SciencePG
Science Publishing Group

Determination of an Optimal Smoothing Technique for Maternal Health Care Statistics (A Case Study of Nakuru County 2012-2016)

Lena Anyango Onyango, Thomas Mageto, Caroline Mugo

Department of Statistics and Actuarial Sciences, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya

Email address:

anyangoonyangolena@gmail.com (L. A. Onyango), ttmageto@jkuat.ac.ke (T. Mageto), cwmugo@jkuat.ac.ke (C. Mugo)

Abstract: One of the Big four agenda is Universal health. This study focused on maternal health. The main aim of maternal health is usually to reduce maternal deaths. One way in aiding to reduce maternal deaths is to forecast maternal deaths using various statistical smoothing techniques. This would enable better future planning for example increase in health facilities. Shapiro-Wilk Normality Test confirmed that there was clear observable difference between the normal distribution and the data. The study hence focused on non-parametric regression methods which include Kernel and Cubic spline smoothing techniques which were applied on maternal health care data. The technique that best dealt with this type of data was identified and used to focus maternal deaths. Selecting an appropriate technique was important to achieve a good forecasting performance. The performance of the two smoothing technique was compared using MSE, MAE and RMSE and the best model identified. In both methods we have smoothing parameters. Selecting smoothing parameter goal is usually to base it on the data. According to the results obtained in the study, it is concluded that Cubic spline smoothing technique which has a lower MSE, MAE and RMSE is better than Kernel based smoothing technique. The statistical software that was used for the analysis was R. The study used maternal health care statistics data for Nakuru County.

Keywords: Smoothing Technique, Cubic Spline, Kernel Smoothing

1. Introduction

Smoothing is a type of data handling technique. In statistics, when data is smoothed, an approximation function is created that usually attempt to capture important patterns in the data. Typically, data smoothing is done to remove noise from a given data set. Data smoothing can be employed to enable better prediction. Maternal health includes the dimensions of health care in family planning, preconception, prenatal and postnatal treatment to reduce maternal morbidity and mortality [1]. Some elementary data smoothing techniques such as visual inter-polation, averaging and mathematical interpolation emerged during the eighteenth century [2]. At that time, smoothing techniques consisted of simple interpolation of the data and eventually evolved into more complex modern methods such as Cubic splines smoothing technique. Kernel density estimation has been previously discussed utterly, giving details about assumptions on the kernel weight, properties of the estimator such as bias and variance [3]. Spline smoothing in some sense corresponds approximately to bandwidth smoothing by a kernel method depending on the local design point density [4]. Consideration of kernel smoothing methods shows that there are desirable properties in how the effective local bandwidth acts in spline smoothing. The smoothing parameter's value for a curve fitting can be chosen by minimizing the expected prediction error [5]. To address the extreme handling of data, smoothing techniques should be used to achieve an accurate result in making predictions. Two smoothing techniques applied on maternal health data is explored and the technique that best deals with this type of data is identified. Usually there are many smoothing techniques available, and selecting the appropriate technique is an important issue to achieve a good forecasting

performance. The Kernel based and Cubic splines smoothing methods are studied. Some error measures for example, mean absolute deviation, root mean squared error, and mean square error are calculated for above smoothing techniques to identify the best method.

2. Literature Review

In Kernel Smoothing Technique, Cross-validation with Kullback-Leibler loss function has been employed to the choosing of a smoothing parameter in the kernel method of density estimation [6]. A framework for this problem is constructed and used to derive an alternative method of cross-validation, based on integrated squared error, recently also proposed [7]. Hall has established the consistency and asymptotic optimality of the new method [8]. Kernel density plays a very important function in statistics. A wrong choice of bandwidth may lead to a bad estimate [9]. Kernel density estimation disadvantage is that it is normally unbiased. Cross-validation is a very good method when it comes to selecting of parameter. For regression function estimation of the Epanechnikov kernel which has the property of optional kernel and which is commonly used in practice is preferred [10]. For the nonparametric estimation of regression functions with a one-dimensional design parameter, a new kernel estimate is defined [11].

