Review Article

A NONPARAMETRIC APPROACH TO VARIANCE FUNCTION ETIMATION: AN APPLICATION TO NAIROBI'S STOCK EXCHANGE MARKET

Abstract

We consider the analysis of a method for fitting regression models to data in the field which exhibit nonconstant variances. We focus on various approaches and methods of estimating the nonconstant variances. We discuss the nonparametric approach which includes the smoothing methods and the selection of the optimal bandwidth. Generally, the principal problems of interest are the choice of the smoothing method and the selection of the bandwidth (R Dennis Cook and Sanford Weisberg 2009). We compare the two mostly used smoothing techniques; the Kernel and the Spline. We illustrate the three smoothing techniques from the data obtained from Nairobi stock exchange market and found that the Kernel smoother produces the best estimate since its variance is less than that of the Spline smoother.

Key words: Kernel smoother, Spline, smoothing, Bandwidth.

1. Introduction

The breakdown of fixed exchange rate system has sharply increased financial risks in financial institutions. Recent finance disasters in trade portfolios like the national bank of Kenya for example, have underlined the need for accurate financial risk measures in institutions such as banks and investment firms. The nature of financial risks has changed with time and therefore the method to measure them must adapt to recent experience. It is in this context that quantitative measures have become vital in the management for both internal and external requirements parallel with others models of returns.

Due to globalization, which has resulted to a fast financial world, there is motivation to develop efficient and effective risk measures which will respond to news just like the other forecasts and must be easy to understand even when the situation is complex. Despite the simplicity of the risk, its management has remained a challenging statistical problem partly because it depends on the joint distribution of the portifolio returns which typically changes overtime. It is for this reason that we estimate the variance function of shares volume of Nairobi stock exchange market. The purpose of the paper is to provide financial managers and shareholders with a non technical and flexible model for market-to-market reporting.

The study has a variance model that will help financial managers and shareholders in the following ways:-Information reporting, Resource allocation and Performance evaluation.

Variance function can be used to adjust performance for risk. This is essential in a trading environment where traders have a natural tendency to take on extra risk. Risk capital charges based on variance function provide correct incentives to traders.

In many applications, it is a priori unclear how the variance function should be specified. However efficient inference for the regression parameters relies on correct variance

functions. When the model is mispecified the resulting model fit can be biased, and the possibility for making wrong inferences exist. Even in the most common form of nonparametric regression where the mean is left unspecified, it is common to assume that the observations are uncorrelated, which can be viewed as a "parametric" assumption on the distribution of the errors. Violation of that assumption has a serious effect on the bandwidth for estimating that mean function (Hardle, W (1993).

THE MODEL

Lets consider the data points $\{(X_i, Y_i)\}_{i=1}^{n}$ *where* X_i is from uniform distribution and $Y_i = m(X_i) + \varepsilon_i, \varepsilon_i \sim N(0, \sigma^2)$. The X-variate is uniformly stochastically distributed over the unit interval and the observation errors ε_i have the standard normal distribution. The errors and the design points are assumed to be mutually uncorrelated that is $E[X_i \varepsilon_i] = 0 \quad \forall_i$

Model 1 (Parametric regression model)

$Y_i = f(x_i, \beta) + \varepsilon_i$, where $f(.)$ is a known function, β is the unknown parameter to be estimated, and the errors ε_i are independent and identically distributed, such that $E[\varepsilon_i] = 0$ and $E[\varepsilon_i^2] = \sigma^2 > 0 \quad \forall_i$ for constant variance.

Model 2 (Non-parametric regression model)

$Y_i = m(X_i) + \varepsilon_i$, $m(.)$ is unknown function to be estimated and the errors ε_i are independent, satisfying $E[\varepsilon_i] = 0$ and $E[\varepsilon_i^2] = \sigma_i^2 > 0$ (non-constant variance).

We therefore consider the regression curve $m(x) = 1 - x + e^{-200\left(x-\frac{1}{2}\right)^2}$ and the structures of the variance function (variance of the observations is a function of the mean).

