

Consistency of an Estimator for Change Point in Volatility of Financial Returns

Josephine Njeri Ngure¹, Anthony Gichuhi Waititu², Simon Maina Mundia³

¹ Pan African University Institute for Basic Sciences, Technology and Innovation

² Department of Statistics and Actuarial Sciences, Jomo Kenyatta University of Agriculture and Technology

³ Department of Statistics and Actuarial Sciences, Dedan Kimathi University of Technology

Correspondence: Josephine Njeri Ngure, Pan African University Institute for Basic Sciences, Technology and Innovation

Received: December 27, 2020 Accepted: January 25, 2021 Online Published: January 26, 2021

doi:10.5539/jmr.v13n1p56 URL: <https://doi.org/10.5539/jmr.v13n1p56>

Abstract

A non parametric Auto-Regressive Conditional Heteroscedastic model for financial returns series is considered in which the conditional mean and volatility functions are estimated non-parametrically using Nadaraya Watson kernel. A test statistic for unknown abrupt change point in volatility which takes into consideration conditional heteroskedasticity, dependence, heterogeneity and the fourth moment of financial returns, since kurtosis is a function of the fourth moment is considered. The test is based on L_2 norm of the conditional variance functions of the squared residuals. A non-parametric change point estimator in volatility of financial returns is further obtained. The consistency of the estimator is shown theoretically and through simulation. An application of the estimator in change point estimation in volatility of United States Dollar/Kenya Shilling exchange rate returns data set is made. Through binary segmentation procedure, three change points in volatility of the exchange rate returns are estimated and further accounted for.

Keywords: non parametric, kernel, volatility

1. Introduction

A change point is a situation where an entire data set is no longer characterized by the same underlying process. Instead, the observations have two or more distinct segments and each segment has a unique underlying process. The segments could have different variances or different mean. Change points are as a result of an observed event or an unobserved combination of factors. These include among others financial liberalization of emerging markets and integration of world equity markets, changes in exchange rate regimes from a fixed exchange rate regime to a floating exchange rate regime and introduction of a single currency like the Euro in Europe. Synonyms for change point include probabilistic diagnostics and disorder problems.

In financial instruments, volatility is rarely constant but instead behaves like a jump processes where it varies over time creating temporal clusters due to volatility clustering. Volatility clustering means large or small price movements tend to be followed by similarly large or small price movements respectively on consecutive hours, days, weeks or other time durations. The clustering of volatility brings about the presence of structural breaks or change points which the standard GARCH model fails to accommodate leading to overestimation of the degree of volatility persistence.

Unlike previous research which considers change points in identical or independent variables or change point in the mean of variables Csorgo and Horváth (1997), Shao and Zhang (2010), the concern here is on dependent heterogeneous processes with finite fourth moment since kurtosis is a function of the fourth moment. The change is assumed to be abrupt (as though it occurs completely between two observations) thus creating piecewise stationary segments and not smooth or gradual which could create locally stationary segments. The change point location is not known a priori as is common with most previous studies. Due to the fact that inferences based on non-parametric approaches are robust against model misspecification, non-parametric models are able to avoid misspecification problem commonly encountered in parametric approaches. Hence, kernel estimator of conditional mean function is employed to estimate the conditional mean function and the conditional variance function in financial returns. Change point estimation is further done on the conditional variance function of the returns series.

The first published article in change point analysis was by Page (1954). He considered testing for a potential single change point in mean for independent and identically distributed normal random variables 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, parametric test statistics which are based on the likelihood

ratio statistic and estimation done using maximum likelihood method Csorgo and Horváth (1997) and nonparametric test statistics and other assumptions being put into consideration. An extensive literature on change point detection and estimation in independent and identically distributed random variables is by Csorgo and Horváth (1997) where mostly a change in mean is considered, or in martingale difference sequences as in Bai (1994) or even change in variance under the assumption that financial returns are independent and identically distributed random variables Inclan and Tiao (1994) which is not a fact when one is working with real financial data. It has over time been done in time series like Chen et al. (2005), Ross (2013), Aminikhanghahi and Cook (2017) under different assumptions on dependence, heterogeneity among others Gerstenberger (2018).

The change point detection and estimation methods can be classified into offline (fixed sample or retrospective) and online (sequential) where the classification depends on the sample acquisition approach Brodsky and Darkhovsky (2013), Csorgo and Horváth (1997) and Chen and Gupta (2011). In offline approach, the data set is first observed, then the detection and further estimation of the change-point is done by looking back in time to recognize where the change occurred. These include among others Lee et al. (2003) where a change point in variance in a non-parametric time series regression model with a strong mixing error process using cumulative sum of squares approach introduced by Inclan and Tiao (1994) is done, Sansó et al. (2020) in change point detection in unconditional variance of financial time series. In sequential approaches, new data is continually arriving and is analyzed adaptively. The goal of on-line change point detection is to detect changes as quickly as possible, while keeping the number of false alarms as low as possible. On-line approaches are mostly used in areas such as quality control, financial risk management, allocation of asset or portfolio selection while off-line approaches are used in areas such as genome analysis, linguistics, audiology among others. They include research by Berkes et al. (2004) for change point detection in GARCH(p,q) models, Koubková (2006) under change point in mean. Change point estimation in volatility is applicable in areas such as options pricing and calculation of value-at-risk. The volatility change point estimator is applied to estimate change point in volatility of United States Dollar/Kenya Shilling exchange rate returns and an account for the change points made. A significant improvement in describing time series is expected, if points in time for volatility changes are identified.

2. Method

2.1 Non Parametric Auto-Regressive Conditional Heteroscedastic (NP-ARCH) Model

Let S_t denote the price of some financial instrument observed at equally spaced discrete time points $t = 1, 2, \dots$ and that $X_t = \log S_t - \log S_{t-1}$ is the continuously compounded single-period return on the asset from time $t - 1$ to t . The volatility of the instrument is the standard deviation of these returns. One assumes the existence of a non-parametric and non-linear relationship between X_t and X_{t-i} , for $i = \{1, 2, \dots, d\}$ which is modeled by a non-linear Auto-Regressive process of the form

$$X_t = m(X_{t-1}, X_{t-2}, \dots, X_{t-d}) + u_t \quad \text{for } t = 1, 2, \dots, n \tag{1}$$

where $\{u_t\}$ are innovations which are independent of the past $\{X_t\}$, $m(\cdot)$ is the conditional mean function of the returns in period t given the previous periods X_{t-1}, X_{t-2}, \dots . In non-parametric approaches, $m(\cdot)$ is allowed to be from some flexible class of functions and it is approximated in such a way that the precision of approximation increases with the size of the sample. Since the interest is on the future volatility, by representing the innovations as

$u_t = \sigma(X_{t-1}, X_{t-2}, \dots, X_{t-d})z_t$, Equation (1) is extended to a Non Parametric Auto-Regressive Conditional Heteroscedastic (NP-ARCH) model of the form

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

where $\mathbb{E}(X_t|X_{t-1} = x_1, \dots, X_{t-d} = x_d) = m(X_{t-1}, \dots, X_{t-d})$, is the conditional mean function, $\text{Var}(X_t|X_{t-1} = x_1, \dots, X_{t-d} = x_d) = \mathbb{E}(u_t^2|X_{t-1} = x_1, \dots, X_{t-d} = x_d) = \sigma^2(X_{t-1}, X_{t-2}, \dots, X_{t-d})$ is the conditional variance function of the returns in period t given the past periods. $\{z_t\}$ are independent and identically distributed random variables (errors) which are time invariant with unspecified continuous and positive probability density function f_z , $\mathbb{E}(z_t|\mathcal{F}_{t-1} = X_{t-1}, \dots, X_{t-d}) = 0$, $\text{Var}(z_t|\mathcal{F}_{t-1} = X_{t-1}, \dots, X_{t-d}) = 1$ and independent of X_{t-1}, \dots, X_{t-d} while $\mathbb{E}z_t = 0$, $\mathbb{E}z_t^2 = 1$, $\mathbb{E}(z_t)^4 < \infty$.