Cubic Spline Smoothing Technique; Non-parametric regression using cubic splines is an attractive, flexible and widely applicable approach to curve estimation, [3]. Although the basic idea was formulated many years ago, the method is not as widely known or adopted as perhaps it should be. To a cubic spline operator for smoothing there is normally a given operator of kernel. A Connection between Cubic and Kernel smoothing has been established theoretically and related approximations to a Green's function have also been established [12, 13]. The fundamental ideas behind cubic spline is in engineer's tool used to draw smooth curves through a number of points. This consists of weights attached to flat surface at the points to be connected. A flexible strip is then put across each of these weights, resulting in a pleasingly smooth curves. The mathematical spline is similar in principle [14]. The points in this case are numerical data. Stupp introduces a general method for proving uniformity in kernel type function estimators bandwidth consistency [15].

3. Methodology

3.1. Kernel Smoothing

In smoothing, the data points of a signal are modified so individual points (presumably due to noise) are decreased. A Kernel smoothing refers to a general class of techniques for nonparametric estimation of functions. Closer points are given higher weights. This technique is so useful in the visualization of data. The simplicity of kernel estimators entails mathematical tractability, so one can delve deeply into the properties of these estimators without highly sophisticated mathematics. In summary, kernel smoothing provides simple, reliable and useful answers to a wide range of important problems. The main feature of kernel smoothing is data speaks for itself, meaning that data decides which function best fits. Kernel smoothing also provides a simple way of finding structure in data sets without the imposition of a parametric model. Kernel smoothing has a low standard error and works well for small or large samples.

A kernel smoother generally defines a set of weights $\{W_i(x)\}_{i=1}^n$ for every x and defines

$$\hat{f}(x) = \sum_{i=1}^n W_i(x)\, y_i \qquad (1)$$

In this general definition, most smoothers can be considered kernel smoothers. In action, kernel smoothers have a simple approach to describe the sequence of weights $\{W_i(x)\}_{i=1}^n$ by describing the weight function shape $W_i(x)$ through a density function with a scale parameter that changes the size and the form of the weights that are near x. This shape function is usually referred as a kernel K. The kernel is a continuous, bounded, and symmetrical real function K that integrates into one:

$$\int K(u)\,du = 1 \qquad (2)$$

The weight sequence is therefore defined by;

$$W_{h\,i}(x) = \frac{K\left(\frac{x-x_i}{h}\right)}{\sum_{i=1}^n K\left(\frac{x-x_i}{h}\right)} \qquad (3)$$

For any given scale parameter h we notice that $\sum_{i=1}^n W_{hi}(x) = 1$. The kernel smoother is then defined as before for any x by $\hat{f}(x) = \sum_{i=1}^n W_i(x)\, Y_i$. Usually, kernel smoother defines weights that smoothly decrease as you move away from the target point.

3.1.1. The Bandwidth h

Bandwidth controls the smoothness or roughness of a density estimate. Choosing the bandwidth value h is more important than choosing the kernel density function. Popular kernels used for smoothing include parabolic (Epanechnikov), Tricube and Gaussian kernels.

Table 1. *Popular Kernels.*

Kernel	Efficiency
Epanechnikov	1.0000
Biweight	0.974
Triangular	0.986
Normal	0.951
Uniform	0.930

Nadaraya Kernel Watson is usually given by

$$\hat{Y}(X_0) = \frac{\sum_{i=1}^N K_{h\lambda}(X_0, X_i) Y(X_i)}{\sum_{i=1}^N K_{h\lambda}(X_0, X_i)} \qquad (4)$$

Some particular cases of kernel smoothers are:

3.1.2. Kernel smoother (Gaussian)

The Kernel (Gaussian) is usually mostly used and is given by;

$$K(x^*, x_i) = \exp\left(-\frac{(x^*-x_i)^2}{2b^2}\right) \quad (5)$$

3.1.3. Nearest Neighbor Smoother

The main idea here is that for each and every points X_0, we take m nearest of neighbors and then we estimate the given value of $Y(X_0)$. As before $h_m(X_0) = \left\|X_0 - X_{|m|}\right\|$, where $X_{|m|}$ is the mth closest to X_0 neighbor and

$$D(t) = \begin{cases} \frac{1}{m}, if\ |t| \le 1 \\ 0, otherwise \end{cases} \quad (6)$$

3.2. Cubic Spline Smoothing Technique

Let $\{X_i, Y_i: i = 1, \dots, n\}$ be a set of observations, based on a relationship $Y_i = f(X_i) + \epsilon_i$ where the ϵ_i are independent, zero mean random variables (normally taken to have constant variance).

