Nonparametric Estimation of the Error Functional of a Location-Scale Model

Emmanuel Torsen[1], Peter N. Mwita[2] and Joseph K. Mung'atu[3]

[1]Pan African University, Institute of Basic Sciences, Technology, and Innovation, Kenya
[2]Department of Mathematics, Machakos University, Kenya
[3] Department of Statistics and Actuarial Sciences, Jomo Kenyatta University of Agriculture and Technology, Kenya

ARTICLE HISTORY
Compiled July 15, 2018

ABSTRACT
Two estimators of the distribution of the error term are proposed based on nonparametric regression residuals; considering a heteroscadastic location-scale model where the mean and variance functions are smooth, and the error term is independent of the independent variable. The asymptotic properties of the two estimators: the unconditional cumulative distribution estimator and the conditional cumulative distribution estimator were examined. Simulation study was conducted, the mean square error of the unconditional cumulative distribution estimator was found to be smaller in comparison to its conditional cumulative distribution estimator counterpart. Hence, we recommend the use of the former.

KEYWORDS
Nonparametric Estimation, Residuals, Error Term, Location-Scale Model

1. Introduction

In statistics, the insight to the innovation distribution is crucial, particularly, in VaR. In finance VaR is a single number measuring the risk of a financial position over a specific period.

The problem of estimating the error variance in homoscedastic nonparametric regression models has studied in the literature, see among others for example [2],[10]

[1] studied the nonparametric estimation of the residual distribution, they considered the heteroscedastic regression model (1). Weak convergence of their proposed residual based estimator was examined, extending the classical work of [3] and [8] Applications to prediction interval and goodness-of-fit were also discussed.

In their paper, [7] examined the problem of fitting a known distribution to the error distribution in a class of stationary and ergodic time series models. Where particularly, the authors considered the GARCH and ARMA-GARCH models.

Non-parametric estimation of the distribution function of the error term is the focus of this paper. In order to achieve this, firstly, we obtain non-parametric estimators (9)

E. Torsen. Email: torsen@mautech.edu.ng
P.N. Mwita. Email: petermwita@mks.ac.ke
J.K. Mung'atu. Email: j.mungatu@fsc.jkuat.ac.ke

and (15) of the conditional location and scale functionals and hence, the standardized non-parametric residuals (SNR) (16).

Based on the SNR (16) the proposed estimator of the distribution function of the error term is the empirical distribution function.

2. Method and Estimation

Let $\{Y_i\}$ denote a stochastic process representing the returns on a given stock, portfolio or market index, where $i \in \mathbb{Z}$ indexes a discrete measure of time, and $F(y|x)$ denote the conditional distribution of Y_i given $X_i = x$. The vector $X_i \in \mathbb{R}^d$ normally includes lag returns $\{Y_i\}$, $1 \leq l \leq p$, for some $p \in \mathbb{N}$, as well as other relevant conditioning variables that reflect economic or market conditions.

Here, we assume that processes Y_i admits a location-scale representation given as

$$Y_i = m(X_i) + \sqrt{h(X_i)}\epsilon_i \tag{1}$$

where $m(.)$ is the unknown nonparametric regression curve and $h(.) > 0$ is a conditional scale function representing heteroscedasticity, defined on the range of X_i , ϵ_i is independent of X_i, and ϵ_i is an independent and identically distributed (iid) innovation process with $\mathbb{E}(\epsilon_i) = 0$, $\mathbb{V}ar(\epsilon_i) = 1$ and distribution function F_ϵ is unknown.

From (1);

$$Q_{Y|X}(\tau|x) = m(X_i) + \sqrt{h(X_i)}q(\tau) \tag{2}$$

where $Q_{Y|X}(\tau|x)$ is the conditional $\tau-$quantile associated with $F(y|x)$ and $q(\tau) = F_\epsilon^{-1}(\tau)$ is the $\tau-$quantile associated with unknown F_ϵ.

It follows from (2) that our estimator given by

$$\hat{Q}_{Y|X}(\tau|x) = \hat{m}(X_i) + \sqrt{\hat{h}(X_i)}\hat{q}(\tau) \tag{3}$$

Next, we discussed the estimation of $m(X_i)$, $h(X_i)$ and $q(\tau)$.