Model (2) is a flexible nonparametric time series model because it does not impose any (parametric) particular form on the conditional mean and conditional variance functions. However, in higher dimensions, there is poor performance called curse of dimensionality, which for $d > 2$ the estimation of the functions in Equation (2) is complicated unless one has a very large sample (For a given bandwidth (window size), the higher the dimension, the less the data there is in a neighborhood of a point $x \in \mathbb{R}^d$ with bandwidth b_n). For simplicity, let $d = 1$ so that Equation (2) becomes

$$X_t = m(X_{t-1}) + \sigma(X_{t-1})z_t. \tag{3}$$

Equation (3) generates heavy tailed distributions. To demonstrate this, suppose a simple model $X_t = \sigma(X_{t-1})z_t$ with z_t

assumed to be i.i.d and from a standard normal distribution. This means that $\mathbb{E}(z_t^4) = 3$ and by Jensen’s inequality,

$$\text{kurtosis}(X_t) = \frac{\mathbb{E}(X_t^4)}{[\mathbb{E}(X_t^2)]^2} = \frac{\mathbb{E}[\sigma^4(X_{t-1})z_t^4]}{[\mathbb{E}(\sigma^2(X_{t-1})z_t^2)]^2} = 3 \frac{\mathbb{E}[\sigma^4(X_{t-1})]}{[\mathbb{E}(\sigma^2(X_{t-1}))]^2} \geq 3$$

This heavy tailed-ness feature implied by Equation (3) makes it a suitable model for modeling data with heavy tails like financial returns data. X_t is assumed to be strictly stationary and strong mixing which is a common assumption satisfied by most financial time series Fan and Yao (2008) hence the theorem below is applied.

Theorem 1 Strong mixing condition: Suppose the existence of a probability space (Ω, \mathcal{F}, P) . Let the dependence measure between any two σ fields \mathcal{A} and $\mathcal{B} \subset \mathcal{F}$ as introduced by Rosenblatt (1956) be defined by

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

Now, suppose $\{X_t, t \in \mathbb{Z}\}$ is a two-sided sequence of variables on a given probability space (Ω, \mathcal{A}, P) . For $-\infty \leq j < l \leq \infty$, let $\mathcal{F}_j^l = \sigma(X_t, j \leq t \leq l)$ denote the σ – field of events which has been generated by the family $\{X_t, j \leq t \leq l (t \in \mathbb{Z})\}$ (\mathcal{F}_j^l assembles all the information collected between time j and l). For each $n \in \mathbb{N}$, define the “coefficient of dependence (mixing)” $\alpha(n)$ by,

$$\alpha(n) = \sup_{-\infty < j < \infty} \alpha(\mathcal{F}_{-\infty}^j, \mathcal{F}_{j+n}^\infty)$$

where $\mathcal{F}_{-\infty}^j$ is the σ – field of events contained in the past of the sequence $\{X_t\}$ up to time j and \mathcal{F}_{j+n}^∞ is the σ – field of events contained in the future of the sequence $\{X_t\}$ from time $j + n$ onwards. The sequence of numbers $(\alpha(n), n \in \mathbb{N})$ is non-increasing (decreasing) in n and are non-negative. The sequence $\{X_t, t \in \mathbb{Z}\}$ is therefore said to be “strong mixing” or “ α mixing” if

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

and hence the sequence is asymptotically independent between the past and the future.

For a one-sided sequence $\{X_t, t \geq 1\}$, then one defines $\alpha(n)$ by

$$\alpha(n) = \sup_{1 \leq j < \infty} \alpha(\mathcal{F}_1^j, \mathcal{F}_{j+n}^\infty)$$

A strictly stationary two-sided sequence $\{X_t, t \in \mathbb{Z}\}$ is said to be strongly mixing if

$$\lim_{n \rightarrow \infty} \alpha(n) = 0 \text{ where } \alpha(n) = \alpha(\mathcal{F}_{-\infty}^0, \mathcal{F}_n^\infty)$$

hence the sequence is said to be ergodic.

Let $\{(X_t, X_{t-1}) \in \mathbb{R} \times \mathbb{R} : t = 1, 2, \dots, n\}$ be a sample of size $n \in \mathbb{N}$, $K(\cdot) : \mathbb{R} \rightarrow \mathbb{R}$ be a kernel function which is a bounded continuous function on \mathbb{R} satisfying the assumption of normalization $\int K(u)du = 1$, which ensures that the method of kernel density estimation results in a probability density function, symmetry about zero $K(u) = K(-u) \forall u$ implying that all the odd moments are zero, $K(u) \geq 0 \forall u$ implying that $K(u)$ is a probability density function, $\int u K(u)du = 0$, and $u_2 = \int u^2 K(u)du < \infty$ and b_n a positive real-valued number, called a bandwidth or smoothing parameter. The kernel density estimator is defined by the kernel density estimator is defined by

$$\hat{f}(x) = \frac{1}{(n-1)b_n} \sum_{t=2}^n K\left(\frac{X_{t-1} - x}{b_n}\right) \text{ for } x \in \mathbb{R} \tag{4}$$

$m(x)$ and $\sigma(x)$ are estimated by non-parametric technique. The assumption is that they are smooth functions and that X_{t-1} has a density function $f(x), x \in [-1, 1]$ which is actually the support of the second order Epanechnikov kernel function

$$K(u) = \begin{cases} \frac{3}{4}(1 - u^2) & \text{support } |u| \leq 1 \\ 0 & \text{otherwise} \end{cases} \text{ where } u = \left(\frac{X_{t-1} - x}{b_n}\right). \tag{5}$$

which is employed when estimating the regression function since it is the most efficient in minimizing the Mean Integrated Squared Error (MISE) and is therefore optimal putting in mind that the choice of the kernel is not as important as the choice of the bandwidth. The non-parametric estimator of the regression function $m(x) = \mathbb{E}[X_t|X_{t-1} = x]$ is obtained by

$$\hat{m}(x) = \begin{cases} \frac{\sum_{t=2}^n K\left(\frac{X_{t-1} - x}{b_n}\right)X_t}{\sum_{t=2}^n K\left(\frac{X_{t-1} - x}{b_n}\right)}, & \text{if } \hat{f}(x) \neq 0 \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

where $K(\cdot)$ is a kernel function and b_n is the bandwidth. $\hat{m}(x)$ is called a kernel estimator or the Nadaraya Watson kernel estimator developed independently by Nadaraya (1964) and Watson (1964).

The estimator of the conditional variance function $\sigma^2(x) = \text{Var}(X_t|X_{t-1} = x)$ which is obtained by using the estimated residuals $\mathbb{E}\left\{\left(X_t - m(X_{t-1})\right)^2 | X_{t-1} = x\right\} = \sigma^2(x)$ is given by

$$\hat{\sigma}^2(x) = \begin{cases} \frac{\sum_{t=2}^n G\left(\frac{X_{t-1}-x}{g_n}\right)\left(X_t - \hat{m}(X_{t-1})\right)^2}{\sum_{t=2}^n G\left(\frac{X_{t-1}-x}{g_n}\right)}, & \text{if } \hat{f}(x) \neq 0 \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

which is still a local constant smoother (kernel estimator) where $G(\cdot)$ is a kernel function and g_n is the bandwidth. A residual based estimator of $\hat{\sigma}^2(x)$ is able to overcome the bias problem of the method of Härdle and Tsybakov (1997) who used a direct estimator

$$\sigma(x) = m_2(x) - m^2(x) \tag{8}$$

where $m_2(x) = \mathbb{E}\left(X_t^2 | X_{t-1} = x\right)$ and $m(x) = \mathbb{E}\left(X_t | X_{t-1} = x\right)$, and which is also likely to produce negative estimate of the volatility function especially when different bandwidth parameters are used to estimate $m_2(x)$ and $m(x)$. Also, a residual based estimator is able to reduce the variance of the difference based estimators. The second order Gaussian Kernel function

$$G(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right) \text{ for } u = \left(\frac{X_{t-1} - x}{g_n}\right), \quad -\infty < u < \infty \tag{9}$$

is employed when estimating the conditional variance function so as to cater for asymmetric behavior of volatility. g_n is a bandwidth that is different from b_n . Fan and Yao (1998) showed that $\hat{m}(x)$ is a consistent estimator of $m(x)$.

The bandwidth b_n is fixed and does not vary with x . It determines the size of the neighborhood and is chosen dependent on $n \in \mathbb{N}$ in such a way that for a larger sample, it is chosen smaller Härdle (1990). Hence, asymptotically, the bandwidth is a sequence of positive real-valued numbers $(b_n)_{n \in \mathbb{N}}$ with $\lim_{n \rightarrow \infty} b_n = 0$. Due to dependency in the returns data set, the bandwidth is estimated with the aim of minimizing the leave-one-out cross-validation function Härdle (1990).