Case 1

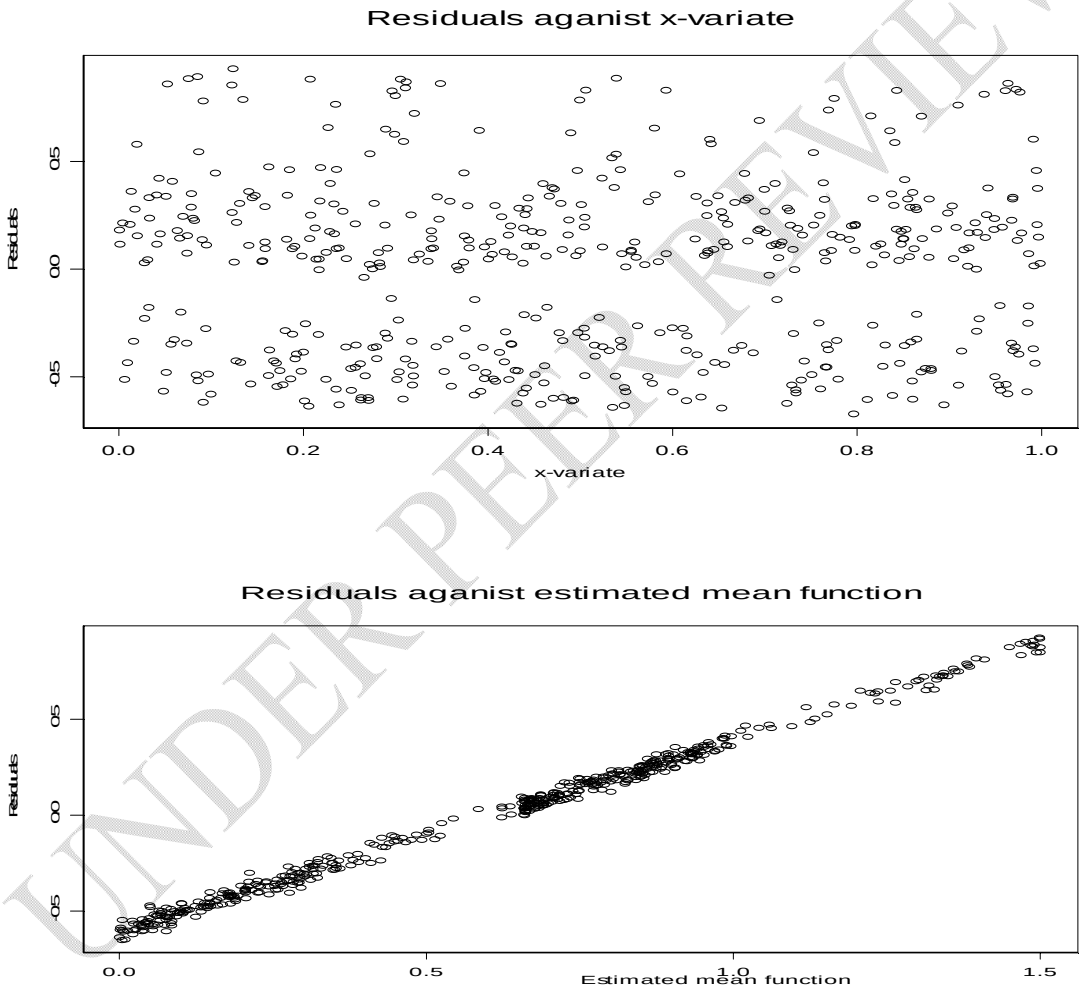
$E[\varepsilon_i] = \text{var}(Y_i) = \text{var}(m(x_i)) = \sigma^2(m(x_i))$ *or* $\sigma_i^2 = m(X_i) + 5$