2.1. Local Linear Estimation (Regression)

In [4], the problem of estimating $m(X)$ in (2) is the same as LLR, estimating the intercept a. Suppose that the second derivative of $m(X)$ exist in a small neighborhood of x, then

$$m(X) \approx m(x) + m'(x)(X - x) \equiv \alpha + \beta(X - x) \tag{4}$$

Now, let us consider a sample $\{X_i, Y_i\}_{i=1}^n$ and LLR: find α and β to minimize

$$\sum_{i=1}^{n}(Y_i - \alpha - \beta(X_i - x))^2 K_1\Big(\frac{x - X_i}{b_1}\Big) \tag{5}$$

Let $\hat{\alpha}$ and $\hat{\beta}$ be the solution to the Weighted Least Square (WLS) problem in (5). Then

$$\hat{\alpha} = \frac{\sum_{i=1}^{n} W_i Y_i}{\sum_{i=1}^{n} W_i} \tag{6}$$

where

$$W(x; b_1) = K_1\Big(\frac{x - X_i}{b_1}\Big)[S_{n,2} - (x - X_i)S_{n,1}] \tag{7}$$

and

$$S_{n,j} = \sum_{i=1}^{n} K_1\Big(\frac{x - X_i}{b_1}\Big)(x - X_i)^j, \quad j = 1, 2 \tag{8}$$

Thus, the LLR estimator for $m(X)$ is defined as

$$\hat{m}(x) = \frac{\sum_{i=1}^{n} W_i Y_i}{\sum_{i=1}^{n} W_i} \tag{9}$$

Which is the first step in our estimation procedure. The second step follows, for the estimation of $h(X)$ in (2), the procedure proposed by [5] is used, and is outlined below: the estimator of $h(x)$ is

$$\hat{h}(x) := \hat{\Gamma} \tag{10}$$

where

$$(\hat{\Gamma}, \hat{\Gamma}_1) = \arg min \sum_{i=1}^{n} (r_i - \Gamma - \Gamma_1(X_i - x))^2 K_2\Big(\frac{X_i - x}{b_2}\Big) \tag{11}$$

Now, with the estimator in (9), we have the sequence of squared residuals $\{r_i = \{Y_i - \hat{m}(x)\}^2\}_{i=1}^{n}$. Therefore,

$$\hat{\Gamma} = \frac{\sum_{i=1}^{n} W_i r_i}{\sum_{i=1}^{n} W_i} \tag{12}$$

where

$$W(x; b_2) = K_2(\frac{x - X_i}{b_2})[S_{n,2} - (x - X_i)S_{n,1}] \tag{13}$$

and

$$S_{n,j} = \sum_{i=1}^{n} K_2\Big(\frac{x - X_i}{b_2}\Big)(x - X_i)^j, \quad j = 1, 2 \tag{14}$$

Hence, the smooth estimator for $h(X)$ is;

$$\hat{h}(x) = \frac{\sum_{i=1}^{n} W_i r_i}{\sum_{i=1}^{n} W_i} \tag{15}$$

The estimators in (9) and (15) are then used to get a sequence of Standardized Nonparametric Residuals (SNR) $\{\hat{\epsilon}_i\}_{i=1}^{n}$, where;

$$\hat{\epsilon}_i = \begin{cases} \dfrac{Y_i - \hat{m}(X)}{\sqrt{\hat{h}(X)}}, & \text{if } \hat{h}(X) > 0 \\ 0, & \text{if } \hat{h}(X) \leq 0 \end{cases} \tag{16}$$

In the third step, we use these SNR to obtain the conditional cumulative density estimator of F_ϵ and then set;

$$\hat{F}_{n,z}(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(z_i \leq x), \quad putting \quad z_i = \hat{\epsilon}_i \tag{17}$$

Which is the unconditional cumulative distribution estimator for $F_{n,z}(x)$, the UCDF of the error innovation.

2.2. Assumptions

A: Bandwidth

(1) $b \longrightarrow 0$, as $n \longrightarrow \infty$
(2) $nb \longrightarrow \infty$, as $n \longrightarrow \infty$

B: Kernel

(1) K has compact support
(2) K is symmetric
(3) K is Lipschitz continuous
(4) K is $\int_{-\infty}^{\infty} K(u)du = 1$ and $\int_{-\infty}^{\infty} uK(u)du = 0$ with $\mu_2(K) = \int_{-\infty}^{\infty} u^2 K(u)du$ and $R(K) = \int_{-\infty}^{\infty} K(u)^2 du$ being the second moment (Variance) and Roughness of the kernel function respectively.
(5) K is bounded and there is $\bar{K} \in \mathbb{R}$, with $K(u) \leq \bar{K} < \infty$ and $K(u) \geq 0, \forall u \in \mathbb{R}$

C:

$$(1 + b^2 C)^{-1} \approx (1 - b^2 C), \quad as \quad b \longrightarrow 0$$

Definition: ($\alpha - mixing$ **or Strong mixing)** Let $\mathcal{F}_k^l$ be the $\sigma - algebra$ of events generated by $\{Y_i, k \leq i \leq l\}$ for $l > k$. The $\alpha - mixing$ coefficient introduced by Rosenblatt (1956) is defined as

$$\alpha(k) = \sup_{\mathcal{A} \in \mathcal{F}_1^i, \mathcal{B} \in \mathcal{F}_{i+k}^{\infty}} |\mathbb{P}(AB) - \mathbb{P}(A)\mathbb{P}(B)|.$$

The series is said to be $\alpha - mixing$ if

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

The dependence described by the $\alpha - mixing$ is the weakest as it is implied by other types of mixing.