2.2 Non-Parametric Auto-Regressive Conditional Heteroscedastic Model Under Structural Change

Under the null hypothesis H_0 , Equation (3) with no change in volatility, is written as

$$X_t = m(X_{t-1}) + \sigma_t(X_{t-1})z_t \quad t = 1, 2, 3, \dots, n \tag{10}$$

This in turn implies that

$$\mathbb{E}\{X_t - m(X_{t-1})\}^2 = \sigma_t^2(X_{t-1}) \tag{11}$$

where $\sigma_t^2(X_{t-1})$ shall be denoted by $\sigma_{(1)}^2(X_{t-1})$ for $t = 1, 2, \dots, n$.

The data structure having changed at a certain point in time means that using one regression model to study the data will obviously leave the data unfitted or poorly explained by the regression model. This implies that the model valid near $t = 1$ is not valid near $t = n$ due to the presence of an unknown change point. Hence, in the presence of an unknown change point in volatility, $\tau \in [2, n - 1]$, Equation (3) becomes

$$X_t = m(X_{t-1}) + \sigma_t(X_{t-1})z_t \implies \mathbb{E}\{X_t - m(X_{t-1})\}^2 = \sigma_t^2(X_{t-1}) \tag{12}$$

where the alternative hypothesis H_A becomes

$$\sigma_t^2(X_{t-1}) := \begin{cases} \sigma_{(1)}^2(X_{t-1}) & \text{for } t = 1, 2, \dots, \tau \\ \sigma_{(2)}^2(X_{t-1}) & \text{for } t = \tau + 1, \dots, n \end{cases} \tag{13}$$

The model having instability in the conditional variance function is alternatively referred to as the model with non-stationary variances.

2.3 Volatility Change Point Test Statistic and Estimator

Define the residuals obtained from non-parametric estimation of conditional mean function and standardized using the conditional variances obtained from the conditional variance function as $\hat{\varepsilon}_t = \frac{X_t - \hat{m}(X_{t-1})}{\hat{\sigma}(X_{t-1})}$ where $\hat{m}(\cdot)$ is the Nadaraya Watson estimator of the unknown regression function $m(\cdot)$.

The sums of squared residuals among the sample segments (partial sums) are defined as

$$\varepsilon_n = \sum_{t=1}^n \hat{\varepsilon}_t^2 \quad \varepsilon_\tau = \sum_{t=1}^\tau \hat{\varepsilon}_t^2 \quad \varepsilon_{n-\tau} = \sum_{t=\tau+1}^n \hat{\varepsilon}_t^2 \tag{14}$$

Define $\bar{\varepsilon}_{1,\tau}$ as the mean of the first τ squared residuals and $\bar{\varepsilon}_{\tau+1,n}$ as the mean of the last $n - \tau$ squared residuals as

$$\bar{\varepsilon}_{1,\tau} = \frac{\varepsilon_\tau}{\tau} = \frac{1}{\tau} \sum_{t=1}^{\tau} \hat{\varepsilon}_t^2 \quad \bar{\varepsilon}_{\tau+1,n} = \frac{\varepsilon_{n-\tau}}{n-\tau} = \frac{1}{n-\tau} \sum_{t=\tau+1}^n \hat{\varepsilon}_t^2 \tag{15}$$

The estimate of the variance of the first τ observations is obtained by $\frac{\varepsilon_\tau}{\tau}$ while the estimate of the variance of the last $n - \tau$ observations is given by $\frac{\varepsilon_{n-\tau}}{n-\tau}$ once the jump point τ has been estimated. The change point estimator $\hat{\tau}$ of the unknown change point position τ is obtained through minimization of the residual sum of squares among all the possible segments of the sample as

$$\begin{aligned} \hat{\tau} &= \arg \min_{\tau} \left(\sum_{t=1}^{\tau} (\hat{\varepsilon}_t^2 - \bar{\varepsilon}_{1,\tau})^2 + \sum_{t=\tau+1}^n (\hat{\varepsilon}_t^2 - \bar{\varepsilon}_{\tau+1,n})^2 \right) \\ &= \arg \min_{\tau} V_{\tau}^2 \end{aligned} \tag{16}$$

where $V_{\tau}^2 = \sum_{t=1}^{\tau} (\hat{\varepsilon}_t^2 - \bar{\varepsilon}_{1,\tau})^2 + \sum_{t=\tau+1}^n (\hat{\varepsilon}_t^2 - \bar{\varepsilon}_{\tau+1,n})^2$. Further, suppose $\bar{\varepsilon}_{1,n} = \frac{1}{n} \sum_{t=1}^n \hat{\varepsilon}_t^2$ is overall mean of the squared residuals. Bai (1994), shows that for each $\tau \in \{2, 3, \dots, n - 1\}$, the following relation holds

$$V_{\tau}^2 = \sum_{t=1}^n (\hat{\varepsilon}_t^2 - \bar{\varepsilon}_{1,n})^2 - n(D_{\tau}^n)^2 \tag{17}$$

This implies that the minimum of V_{τ}^2 will occur when D_{τ}^n is maximum. Hence,

$$\hat{\tau} = \arg \min_{\tau} V_{\tau}^2 = \arg \max_{\tau} (D_{\tau}^n)^2 = \arg \max_{\tau} |D_{\tau}^n| \tag{18}$$

where considering a weighted l_2 norm,

$$l_2 \left(\sigma_{(1)}^2(X_{t-1}) - \sigma_{(2)}^2(X_{t-1}) \right) = \left(\sum_{\tau=1}^n w_{\tau} |\varepsilon_{\tau} - \varepsilon_{n-\tau}|^2 \right)^{\frac{1}{2}}$$

one obtains

$$D_{\tau}^n = \left(\frac{\tau}{n} \left(1 - \frac{\tau}{n} \right) \right)^{\frac{1}{2}} \left| \frac{1}{\tau} \sum_{t=1}^{\tau} \hat{\varepsilon}_t^2 - \frac{1}{n-\tau} \sum_{t=\tau+1}^n \hat{\varepsilon}_t^2 \right| \tag{19}$$

which is a non-parametric statistic for change point detection in volatility. The non-parametric statistic so obtained does not utilize any a priori information about the data. This test statistic suitable for processes which are not independent identically distributed as opposed to Inflan and Tiao (1994) test.

A good choice of the estimator for the change point τ is where the test statistic has a global maximum since the maximum usually occur in the area of the “true” change point. This means that all possible time points are taken to account and hence the use of maximum statistics. Thus the estimator for unknown change point in volatility of financial returns is the function

$$\hat{\tau} = \arg \max_{\tau} |D_{\tau}^n| \tag{20}$$

2.3 Consistency of the Volatility Change Point Estimator

When proving consistency of the volatility change point estimator, one needs to show that the volatility change point fraction $\frac{\hat{\tau}}{n} = \hat{k}$ is consistent for the “true” change point fraction $\frac{\tau}{n} = k^*$ under H_A but not showing that the indexes are consistent Truong et al. (2018). Distance between a true break point index and its estimated counterpart $|\hat{\tau} - \tau|$ never converge to zero and thus the need to consider the break point fraction instead of the indexes themselves. Hence, one needs to show that the relative error $\frac{|\hat{\tau} - \tau|}{n}$ decreases with the sample size n meaning that $\hat{k} \xrightarrow{p} k^*$ as the sample size diverges, The results on consistency of the volatility change point estimator are summarized in Theorem 2 and Theorem 3 below.

Theorem 2 Let τ be the “true” location (position) of the change point under the alternative hypothesis

$$\sigma_t^2(X_{t-1}) = \begin{cases} \sigma_{(1)}^2(X_{t-1}) & \text{for } t \leq \tau \\ \sigma_{(2)}^2(X_{t-1}) & \text{for } t > \tau \end{cases}$$

Let $\hat{\tau}$ be the estimate of τ given by

$$\hat{\tau} = \arg \max_{\tau} |D_t^n|$$

Then $\hat{\tau}$ is consistent for the “true” break point τ Chen and Gupta (2001).

In proving the consistency of an estimator that has been obtained by maximizing an objective function, one needs to show that the objective function $|D_t^n|$ with integer-valued variable τ , $\tau = 2, \dots, n - 1$ as its argument, is uniformly close to its mean function (the expected value) and that this mean function has a maximum which is unique.

Theorem 3 below for the convergence in probability for the change point fraction \hat{k} is the main result in proving the consistency of the estimated change point fraction.