The estimate of the cubic smoothing spline is known as the minimizer (over the twice-differentiable function class)

$$\sum_{i=1}^{n}\{Y_i - \hat{f}(x_i)^2 + \lambda \int \widehat{f''}(x)^2\ dx \quad (7)$$

$\lambda \ge 0$ usually regulates trading-off among data fineness and roughness of the function estimate. This is often calculated by cross-validation or restricted marginal likelihood (REML) which takes advantage of the relation between spline smoothing and Bayesian estimation. Many methods to select parameters of smoothing such as cross validation(CV), generalized maximum likelihood (GML), generalized cross-validation (GCV) and unbiased risk (UBR) have been established under the assumption of observations being independent.

3.2.1. Derivation of the Cubic Smoothing Spline

We can fit a smoothing spline in two steps:
1. Firstly, we derive the values $\hat{f}(x_i): i = 1, \dots, n$
2. From these values obtained, derive $\hat{f}(x)$ for each and every x

Then, treat the second step first.

Given the vector

$$\hat{m} = (\hat{f}(x_1), \dots, \hat{f}(x_n))^T \quad (8)$$

of the values which are fitted, the spline criterion of a sum squares is normally fixed. It only remains to minimize $\int \widehat{f''}(x)^2 dx$. This spline interpolation is a linear function given as:

$$\hat{f}(x) = \sum_{i=1}^{n}\hat{f}(x_i)\ f_i(x) \quad (9)$$

where $f_i(x)$
are a spline-based function set. The roughness penalty therefore has the form

$$\int \widehat{f''}(x)^2 dx = \hat{m}^T \hat{m} \quad (10)$$

Where the A elements are

$$\int f''_i(x)f''_j(x)\ dx \quad (11)$$

The basic functions, and therefore the matrix A, depend on the predictor variables x_i configuration, but not on the Y_i or $\hat{m}$ responses. Return to the first step. The penalized sum-of-squares is given as;

$$\{Y - \hat{m}\}^T\{Y - \hat{m}\} + \lambda\ \hat{m}^T A\hat{m}, \quad (12)$$

Where $Y = (Y_1, \dots, Y_n)^T$. Minimizing over $\hat{m}$ by differentiating against $\hat{m}$. This results in:

$$-2\{Y - \hat{m}\} + 2\ \lambda\ A\hat{m} = 0 \text{ and } \hat{m} = (I + \lambda\ A)^{-1}Y \quad (13)$$

Smoothing spline model is given by:

$$y_i = f(t_i) + \epsilon_i, i = 1, \dots, n, t_i \epsilon\ [0,1] \quad (14)$$

where $\epsilon = (\epsilon_1, \dots, \epsilon_n)' \sim N(0, \sigma^2 W^{-1})$ and σ^2 is unknown. Assume $f \in W_2^m$, where
$W_2^m = \{f: f^{(v)}$ absolutely continuous, v=0,…, m-1,

$$\int_0^1 (f^{(m)}(t))^2 dt < \infty\} \quad (15)$$

$\hat{f}_\lambda$ the minimizer is

$$\min_{f \in W_2^m}\{\frac{1}{n}(y - f)'W(y - f) + \lambda \int_0^1 \left(f^{(m)}(t)\right)^2 dt\} \quad (16)$$

Where $y = (y_1, \dots, y_n)'$ and $f = (f(t1), \dots, f(tn))'$. Parameter λ governs the relationship between fitness and smoothness of the estimate and is normally called the smoothing parameter.

Given $a < t_1 < t_2 < \cdots < t_n < b$, an operator g is a cubic if:
1. On one interval $(a, t_1), (t_1, t_2), \dots, (t_n, b), g$ is a cubic polynomial.
2. The polynomial pieces fit together at points t_i (called knots) such that g itself and its first and second derivative are continuous at each t_i and hence on the whole$[a, b]$.