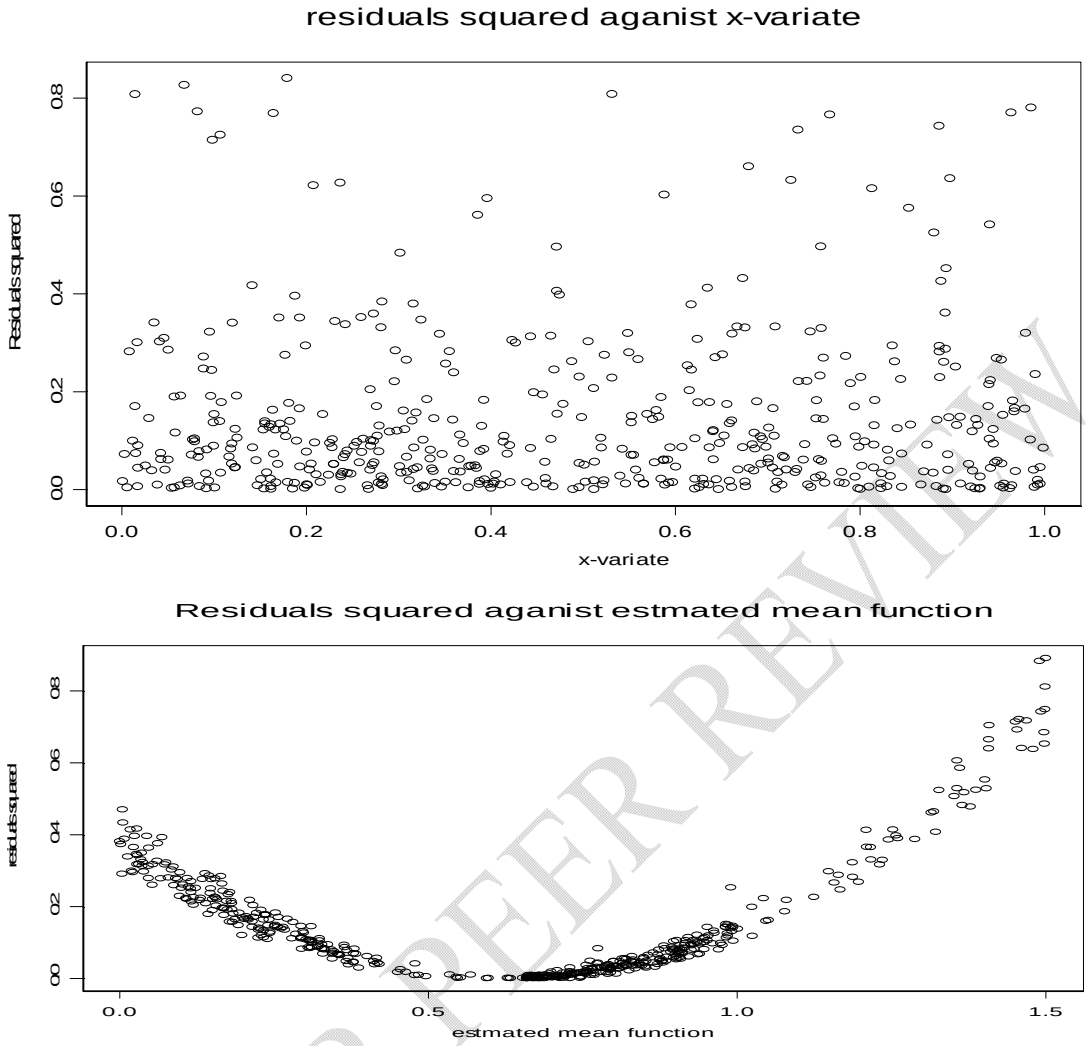
Case 2

$\sigma_i^2 = X_i^2 + 5$

In case 1, the variance function is mean dependent, whereas it depends on the design points in case 2.

Image 1: DATA SIMULATION USING CASE 1 AND 2 AND SCATTER PLOTS OF RESIDUALS AGANIST X-VARIATE AND RESIDUALS AGANIST $\hat{m}(x)$





2. Basic idea of smoothing

If m(.) is believed to be smooth, then the observation at x_i, near x should contain information about the value of m(.) at x. Thus, it should be possible to use something like local average of data near x to construct an estimator of m(x). Smoothing of a data set $\{x_i, y_i\}$, i = 1, 2, 3……n involves the approximation of the mean response curve m(.) in the regression relationship. The function of interest could be the regression curve itself. In the trivial case in which m(.) is a constant, estimation of m(.) reduces to the point of location, since an average over the response variables y yields an estimate of m(.). In practical studies though it is unlikely that the regression curve is constant. Rather the assumed curve is modeled as a smooth continuous function of a particular structure which is nearly constant in small neighborhoods around x (Jobson. J.D and Fuller, W.A

1980). It is not easy to judge from looking even at a two-dimensional scatter plot whether a regression curve is locally constant. A reasonable approximation to the regression curve m(.) will therefore be any representative point close to the centre of this band of response variables (Carrlo, R.J and Ruppert, D 1987). A quite natural choice is the mean of the response variables, near a point x. The local average should be constructed in such a way that it is defined only from observations in a small neighborhood around x, since y – observations from points far away from x will have, in general very different mean values.