Theorem 3 Consider a sample of squared residuals $\hat{\varepsilon}_1^2, \hat{\varepsilon}_2^2, \dots, \hat{\varepsilon}_n^2$ satisfying the alternative change point hypothesis and the change point estimator $\hat{\tau}$ given in (18). If the sequences $\{\hat{\varepsilon}_{1,t}^2, t \in \mathcal{Z}\}$ and $\{\hat{\varepsilon}_{2,t}^2, t \in \mathcal{Z}\}$ satisfy

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

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

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

where $0 < B < \infty$ (is a positive constant) and $k^* = \frac{\tau}{n}$.

To proof Theorem 3 above, suppose that $\{\hat{\varepsilon}_{1,t}^2, t \in \mathcal{Z}\}$ and $\{\hat{\varepsilon}_{2,t}^2, t \in \mathcal{Z}\}$ are two conditional heteroscedastic processes of regression residuals which can be described as pre-break and post-break sub-samples respectively. Further, suppose that one obtains a sample $\hat{\varepsilon}_1^2, \hat{\varepsilon}_2^2, \dots, \hat{\varepsilon}_n^2$ such that

$$\hat{\varepsilon}_t^2 := \begin{cases} \hat{\varepsilon}_{1,t}^2 & \text{if } 1 \leq t \leq \tau \\ \hat{\varepsilon}_{2,t}^2 & \text{if } \tau < t \leq n \end{cases} \tag{21}$$

Also, assume that the two sequences of residuals $\{\hat{\varepsilon}_{1,t}^2, t = 1, 2, \dots, \tau\}$ and $\{\hat{\varepsilon}_{2,t}^2, t = \tau + 1, \dots, n\}$ have different conditional variance functions such that

$$\mathbb{E}(\hat{\varepsilon}_{i,t}^2) := \begin{cases} \sigma_{(1)}^2(X_{t-1}) & \text{for } 1 \leq t \leq \tau \text{ when } i = 1 \\ \sigma_{(2)}^2(X_{t-1}) & \text{for } \tau < t \leq n \text{ when } i = 2 \end{cases} \tag{22}$$

where $\tau = k^*n$ and $0 < k^* < 1$. The conditional variance function is assumed to change at time $t = \tau$. One assumes that $\hat{\varepsilon}_{1,t}^2$ is a process which is independent from $\hat{\varepsilon}_{2,t}^2$ thus resulting in a discontinuity in the conditional variance function of the residuals at τ . Since the returns are dependent due to the fact that they are obtained from a single instrument, the change in the conditional variance function will result in the observations having starting values not out of stationary models. Hence, one assumes that the observations after the change are based on a time series which is stationary.

The disordered sequence $\{\hat{\varepsilon}_t^2\}$ which has a change point is no longer a stationary sequence and τ is referred to as the change point while $\hat{\tau}$ is the change point estimator of τ .

$|\mathbb{E}(D_t^n)|$ will achieve its maximum at $t = \tau$ and $\mathbb{E}(D_t^n)$ has a unique maximum at τ and also that $(D_t^n - \mathbb{E}(D_t^n))$ is uniformly small in τ for large n . The first sum shall thus be weighted by $\frac{\tau}{n}$ converging to k^* which is the percentage of chronologically ordered observations before the change point while the second sum is weighted by $1 - \frac{\tau}{n}$ converging to $1 - k^*$ which is the percentage of the chronologically ordered observations after the change. This results to

$$\mathbb{E}(D_t^n) = \begin{cases} \Delta k^{\frac{1}{2}} (1 - k^*)^{\frac{1}{2}} & \text{if } t \leq \tau \\ \Delta k^{*\frac{1}{2}} (1 - k)^{\frac{1}{2}} & \text{if } t > \tau \end{cases} \tag{23}$$

where $1 < \tau < n$, is parametrized as $\tau = [nk^*]$ with $k^* \in (0, 1)$.

Consideration of both cases when $t \leq \tau$ and when $t > \tau$ is made even though there is symmetry.

Thus

$$\mathbb{E}(D_\tau^n) = \Delta k^{*\frac{1}{2}} (1 - k^*)^{\frac{1}{2}} \tag{24}$$

From Equations (23) and (24) one obtains

$$|\mathbb{E}(D_\tau^n)| - |\mathbb{E}(D_t^n)| = \begin{cases} |\Delta| k^{*\frac{1}{2}} (1 - k^*)^{\frac{1}{2}} - |\Delta| k^{\frac{1}{2}} (1 - k)^{\frac{1}{2}} = |\Delta| (1 - k^*)^{\frac{1}{2}} (k^{*\frac{1}{2}} - k^{\frac{1}{2}}) & \text{if } t \leq \tau \\ |\Delta| k^{*\frac{1}{2}} ((1 - k^*)^{\frac{1}{2}} - |\Delta| k^{*\frac{1}{2}} (1 - k)^{\frac{1}{2}}) = |\Delta| k^{*\frac{1}{2}} ((1 - k^*)^{\frac{1}{2}} - (1 - k)^{\frac{1}{2}}) & \text{if } t > \tau \end{cases} \tag{25}$$

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

$$k^{\frac{1}{2}} - k^{\frac{1}{2}} \geq \frac{1}{2} k^{*\frac{-1}{2}} (k^* - k) \tag{26}$$

If $k > k^*$, then $1 - k < 1 - k^*$ and Equation (26) above yields

$$(1 - k^*)^{\frac{1}{2}} - (1 - k)^{\frac{1}{2}} \geq \frac{1}{2} (1 - k^*)^{-\frac{1}{2}} (k - k^*) \tag{27}$$

Combining Equations (25), (26) and (27), one obtains

$$|\mathbb{E}(D_\tau^n) - \mathbb{E}(D_t^n)| \geq |\Delta| \frac{1}{2} (k^*)^{-\frac{1}{2}} (1 - k^*)^{-\frac{1}{2}} \min\{k^*, 1 - k^*\} |k^* - k| \tag{28}$$

Also

$$D_t^n - D_\tau^n = [D_t^n - \mathbb{E}(D_t^n) + \mathbb{E}(D_t^n)] - [D_\tau^n - \mathbb{E}(D_\tau^n) + \mathbb{E}(D_\tau^n)] \tag{29}$$

which from the triangle inequality and the relation that $|a| = |a - b + b| \leq |a - b| + |b|$ yields

$$\begin{aligned} |D_t^n| - |D_\tau^n| &\leq |D_t^n - \mathbb{E}(D_t^n)| + |\mathbb{E}(D_t^n)| + |D_\tau^n - \mathbb{E}(D_\tau^n)| - |\mathbb{E}(D_\tau^n)| \\ &\leq 2 \left(\max_\tau |D_t^n - \mathbb{E}(D_t^n)| \right) + |\mathbb{E}(D_t^n)| - |\mathbb{E}(D_\tau^n)| \end{aligned} \tag{30}$$

Taking $|\mathbb{E}(D_\tau^n) - \mathbb{E}(D_t^n)|$ to the left hand side of the inequality will imply that

$$|\mathbb{E}(D_\tau^n) - \mathbb{E}(D_t^n)| \leq 2 \left(\max_\tau |D_t^n - \mathbb{E}(D_t^n)| \right) + |D_\tau^n| - |D_t^n| \tag{31}$$

Therefore, from Equations (28) and (31) the following is obtained

$$\begin{aligned} |\Delta| \frac{1}{2} (k^*)^{-\frac{1}{2}} (1 - k^*)^{-\frac{1}{2}} \min\{k^*, 1 - k^*\} |k^* - k| &\leq |\mathbb{E}(D_\tau^n) - \mathbb{E}(D_t^n)| \\ &\leq 2 \left(\max_\tau |D_t^n - \mathbb{E}(D_t^n)| \right) + |D_\tau^n| - |D_t^n| \end{aligned} \tag{32}$$

Replacing k by \hat{k} in Equation (30) above and noting that $|D_\tau^n| \leq |D_t^n|$ one obtains

$$|\Delta| \frac{1}{2} (k^*)^{-\frac{1}{2}} (1 - k^*)^{-\frac{1}{2}} \min\{k^*, 1 - k^*\} |k^* - \hat{k}| \leq 2 \left(\max_\tau |D_t^n - \mathbb{E}(D_t^n)| \right) \tag{33}$$

Considering D_t^n as given in Equation (19), $\max_\tau |D_t^n - \mathbb{E}(D_t^n)|$ needs to be estimated.