3. Asymptotic Properties of the Estimator

3.1. Asymptotic Properties of the Unconditional Cumulative Distribution Estimator

Given the observations $\{X_i, Y_i\}_{i=1}^n$, we compute the mean and variance of (17) to show the Asymptotic Properties;

Mean:

$$\begin{aligned}
\mathbb{E}[\hat{F}_{n,z}(x)] &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(z_i \leq x)\right] \\
&= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\left(\frac{Y_i - \hat{m}(X)}{\sqrt{\hat{h}}(X)} \leq x\right)\right] \\
&= \frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n} \mathbb{1}\left(\frac{Y_i - \hat{m}(X)}{\sqrt{\hat{h}}(X)} \leq x\right)\right] \\
&= \frac{1}{n}\sum_{i=1}^{n} \mathbb{P}\left(\frac{Y_i - \hat{m}(X)}{\sqrt{\hat{h}}(X)} \leq x\right) \\
&= \frac{1}{n}.n\mathbb{P}\big(z_i \leq x\big) \\
&= F_z(x)
\end{aligned} \tag{18}$$

Remark: $\hat{F}_{n,z}(x)$ is an unbiased estimator of $F_z(x)$.

Variance:

$$
\begin{aligned}
Var[\hat{F}_{n,z}(x)] &= Var\left[\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(z_i \le x)\right] \\
&= Var\left[\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}\left(\frac{Y_i - \hat{m}(X)}{\sqrt{\hat{h}}(X)} \le x\right)\right] \\
&= \frac{1}{n^2}\sum_{i=1}^{n} Var\left[\mathbb{1}\left(\frac{Y_i - \hat{m}(X)}{\sqrt{\hat{h}}(X)} \le x\right)\right] \\
&= \frac{1}{n} Var\left(\frac{Y_i - \hat{m}(X)}{\sqrt{\hat{h}}(X)} \le x\right) \\
&= \frac{1}{n}\left\{\mathbb{E}\left[\mathbb{1}^2\left(\frac{Y_i - \hat{m}(X)}{\sqrt{\hat{h}}(X)} \le x\right)\right] - \left(\mathbb{E}\left[\mathbb{1}\left(\frac{Y_i - \hat{m}(X)}{\sqrt{\hat{h}}(X)} \le x\right)\right]\right)^2\right\} \\
&= \frac{1}{n}\left\{\mathbb{E}\left[\mathbb{1}\left(\frac{Y_i - \hat{m}(X)}{\sqrt{\hat{h}}(X)} \le x\right)\right] - \big(F_z(x)\big)^2\right\} \\
&= \frac{1}{n}\left\{F_z(x) - \big(F_z(x)\big)^2\right\} \\
&= \frac{1}{n}F_z(x)\big[1 - F_z(x)\big] \\
&= \sigma_z^2
\end{aligned}
\tag{19}
$$

Hence, by central limit theorem;

$$
\sqrt{n}\big(\hat{F}_{n,z}(x) - F_z(x)\big) \xrightarrow{d} \mathbf{N}(0, \sigma_z^2) \tag{20}
$$

Mean Square Error:

$$
\begin{aligned}
MSE[\hat{F}_{n,z}(x)] &= \mathbb{E}\big[\hat{F}_{n,z}(x) - F_z(x)\big]^2 \\
&= Var\big(\hat{F}_{n,z}(x)\big) + \big[Bias(\hat{F}_{n,z}(x))\big]^2 \\
&= Var\big(\hat{F}_{n,z}(x)\big) + \big[0\big]^2 \\
&= Var\big(\hat{F}_{n,z}(x)\big) \\
&= \frac{1}{n}F_z(x)\big[1 - F_z(x)\big] \\
&= \sigma_z^2
\end{aligned}
\tag{21}
$$

