variable (Food shares). Households size and adult variables have less effect on the dependent varia ble.

### 4.2 Estimated functional form of food shares in Tanzania

The figures below are the estimated function form of food shares in Tanzania using nonpa rametric kernel density estimation for Gaussian kernel function and Epanechnikov kernel density function together with LSCV driven-data automatic bandwidth selection method p resent in computer package R together with "np" package.

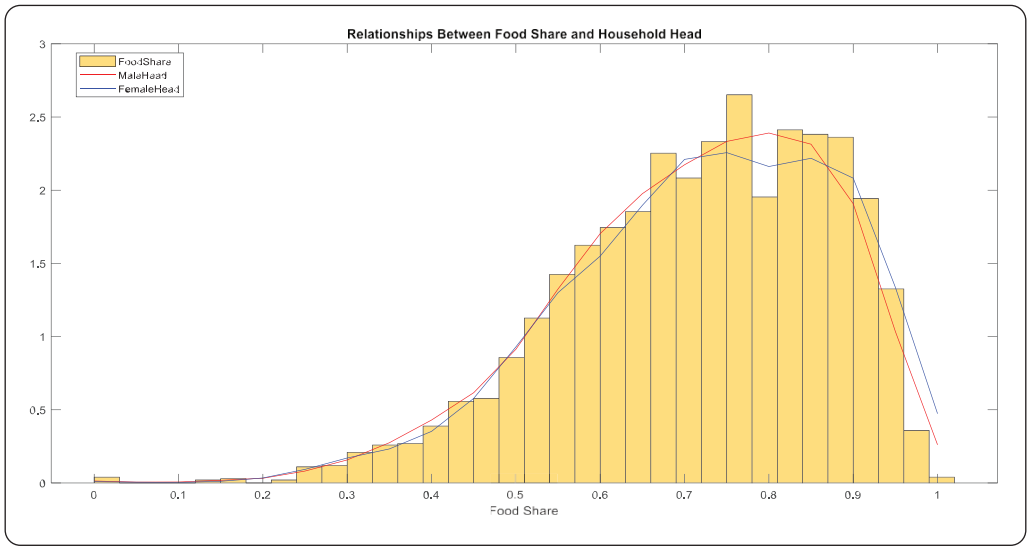


**Figure 1: Nonparametric estimation of food shares for male/female-headed households by using Gaussian kernel density function and Epanechnikov kernel density function**

**Source:** Field Data (2024)

Figure 1 shows the result of nonparametric kernel density estimation (an automatic data-driven bandwidth-the least-squares cross-validation data-driven bandwidth (LSCV (h)) = 0.087, scale factor 0.935 and Gaussian kernel density and Epanechnikov kernel density, estimated via 3344 replications). The function form depicts that food shares for male-headed households are higher than for female-headed households.

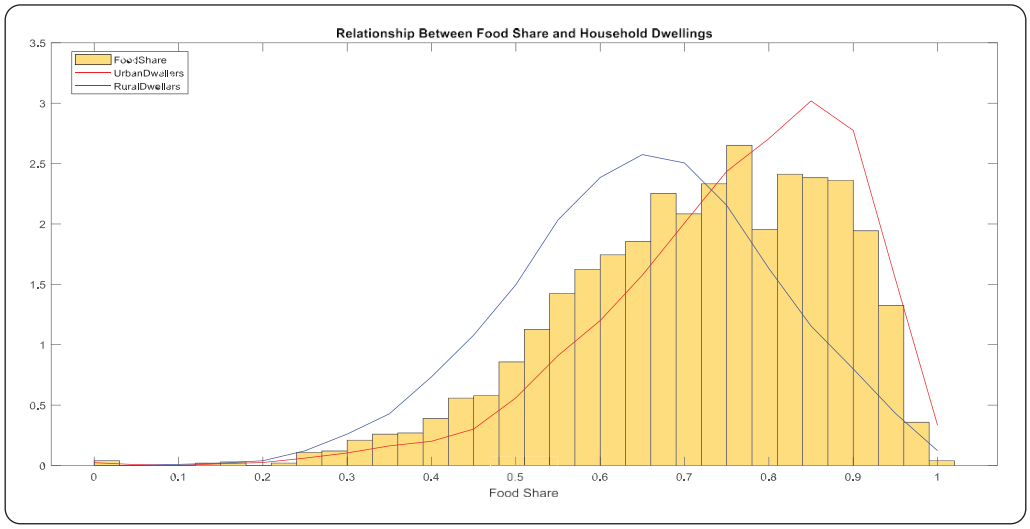


**Figure 2: Nonparametric estimation of food shares for urban/rural households by using Gaussian kernel Density function and Epanechnikov kernel density function.**

**Source:** Field Data (2024)

Figure 2 shows the results of nonparametric kernel density estimation (with a data-driven bandwidth-the least-squares cross-validation bandwidth (LSCV (h)) = 0.609 and Scale factor = 0.708, estimated via 3344 replications). The function form depicts that food shares for urban households are higher than for rural households and also, as the urban/rural dwellers decrease, the food shares increase.