Writing

$$\begin{aligned} |D_t^n - \mathbb{E}(D_t^n)| &= \frac{1}{n} \left| \frac{\sqrt{n-\tau}}{\sqrt{\tau}} \sum_{i=1}^{\tau} [\hat{\epsilon}_i^2 - \mathbb{E}(\hat{\epsilon}_i^2)] - \frac{\sqrt{\tau}}{\sqrt{n-\tau}} \sum_{i=\tau+1}^n [\hat{\epsilon}_i^2 - \mathbb{E}(\hat{\epsilon}_i^2)] \right| \\ &= \frac{1}{n} \left| \frac{\sqrt{n-\tau}}{\sqrt{\tau}} \frac{\sqrt{n-\tau}}{\sqrt{n-\tau}} \sum_{i=1}^{\tau} [\hat{\epsilon}_i^2 - \mathbb{E}(\hat{\epsilon}_i^2)] - \sqrt{\tau} \frac{1}{\sqrt{n-\tau}} \frac{\sqrt{\tau}}{\sqrt{\tau}} \sum_{i=\tau+1}^n [\hat{\epsilon}_i^2 - \mathbb{E}(\hat{\epsilon}_i^2)] \right| \\ &= \frac{1}{n} \left| \frac{(n-\tau)}{\sqrt{\tau} \sqrt{(n-\tau)}} \sum_{i=1}^{\tau} [\hat{\epsilon}_i^2 - \mathbb{E}(\hat{\epsilon}_i^2)] - \frac{\tau}{\sqrt{n-\tau} \sqrt{\tau}} \sum_{i=\tau+1}^n [\hat{\epsilon}_i^2 - \mathbb{E}(\hat{\epsilon}_i^2)] \right| \end{aligned} \tag{34}$$

Opening the brackets

$$|D_t^n - \mathbb{E}(D_t^n)| = \frac{1}{n \sqrt{n-\tau} \sqrt{\tau}} \left| n \sum_{i=1}^{\tau} [\hat{\epsilon}_i^2 - \mathbb{E}(\hat{\epsilon}_i^2)] - \tau \sum_{i=1}^{\tau} [\hat{\epsilon}_i^2 - \mathbb{E}(\hat{\epsilon}_i^2)] - \tau \sum_{i=\tau+1}^n [\hat{\epsilon}_i^2 - \mathbb{E}(\hat{\epsilon}_i^2)] \right| \tag{35}$$

Which implies

$$|D_t^n - \mathbb{E}(D_t^n)| = \frac{1}{n \sqrt{n-\tau} \sqrt{\tau}} \left| n \sum_{i=1}^{\tau} [\hat{\epsilon}_i^2 - \mathbb{E}(\hat{\epsilon}_i^2)] - \tau \sum_{i=1}^n [\hat{\epsilon}_i^2 - \mathbb{E}(\hat{\epsilon}_i^2)] \right| \tag{36}$$

Distributing $\frac{1}{n}$ and introducing the magnitude in Equation (36) yields

$$\begin{aligned} |D_t^n - \mathbb{E}(D_t^n)| &\leq \frac{1}{\sqrt{n-\tau}\sqrt{\tau}} \left| \sum_{i=1}^{\tau} [\hat{\epsilon}_i^2 - \mathbb{E}(\hat{\epsilon}_i^2)] \right| + \frac{1}{\sqrt{n-\tau}\sqrt{\tau}} \frac{\tau}{n} \left| \sum_{i=1}^n [\hat{\epsilon}_i^2 - \mathbb{E}(\hat{\epsilon}_i^2)] \right| \\ &\leq \frac{1}{\sqrt{n-\tau}\sqrt{\tau}} \left| \sum_{i=1}^{\tau} [\hat{\epsilon}_i^2 - \mathbb{E}(\hat{\epsilon}_i^2)] \right| + \frac{1}{\sqrt{n-\tau}\sqrt{\tau}} \left| \sum_{i=1}^{\tau} [\hat{\epsilon}_i^2 - \mathbb{E}(\hat{\epsilon}_i^2)] \right| \end{aligned} \tag{37}$$

on adding the two parts

$$\leq \frac{2}{\sqrt{n-\tau}\sqrt{\tau}} \left| \sum_{i=1}^{\tau} [\hat{\epsilon}_i^2 - \mathbb{E}(\hat{\epsilon}_i^2)] \right|$$

where one concludes that

$$\max_{\tau} |D_t^n - \mathbb{E}(D_t^n)| \leq 2 \max_{\tau} \frac{2}{\sqrt{n-\tau}\sqrt{\tau}} \left| \sum_{i=1}^{\tau} [\hat{\epsilon}_i^2 - \mathbb{E}(\hat{\epsilon}_i^2)] \right| \tag{38}$$

A general Hajek Renyi type inequality for dependent processes for the maximum of weighted sums in Theorem 4 below is applied Kokoszka et al. (2000).

Theorem 4 Let Y_1, Y_2, \dots, Y_n be any random variables with finite second moments and suppose b_1, b_2, \dots, b_n are any non-negative constants (a decreasing positive sequence of constants). Then for any $\epsilon > 0$,

$$\begin{aligned} \epsilon^2 P \left\{ \max_{m \leq t \leq n} b_{\tau} \left| \sum_{i=1}^t Y_i \right| > \epsilon \right\} &\leq b_m^2 \sum_{i,j=1}^m \mathbb{E}(Y_i Y_j) + \sum_{\tau=m}^{n-1} |b_{\tau+1}^2 - b_{\tau}^2| \sum_{i,j=1}^{\tau} \mathbb{E}(Y_i Y_j) \\ &\quad + 2 \sum_{\tau=m}^{n-1} b_{\tau+1}^2 \mathbb{E} \left(|Y_{\tau+1}| \left| \sum_{i=1}^{\tau} Y_i \right| \right) + \sum_{\tau=m}^{n-1} b_{\tau+1}^2 \mathbb{E} Y_{\tau+1}^2 \end{aligned}$$

Setting $m = 1, Y_i = \hat{\epsilon}_i^2 - \mathbb{E}(\hat{\epsilon}_i^2)$ and $b_1, b_2, \dots, b_n = \frac{1}{\sqrt{n-\tau}\sqrt{\tau}}$ by Cauchy Schwarz inequality,

$$\begin{aligned} \epsilon^2 P \left\{ \max_{\tau} \frac{1}{\sqrt{n-\tau}\sqrt{\tau}} \left| \sum_{i=1}^{\tau} [\hat{\epsilon}_i^2 - \mathbb{E}(\hat{\epsilon}_i^2)] \right| > \epsilon \right\} &\leq \frac{1}{(n-\tau)\tau} \sum_{i,s=1}^{\tau} \mathbb{E}((\hat{\epsilon}_i^2 - \mathbb{E}(\hat{\epsilon}_i^2))((\hat{\epsilon}_s^2 - \mathbb{E}(\hat{\epsilon}_s^2))) \\ &\quad + \sum_{i=1}^n \left| \frac{1}{(n-\tau+1)(\tau+1)} - \frac{1}{(n-\tau)\tau} \right| \\ &\quad \sum_{i,s=1}^{\tau} \mathbb{E}((\hat{\epsilon}_i^2 - \mathbb{E}(\hat{\epsilon}_i^2))((\hat{\epsilon}_s^2 - \mathbb{E}(\hat{\epsilon}_s^2))) \\ &\quad + 2 \sum_{i=0}^{n-1} \frac{1}{(n-\tau)\tau} \mathbb{E} \left(\left| \hat{\epsilon}_{i+1}^2 - \mathbb{E}(\hat{\epsilon}_{i+1}^2) \right| \left| \sum_{s=1}^i (\hat{\epsilon}_s^2 - \mathbb{E}(\hat{\epsilon}_s^2)) \right| \right) \\ &\quad + \sum_{i=0}^{n-1} \frac{1}{(n-\tau)\tau} \mathbb{E}(\hat{\epsilon}_{i+1}^2 - \mathbb{E}(\hat{\epsilon}_{i+1}^2))^2 \\ &\leq \frac{2}{(n-\tau)\tau} \sum_{i=0}^{n-1} \mathbb{E} \left(\left| \hat{\epsilon}_{i+1}^2 - \mathbb{E}(\hat{\epsilon}_{i+1}^2) \right| \left| \sum_{s=1}^i (\hat{\epsilon}_s^2 - \mathbb{E}(\hat{\epsilon}_s^2)) \right| \right) \\ &\quad + \frac{1}{(n-\tau)\tau} \sum_{i=0}^{n-1} \mathbb{E}(\hat{\epsilon}_{i+1}^2 - \mathbb{E}(\hat{\epsilon}_{i+1}^2))^2 \\ &\leq \frac{2}{(n-\tau)\tau} \sum_{i=0}^{n-1} (\text{var}(\hat{\epsilon}_{i+1}^2))^{\frac{1}{2}} \left(\sum_{s,s'=1}^i \text{cov}(\hat{\epsilon}_s^2 \hat{\epsilon}_{s'}^2) \right)^{\frac{1}{2}} \\ &\quad + \frac{1}{(n-\tau)\tau} \sum_{i=0}^{n-1} \text{var}(\hat{\epsilon}_{i+1}^2) \\ &= \frac{B}{(n-\tau)\tau} \leq \frac{B}{\sqrt{n-\tau}\sqrt{\tau}} \leq \frac{B}{\sqrt{n}} \end{aligned} \tag{39}$$

