Original Research Article

Change point estimation in volatility of a time series with application to USD/KES exchange rate data set

Josephine Njeri Ngure[1*], Anthony Gichuhi Waititu[2], Simon Maina Mundia[3]

[1]Pan African University Institute for Basic Sciences, Technology and Innovation,
[2]Department of Statistics and Actuarial Sciences.
[3]Department of Statistics and Actuarial Sciences.
Email: ngurejosephine@gmail.com , awaititu@jkuat.ac.ke , simon.maina@dekut.ac.ke

Abstract

Background: A major problem, which a scientist is likely to encounter when analyzing data, is lack of homogeneity in the stochastic structure of the data. That is, there may be non-stationarity in the conditional variance function of the data. Undetected discontinuities within the structure of the data easily make the results of the analyses to be invalidated. Detection of structural changes in volatility of a time series is important in understanding volatility dynamics and the stylized facts observed in financial data. In this work, the conditional mean function of returns is estimated using Nadaraya Watson kernel estimator, residuals extracted and the conditional variance function estimated. A Kolmogorov-Smirnov type estimator for change point estimation in volatility of a time series is developed and applied to squared residuals. Theoretically and through simulations, it is shown that the estimator is consistent, depends on the size of the sample and the magnitude and the location of the change. The developed estimator is then used to estimate change points in USD/KSH exchange rate data set and two change points are identified corresponding to 11th May 2011 and 22nd February 2012.
Keywords: CUSUM, GARCH ICSS, Kernel, Nonparametric, NPCPM

***Address for Correspondence:**
Josephine Njeri Ngure, Pan African University Institute for Basic Sciences, Technology and Innovation.
Email: ngurejosephine@gmail.com
Received Date: 20/05/2020 Revised Date: 19/06/2020 Accepted Date: 10/07/2020

Access this article online

Quick Response Code:

Website: www.statperson.com

Accessed Date: 14 September 2020

INTRODUCTION

When a scientist seeks to analyze data, non-homogeneity in the stochastic structure of the data poses a challenge since the data may have some forms of discontinuity in terms of changing autocorrelation or mean or variances among others. This phenomena is referred to as structural breaks or structural changes and the associated statistical methodology is referred to as change point analysis. Change point analysis is of utmost interest because any undetected discontinuities within the structure of the data can easily lead to invalidation of any analytical results. Predictions and statistical inferences will be invalidated if changes in regimes during data collection stage (period) are ignored or not taken into consideration. A financial time series setting in which the data is sequentially ordered (measurements in time) is assumed, and an account of an unknown break point through a break point model is done. This is because many change point times usually go unrealized and undocumented. The change point is assumed to be abrupt as if occurring between any two observations and not gradual or smooth. The change point is chosen with the intention of maximizing the separation between two segments. Thus, the goal of change point detection and estimation is to recover these segments as accurately as possible. Finally, a mathematical model is created for each segment of homogeneity (Darkhovksy and Brodsky, 1986). Change point analysis is performed in either online (sequential) or offline setup (posterior). In online approach, one observes a sequence of time ordered observations sequentially from a given process and assume

How to site this article: Josephine Njeri Ngure, Anthony Gichuhi Waititu, Simon Maina Mundia. Change point estimation in volatility of a time series with application to USD/KES exchange rate data set. *International Journal of Statistika and Mathemtika*. August to October 2020; 35(1): 01-12. http://www.statperson.com

that at first the observations have the same distribution, where the process is said to be in control, or the process has changed at some unknown point τ ,where it is believed to have gotten out of control. In the offline approach, retrospective analysis of the entire series is performed. The offline algorithms consider the entire data set at once and look back in time to recognize where the change occurred. Change point detection approaches can be multiple change points where there is more than one change point or a single change point. The problem may also be parametric where the model generating the change points and its parameters are known or can be determined, or non-parametric among other scenarios. In stock returns data , volatility behaves more like a jump process where it fluctuates around some value for an extended time period before an abrupt change then after it fluctuates around another new value for some other extended time period (Tsay, 2005). The standard GARCH model fails to contain the possibility of these sudden jumps, and thus the degree of long term volatility persistence tends to be over-estimated (persistence in volatility tends to increase with the length of the sample often due to a switch in regimes somewhere in the sample (Diebold,1986). Hence, ignoring structural changes in financial returns series yields spurious persistence in volatility (close to unit root). The first published article concerning change points analysis was done by (Page, 1954) who considered testing for a potential single change point for data from a common parametric distribution motivated by a quality control setting in manufacturing. Since then, change point analysis has developed rapidly with considerations on either multiple change point detection and estimation, different types of data and other assumptions being put into consideration. The problem of change point in mean of observations was considered by (Darkhovksy and Brodsky, 1986) using the generalized variant of the Kolmogorov-Smirnov test statistic to check the equality of distribution. (Chen *et al.*, 2005) proposed a procedure that was able to combine the least squares approach which does not require specific forms of the marginal or the transitional density functions to estimate the change points in the conditional variance (volatility) of a non-parametric model of time. The proposed test was consistent and more powerful than the non-parametric ones already existing tests in literature. Finally, the practicality of the methods was by application to the Hong Kong stock market index (HSI) series under the independence assumption although it is well known that stock returns especially absolute and squared returns are auto correlated and there is dependency. Modeling of financial volatility in the presence of abrupt changes was done by (Ross, 2013) where the author incorporated the Iterative Cumulative Sum of Squares (ICSS) and Generalized Auto-Regressive Conditional Heteroscedastic (GARCH) algorithm to detect changes in volatility of financial returns. This was based on the assumption that financial returns follow a Gaussian distribution and thus the algorithm was found to produce very many spurious jump points if this assumption is violated. Also, ICSS algorithm worked best with weekly returns and detected spurious change points with daily returns. This method was replaced with Non Parametric Change Point Method (NPCPM) GARCH algorithm based on ranking of the observations raising the question on whether ranking affected the skewness of the observations. (Irungu, 2018) in her thesis considered change point in GARCH models with a restriction from GARCH(1,1) to GARCH (p, q) by use of autocorrelation function. This raised the simple question of whether the estimator would be used to estimate a change point from GARCH (p, q) to GARCH (p^*, q^*). The fitted GARCH model upon change point detection gave higher prices compared to Black Scholes option pricing model when the option is out of the money indicating that volatility dynamics affect the prices of the option. The Cumulative Sum (CUSUM) approach to change point detection and estimation has proved not to be robust to outliers and hence not a good approach to be applied in financial returns. This is because, returns have heavy tails due to the presence of outliers. Traditional methods of change point detection assume knowledge of the distribution before and after the change point like Exponential Weighted Moving Average (EWMA) and Generalized Likelihood Ratio (GLR) test. Though non-parametric methods of change point analysis have often been done by ranking of the observations, the rising question is whether ranking affects the skewness of the observations. Hence a non-parametric method in which ranking is not done is adopted in this paper in order to maintain the skewness property of the observations. Besides, a non-parametric Auto-Regressive Conditional Heteroscedastic model for the data set is assumed to avoid model misspecification associated with wrong parameter estimation. Daily returns are opted for accurate volatility modelling and a change point is estimated. The consistency of the change point estimator is shown through simulation. A significant improvement in describing a time series is expected once the change point in volatility is identified.