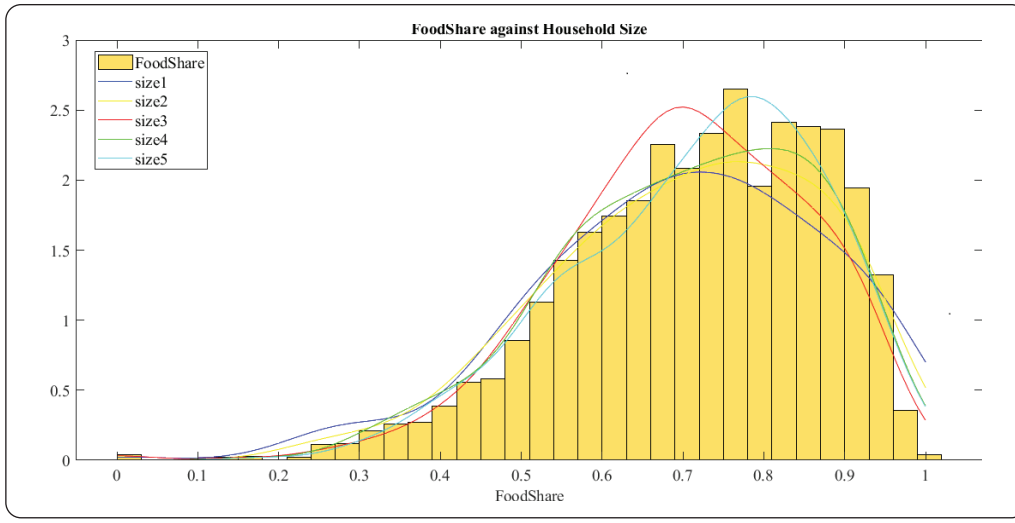


**Figure 3: Nonparametric estimation of food shares and households size dwellings by using Gaussian kernel density function and Epanechnikov kernel density function.**

**Source:** Field Data (2024)

Figure 3 shows the results of nonparametric kernel density estimation (with a data-driven bandwidth-the least-squares cross-validation bandwidth (LSCV (h)) = 0.454 and Scale Factor = 0.668, estimated via 3344 replications). The function form depicts that food shares for household size increases as the number of dwellers decreases and also, as the household size decreases, the food shares increases.

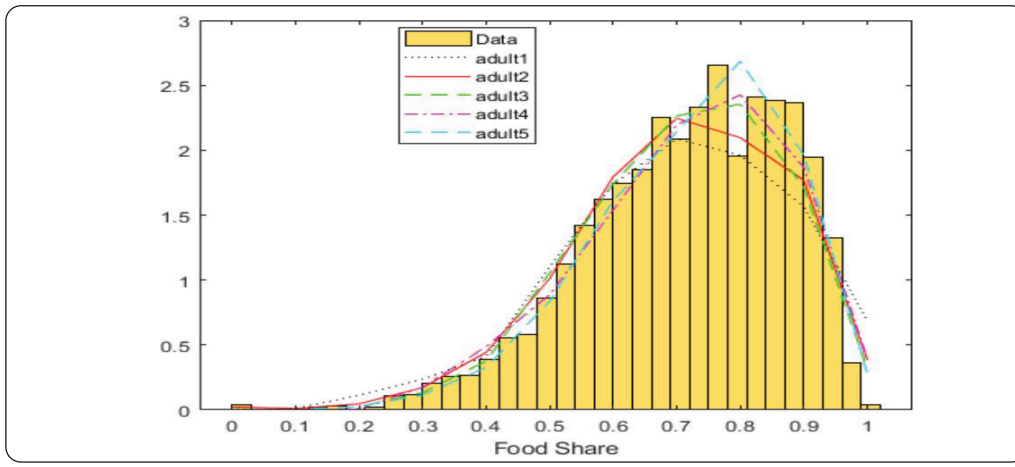


**Figure 4: Nonparametric estimation of food shares and adult equivalents dwellings by using Gaussian and Epanechnikov kernel density function**

**Source:** Field Data (2024)

Figure 4 shows the results of nonparametric kernel density estimation (with a data-driven bandwidth-the least-squares cross-validation bandwidth (LSCV (h)) = 0.967 and Scale factor = 0.604, estimated via 3344 bootstrap replications). The function form depicts that food shares for the adult equivalent increases as the number of dwellers decreases.