The local averaging procedure can be viewed as the basic idea of smoothing. More formally this procedure is defined as

$$\hat{m}(x) = n^{-1}\sum_{i=1}^{n} w_{ni}(x)y_i,$$

Where $\{w_{ni}(x)\}$, i = 1,2……n donates a subsequences of weights which may depend on the whole vector $\{x_i\}$,i= 1 , 2,……..n. Every smoothing method to be described in this work is, at least asymptotic, of the form $\hat{m}(x) = n^{-1}\sum_{i=1}^{n} w_{ni}(x)y_i$ .Quite often the regression estimator $\hat{m}(x)$ is just referred to as a smoother and the outcome of smoothing procedure is simply called the smooth.

Special attention has to be paid to the fact that smoothers, by definition average over observations with different mean values. The amount of averaging is controlled by the weight sequence $\{w_{ni}(x)\}$, i = 1, 2,… ,n, which is tuned by a smoothing parameter. This smoothing parameter regulates the size of the neighborhood around x meaning a local average over too large a neighborhood would cast away the good with the bad (A Ian McLeod 1998). In this situation an extremely oversmoothed curve would be produced, resulting in a biased estimate $\hat{m}(x)$. Defining the smoothing parameter so that it corresponds to a very small neighborhood would not shift the chaff from the wheat. Only a small number of observations would contribute nonnegligible to the estimate $\hat{m}(x)$ at x making it very rough and wiggly. In this case the variability of $\hat{m}(x)$ would be inflated. Finding the choice of smoothing parameter that balances the trade-off between over smoothing and under smoothing is called the smoothing parameter selection problem (Sanford Weisberg 2005).

3. A comparison of Kernel and Spline smoothers

Kernel function K (.)

One of the most active research areas in statistics in the last 20 years has been the search for a method to find the "optimal" bandwidth for a smoother. There are now a great number of methods to do this although none of them is fully satisfactory. Here we present the comparative of two commonly used and easy to implement smoothers. The Kernel and the cubic spline.

1-dimensional kernel function takes the form $K_h(x - X_i) = \frac{1}{h}K\left(\frac{x - X_i}{h}\right) = K(u)$.

It is a type of local smoother which assigns weights to the observations X_i. The weights decreases with the distance between the point of estimation x and X_i, i=1, 2, 3…n.

Various forms of kernel form of kernel exist including uniform, triangle, Epenechnikov, Bisquare and Gaussian among others. Among these kernel function, Gaussian has infinite support while all the others are bounded in [-1,1] (Carrlo, R.J and Ruppert, D 1987). In this study we use the Gaussian kernel which takes the form $K(u) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}u^2\right]$ where $u = \frac{x - X_i}{h}$. The parameter h is called the bandwidth which determine how large neighbourhood of the target point x, is used in estimation. A large bandwidth generates a smooth curve but with a high possibility of obscuring the interesting structures. A very small bandwidth generates a wigglier curve.

However, both practical and theoretical considerations limit the choice of the kernels. Given the bivariate data $(x_i, y_i)_{i=1}^n$, the smoothed value $\hat{m}(.)$ produced by a kernel function K(.) can be given as

$$m(.) = \frac{n^{-1}\sum_{I=1}^{N} K\left(\frac{x - x_i}{h}\right)Y}{n^{-1}\sum_{i=1}^{N}\left(\frac{x - x_i}{h}\right)} \quad , \quad 0<x, x_i<1, \; i=1,...,N$$

(Elizbar A Nadaraya 1964)

Cubic spline

A common measure of "fidelity to the data" for a curve g is the residuals sum of squares $\sum_{i=1}^{n}(y_i - g(x_i))^2$ if g is allowed to be any curve unrestricted in functional form. Then this distant measure can be reduced to zero by any g that interpolates the data. Such a curve would not be acceptable on the ground that it is unique and that it is too wiggly for a structure oriented interpolation. The spline smoothing approach avoids this implausible interpolation of the data by quantifying the competition between the aims to produce a curve without too much rapid local variation (Carrlo, R.J and Ruppert, D 1987). There are several ways to quantify local variation. One could define measure of roughness based, for instance, on the first, second and so forth derivatives. In order to explicate the main ideas the integrated squared deviation is most convenient that is the roughness penalty

$$\int (g''(x))^2 dx$$

is used here to quantify local variation. Using these measures we define the weighted sum

$$s_\lambda(g) = \sum_{i=1}^{n}(y_i - g(x_i))^2 + \lambda \int (g''(x))^2 dx$$

where λ denotes a smoothing parameter. The smoothing parameter λ represents the rate of exchange between residuals errors and roughness of the curve g. The problem of minimizing $s_\lambda(.)$ over the class of all twice differentiable functions on the interval

$[a,b]=\lfloor x_{(1)}, x_{(n)} \rfloor$ has a unique solution $\hat{m}_{\lambda}(x)$ which is defined as the cubic spline. The observations considered are in a small neighborhood of x since y observations far away from x will have, in general very different mean values. In obtaining the smooth curve the selection of the kernel function is not enough but rather the consideration of the bandwidth h is equally very important (Silverman, Bernard W 1984).