Section 2: NON-PARAMETRIC ARCH (NP-ARCH) MODEL

For any given time series, the features of interest are analyzed using non-parametric techniques. The characteristic under investigation is allowed to have a form, which is approximated with increasing precision, as the sample size grows unbounded. Conditional variances or conditional quartiles are required if interval forecasts or estimates of future volatility are desired as shall be necessary in this work. Let

$$X_t = log\left(\frac{S_t}{S_{t-1}}\right) \quad (1)$$

be the return series at equally spaced discrete time points for $t = 1,2, \dots T$ and $X_{t-1}, X_{t-2}, \dots, X_{t-d}$ be the return processes at any time periods less than t, $S_t$is the price process of the stock in period t for $t = 1,2, \dots T$. Assume that there is a non-parametric and non-linear relationship between X_t and X_{t-i} modelled by a non-linear auto regressive process of the form

$$X_t = m(X_{t-1}, X_{t-2}, \dots, X_{t-d}) + u_t \qquad t = 1,2, \dots, T \quad (2)$$

where u_ta series of innovations (random shocks) which is independent of $X_{t-1}, \dots, X_{t-d}$ satisfying

$$E(u_t|X_{t-1}, X_{t-2}, \dots, X_{t-d}) = 0 \quad (3)$$

$m(.)$ is the conditional mean (smooth) function in period t given past time periods $X_{t-1}, X_{t-2}, \dots, X_{t-d}$. Since in many situations point forecasting is too limited an objective, and the future volatility and higher order moments are of interest in addition to the conditional mean, the following representation of the innovation u_t holds

$$u_t = \sigma(X_{t-1}, X_{t-2}, \dots, X_{t-d})z_t \quad (4)$$

and thus equation (2) above is extended to a more general Non-Parametric Autoregressive Conditional Heteroskedastic (NP-ARCH) model

$$X_t = m(X_{t-1}, X_{t-2}, \dots, X_{t-d}) + \sigma(X_{t-1}, X_{t-2}, \dots, X_{t-d})z_t \quad (5)$$

$E(X_t|X_{t-1} = x_1, X_{t-2} = x_2, \dots, X_{t-d} = x_d) = m(x_{t-1}, x_{t-2}, \dots, x_{t-d})$ is the conditional mean function of the returns

variance$(X_t|X_{t-1} = x_1, X_{t-2} = x_2, \dots, X_{t-d} = x_d) =$

$E(\mu_t^2|X_{t-1} = x_1, X_{t-2} = x_2, \dots, X_{t-d} = x_d)\sigma^2(x_{t-1}, x_{t-2}, \dots, x_{t-d})$ is the conditional variance (smooth) function of the returns,

z_t is an independent and identically distributed sequence of random variables with

$$E(z_t|X_{t-1}, \dots, X_{t-d}) = 0 \quad (6)$$

$$\text{Variance}(z_t|X_{t-1}, \dots, X_{t-d}) = 1 \quad (7)$$

and independent of $X_{t-1}, X_{t-2}, \dots, X_{t-d}$ with from Gaussian distribution.

Equation (5) is the most flexible non-parametric time series model because it does not impose any (parametric) particular form on the conditional mean and conditional variance functions. Due to curse of dimensionality problem, whereas the dimension d grows the statistical and computational inefficiency comes in, d is set to one in this work so that equation (5) above becomes

$$X_t = m(X_{t-1}) + \sigma(X_{t-1})z_t \quad (8)$$

The Nadaraya Watson kernel regression estimator was first proposed independently by (Nadaraya, 1964) and (Watson, 1964) and hence the estimates of the functions $m(x)$ and $\sigma^2(x)$ are obtained by using Nadaraya Watson estimator of the unknown regression function (conditional mean at the evaluation points) and its properties such that

$$\hat{m}(x) = \frac{\sum_{t=2}^{T} K\left(\frac{X_{t-1}-x}{b_x}\right) X_t}{\sum_{t=2}^{T} K\left(\frac{X_{t-1}-x}{b_x}\right)} \tag{9}$$

is the conditional mean function at evaluation points, while,

$$\hat{\sigma}^2(x) = \frac{\sum_{2}^{T} K\left(\frac{X_{t-1}-x}{b_x}\right)\left(X_t - \hat{m}(X_{t-1})\right)^2}{\sum_{t=2}^{T} K\left(\frac{X_{t-1}-x}{b_x}\right)} \tag{10}$$

is the conditional variance function at evaluation points.

(Fan and Yao, 1998) showed that $\hat{m}(x)$ is a consistent estimator of $m(x)$. $K(.): \mathbb{R} \to \mathbb{R}$ is a kernel function, which is continuous, symmetric, integrating to one with bounded support $[-1,1]$ in that the estimator only uses the observations in the interval $(x - b_x, x + b_x)$ and b_x is the bandwidth parameter or the tuning parameter. A residual based estimator of $\hat{\sigma}^2(x)$is able to overcome the bias problem of (Härdle &Tsybakov, 1997). The bandwidth (smoothing parameter) controls the level of neighboring such that for a given kernel function K and a fixed x, observations $(X_{t-1}, X_t$with X_{t-1} far from $x$are given more weights as b_x increases. This means that the larger the bandwidth is chosen, the less the mean function $\hat{m}(x)$ is changing with x. Hence, the degree of smoothness of the conditional mean function increases with the bandwidth. Therefore, it means that a weighted average of the observations is used as an estimator for the conditional mean function. The estimators of the mean function and conditional variance function have shown to be strongly consistent and asymptotically normal for α mixing observations. The Epanechnikov kernel is employed since it is the most efficient in minimizing the Mean Integrated Squared error putting in mind that the choice of the kernel is not as important as the choice of the bandwidth (Altman, 1992) (this does not mean the choice of the kernel is disregarded).