3.2.2. Specification of Cubic Spline

Cubic spline is specified as below;

$$g(t) = d_i(t - t_i)^3 + c_i(t - t_i)^2 + b_i(t - t_i) + a_i, \text{ for } t_i \le t \le t_{i+1} \quad (17)$$

3.3. Choosing the Smoothing Parameter

There are two different philosophical approaches: Subjective choice; Automatic method - chosen by data - (Cross-validation; Generalized cross-validation)

3.3.1. Cross Validation

Used to select value of λ

$$\min_\alpha CV(\alpha) = \frac{1}{n}\sum_{i=1}^{n}\{Y_i - \hat{g}^{(-1)}(t_i; \alpha)\}^2 = \frac{1}{n}\sum_{i=1}^{n}(\frac{Y_i - \hat{g}^{(-1)}(t_i)}{1 - A_{ii}(\alpha)})^2 \quad (18)$$

if $\hat{g}$ is the spline smoother with α

3.3.2. Generalized Cross-Validation

$$\min_\alpha GCV(\alpha) = \frac{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{g}(t_i))^2}{[1 - n^{-1}tr A(\alpha)]^2} \frac{n \times sum\ of\ squares\ of\ all\ the\ residuals}{(equivalent\ df)^2} \quad (19)$$

3.4. Comparison of Models Performance

Three different prediction criteria are used to compare the performance of cubic spline and kernel model. These include Mean Square Error (MSE), Mean Absolute Deviation (MAD) and Root Mean Squared Error (RMSE). These criteria are defined as follows:

$$MSE = \frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{y}_t)^2 \qquad (20)$$

$$or\ RMSE = \sqrt{MSE} \qquad (21)$$

$$MAE = \frac{1}{T}\sum_{t=1}^{T}|y_t - \hat{y}_t| \qquad (22)$$

4. Data Description and Results

The study used maternal health care statistics data for Nakuru County. The variables for this data included:

Table 2. Variables of the data.

Dependent Variable	Independent Variable	
Total Maternal Deaths	Total Skilled Delivery	New Antenatal Clients
	Pregnant women completing four antenatal visits.	

4.1. Test for Normality

Table 3. Testing for normality.

Shapiro-Wilk Normality Test	
Variables	P-value
Total Maternal Deaths	0.00804
Total Skilled Delivery	0.006968
Pregnant women completing four antenatal visits	0.01354
New Antenatal Clients	0.1722

The P-value is very low meaning that if the data was normally distributed, there would be very little chance of seeing the same sample from such a distribution. Clearly the P-Values < 0.05 suggesting strong evidence of non-normality and so non-parametric test should be used. Shapiro-Wilk normality test was used in conjunction with both Q-Q plot and Histogram to check normality of the data.
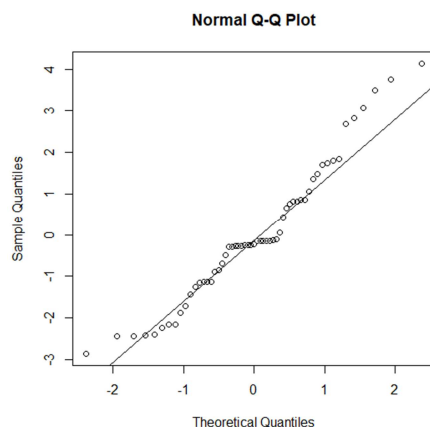
Normal Q-Q plot

Figure 1. Q-Q plot.

Q-Q plot is an alternative graphical method of assessing normality to the histogram and is easier to use when there are small sample sizes. The plot helps us to understand the data set's distribution. The scatter normally compares the data to a standard normal distribution. The scatter is not as close to the line with clear pattern coming off the line. By the heavy tail, the data is obviously not normal. The data is therefore not assumed to be distributed normally.