### 4.3 Discussion of the findings

The objective of the study was to estimate the density function form of food shares of households in Tanzania and to find out the influence of gender, location and households size of food consumption in Tanzania using the least square cross-validation bandwidth selection method across Gaussian and Epanechnikov kernel functions. The LSCV depict that was efficient and optimal for density function form of food shares for male/female headed household, urban/rural household, followed by household size (dwellers) and adult equivalent household. The density function of food shares of male headed household was high than female headed household, urban household depicted high food shares in relation to rural household, and food shares increases as household sizes increases. LSCV was optimal because the MISE and ISE values were low resulting in efficiency and bias-variance trade-off. The LSCV bandwidth had a high convergence rate and consistency resulting in an efficiency density function. The Mean Square Error (MSE), Weighted Mean Absolute Error (MAE), Standard Errors (SE), Coefficients of determination (R-squared), Mean Absolute Percentage Error (MAPE), and Integrated Mean Square Errors (IMSE) depicted high variance and low bias resulting in low efficiency.

## 5.0 Conclusion and Recommendations

### 5.1 Conclusion

The least-square cross-validation data-driven automatic bandwidth selection method was efficient and optimal for both Gaussian and Epanechnikov kernel functions. The density function form of consumption of food shares of male/female-headed households, urban/rural households, household size/dwellers, and adult equivalent households was consistent and efficient. Food shares were significantly different for male/female households, urban/rural areas households, household size, and adult equivalent households. Hence, least square cross-validation as an automatic data-driven bandwidth selection method is the optimal, robust, and universal method for density smoothing parameter selection.

### 5.2 Recommendations

Key findings suggest that using a nonparametric kernel density estimation method with least squares cross-validation (LSCV) data-driven bandwidth selection method together with Gaussian kernel function, or Epanechnikov kernel function, the food shares for male-headed household is higher compared with the female-headed household, urban household revealed higher food shares than rural household, food shares for household sizes increases as the number of dwellers increases and the food shares for adult equivalents of household is optimum high as the number of dwellers increases..

Based on the paper's findings, we recommend the use of a nonparametric kernel density estimation method and Gaussian kernel function or Epanechnikov kernel function which considerably has fewer assumptions and does not involve model specification of

an unsuitable function form for relationship between an independent variable and the dependent variable. However, currently, nonparametric kernel density estimation methods are not used rarely, because need computational ability. Also, it is recommended that, the selection of bandwidth (smoothing, or window parameter) that trade-off between variance and bias of an estimator remain an arbitrary investigator's ability or depend on a particular situation, sample size of data sets and true density.

### 5.3 Area for further research

This paper open room and provide a prospect way for further study on application of nonparametric kernel density estimation techniques for consumption data analysis.

Consequently, this paper has given a room for further study on various applications of nonparametric density estimation methods in cross-section data, unbalanced data, time series data, panel data, and longitudinal data in solving statistical, economical and econometric problems and real situations, since the current paper came up with findings from only one area of analyzing consumption of food shares data using nonparametric kernel density estimation together with LSCV data-driven bandwidth selection method.

## References

Borrajo, M. I., Gonzalez- Manteiga, W., & Martinez- Miranda, M.D. (2017). Bandwidth selection for kernel density estimation with length-biased data. *Journal of Nonparametric Statistics*, 29(5), 636-668.

Bowman, A. (1984). An alternative method of cross-validation for the smoothing of density Estimates, *Biometrika*, 71(2), 353-360.

Cai, Z., Gu, J. & Li, Q. (2009). Some recent developments in nonparametric econometrics, Advances in Econometrics. Cambridge University Press. London.

Chen, Y. C., (2018). Modal regression using kernel density estimation: A review. *Wiley Interdisciplinary Reviews: Computatonal Statistics,* 10(2), 1431-1455

Deheuvels, P. (1977). Estimation Nonparametric de la Densite' par Histogrammes Generalizes. *Revue de Statistique Appliqué,* 25(3), 5-42.