The comparison is preformed on a simulated data set. When we look at the kernel smoothing, a variety of kernel functions have been studied. Kernel smoothing technique is one of the simplest estimation which is straight forward to implement without further mathematical knowledge and it is understandable on an intuitive level. The decision about the right amount of smoothing is crucial. The challenge in smoothing is to choose the best bandwidth that balances the desire to reduce the variance of the estimator (which needs lots of data points) yet capture significant small-scale features in the underlying distribution (which needs a narrow bandwidth).
Every smoothing method has to be tuned by some smoothing parameter which balances the degree of fidelity to the data against the smoothness of the estimated curve. A choice of the smoothing parameter has to be made in practice and controls the performance of the estimators. One thing that has to be noted here is that the user of a nonparametric smoothing technique should be aware that the final decision about an estimated regression curve is partly subjective since even asymptotically optimal smoothers contain a considerable amount of noise that leaves space for subjective judgment. It is therefore of great importance to make such a decision in interaction with the data, which means that ideally one should have a computer resource with some sort of interactive graphical display. Our main interest here is to show the best of these functions more specifically when we are considering kernel to be normal –the Gaussian density given by

$$\mathrm{K(x)}=\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}x^{2}\right] \qquad |x|<1$$

Several procedures of obtaining the bandwidth have been studied and our choice for the best bandwidth is to come up with several plots and select the bandwidth which outperforms the rest. The selection criterion here has been made fast and easy by using computer software S-plus where we have come up with a program which will perform this task. In our investigation here we compare the techniques of smoothing by using data obtained from Nairobi stock exchange. The best smooth parameter is therefore found by looking at the curves plotted using this program and hence choosing the one that adequately fits the data (that is the one which is optimal). In this data set, we vary smoothing parameters until we get the best. The best smoothing parameter that we get here, can be used straight forward to smooth any given data of any size instead of using trial and error method to get the best smoothing parameter.

6.Choosing the smoothing parameter

The problem of deciding how much to smooth is of great importance in non parametric regression. Hence in this section we will be focusing on finding a good way of choosing the smoothing parameter of various smoothing methods. The conditions required for a bandwidth selection rule to be “good” are, first of all it should have theoretical desirable properties secondly it has to be applicable in practice. Regarding the first condition there

have been a member of criteria proposed that measure in one way or another how the estimate is to the true curve.
Before embarking on technical solutions of the problem it is worth noting that a selection of the smoothing parameter is always related to a certain interpretation of the smooth (Norman R Draper and Harry Smith 1998).. If the purpose of smoothing is to increase the "signal to noise ratio" for presentation or to suggest a simple (parametric) model, then a slightly "over smoothed" curve with a subjectively chosen smoothing parameter might be desirable. On the other hand, when the interest is purely in estimating the regression curve itself with an emphasis on local structures then a slightly "under smoothed" curve may be appropriate. However, a good automatically selected parameter is always a useful starting (view) point. An advantage of automatic selection of the bandwidth for kernel smoothers is that comparison between laboratories can be made on the basis of a standardized method. Another advantage of the same lies in the application of additive models for investigation of high-dimensional regression data.

We compare the two smoothing techniques using simulated data sets.
Some of the smoothing techniques are;

(i) Kernel smoothing technique
(ii) Spline smoothing technique

Figure 1: shows a simulated data of size n =1000 data points with a kernel smoothing technique with smoothing parameters: h = 0.004, 0.175, 0.3 and 0.7. The blue curve whose smoothing parameter h=0.175 has the optimal bandwidth. It outperforms the rest as the best smoothing parameter. Smoothing parameters less than 0.175 would under smooth the data and smoothing parameters greater than 0.175 would over smooth the data.