Histogram;

Plotting the variable of interest histogram provides an indication of the distribution shape. The histogram is smoothed by a density curve, and is usually added to the graph. The histogram shows the data is clearly skewed so no parametric test should be carried out using this data.
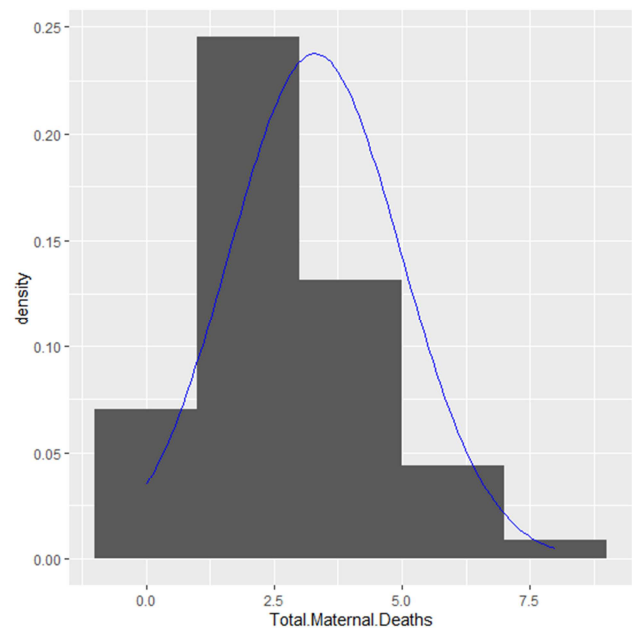

Figure 2. Plot of Histogram with a density curve.

4.2. Fitting Cubic Splines to the Data

A smoothing parameter is required to be chosen in cubic splines. Smoothing parameter controls the trade-off between data accuracy and portion estimate roughness. In this study, cross validation was used to choose the smoothing parameter.

Table 4. Results from the fitted cubic model.

Variables	Estimate	Pr (>|t|)
TSD	-1.24378	0.0329
NAC	-0.08144	0.00958
PWCV	-2.01392	0.0413

β_1 Is -1.24378, a unit increase in Total Skilled Delivery causes a decrease in Total Maternal Death by 1.24378. β_2 Is -0.08144, a unit increase in New Antenatal Clients causes a decrease in Total Maternal Death by 0.08144. β_3 is -2.01392, a unit increase in Pregnant women completing four antenatal visits causes a decrease in Total Maternal Death by 2.01392. The p-value of Pregnant women completing four antenatal visits is (0.0413) of Total Skilled Delivery is (0.0329) of New

Antenatal Clients is (0.00958) are all < 0.05 hence all the variables in the model are significant.
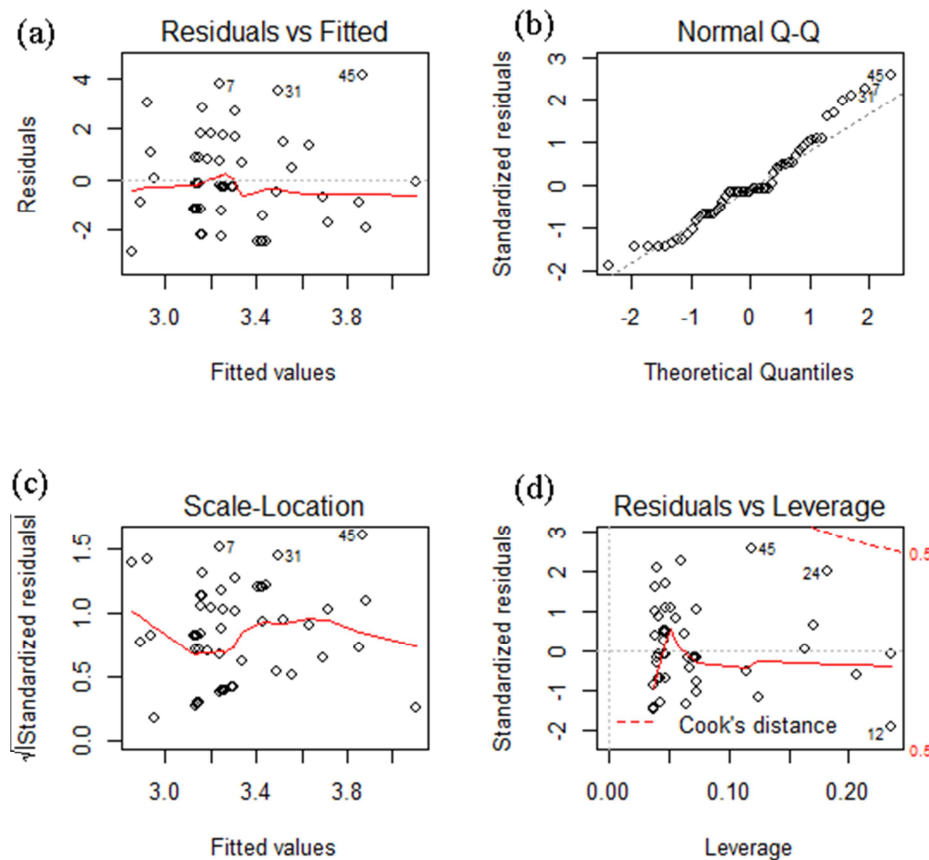
4.2.1. Diagnostic Plots