Doucet, A., de Freitas, N. & Gordon, N. (2001). Sequential Monte Carlo Methods in Practice. Springer, New York. *Econometrica,* 56(5), 931-954*.*

Hall, P. & Marron, J. S. (1992). Smoothed cross-validation*. Probability Theory and Related Fields.* 92(1), 1-20.

Harpole, J. K., Woods, C. M., Robebaugh, T. L., Levinson, C. A., & Lenze, E.J. (2014). How Bandwidth selection algorithms impact exploratory data analysis using kernel density estimation. *Psychological Methods*, 19(3), 428-443.

Hardle, P., Marron, J., and Park, B., U. (1992). Smoothed Cross-validation Probability. *Theory And Related Field*, 92(2), 1-20.

Hardle, W., Muller, M., Sperlich, S. & Werwatz, A. (2005). *Nonparametric and Semi- parametric Models- An introduction*. Spring Verlag. Berlin Heidelberg.

Hart, J. D., & Li, Q. (2006). *Nonparametric Smoothing and Lack-of-Fit Tests*. New York: Springer Verlag.

Hayfield T. & Racine, J, S. (2008). Nonparametric in Econometrics: The np package. *Journal of Statistical Software*, 76(1), 15-18.

Heidenreich, N. B., Schindler, S., and Sperlich, S. (2013). *Bandwidth selection for kernel density Estimation: A review of fully automatic selectors*. AStA *Advances in Statistical Analysis*, 97(4), 403-433.

Henderson, D., R. J. Carroll, & Q. Li (2006). *Nonparametric estimation and testing of fixed Effects panel data models*. Unpublished manuscript, Texas A & M University.

Gramacki, A., (2018). *Nonparametric Kernel Density Estimation and Its Computational Aspects*. Springer International Publishing AG, Gewerbestrasse 11, 6330 Cham, Switzerland.

Jones, M.C., Marron J.S., & Sheather. J. (1996). A brief survey of bandwidth selection for density Estimation. *Journal of the American Statistical Association*. 91(7), 401- 407.

Li, Q. & J. S. Racine (2004). Cross-validated local linear nonparametric regression. *Statistical Sinica* 14(2), 485–512.

Lin, X. & Carroll, R. J. (2000). Nonparametric function estimation for clustered data when the predictor is measured without/with error. *Journal of the American Statistical Association*, 54(3), 98-101.

National Bureau of Statistics (2016). *Tanzania National Panel Survey Report (NPS)-Wave 4, 2014-2015*, Dar es Salaam, Tanzania.

Park, B. U., Jeong, S. O. & Jones, M. C. (2003). Adaptive variable location kernel density Estimators with good performance at boundaries. *Journal of American Statistical Association*, 85(8), 92-99

Racine, J. S. (1997). Consistent significance testing for nonparametric regression. *Journal of Business and Economic Statistics,* 15(3), 369–379.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, 67(4), 390-402.

Rudemo, M. (1982). Empirical choice of histograms and kernel density estimators. *Scandinavian Journal of Statistics*, 9(8), 65-78

Sheather, S. J. (2004). Density estimation. *Statistical science*, 19(7), 588-597.

Sheather, S.J. & Jones, M.C. (1991). A reliable data-based bandwidth selection method for kernel Density estimation. *Journal of the Royal Statistical Society*, 53(4), 683-690

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

Simonoff, J. S. (1996). *Smoothing Methods in Statistics*. Springer, New York.

Scott, D., & Terrell, G. (1987). Biased and unbiased cross-validation in density estimation. *Journal of American Statistical Association*, 82(2), 1131-1146.

Stone, C. J. (1984). An asymptotically optimal window selection rule for kernel density estimation. *Annuals of Statistics*, 12(6), 1285-1297.

Taylor, C.C. (1989). Bootstrap Choice of the Smoothing Parameter in Kernel Density Estimation. *Biometrika*, 76(5), 705-712.

Terrell, G. R., & Scott, D.W., (1985). Over smoothed Nonparametric Density Estimation*. Journal of the American Statistical Association*, 80(3), 209-214.

Terrell, G.R (1990).The Maximal Smoothing Principle in Density Estimation. *Journal of the American Statistical Association*, 85(9), 470-477.

Ullah, A., Wan, A. & Chaturvedi, A. (2005). *Handbook of Applied Econometrics and Statistical inference*. Marcel Dekker.

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

Wilcox, R. (2004). Kernel density estimation. An approach to understanding how groups differ. Understanding statistics, 3(4), 333-348.