When a kernel estimator is applied to dependent data, for example in financial time series returns data, then it is affected only by the dependence among the observations in a small window and not by that between all data. This fact therefore reduces the dependence between the estimates so that most of the techniques developed for independent data are applicable as well. This is referred to as the **Whitening by window principle** (Härdle, 1990). The memory of the underlying process decreases with distance between events and that the rate of decay can be estimated by the strong or α mixing condition stated below

α **(strong) mixing condition**: *Suppose the existence of a probability space* (Ω, F, P). *Let the dependence between two* σ *fields* $\mathcal{A}$ *and* $\mathcal{B} \subset F$ *be defined by*

$$\alpha(\mathcal{A}, \mathcal{B}) := \sup_{A \in \mathcal{A}, B \in \mathcal{B}} |p(A \cap B) - p(A)p(B)|$$

Suppose now $X := (X_t, t \in \mathbb{Z})$ *is a stationary two-sided sequence of random variables on a given probability space* (Ω, F, P). *For*$-\infty \le j < l \le \infty$, *let* $F_j^l = \sigma(X_t, j \le t \le l)$ *denote the* σ *−field of events which have been generated by the random variables* $X_t, j \le t \le l)$. F*or each* $\in N$ *, define the "coefficient of dependence (mixing)"*

$$\alpha(n) = \alpha(X, n): \sup_{-\infty \le j \le \infty} \alpha\left(F_{-\infty}^{j}, F_{j+n}^{\infty}\right).$$

The sequence of numbers $(\alpha(n), n \in N)$ *is non-increasing*. (Rosenblatt, 1956), *the random sequence* X_t *is thus said to be "strongly mixing" or* "α *−mixing" if*

$$\lim_{n\to\infty} \alpha(n) = 0$$

Actually, strong mixing sequences which include the Auto Regressive Moving Average (ARMA) processes, GARCH processes, m-dependent processes and ergodic markovian processes are asymptotically independent meaning that the statistical dependence between X_j and X_l tends to zero as the time interval $|j-l|$ increases.

Theorem 1: ***(Decay of correlations):*** *A stationary system is said to be mixing if and only if*

$$\lim_{l\to\infty} covariance\left(f(X_j), g(X_l)\right) = 0$$

for all bounded observables f, g.

$Covariance\ (X_j, X_l) \to 0$ as $l \to \infty$ for a sequence which is stationary and strong mixing.

Lemma 1: *For any time series X_t the following holds*

- *X_t is ergodic. This implies "average asymptotic independence" but not final total independence.*
- *X_{-t} is a mixing time series with similar mixing rates.*
- *If a strictly stationary process is mixing, it must be ergodic and hence mixing implies ergodicity.*

(Killick et al, 2012) *gives conditions on m, σ and the innovations that imply geometric ergodicity of $\{X_t\}$*

The second term of equation (8)

$$X_t = \sigma(X_{t-1})z_t \tag{11}$$

generates heavy tailed distributions which is shown by application of Jensen's inequality, with z_t having a standard normal distribution

$$\text{Kurtosis } (X_t) = \frac{E(X_t^4)}{\left[E(X_t^2)\right]^2} = \frac{E\left[\sigma^4(X_{t-1})z_t^4\right]}{\left[E\left(\sigma^2(X_{t-1})z_t^2\right)\right]^2} = 3\frac{E[\sigma^4(X_{t-1})]}{\left[E\left(\sigma^2(X_{t-1})\right)^2\right]} \geq 3 \tag{12}$$

This heavy tailed-ness feature implied by equation (8) makes it a successful model for modelling data, which exhibit heavy tails for example in financial returns data set.

Jensen's Inequality: (Pishro-Nik, 2016). *This inequality is stated by remembering that the variance of any random variable X is a positive value. That is*

$$Variance(X) = E(X^2) - (EX)^2 \geq 0$$

Thus $E(X^2) \geq (EX)^2$. Suppose $g(.): R \to R$ is a convex function for example $g(x) = x^2$. Also, assume that the expectations of X and $g(X)$ exist (is finite). Thus, the Jensen's inequality states that for any convex function g, the following relation holds

$$E[g(X)] \geq g(E[X]) .$$

Section 3: CHANGE POINT TEST STATISTIC

When doing change point analysis, the point of interest is to make a decision on whether the observations under study follow one model or if there is at least one moment in time (one time point) where the model undergoes a change.

Without a change point in volatility, equation (8) is re-written as

$$X_t = m(X_{t-1}) + \sigma_t(X_{t-1})z_t \quad t = 1,2,3,\ldots,T \tag{13}$$

where $m(x)$and $\sigma(x)$are the regression function (conditional mean) and volatility function respectively while z_t is the regression error which is assumed to be independent of X_{t-1} and fulfilling $E(z_t) = 0$, $E(z_t^2) = 1$, and $m_4 = E\{z_t^4\} < \infty$. This in turn implies that

$$E\{X_t - m(X_{t-1})\}^2 = \sigma_t^2(X_{t-1}) \tag{14}$$

Where $\sigma_t^2(X_{t-1})$is denoted by $\sigma_{(1)}^2(X_{t-1})$ for $t = 1,2,3,\dots,T$.
Equation (8) with a single change point in volatility is re-written as

$$X_t = m(X_{t-1}) + \sigma_t(X_{t-1})z_t \rightarrow E\{X_t - m(X_{t-1})\}^2 = \sigma_t^2(X_{t-1}) \tag{15}$$

Whereby in the presence of an unknown change point position $\tau \in \{2,3,\dots,T-1\}$

$$\sigma_t^2(X_{t-1}); = \begin{cases} \sigma_{(1)}^2(X_{t-1})\ for\ t = 1,2\dots,\tau \\ \sigma_{(2)}^2(X_{t-1})\ for\ t = \tau+1,\dots,T \end{cases} \tag{16}$$