Figure 3. *Diagnostic plots.*

4.2.2. Explanation of the Diagnostic Plots

Plot (a) is relatively shapeless without clear pattern of the data, no obvious outliers and it is generally symmetrically distributed around the 0 line without particularly large residuals. The residuals get larger as we move left to right and there are a few potential outliers. Hence the assumption whether the relationship between the variables of this study being linear is true. Plot (b) shows if residuals are normally distributed. We see that these residuals are not lined up well on the straight dashed line. Hence since the points do not form a line that is roughly straight, the assumption of the dependent variable being normally distributed is not true. In Plot (c) Scale-Location is also called Spread-Location. In the plot, horizontal line with randomly spread points is noticed. Hence the assumption of homoscedastic is true. In plot (d) outlying values at the upper right corner or at the lower right corner are checked if they exist. This plot helps us to find influential cases if any. Not all outliers are influential. In our case there is no influential case or cases. The red dashed line are not so much seen because all cases are well inside the red-dashed line.

4.3. Fitting Kernel Smoothing Technique to the Data

The kernel-based smoothing technique has been fitted to the data given in this section. Data speak for themselves is the main feature of kernel smoothing, which means that the data determines what function fits best. Kernel smoothing also provides a simple way of finding structure in data without the imposition of parametric model. Bandwidth or the smoothing parameter was chosen by cross validation. In this study Gaussian kernel, one of the most commonly used kernels was used.

Table 5. *Results from fitted kernel model.*

Bandwidth For Total Skilled Delivery	Bandwidth For New Antenatal Clients	Bandwidth for pregnant women completing 4 antenatal visits
1546.99	1189.22	1386.53

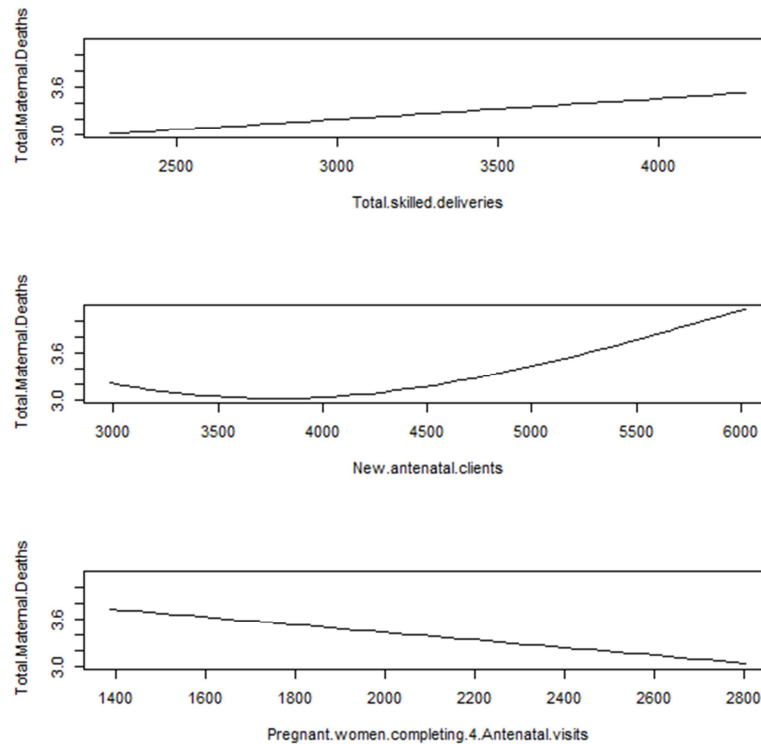The bandwidths were found to be 1546.99, 1189.22 and 1386.53 respectively.