Kernel smoothing technique

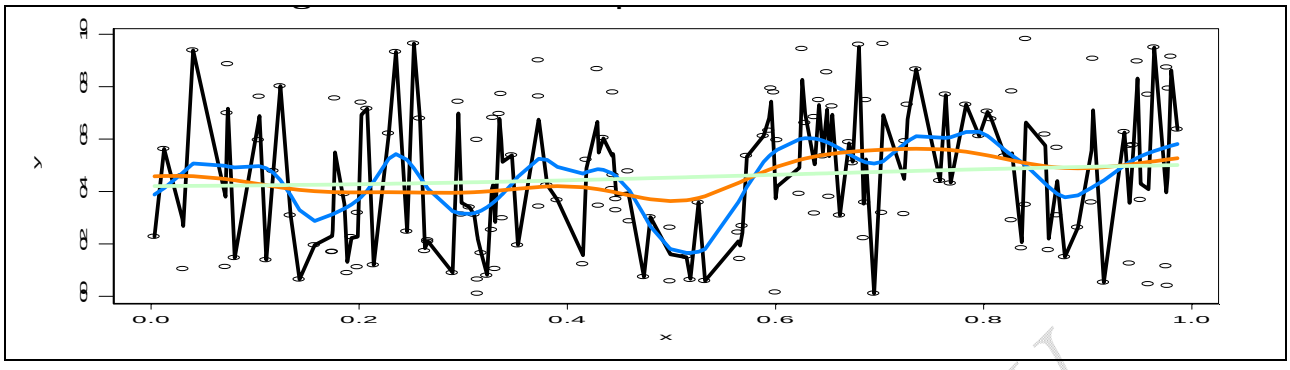


Figure 2: shows a simulated data of 100 data points with a kernel smoothing technique with smoothing parameters: h = 0.004, 0.175, 0.3 and 0.7. The blue curve whose smoothing parameter h=0.175, outperforms the rest as the best smoothing parameter in this data. Smoothing parameters less than 0.175 would under smooth the data and smoothing parameters greater than 0.175 would over smooth the data. In the case of an extremely over smooth, the curve produced would result in the bias estimate $\hat{m}$.