This indicates presence of a structural change in conditional variance function (volatility) of the returns with time where τ is an unknown change point location (position) and $\tau + 1$ is where the change is starting. The intention is to locate (estimate) the change point position τ on the assumption that $z_t \sim N(0,1)$ and it is time invariant.
Hence, the hypothesis of equality of the conditional variances (volatility) of the returns is stated in terms of the conditional variance function as

$$H_0; = \sigma_t^2(X_{t-1}) = \sigma_{(1)}^2(X_{t-1})\ for\ t = 1,2,3,\dots,T.$$

While the at most one change point alternative states that for $\tau \in \{2,3,\dots,T-1\}$

$$H_A; \begin{cases} \sigma_t^2(X_{t-1}) = \sigma_{(1)}^2(X_{t-1})\ for\ t = 1,2\dots,\tau \\ \sigma_t^2(X_{t-1}) = \sigma_{(2)}^2(X_{t-1})\ for\ t = \tau+1,\dots,T \end{cases}$$

This implies $1 < \tau < T$ are candidate change points (Brodsky and Darkhovsky, 2013). Suppose the residuals obtained from the non-parametric estimation of conditional mean function and standardized using the conditional variances obtained from conditional variance function are defined as

$$\hat{\varepsilon}_t = \frac{X_t - \hat{m}(X_{t-1})}{\hat{\sigma}(X_{t-1})}, t = 1,2,\dots,T \tag{17}$$

where $\hat{m}(.)$ is the estimator of the conditional mean function as in equation (9). The interest is to test non-parametrically the hypothesis above by defining the partial sum of the squared residuals across all possible sample segments as

$$\varepsilon_T = \sum_{t=1}^{T} \hat{\varepsilon}_t^2,\ \varepsilon_\tau = \sum_{t=1}^{\tau} \hat{\varepsilon}_t^2, \varepsilon_{\tau+1} = \sum_{t=\tau+1}^{T} \hat{\varepsilon}_t^2 \tag{18}$$

When dealing with squared residuals, it is important to ensure that the sum of squared residuals is minimized. Define $\acute{\varepsilon}_{1,\tau}$ as the mean of the first τ residuals and $\acute{\varepsilon}_{\tau+1,T}$ as the mean of the last $T-\tau$ residuals written as

$$\acute{\varepsilon}_{1,\tau} = \frac{\varepsilon_\tau}{\tau} = \frac{1}{\tau}\sum_{t=1}^{\tau} \hat{\varepsilon}_t^2 \tag{19}$$

$$\acute{\varepsilon}_{\tau+1,T} = \frac{\varepsilon_{\tau+1}}{T-\tau} = \frac{1}{T-\tau}\sum_{t=\tau+1}^{T} \hat{\varepsilon}_t^2 \tag{20}$$

The change point estimator $\hat{\tau}$ of the unknown change point position τ is defined by

$$\hat{\tau} = argmin\left(\sum_{t=1}^{\tau}\left(\hat{\varepsilon}_t^2 - \acute{\varepsilon}_{1,\tau}\right)^2 + \sum_{t=\tau+1}^{T}\left(\hat{\varepsilon}_t^2 - \acute{\varepsilon}_{\tau+1,T}\right)^2\right) \tag{21}$$

$$\hat{\tau} = argmin\ V_\tau^2$$

Where $V_\tau^2 = \left(\sum_{t=1}^{\tau}\left(\hat{\varepsilon}_t^2 - \acute{\varepsilon}_{1,\tau}\right)^2 + \sum_{t=\tau+1}^{T}\left(\hat{\varepsilon}_t^2 - \acute{\varepsilon}_{\tau+1,T}\right)^2\right)$. This shows that the change point is estimated by minimizing the sum of squared residuals among the possible sample segments.
Suppose $\acute{\varepsilon}_{1,T} = \frac{1}{T}\sum_{t=1}^{T} \hat{\varepsilon}_t^2$ is the overall mean of the squared residuals. From (Bai, 1994) for each $\tau \in \{2,3,\dots,T-1\}$ the following relation holds

$$\sum_{t=1}^{T}\left(\hat{\varepsilon}_t^2 - \acute{\varepsilon}_{1,T}\right)^2 - T(D_t^T)^2 = V_\tau^2 \tag{22}$$

This implies that the minimum of V_t^2 will occur when D_t^T is maximum.

$$\hat{\tau} = arg\ \min_\tau(V_t^2) = arg\ \max_\tau(D_t^T)^2 = arg\ \max_\tau|D_t^T| \tag{23}$$

To obtain D_t^T , define the partial sum of the squared residuals among all possible sample segments as in equation (18) and assume the possibility of quantifying the deviation between $\sigma_{(1)}^2(X_{t-1})$ and $\sigma_{(2)}^2(X_{t-1})$ written as $l_p(\sigma_{(1)}^2(X_{t-1}) - \sigma_{(2)}^2(X_{t-1}))$. The concern is to maximize the distance between the minimized sums of squared residuals where for $p \geq 1$

$$l_p\left(\sigma_{(1)}^2(X_{t-1}) - \sigma_{(2)}^2(X_{t-1})\right) = \left(\sum_{t=1}^{T} w_\tau|\varepsilon_\tau - \varepsilon_{\tau+1}|^p\right)^{\frac{1}{p}} \tag{24}$$

Motivated by the l_p norm and properties of l_p space, the change point statistic is developed by setting $p = 2$. By application of the reverse triangle inequality (which gives the lower bound instead of the upper bound) and which follows from the regular triangle inequality one obtains

$$l_p\left(\sigma_{(1)}^2(X_{t-1}) - \sigma_{(2)}^2(X_{t-1})\right) = \left(\sum_{t=1}^{T} w_\tau|\varepsilon_\tau - \varepsilon_{\tau+1}|^2\right)^{\frac{1}{2}}$$

$$=\left(\sum_{t=1}^{T} w_\tau^{\frac{1}{2}}\left|\varepsilon_\tau-\varepsilon_{\tau+1}\right|\right)$$

$$=E\left(w_\tau^{\frac{1}{2}}\left|\varepsilon_\tau-\varepsilon_{\tau+1}\right|\right)\geq w_\tau^{\frac{1}{2}}\left|\mathbb{E}(\varepsilon_\tau)-\mathbb{E}(\varepsilon_{\tau+1})\right| \quad (25)$$