Figure 4. *Plot of the fitted kernel model.*

From the first plot, we see that when TSD increased, the TMD increased. This is contrary to our expectation. There might be another factor contributing to increase in TMD even with an increased TSD. From the second plot, we see that when NAC were 3000, TMD decreased up to when it was 4000 but after that the number of TMD increased. This may be due to constant number of caregivers. Other people might not have been attended to due to large number and hence the cause of increased TMD. From the third plot, we see that as the number of PWCV increased, TMD was decreasing. This is true since those who were finishing all the four antenatal visits were taking into consideration all the precautions during pregnancy.

4.4. Comparing the Performance of Cubic Spline and Kernel Smoothing Technique

Table 6. *Results.*

Performance Criteria	Cubic Spline Technique	Kernel Regression Technique
MSE	2.932502	3.28672
RMSE	1.712455	1.81293
MAE	0.2234266	0.2430285

According to Table six it is shown that performance criteria values obtained by cubic spline technique are smaller than results of the kernel regression technique. Hence, it can be said that cubic spline smoothing technique is better than kernel smoothing technique.

4.5. Forecasting Monthly Maternal Deaths Using the Best Smoothing Technique

Cubic Spline smoothing technique was used to forecast

monthly maternal deaths.

4.5.1. Results of forecasting using Cubic Spline Smoothing Technique

Table 7. *Forecasting monthly maternal deaths.*

Months	Point Forecast	Lo 95	Hi 95
58	2.699412	2.306904	3.091920
59	2.773840	2.300223	3.247457
60	2.848269	2.269489	3.427048
61	2.922697	2.218973	3.626421
62	2.997125	2.152031	3.842220
63	3.071554	2.071051	4.072056
64	3.145982	1.977721	4.314244
65	3.220411	1.873258	4.567563
66	3.294839	1.758574	4.831104
67	3.369267	1.634376	5.104159
68	3.443696	1.501227	5.386165
69	1.359591	3.518124	5.676657

4.5.2. Explanation

Lo95 and Hi95 are the minimum and maximum boundaries of about 95 percent, respectively. That means we are 95 percent confident that the true population forecast is somewhere between our Lo95 and Hi95. Since we are concerned with forecast values, this would in effect be a case of an interval of expectation rather than an interval of confidence. The interval reflects the set of possible values that we plan to find at some future point in time for a prediction interval. In the table above, the 95% forecast interval informs us that there is a 95% chance that the future observation value will fall somewhere between the minimum and maximum limits. More precisely, a 95% prediction interval informs us that if we calculate 95% prediction

intervals across repeated samples, the future observation value should fall on about 95% of these samples within the minimum and maximum bound.

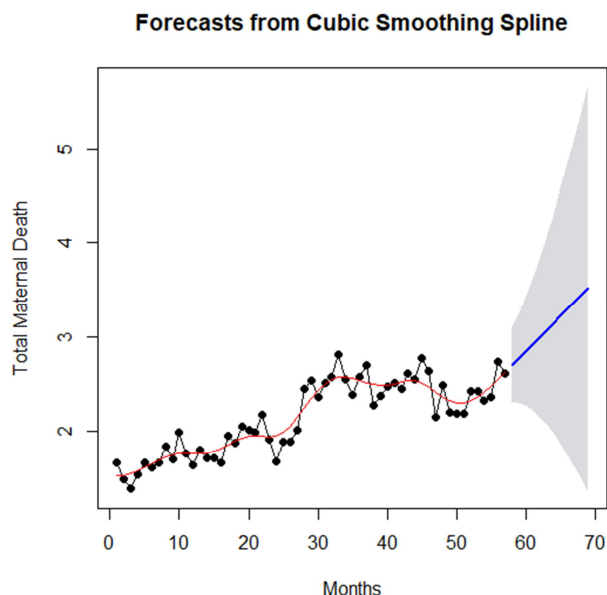From the graph below we see that Total Maternal Deaths will be increasing.