This in turn implies that

$$\epsilon^2 P \left\{ \max_{\tau} \frac{1}{\sqrt{n-\tau}\sqrt{\tau}} \left| \sum_{t=1}^{\tau} [\hat{\epsilon}_t^2 - \mathbb{E}(\hat{\epsilon}_t^2)] \right| > \epsilon \right\} \leq \frac{B}{\sqrt{n}} \tag{40}$$

which further implies that

$$P \left\{ \max_{\tau} \frac{1}{\sqrt{n-\tau}\sqrt{\tau}} \left| \sum_{t=1}^{\tau} [\hat{\epsilon}_t^2 - \mathbb{E}(\hat{\epsilon}_t^2)] \right| > \epsilon \right\} \leq \frac{B}{\sqrt{n\epsilon^2}} \tag{41}$$

substituting this result into Equation (33) yields

$$P \left\{ |\Delta| \bar{k} |k^* - \hat{k}| > \epsilon \right\} \leq P \left\{ \max_{\tau} \frac{1}{\sqrt{n-\tau}\sqrt{\tau}} \left| \sum_{t=1}^{\tau} [\hat{\epsilon}_t^2 - \mathbb{E}(\hat{\epsilon}_t^2)] \right| > \epsilon \right\} \leq \frac{B}{\sqrt{n\epsilon^2}} \tag{42}$$

where $\bar{k} = \frac{1}{2} (k^*)^{-\frac{1}{2}} (1 - k^*)^{-\frac{1}{2}} \min \{k^*, 1 - k^*\}$, and B is some positive constant which further proves Theorem 3 which was initially stated as

$$P \left\{ |k^* - \hat{k}| > \epsilon \right\} \leq \frac{B}{\epsilon^2 \Delta^2 \sqrt{n}} \tag{43}$$

From Equation (43), it can be seen that as $n \rightarrow \infty$, then $\frac{B}{\epsilon^2 \Delta^2 \sqrt{n}} \rightarrow 0$ which helps complete the proof for consistency of the change point estimator. One can observe that consistency of the change point estimator depends on the magnitude of change due to the fact that Δ is kept on the right hand side.

3. Results

3.1 A Simulation Study on Consistency of the Change Point Estimator

The performance of the change point estimator with respect to sample size, change point location and the magnitude of change is investigated by simulating an ARMA(1,1) ARCH(1) model is simulated with a single change point in volatility function as

$$\sigma_t^2(X_{t-1}) = \begin{cases} (0.1 + \Delta) + 0.1\epsilon_{t-1}^2 & \text{for } t = 1, 2, \dots, \tau \\ 0.1 + 0.1\epsilon_{t-1}^2 & \text{for } t = \tau + 1, \dots, n \end{cases} \tag{44}$$

$$X_t = 0.35X_{t-1} + \epsilon_t + 0.4\epsilon_{t-1}, \quad \epsilon_t = \sigma_t(X_{t-1})z_t.$$

Where $\Delta = \{0.3, 0.5, 0.7\}$ and the sample sizes $n = \{500, 1000, 2000, 4000\}$ and the true change point instants being $\tau = \left\{ \frac{1}{3}n, \frac{1}{2}n, \frac{2}{3}n \right\}$. A table after 1000 bootstrap samples for each sample size is created. In each simulation, the estimates of the change point highly depended on the locations of the change points, the magnitude of change and the sample size just as per Equation (40) above. As the sample size grew unbounded, the error of estimation was decreasing to zero.

Table 1. Numerical results on consistency of the change point estimator

		$\Delta = 0.3$		$\Delta = 0.4$		$\Delta = 0.5$	
n	τ	$\hat{\tau}$	$\frac{ \hat{\tau}-\tau }{n} \rightarrow 0$	$\hat{\tau}$	$\frac{ \hat{\tau}-\tau }{n} \rightarrow 0$	$\hat{\tau}$	$\frac{ \hat{\tau}-\tau }{n} \rightarrow 0$
500	166	164	0.00402	165	0.00201	165	0.00201
	249	247	0.00402	248	0.00201	248	0.00201
	332	330	0.00402	330	0.00402	331	0.00201
1000	332	330	0.00200	331	0.00100	331	0.00100
	499	498	0.00100	498	0.00100	498	0.00100
	666	664	0.00200	665	0.00100	665	0.00100
2000	666	663	0.00150	665	0.00050	665	0.00050
	999	997	0.00100	998	0.00050	998	0.00050
	1332	1330	0.00100	1331	0.00050	1331	0.00050
4000	1332	1331	0.00025	1331	0.00025	1331	0.00025
	1999	1997	0.00025	1998	0.00025	1998	0.00025
	2666	2665	0.00025	2665	0.00025	2665	0.00025

From the sequence of the estimates $\hat{\tau}$, it was observed that the estimator seemed to work better in larger samples than in smaller samples and in samples with larger magnitude of shifts than those with smaller magnitude of shift as evident from the table 1. Also, the bias of estimation, $\frac{|\hat{\tau}-\tau|}{n}$ is decreasing as the sample size increases and as the magnitude of the shift

Δ grows, which further verifies that the estimator $\hat{\tau}$ is consistent. The accuracy of the change point detection is seen to increase with the sample size.

3.2 Sampling Distribution of the Change Point Estimator

Besides being a function, the change point estimator $\hat{\tau}$ is a random variable. As a random variable, an estimator will itself follow some probability distribution, which is called the sampling distribution of the estimator. A distribution of all change points thus becomes the sampling distribution. The plots of the histograms of estimated change points when the “true” change point is located at $\frac{1}{3}n, \frac{1}{2}n, \frac{2}{3}n$ for a sample size of $n = 4000$ after 500 bootstrap samples of

$$\sigma_t^2(X_{t-1}) = \begin{cases} 0.6 + 0.1\epsilon_{t-1}^2 & \text{for } t = 1, 2, \dots, \tau \\ 0.1 + 0.1\epsilon_{t-1}^2 & \text{for } t = \tau + 1, \dots, n \end{cases} \tag{45}$$

$$X_t = 0.35X_{t-1} + \epsilon_t + 0.4\epsilon_{t-1}, \quad \epsilon_t = \sigma_t(X_{t-1})z_t.$$

where z_t is a sequence of i.i.d random variables symmetric around zero from a normal distribution and $\Delta = 0.3$ are as shown below in figures 1, 2 and 3 respectively.

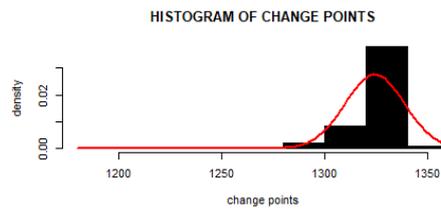


Figure 1. Sampling distribution of $\hat{\tau}$ when $\tau = \frac{1}{3}n = 1332$

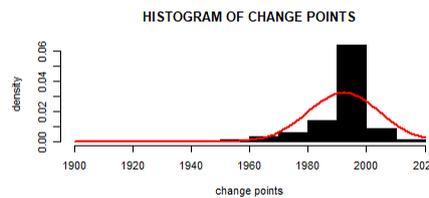


Figure 2. Sampling distribution of $\hat{\tau}$ when $\tau = \frac{1}{2}n = 1999$

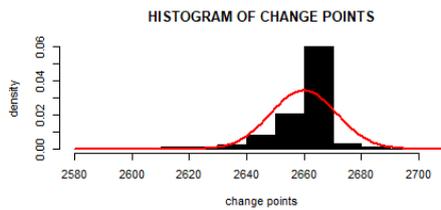


Figure 3. Sampling distribution of $\hat{\tau}$ when $\tau = \frac{2}{3}n = 2666$

Figures 1, 2 and 3 show that the sampling distribution of $\hat{\tau}$ is skewed to the left and the degree of the skewness depends on the position of the change point in that it increases as the change point moves to the right hand side of the sample. The distribution is seen to have a long slowly decaying tail (leptokurtic).