Therefore, the change point estimator is developed from a process generated by $l_2\left(\sigma_{(1)}^2(X_{t-1})-\sigma_{(2)}^2(X_{t-1})\right)$ as

$$w_\tau^{\frac{1}{2}}\left|\mathbb{E}(\varepsilon_\tau)-\mathbb{E}(\varepsilon_{\tau+1})\right| = w_\tau^{\frac{1}{2}}\left|\frac{1}{\tau}\sum_{t=1}^{\tau}\hat{\varepsilon}_t^2-\frac{1}{T-\tau}\sum_{t=\tau+1}^{T}\hat{\varepsilon}_t^2\right| \quad (26)$$

$w_\tau^{\frac{1}{2}}$ is a weight function, which is measurable and which depends on the sample size T and the change point position τ . It gives the sensitivity of the test statistic against different alternatives in the sense of the position of change. The weight function is chosen so that it satisfies the condition that

$$\sum_{t=1}^{T}\hat{\varepsilon}_t^2=\frac{\tau}{T}\sum_{t=1}^{T}\hat{\varepsilon}_t^2\rightarrow\frac{1}{T}\left(\sum_{t=1}^{\tau}\hat{\varepsilon}_t^2-\frac{\tau}{T}\sum_{t=1}^{T}\hat{\varepsilon}_t^2\right)=0 \quad (27)$$

This is obtained from the relation

$$w_\tau\left(\frac{1}{\tau}\sum_{t=1}^{\tau}\hat{\varepsilon}_t^2-\frac{1}{T-\tau}\sum_{t=\tau+1}^{T}\hat{\varepsilon}_t^2\right)=\frac{1}{T}\left(\sum_{t=1}^{\tau}\hat{\varepsilon}_t^2-\frac{\tau}{T}\sum_{t=1}^{T}\hat{\varepsilon}_t^2\right)$$

$$=\frac{1}{T}\left(\sum_{t=1}^{\tau}\hat{\varepsilon}_t^2-\frac{\tau}{T}\sum_{t=1}^{\tau}\hat{\varepsilon}_t^2-\frac{\tau}{T}\sum_{t=\tau+1}^{\tau}\hat{\varepsilon}_t^2\right) \quad (28)$$

Simplifying equation (28) by introducing $\frac{\tau}{\tau}$ and $\frac{T-\tau}{T-\tau}$, the test statistic for change point detection as from equation 26 becomes

$$D_t^T=\left(\frac{\tau}{T}\left(1-\frac{\tau}{T}\right)\right)^{\frac{1}{2}}\left|\frac{1}{\tau}\sum_{t=1}^{\tau}\varepsilon_t^2-\frac{1}{T-\tau}\sum_{t=\tau+1}^{T}\varepsilon_t^2\right| \quad (29)$$

This is a Kolmogorov Smirnov (KS) type test statistic. The KS type test statistic does not utilize any a priori information about the data and hence it is non-parametric.

Section 3.1: Single change point estimator

The KS type estimator of change point will be the point where the KS type test statistic has its global maximum. This is because, the global maximum will often occur at the area of true change point (the point where there is maximum distance between the average of the squared residuals). Hence, a good choice of the estimator for the time of change is as given as

$$\hat{\tau}=\underset{\tau}{argmax}\left|D_t^T\right| \quad (30)$$

The estimate $\hat{\tau}$ is the point at which there is maximal sample evidence for a break in the squared residual process. In the presence of a single break it shall be shown that $\hat{\tau}$ is a consistent estimator of the true change point position τ

Section 4: CONSISTENCY OF THE CHANGE POINT ESTIMATOR

Here, one needs to show that the change point estimator $\hat{\tau}$ approaches the true change point position τ under the alternative hypothesis as the sample size grows unbounded. Consistency results in change point problems only deal with change point fractions and not the time indexes themselves. Distance between a true break point index and its estimated counterpart usually never converge to zero even for the simple models. Hence, the distance between the true change point index and its estimated counterpart is obtained, then normalized by the size of sample (Truong, Oudre and Vayatis, 2018). For a consistent estimator, this ratio decreases to zero as the size of the sample grows unbounded meaning $\frac{|\hat{\tau}-\tau|}{T}\rightarrow 0$ as $n\rightarrow\infty$. Define $\hat{k}=\frac{\hat{\tau}}{T}$ which is the fraction of the change point estimate and the "true" change point fraction $k^*=\frac{\tau}{T}$. Then under the alternative hypothesis, $\hat{k}\xrightarrow{p}k^*$ which in turn implies that $\hat{\tau}$ is consistent for τ .

Theorem 2: *Consider a sample of squared residuals $\hat{\varepsilon}_1^2,\hat{\varepsilon}_2^2,\ldots\hat{\varepsilon}_T^2$ satisfying the alternative hypothesis and the change point estimator $\hat{\tau}$ given in equation (30) above. If the sequences $\{\hat{\varepsilon}_{1,t}^2, t\in\mathbb{Z}\}$ and $\{\hat{\varepsilon}_{1,t}^2, t\in\mathbb{Z}\}$ satisfy*

$$\Delta:=\sigma^2{}_{(1)}(X_{t-1})-\sigma^2{}_{(2)}(X_{t-1})\neq 0$$

Where $\Delta<\infty$ denotes the finite magnitude of the jump in the conditional variance function then for $\hat{k}=\frac{\hat{\tau}}{T}$

$$P\left\{\left|\hat{k}-k^*\right|>\epsilon\right\}\leq\frac{B}{\epsilon^2\Delta^2}T^{-\frac{1}{2}}$$

Where $0<B<\infty$ (is a positive constant) and $k^*=\frac{\tau}{T}$ (Kokoszka and Leipus, 2000).

To prove the theorem above, start by supposing that $\{\hat{\varepsilon}^2_{1,t}, t \in \mathbb{Z}\}$ and $\{\hat{\varepsilon}^2_{2,t}, t \in \mathbb{Z}\}$ are two conditional heteroscedastic processes of residuals. Suppose that one obtains a sample $\hat{\varepsilon}^2_1, \hat{\varepsilon}^2_2, \dots, \hat{\varepsilon}^2_T$ such that

$$\hat{\varepsilon}^2_t := \begin{cases} \hat{\varepsilon}^2_{1,t} \; if \; 1 \le t \le \tau \\ \hat{\varepsilon}^2_{2,t} \; if \; \tau < t \le T \end{cases} \tag{31}$$

Also, assume that the two sequences have different conditional variance functions such that

$$\mathbb{E}(\hat{\varepsilon}^2_{i,t}) := \begin{cases} \sigma^2{}_{(1)}(X_{t-1}) \; for \; 1 \le t \le \tau \; when \; i = 1 \\ \sigma^2{}_{(2)}(X_{t-1}) \; for \; \tau < t \le T \; when \; i = 2 \end{cases} \tag{32}$$

where $\tau = [k^*T]$, $k^* \in (0,1)$ and $[.]$ represents the largest integer less than or equal to its argument. One needs to argue that the objective function D^T_t converges in probability to a non-stochastic function of parameters which has a unique global maximum at τ and that $(D^T_t - E(D^T_t)$ is small in τ for large T.