Figure 5. *Plot of Forecasts from Cubic Smoothing Spline.*

5. Conclusion

In non-parametric regression methods, detailed results can be found between explanatory and response variables. In this study, the main objective was to determine the best smoothing technique for maternal health care data. The specific objectives were; to fit Cubic splines and Kernel smoothing techniques to the data; to compare the performances of Cubic splines and Kernel smoothing technique and lastly to forecast maternal deaths using the best smoothing technique. Shapiro-Wilk Normality test confirmed that there was clear observable difference between the normal distribution and the data. Cubic splines model and Kernel model was fitted to the data. Three different prediction criteria were employed to enable comparison of the performance of these models. These are MSE, RMSE and MAE which are chosen as a measure of fit since Cubic spline and Kernel based model are estimated by using the sum of squared error. In these prediction criteria, the smaller the value the better the model. Lower values indicates a better fit. Cubic spline model indicated a good performance since it had a lower MAE, RMSE and MSE as compared to Kernel based model. Cubic spline gave a good forecasting performance which would enable future planning for example increase of skilled personnel, increase of health facilities for the ultimate goal of maternal deaths reduction.

References

[1] C. and K., "Government printer," *Kenya: Nairob,* 2010.

[2] B. and H. J., "Kernel estimators of regression functions," *Advances in econometrics: Fifth world congress,* vol. 1, pp. 99--144, 1987.

[3] S. and. B. W., "Some aspects of the spline smoothing approach to non-parametric regression curve fitting," *Journal of the Royal Statistical Society: Series B (Methodological,* vol. 47, pp. 1--21, 1985.

[4] S. B. W. O. and O., "Spline smoothing: the equivalent variable kernel method," *The Annals of Statistics,* vol. 12, pp. 898--916, 1984.

[5] M. and C., "Choosing a smoothing parameter for a curve fitting by minimizing the expected prediction error," *Acta Universitatis Apulensis, Mathematics-Informatics,* vol. 5, pp. 91--96, 2003.

[6] Bowman, W. Adrian, H. Peter, T. and D., "Cross-validation in nonparametric estimation of probabilities and probability densitie," *Biometrika,* vol. 71, pp. 341--351, 1984.

[7] R. and M., "Empirical choice of histograms and kernel density estimators," *Scandinavian Journal of Statistics},* pp. 65--78, 1982.

[8] H. P. a. S. S. J. a. J. M. a. M. and J.. S., "On optimal data-based bandwidth selection in kernel density estimation," *Biometrika,* vol. 78, pp. 263--269, 1991.

[9] Hardle, M. and J., "Bootstrap simultaneous error bars for nonparametric regression," *The Annals of Statistics,* pp. 778--796, 1991.

[10] G. J. a. L. and H., "Chiral perturbation theory: expansions in the mass of the strange quark," *Nuclear Physics B,* vol. 250, pp. 465--516, 1985.

[11] G. T. a. M. and H.-G., "Kernel estimation of regression functions," *Smoothing techniques for curve estimation,* pp. 23--68, 1979.

[12] N.. K. a. S. G. H. O. and O., "Serial and parallel processing of visual feature conjunctions," *Nature,* vol. 320, pp. 264--265, 1986.

[13] M. K. O. and O., "A comparison of a spline estimate to its equivalent kernel estimate," *The Annals of Statistics,* vol. 19, pp. 817--829, 1991.

[14] S. L. a. P. R. L. a. B. and G. E., "Harmonic splines for geomagnetic modelling," *Physics of the Earth and Planetary Interiors,* vol. 28, pp. 215--229, 1982.

[15] S.. R. a. M. W. P. a. V. D. B.. M. J. a. W. M. a. F. B. a. T. M. J. K. a. B.. A. A. a. M. C. a. B. and U. a. o., "Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma," *New England Journal of Medicine,* vol. 352, pp. 987--996, 2005.