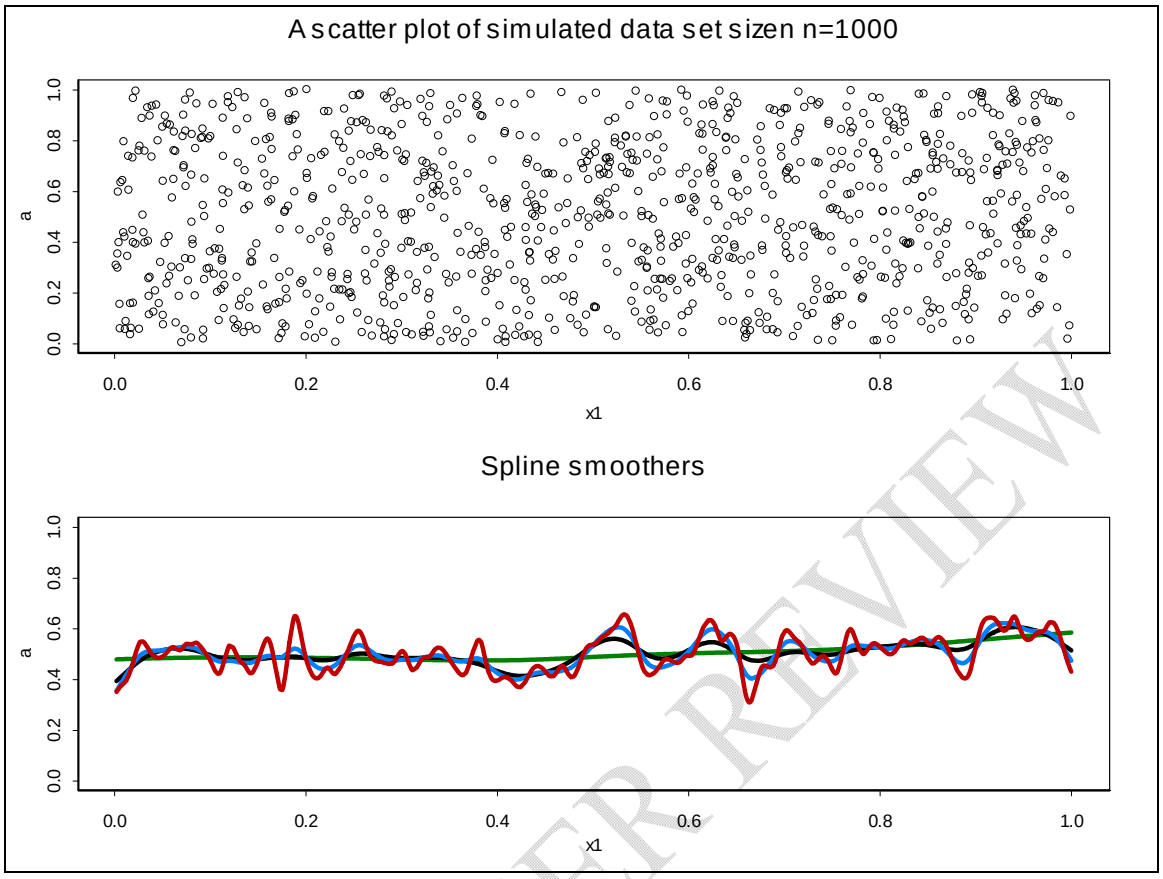


Figure 3: shows a simulated data of size n =1000 data points with a spline smoothing technique with smoothing parameters: λ = 5, 20, 35, 70. The blue curve whose smoothing parameter (bandwidth) λ = 35 has the optimal bandwidth. It looks fairly smooth and hence it outperforms the rest as the best smoothing parameter.

6. Empirical study (real data)

Here, we assume that the variance function is completely unspecified. The non-parametric approach in estimating a regression curve has its main purposes which includes:-

- Providing a versatile method of exploring the general relationship between two variables.
- Giving predictions of observations yet to be made without reference to a fixed parametric model.
- Providing a foot for finding spurious observations by studying the influence of isolated points.
- Constitutes a flexible method of substituting for missing values or interpolation between adjacent x-values.
- By the nature of flexibility the nonparametric method is helpful in preliminary and exploratory statistical analysis of a data set (J Brian Gray2000).

We specifically explore the possibility of situations where nothing is known about the variance function in heteroscedastic regression problems except that it is a smooth function of the design or mean response.

We denote the variance function by $V\left(x_i, \hat{\beta}\right)$ or V $\{c_i\}$. In estimating the variance function we use residuals. The residuals are defined as $y_i - f\left(x_i, \hat{\beta}\right)$.

Then the expectation of the squared residuals gives the estimate of the variance function given by

$E(r_i^2) = E\left[y_i - f\left(x_i, \beta\right)\right]^2 \cong V\left(x_i, \beta\right)$ (A Ian McLeod 1998)

We can also have the model in the design alone which is defined as

$$Var(y_i) = \sigma_i = V\left(c_i\right)$$

where V(.) is unknown

and $\{c_i\}$ is a set of identically and independent distributed random variables independent of $\{e_i\}$.In practical studies, to achieve the smooth conditions we use large data sets with the help of graphical enhancements and smoothing techniques.

We will use the kernel and cubic spline smoothers to portray the approach. A variety of kernel functions have been studied. However, both practical and theoretical considerations limit the choice of these kernels.

7.Example

We illustrate the two smoothing techniques from data obtained from the Nairobi stock exchange. We want to study the share volume in successive months over a period of 5 years 4 months of Kenya commercial bank. We will construct a regression model that relates to time (x_i) and (y_i) the share volume. The list of the variables are time, share volume in Kenyan shilling, residuals and residuals squared.