The first sum is weighted by $\frac{\tau}{T}$ denoted by k^* which is the percentage of the observations before the change point while the second sum is weighted by $1 - \frac{\tau}{T}$ denoted by $1 - k^*$ which is the percentage of the observations after the change point. One needs to note that $|\, E(D^T_t)|$ will achieve its maximum at $t = \tau$ which results to

$$E(D^T_t) = \begin{cases} \Delta k^{\frac{1}{2}}(1-k^*)^{\frac{1}{2}} \; if \; t \le \tau \\ \Delta k^{*\frac{1}{2}}(1-k)^{\frac{1}{2}} \; if \; t > \tau \end{cases} \tag{33}$$

Thus

$$E(D^T_\tau) = k^{*\frac{1}{2}}(1-k^*)^{\frac{1}{2}} \tag{34}$$

From equations (32) and (33) one obtains

$$|ED^T_\tau| - |E(D^T_t| = \begin{cases} |\Delta|(1-k^*)^{\frac{1}{2}} \left(k^{*\frac{1}{2}} - k^{\frac{1}{2}}\right) \; if \; t \le \tau \\ |\Delta| k^{*\frac{1}{2}}((1-k^*)^{\frac{1}{2}} - (1-k)^{\frac{1}{2}}) \; if \; t > \tau \end{cases} \tag{35}$$

If $k \le k^*$, then by the mean value theorem

$$k^{*\frac{1}{2}} - k^{\frac{1}{2}} \ge \frac{1}{2} k^{*-\frac{1}{2}}(k^* - k) \tag{36}$$

If $k > k^*$, then $1 - k < 1 - k^*$and equation (36) above yields

$$(1-k^*)^{\frac{1}{2}} - (1-k)^{\frac{1}{2}} \ge \frac{1}{2}(1-k^*)^{-\frac{1}{2}}(k - k^*) \tag{37}$$

Combining equations 35, 36 and 37 one obtains

$$|ED^T_\tau| - |E(D^T_t)| \ge |\Delta| \frac{1}{2} k^{*-\frac{1}{2}}(1-k^*)^{-\frac{1}{2}} \min\{k^*, 1-k^*\}|k^* - k| \tag{38}$$

Also,

$$D^T_t - D^T_\tau = [D^T_t - E(D^T_t) + E(D^T_t)] - [D^T_\tau - E(D^T_\tau) + E(D^T_\tau)] \tag{39}$$

Which yields

$$|D^T_t| - |D^T_\tau| \le |D^T_t - E(D^T_t)| + |E(D^T_t)| + |D^T_\tau - E(D^T_\tau)| - E(D^T_\tau)$$
$$\le 2 \max_\tau |D^T_t - ED^T_t|) + |ED^T_t| - |ED^T_\tau| \qquad (40)$$

Therefore from equations 38 and 40, the following is obtained

$$|\Delta| \frac{1}{2} k^{*-\frac{1}{2}}(1-k^*)^{-\frac{1}{2}} \min\{k^*, 1-k^*\} |k^* - k| \le |ED^T_\tau| - |ED^T_t|$$
$$\le 2\,(\max_\tau |D^T_t - ED^T_t|) + |D^T_\tau| - |D^T_t| \tag{41}$$

Replacing k by $\hat{k}$ in equation 41 above and noting that $|D^T_\tau| \le D^T_{\hat{\tau}}$ one obtains

$$|\Delta| \frac{1}{2} k^{*-\frac{1}{2}}(1-k^*)^{-\frac{1}{2}} \min\{k^*, 1-k^*\}|k^* - k| \le 2\,(\max_\tau |D^T_t - ED^T_t|) \tag{42}$$

Algebraic calculations with application of Hajek and Renyi Inequality for dependent variables stated in (Kokoszka and Leipus, 2000) establishes the consistency of the change point fraction hence proving theorem 2.

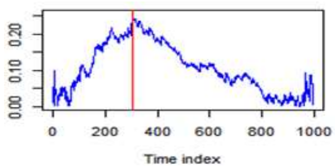
Section 5: EMPIRICAL STUDY

Subsection 5.1: Simulation study

The performance of the estimator is investigated by simulating an ARMA (1, 1) ARCH (1) model by supposing the model has a single change point in the volatility function defined as

$$X_t = 0.35X_{t-1} + \varepsilon_t + 0.4\varepsilon_{t-1}$$
$$\sigma_t^2(X_{t-1}) = \begin{cases} 1.5 + 0.1\varepsilon_{t-1}^2 \; for\; t = 1,2,\ldots,\tau \\ 1 + 0.1\varepsilon_{t-1}^2 \; for\; t = \tau+1,\ldots,T \end{cases} \tag{43}$$

ε_t=$\sigma_t(X_{t-1})z_t$ and $z_t \sim i.i.dN(0,1)$. At first 1000 returns with the change point fixed at $\tau = 330$ which is $\frac{1}{3}T$ were simulated. Meaning model one is assumed for $t \in [1{:}\,330]$ while model two was assumed for $\in [331{:}\,1000]$. The estimate of τ was 314. The change point was further fixed at $\tau = 500$ which is $\frac{1}{2}T$. Meaning model one was assumed for $t \in [1{:}\,500]$ while model two assumed for $t \in [501{:}\,1000]$. The estimate of τ was 492. Finally, the change point was fixed at $\tau = 670$ which is $\frac{2}{3}T$ meaning model one was assumed for $t \in [1{:}\,670]$ while model two assumed for $\in [671{:}\,1000]$. The estimate of τ was 654. The plots of the test statistic for the three change point estimates are as shown in figures 1, 2, and 3 below respectively.