We weighted (17) to have;

$$
\begin{aligned}
\tilde{F}_{n,z}(x) &= \frac{\sum_{i=1}^{n} K\left(\frac{X_i - x}{b}\right)\mathbb{1}(z_i \leq x)}{\sum_{i=1}^{n} K\left(\frac{X_i - x}{b}\right)} \\
&= \frac{\sum_{i=1}^{n} K\left(\frac{X_i - x}{b}\right) Z_i}{\sum_{i=1}^{n} K\left(\frac{X_i - x}{b}\right)}, \quad \textit{putting} \quad Z_i = \mathbb{1}(z_i \leq x) \\
&= \frac{\sum_{i=1}^{n} K\left(\frac{X_i - x}{b}\right) Z_i}{\hat{f}(x)} \\
&:= \frac{A}{B}
\end{aligned} \tag{22}
$$

where $\hat{f}(x)$ is the estimator of the kernel function. Hence, (22) is the Conditional Cumulative Distribution Estimator (CCDE)

3.2. Asymptotic Properties of Conditional Cumulative Distribution Estimator

We obtained the asymptotic properties of (22). to do so, we assume that the variance of the ratio of the two random variables ($A \quad and \quad B$) exist. Let Assumption A to C hold, then the mean and variance of our estimator are;

Consider the denominator (B),

Mean:

$$
\begin{aligned}
\mathbb{E}[B] &:= \mathbb{E}[\hat{f}(x)] \\
&= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} \frac{1}{b} K\left(\frac{X_i - x}{b}\right)\right] \\
&= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[\frac{1}{b} K\left(\frac{X_i - x}{b}\right)\right] \\
&= \int_{-\infty}^{\infty} \frac{1}{b} K\left(\frac{t - x}{b}\right) f(t)dt
\end{aligned} \tag{23}
$$

using change of variable $u = \frac{t-x}{b}$

$$
\mathbb{E}[\hat{f}(x)] = \int_{-\infty}^{\infty} K(u) f(x + bu) du
$$

We approximate the integral by using second order Taylor's expansion of $f(x + bu)$, for small b.

$$
f(x + bu) = f(x) + f^{(1)}(x) + \frac{1}{2} f^{(2)}(x)(bu)^2 + o(b^2)
$$

$\implies$

$$
\int_{-\infty}^{\infty} K(u) f(x + bu) du = f(x) + \frac{b^2}{2} f^{(2)}(x)\mu_2(K) + o(b^2)
$$

∴

$$\mathbb{E}[\hat{f}(x)] = f(x) + \frac{b^2}{2} f^{(2)}(x)\mu_2(K) + o(b^2) \tag{24}$$

Hence,

$$\begin{aligned} Bias(\hat{f}(x)) &= \mathbb{E}[\hat{f}(x)] - f(x) \\ &= \frac{b^2}{2} f^{(2)}(x)\mu_2(K) + o(b^4) \end{aligned} \tag{25}$$

Remark:

For second order kernels ($\nu = 2$), the bias is increasing in the square of the bandwidth [6].

Variance:

The kernel estimator is a linear estimator and $K\left(\frac{X_i - x}{b}\right)$ is iid,

$$\begin{aligned} Var[B] &:= Var[\hat{f}(x)] \\ &= Var\left[\frac{1}{nb}\sum_{i=1}^{n} K\left(\frac{X_i - x}{b}\right)\right] \\ &= \frac{1}{(nb)^2}\sum_{i=1}^{n} Var\left[K\left(\frac{X_i - x}{b}\right)\right] \\ &= \frac{1}{nb^2} Var\left[K\left(\frac{X_i - x}{b}\right)\right] \\ &= \frac{1}{nb^2}\left\{\mathbb{E}\left[K\left(\frac{X_i - x}{b}\right)^2\right]\right\} - \frac{1}{n}\left\{\frac{1}{b}\mathbb{E}\left[K\left(\frac{X_i - x}{b}\right)\right]\right\}^2 \end{aligned} \tag{26}$$

But the second term; $\frac{1}{b}\mathbb{E}\left[K\left(\frac{X_i - x}{b}\right)\right] = f(x) + o(1)$, which is $O(n^{-1})$. Consider the first term, do change of variable and a first order Taylor's expansion;

$$\begin{aligned} \frac{1}{b}\mathbb{E}\left[K\left(\frac{X_i - x}{b}\right)^2\right] &= \frac{1}{b}\int_{-\infty}^{\infty} K\left(\frac{t - x}{b}\right)^2 f(t)dt \\ &= \int_{-\infty}^{\infty} K(u)^2 f(x + bu)du \\ &= \int_{-\infty}^{\infty} K(u)^2 [f(x) + O(b)]du \\ &= f(x)R(K) + O(b) \end{aligned} \tag{27}$$

Therefore,

$$Var[\hat{f}(x)] = \frac{f(x)R(K)}{nb} + O(n^{-1}) \tag{28}$$

Remark:

The remainder $O(n^{-1})$ is of smaller order than the $O(\frac{1}{nb})$ leading term, since $b^{-1} \longrightarrow \infty$.