3.3 Application to Foreign Exchange Rate Data Set

The existence of a change point in the conditional variance (volatility) of USD/KSH daily returns from 1st January 2010 to 27th November 2020 is investigated where $n = 2839$ historical exchange rates. This means obtaining 2838 continuously compounded returns. Lagging the returns by one implies having 2837 continuously compounded returns at lag one.

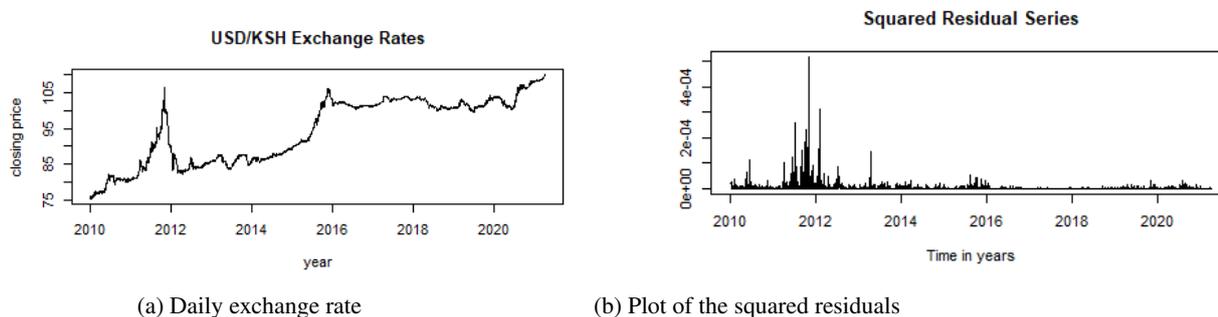


Figure 4. Exchange rate data and squared residual series

The exchange rates are plotted as in Figure 4a above. Since the conditional variances are obtained via the conditional expectation of the squared residuals, the squared residuals plot is as shown in Figure 4b above.

The estimate of the first change point was $\hat{\tau}_1 = 668$ corresponding to 26th July 2012. The data was then split into two segments from [1 : 668] and from [669 : 2736] and change point estimation done on both segments. The second estimate of the change point was $\hat{\tau}_2 = 375$ corresponding to 13th June 2011. From the second segment [669 : 2736], the third change point estimate was at data point 850 meaning that $\hat{\tau}_3 = 1519$ corresponding to 30th October 2015 in the original returns series at lag one.

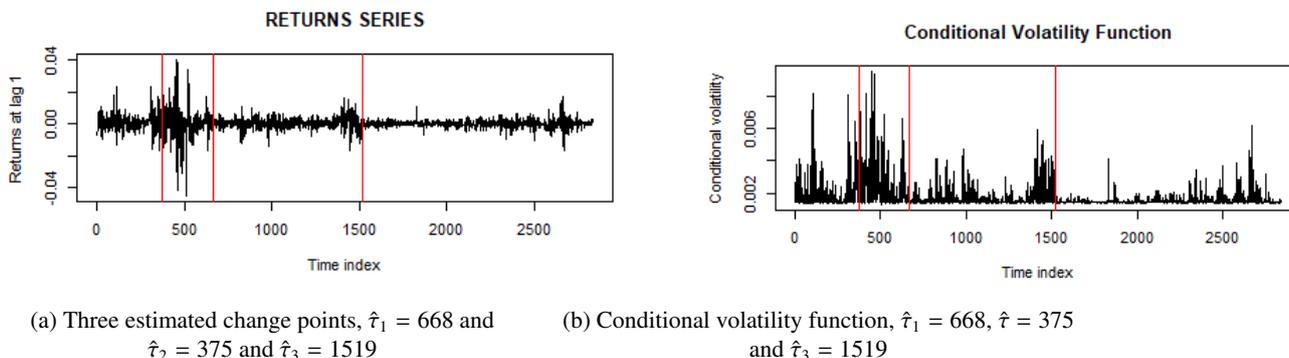


Figure 5. Returns series at lag one and the conditional volatility function with three estimated change points

The returns series and the conditional volatility function with three estimated change points, $\hat{\tau}_1 = 668$ corresponding to 26th July 2012, $\hat{\tau}_2 = 375$ corresponding to 13th June 2011 and $\hat{\tau}_3 = 1519$ corresponding to 30th October 2015 are as in Figures 5a and 5b respectively.

The corresponding plot of the change point statistic with $\hat{\tau}_1 = 668$, $\hat{\tau}_2 = 375$ and at $t = 850$ for segment [668 : 2837] corresponding to $\hat{\tau}_3 = 1519$ are as shown in Figures 6a, 6b and 7 respectively.

4. Discussion of Results

From the exchange rate plot there is an increasing trend between January 2010 up to October 2011 where the exchange rate prices were at the peak. The USD/KSH exchange rate depreciated from 83.89 to 101.39 between April 2011 to October 2011. After October 2011, there is a decreasing trend after which the exchange rates started to rise again around the year 2014 up to the year 2016. From the year 2016, the USD/KSH seems not to be varying very much. From the year

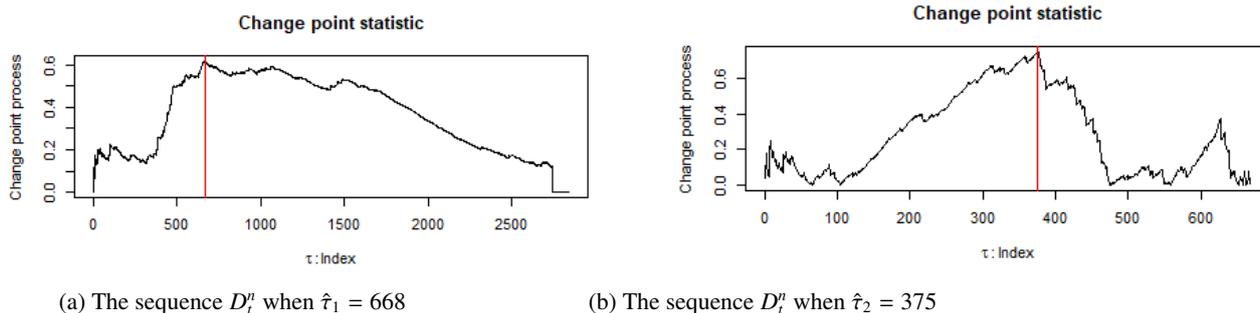


Figure 6

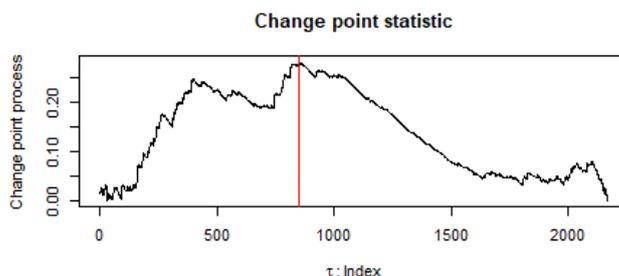


Figure 7. The sequence D_t^n for change point at $t = 850$ for segment $[668 : 2837]$, corresponding to $\hat{\tau}_3 = 1519$ in the original return series at lag one

2020, the USD/KSH exchange rate seems to be taking an upward trend probably due to economic crisis brought about by Covid-19 pandemic like loss of jobs and ban on international travels by governments in most countries of the world.

Accounting for the change point in June 2011, Kenya was importing too much and exporting too little (imbalanced economy) which made it vulnerable to shocks. In 2011, imports soared because of higher costs of fuel and food while exports were stagnant which led to increase in demand for foreign exchange to finance imports. The foreign exchange market witnessed significant volatility between May 2011 and October 2011 as seen in exchange rate plot Figure 4a and the return series 5a reflecting the general volatility in the global financial markets. This resulted in the weakening of the Kenyan Shilling just like other currencies in the region and other global markets whereby the exchange rate for the Kenya shilling against the US dollar depreciated from an average of 84.2 in March 2011 to 101.39 in October 2011 (20.42% depreciation).

Accounting for the change point in volatility of the returns on 26th July 2012, the Kenyan economy experienced slow growth at the beginning of 2012 following high inflation and high interest rates from commercial banks. For the change point on 30th October 2015, the Kenya shilling depreciated against the United States Dollar in the financial year ended June 2015 due to tightening of global financial market conditions and further continued to depreciate in the financial year starting July 2015. The depreciation was further aggregated by increase in food prices due to delayed rains in the same period up to December 2015.