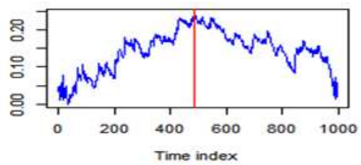


Figure 1: Change point estimated at $\tau = 314$; **Figure 2:** Change point estimated at $\tau = 492$

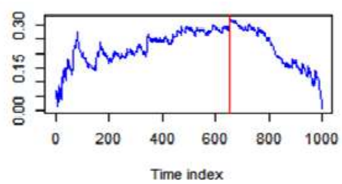
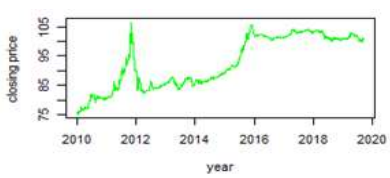


Figure 3: Change point estimated at $\tau = 654$.

Subsection 5.2: Real data set application

Change point estimation is done on USD/KES daily exchange rate data set from 2 January 2010 to 22 March 2019 with $T = 2444$ observations. The exchange rates plot is as in figure 4 below.



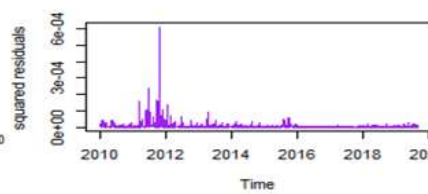
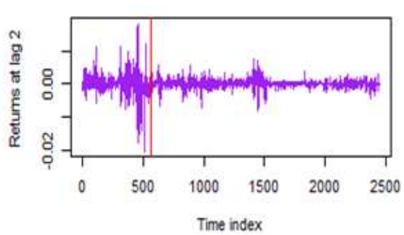


Figure 4: USD /KSH exchange rate; **Figure 5:** The squared residuals plot



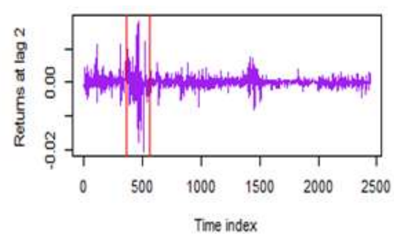
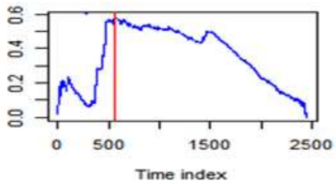


Figure 6: Change point at point 560; **Figure 7:** Change points at point 358 and 560

The returns with two estimated change points, $\hat{\tau}_1 = 560$ corresponding to 22nd February 2012 and $\hat{\tau}_2 = 358$ corresponding to 11th May 2011 is as in figures 6 and 7 above.



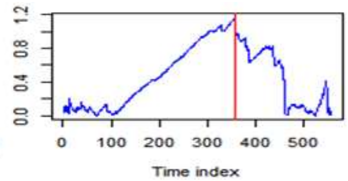


Figure 8: Change point statistic, $\hat{\tau}_1 = 560$; **Figure 9:** Change point statistic, $\hat{\tau}_2 = 358$

The corresponding plots of the change point statistic with change points estimated at $\hat{\tau}_1 = 358$ and $\hat{\tau}_2 = 560$ were as shown in figures 8 and 9 above respectively.

SUBSUBSECTION 5.2.1: RESULTS AND DISCUSSIONS

From the exchange rate plot, figure 4 there is an increasing trend between January 2010 and October 2011 where the exchange rate prices were at the peak. Afterwards, a decreasing trend is observed after which the exchange rates started to rise again. The USD/KSH exchange rate depreciated from 83.89 to 101.39 between April 2011 and October 2011 (Kibiy and Nasieku, 2016). The Report of the parliamentary select committee dated February 2012 showed a combination of economic, human and institutional failures led to the decline of the Kenyan shilling against major world currencies an example of the United States Dollar. Some of the economic causes included a wide current account deficit due to a rise in imports of machinery and transport equipment, which are key inputs for the manufacturing sector, the debt crisis in the Euro zone, large import bill of some non-essential commodities, the Arab Spring which included Islamic and Arab political turbulences. This negatively affected tourism and horticultural export not just in Kenya but also in East Africa as a whole while drought and raising cost of fuel significantly affected the overall industrial and agricultural production. This made the shilling depreciate against major currencies like the US dollar due to reduced foreign currency inflows because of declining exports and tourism activity. The constitutional and human failures included but not limited to policy indecisiveness and inaction by the Central Bank of Kenya, which had the capacity to stop the slide, some banks could have been engaging in speculative activities and inaction by the ministry of Finance, which also had the power to intervene once Central Bank of Kenya failed to do so. For the change point in volatility in February 2012, from the Annual Report and Financial Statements from Bamburi Cement, the economy in Kenya reported to have experienced slow growth at the beginning of 2012 following high inflation and high interest rates from commercial banks. The foreign exchange market witnessed significant volatility between May 2011 and October 2011 as seen in figures 6 and 7 reflecting the general volatility in the global financial markets as well as increase in demand for foreign exchange to finance imports. This result was that the Kenyan Shilling like other currencies in the region and other global markets therefore weakened substantially for example the Kenyan shilling against the US dollar depreciated from an average of 84.2 in March 2011 to 101.39 in October 2011 (20.42% in percentage depreciation).

Subsection 5.3: Simulations on consistency of the change point estimator.

1000 bootstrap samples of ARMA (1,1) ARCH (1) model are simulated under different sample sizes and magnitudes of change by supposing the model has a single change point in the volatility function defined as

$$X_t = 0.35X_{t-1} + \varepsilon_t + 0.4\varepsilon_{t-1}$$

$$\sigma_t^2(X_{t-1}) = \begin{cases} (1+\Delta) + 0.1\varepsilon_{t-1}^2 \; for \; 1 \le t \le \tau^* \\ 1 + 0.1\varepsilon_{t-1}^2 \; for \; \tau^* < t \le T \end{cases} \tag{44}$$

$\varepsilon_t = \sigma_t(X_{t-1})z_t$, where $z_t \sim i.i.dN(0,1)$. $\Delta = \{0.3, 0.5, 0.8\}$ and the sample sizes are $T = \{500, 1000, 2000, 4000\}$. The "true" change point was fixed at $\frac{1}{3}T = 0.3T$, $\frac{1}{2}T = 0.5T$ and $\frac{2}{3}T = 0.67T$. The mode of the various estimates of the change point after 1000 bootstrap samples is obtained as the estimate of the change point. In each simulation, the estimates of the change point highly depended on the sample size, the location of the change point as well as the magnitude of change. The estimates of the change point are more accurate when the magnitude of change is large and as the sample size grows unbounded.