Figure 4

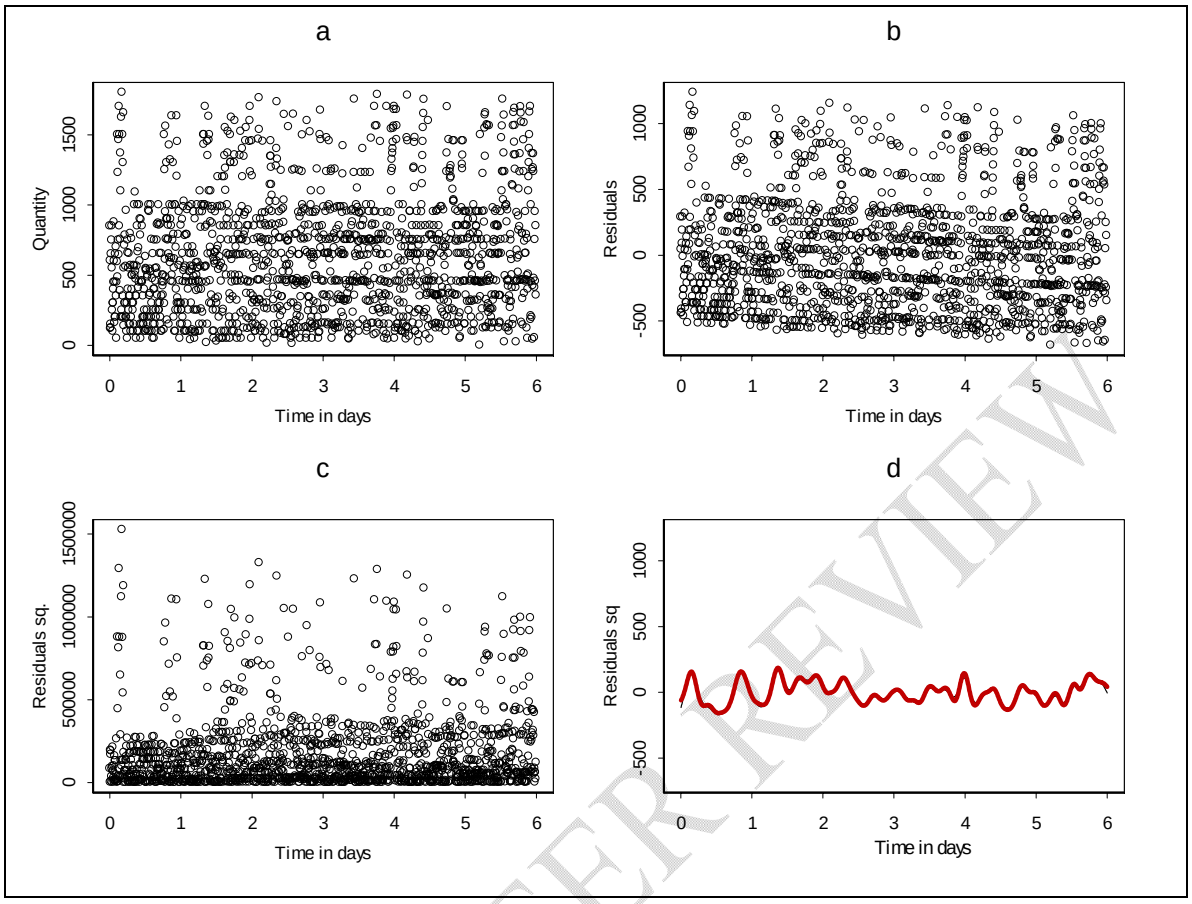


Figure 4: a) shows a scatter plot of share volume against time. b) Shows residuals squared against time. c) Shows residuals squared against time.
d) Shows both kernel and spline smooth of squared residuals against time.

8.Conclusion

Recommendation here is that the Kernel smoother produces the best estimate since its variance is less than that of the Spline smoother. From figure 3(d), we see that the spline has more variability around the middle and this shows that its variance is higher.
There is a great difference along the boundaries (that is the beginning and the end) and we recommend that more research be conducted to find out why there is such discrepancy.

COMPETING INTERESTS DISCLAIMER:

Authors have declared that they have no known competing financial interests OR non-financial interests OR personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

1 A Ian McLeod(1998). *Diagnostic check for monotone spread*. Computational statistics & data analysis, 26(4):437–443.

2 Carrlo, R.J and Ruppert, D (1987). *Variance function estimation*. Journal. of American statistical association **82**, 1079-1091

3 Hardle, W (1993). *Applied Nonparametric Regression*, Cambridge, U.K, Cambridge University press.

4 Sanford Weisberg (2005). *Applied linear regression, volume 528*. John Wiley & Sons.

5 Jobson. J.D and Fuller, W.A (1980). *Least squares estimation when the covariance matrix and the parameter vector are functionally related*. Journal of the American statistical Association

6 Norman R Draper and Harry Smith (1998). *Applied regression analysis*, volume 326. John Wiley & Sons.

7 *Silverman, B.W.[1986]. Density estimation for statistics and data analysis*, Chapman and Hall, London.

8 R Dennis Cook and Sanford Weisberg(2009). *Applied regression including computing and graphics*. John Wiley & Sons.

9 Silverman, Bernard W (1984), *Spline smoothing: the equivalent variable kernel method*, The annals of Statistics, pages 898–916.

10 J Brian Gray(2000). *Applied regression including computing and graphics*.

11 Elizbar A Nadaraya (1964). *On estimating regression*. Theory of Probability & Its Applications, 9(1):141–142