4.2 Conclusion and Recommendations

In this research, a non-parametric procedure for estimating a change point in volatility of financial returns is considered. The procedure allows for change point estimation in sequences with conditional heteroscedastic variances. The consistency of the change point estimator is proven theoretically and shown through simulations. It is observed that the estimator best estimates change point when the change point is fixed around the middle of the sample, when the magnitude of change is large and as the sample size grows. The sampling distribution of the change point estimator is shown through histograms and is found to be negatively skewed and the degree of skewness increases as the change point moves to the right.

The change point estimation approach can be applied to multidimensional non-parametric models of the form $X_t = m(X_{t-1}, \dots, X_{t-d}) + \sigma(X_{t-1}, \dots, X_{t-d})$ where either the volatility function is changing with time. One can also consider a case where both the conditional mean function and the conditional variance function are changing. In such cases, the functions $m(\cdot)$ and $\sigma(\cdot)$ should be estimated using multivariate kernel methods. This should be done carefully due to the curse of

dimensionality problem. Also, situations where there are change points both in conditional mean function and conditional volatility function would be much interesting. The authors aim to obtain the limit distribution of the change point test statistic in a subsequent paper.

Acknowledgements

The authors greatly acknowledge Pan African University Institute for Basic Sciences, Technology and Innovation (PAUSTI) in collaboration with African Development Bank and African Union for funding this research. The USD/KSH exchange rate data used to support the findings of this research is available from The Central Bank of Kenya <https://www.centralbank.go.ke>

References

- Aminikhanghahi, S., & Cook, D. J. (2017). A survey of methods for time series change point detection. *Knowledge and information systems*, 51(2), 339–367.
- Bai, J. (1994). Least squares estimation of a shift in linear processes. *Journal of Time Series Analysis*, 15(5), 453–472.
- Berkes, I., Gombay, E., Horváth, L., & Kokoszka, P. (2004). Sequential change-point detection in garch (p, q) models. *Econometric theory*, 20(6), 1140–1167.
- Brodsky, E., & Darkhovsky, B. S. (2013). *Nonparametric methods in change point problems*, volume 243. Springer Science & Business Media.
- Chen, G., Choi, Y. K., & Zhou, Y. (2005). Nonparametric estimation of structural change points in volatility models for time series. *Journal of Econometrics*, 126(1), 79–114.
- Chen, J., & Gupta, A. K. (2001). On change point detection and estimation. *Communications in statistics-simulation and computation*, 30(3), 665–697.
- Chen, J., & Gupta, A. K. (2011). *Parametric statistical change point analysis: with applications to genetics, medicine, and finance*. Springer Science & Business Media.
- Csorgo, M. and Horváth, L. (1997). *Limit theorems in change-point analysis*. John Wiley & Sons Chichester.
- Fan, J., & Yao, Q. (1998). Efficient estimation of conditional variance functions in stochastic regression. *Biometrika*, 85(3), 645–660.
- Fan, J., & Yao, Q. (2008). *Nonlinear time series: nonparametric and parametric methods*. Springer Science & Business Media.
- Gerstenberger, C. (2018). Robust wilcoxon-type estimation of change-point location under short-range dependence. *Journal of Time Series Analysis*, 39(1), 90–104.
- Härdle, W. (1990). *Applied nonparametric regression*. Number 19. Cambridge university press.
- Härdle, W., & Tsybakov, A. (1997). Local polynomial estimators of the volatility function in nonparametric autoregression. *Journal of econometrics*, 81(1), 223–242.
- Inclan, C., & Tiao, G. C. (1994). Use of cumulative sums of squares for retrospective detection of changes of variance. *Journal of the American Statistical Association*, 89(427), 913–923.
- Kokoszka, P., & Leipus, R., et al. (2000). Change-point estimation in arch models. *Bernoulli*, 6(3), 513–539.
- Koubková, A. (2006). *Sequential change-point analysis*. PhD thesis.
- Lee, S., Na, O., & Na, S. (2003). On the cusum of squares test for variance change in nonstationary and nonparametric time series models. *Annals of the Institute of Statistical Mathematics*, 55(3), 467–485.
- Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability & Its Applications*, 9(1), 141–142.
- Page, E. S. (1954). Continuous inspection schemes. *Biometrika*, 41(1/2), 100–115.
- Pedoe, D. (2013). *Geometry: A comprehensive course*. Courier Corporation.
- Pishro-Nik, H. (2016). *Introduction to probability, statistics, and random processes*. University of Massachusetts.

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.

Ross, G. J. (2013). Modelling financial volatility in the presence of abrupt changes. *Physica A: Statistical Mechanics and its Applications*, 392(2), 350–360.

Sahoo, P., & Riedel, T. (1998). *Mean value theorems and functional equations*. World Scientific.

Sansó, A., Carrion, J., & Aragón, V. (2020). Testing for changes in the unconditional variance of financial time series. *Revista de Economía Financiera*, 4, 32-52.

Shao, X., & Zhang, X. (2010). Testing for change points in time series. *Journal of the American Statistical Association*, 105(491), 1228–1240.

Truong, C., Oudre, L., & Vayatis, N. (2018). A review of change point detection methods. *arXiv preprint arXiv:1801.00718*.

Watson, G. S. (1964). Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 359–372.

Wigren, T. (2015). The cauchy-schwarz inequality: Proofs and applications in various spaces.

Theorem 5 [Taylor’s theorem] It provides an estimate for the error involved in the approximation of functions by polynomials. Let g be an $(n + 1)$ times differentiable function on an interval which is open and which contains the points a and x . Then

$$g(x) = g(a) + g'(a)(x - a) + \frac{g''(a)}{2!}(x - a)^2 + \dots + \frac{g^{(n)}(a)}{n!}(x - a)^n + R_n(x)$$

where $R_n(x) = \frac{g^{(n+1)}(c)}{(n+1)!}(x - a)^{n+1}$ for some number c between a and x . $R_n(x)$ which is the remainder term gives the error in the approximation. If $R_n(x) \rightarrow 0$ as $n \rightarrow \infty$, then one obtains a sequence of better and better approximations of g . When $n = 0$, then the Taylor’s theorem reduces to the Mean Value Theorem which itself is a consequence of Rolle’s theorem Sahoo and Riedel (1998).

Theorem 6 [Rolle’s theorem] If a function g is continuous on the closed interval $[a, b]$ and differentiable on an open interval (a, b) such that $g(a) = g(b)$, then $g'(c) = 0$ for some c with $a \leq c \leq b$ Sahoo and Riedel (1998).

Inequality 1 Jensen’s inequality: This can be demonstrated by remembering that the variance of every random variable X is a positive (or non-negative) value i.e

$$var(X) = \mathbb{E}(X^2) - (EX)^2 \geq 0$$

Thus $\mathbb{E}(X^2) \geq (\mathbb{E}X)^2$.

Suppose $g(\cdot) : \mathbb{R} \rightarrow \mathbb{R}$ is a convex function e.g $g(x) = x^2$. 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 , there is

$$\mathbb{E}[g(X)] \geq g(\mathbb{E}[X])$$

For the proof, Pishro-Nik (2016).

Inequality 2 Regular triangle inequality: This inequality states that for any triangle, the sum of the lengths of any two sides should be greater than or equal to the length of the side remaining. i.e Suppose without loss of generality that we have vectors x and y in L^p and that x is no smaller than y .

$$\|x\| + \|y\| \geq \|x + y\|$$

For the proof, Pedoe (2013) and by considering the Cauchy Schwartz inequality.

Inequality 3 Cauchy Schwartz Inequality: Let x and y be vectors in L^p space. Then we shall have

$$|x \cdot y| \leq \|x\| \|y\|$$

If one of the two vectors is zero, then both sides are zero. So one may assume that both x and y are non-zero so as to proof the inequality.

Also, for any $a_1, a_2, \dots, a_n, b_1, b_2, \dots, b_n \in \mathbb{R}$ it holds $|\sum a_i b_i| \leq (\sum_{i=1}^n a_i^2)^{\frac{1}{2}} (\sum_{i=1}^n b_i^2)^{\frac{1}{2}}$.

Also, for any random variables X and Y , it holds $\mathbb{E}[|XY|] \leq \mathbb{E}[X^2]^{\frac{1}{2}} \mathbb{E}[Y^2]^{\frac{1}{2}}$ Wigren (2015).

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).