Table 1: Numerical results for consistency of the change point estimator

T	τ	$\hat{\tau}$ when $\Delta = 0.3$	$\frac{\hat{\tau}-\tau}{T} \to 0$	$\hat{\tau}$ when $\Delta = 0.5$	$\frac{\hat{\tau}-\tau}{T} \to 0$	$\hat{\tau}$ when $\Delta = 0.8$	$\frac{\hat{\tau}-\tau}{T} \to 0$
500	$0.3T$ =166	149	0.03407	159	0.01406	164	0.00402
	$0.5T = 249$	243	0.01205	246	0.00602	246	0.00602
	$0.67T = 332$	325	0.01406	326	0.01205	329	0.00803

1000	$0.3T$ =332	323	0.00902	326	0.00601	329	0.00301
	$0.5T = 499$	490	0.00902	496	0.00301	496	0.00100
	$0.67T = 666$	655	0.01102	662	0.00401	663	0.00301
2000	$0.3T$ =666	657	0.00551	660	0.00300	664	0.00100
	$0.5T = 999$	993	0.00300	997	0.00100	998	0.00050
	$0.67T = 1332$	1325	0.00350	1326	0.00300	1330	0.00100
4000	$0.3T$ =1332	1322	0.00250	1328	0.00100	1328	0.00100
	$0.5T = 1999$	1994	0.00125	1996	0.00075	1998	0.00025
	$0.67T = 2666$	2657	0.00225	2659	0.00018	2663	0.00075

SECTION 6: CONCLUSION AND RECOMMENDATIONS

In this paper, a non-parametric procedure to estimate change point in volatility of a time series has been proposed. The consistency of the change point estimator is proven both theoretically and through simulations. The change point estimator is able to better estimate the change point when it is around the middle of the series, and in larger samples than in smaller samples. The estimator seemed to work better in samples with large magnitude of shifts than in samples with small magnitudes of shifts. Non-parametric modeling is important in finance and non-parametric estimators are very powerful in distinguishing among many models like derivative pricing models. One can easily extend the method to multidimensional non-parametric models (models of higher dimension) of the form $X_t = m(X_{t-1}, X_{t-2}, \dots, X_{t-d}) + \sigma(X_{t-1}, X_{t-2}, \dots, X_{t-d})z_t$ where the regression function and the conditional variance functions should be estimated using multivariate kernel methods. $m\,(.)\,, \sigma(.)$ are multiple variable d functions, while at the same time being careful on how to deal with the curse of dimensionality problem which may lead to poor performance in higher dimensional regression problem, since for $d > 2$, the subspace of R^{d+1} spanned by the data is rather empty. A case where both the conditional mean function and the conditional variance function are changing is a future research problem.

*The authors greatly acknowledge Pan African University Institute for Basic Sciences, Technology, and Innovation (PAUSTI) in collaboration with World Bank through African Union for funding this research.

The USD/KSH exchange rate data set used to support the findings of this study is available from The Central Bank of Kenya https://www.centralbank.go.ke.

REFERENCES

1. Altman, N, S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, *46*(3), 175–185.
2. Bai, J. (1994). Least squares estimation of a shift in linear processes. *Journal of Time Series Analysis*, *15*(5), 453–472.
3. Brodsky, E., and Darkhovsky, B. S. (2013). *Nonparametric methods in change point problems* (Vol. 243). Springer Science and Business Media.
4. Chen, G., Choi, Y. K., and Zhou, Y. (2005). Nonparametric estimation of structural change points in volatility models for time series. *Journal of Econometrics*, *126*(1), 79–114.
5. Darkhovsky, B. S., and Brodsky, B. E. (1986). Nonparametric method of the quickiest detection of a change in mean of a random sequence. *Theory of Probability and Applications*, *32*(4), 703–711.
6. Diebold, F. X. (1986). *Temporal aggregation of ARCH processes and the distribution of asset returns* (No.200). Board of Governors of the Federal Reserve System (US).
7. Fan, J., andYao, Q. (1998). Efficient estimation of conditional variance functions in stochastic regression. *Biometrika*, *85*(3), 645–660.
8. Härdle, W. (1990). *Applied nonparametric regression* (No 19). Cambridge university press.
9. Härdle, W., and Tsybakov, A. (1997). Local polynomial estimators of the volatility function in nonparametric autoregression. *Journal of Econometrics*, *81*(1), 223–242.
10. Irungu, I. W. (2018). *Manhattan Distance for Model Order Change- Point Estimation in GARCH Models*. Thesis JKUAT.
11. Kibiy, J., and Tabitha , N. (2016). Determinants of Exchange Rate Volatility of the Kenyan Shilling Against World Major Currencies. *International Journal of Social Sciences and Information Technology*, *2*(11), 1181–1198.
12. Killick, R., Fearnhead, P., and Eckley, I. A. (2012). Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, *107*(500), 1590–1598.
13. Kokoszka, P., and Leipus, R.. (2000). Change-point estimation in ARCH models. *Bernoulli*, *6* (3), 513–539.
14. Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability and Its Applications*, *9* (1), 141–142.
15. Page, E. S. (1954). Continuous inspection schemes. *Biometrika*, *41*(1/2), 100–115.
16. Pishro-Nik, H. (2016). *Introduction to probability, statistics, and random processes*.
17. Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proceedings of the National Academy of Sciences of the United States of America*, *42*(1), 43.
18. Ross, G. J. (2013). Modelling financial volatility in the presence of abrupt changes. *Physica A: Statistical Mechanics and Its Applications*, *392*(2), 350–360.
19. Truong, C., Oudre, L.,and Vayatis, N. (2018). A review of change point detection methods. *ArXiv Preprint ArXiv:1801.00718*.
20. Tsay , R. S (2005). *Analysis of Financial Time Series,* New York: Wiley.
21. Watson, G. S. (1964). Smooth regression analysis.

Sankhy{\=a}: The Indian Journal of Statistics, Series A, 359–372.

Source of Support: None Declared
Conflict of Interest: None Declared