Now, the numerator (A):

Mean:

$$
\begin{aligned}
\mathbb{E}[A] &:= \mathbb{E}\left[\frac{1}{b}\sum_{i=1}^{n} K\Big(\frac{x-X_i}{b}\Big)Z_i\right] \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{v}{b}K\Big(\frac{x-u}{b}\Big)f(u,v)dudv \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} vK(t)f(x-bt,v)dtdv
\end{aligned} \tag{29}
$$

but

$$f(v|x-bt) = \frac{f(x-bt,v)}{f(x-bt)}$$

hence

$$f(x-bt,v) = f(v|x-bt)f(x-bt)$$

Therefore,

$$
\begin{aligned}
\mathbb{E}[A] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} vK(t)f(v|x-bt)f(x-bt)dtdv \\
&= \int_{-\infty}^{\infty} vK(t)f(x-bt)\Big\{\int_{-\infty}^{\infty} f(v|x-bt)dv\Big\}dt \\
&= \int_{-\infty}^{\infty} vK(t)f(x-bt)F_z(x-bt)dt
\end{aligned} \tag{30}
$$

By second order Taylor's expansion of $f(x-bt)$ and $F_z(x-bt)$, we have

$$f(x-bt) = f(x) + f^{(1)}(x)bt + \frac{1}{2}f^{(2)}(x)(bt)^2 + o(b^2)$$

$$F_z(x-bt) = F_z(x) + F_z^{(1)}(x)bt + \frac{1}{2}F_z^{(2)}(x)(bt)^2 + o(b^2)$$

Therefore,

$$\mathbb{E}\left[\frac{1}{b}\sum_{i=1}^{n} K\Big(\frac{x-X_i}{b}\Big)Z_i\right] = f(x)F_z(x)+b^2\mu_2(K)\Big[f^{(1)}(x)F_z^{(1)}(x)+\frac{1}{2}f^{(2)}(x)F_z^{(2)}(x)+\frac{1}{2}f(x)F_z^{(2)}(x)+o(b^2)\Big]$$

Hence,

$$
\begin{aligned}
\mathbb{E}[\tilde{F}_{n,z}(x)] &= \frac{\mathbb{E}\Big[\int_{-\infty}^{\infty}\hat{f}(x,Z)dxdZ\Big]}{\mathbb{E}[\hat{f}(x)]} \\
&\approx \frac{f(x)\Big\{F_z(x)+b^2\mu_2(K)\Big[f^{-1}(x)f^{(1)}(x)F_z^{(1)}(x)+\frac{1}{2}f^{-1}(x)f^{(2)}(x)F_z(x)+F_z^{(2)}(x)\Big]\Big\}}{f(x)\Big[1+\frac{b^2}{2}\mu_2(K)f^{-1}(x)f^{(2)}\Big]} \\
&= F_z(x)+\frac{b^2}{2}\mu_2(K)\Big[F_z^{(2)}(x)+2f^{-1}(x)f^{(1)}(x)F_z^{(1)}(x)F_z(x)\Big]
\end{aligned} \tag{31}
$$

by Assumption C, for the term in the denominator.

Hence,

$$
Bias[\tilde{F}_{n,z}(x)] = \frac{b^2}{2}\mu_2(K)\Big[F_z^{(2)}(x)+2f^{-1}(x)f^{(1)}(x)F_z^{(1)}(x)F_z(x)\Big] \tag{32}
$$

If the first derivative of the probability distribution function of x is zero ($f^{(1)}(x)=0$) as often assumed by some authors, this bias yield that of a fixed design [11];

$$
Bias[\tilde{F}_{n,z}(x)] \approx \frac{b^2}{2}\mu_2(K)F_z^{(2)}(x) \tag{33}
$$

The variance of our estimator (22) is obtained using the following approximation [11, 12],

$$
Var\Big(\frac{A}{B}\Big) \approx \frac{[\mathbb{E}(A)]^2}{[\mathbb{E}(B)]^2}\left\{\frac{Var(A)}{[\mathbb{E}(A)]}-\frac{2Cov(A,B)}{[\mathbb{E}(A)][\mathbb{E}(B)]}+\frac{Var(B)}{[\mathbb{E}(B)]}\right\} \tag{34}
$$

Now,

$$
\begin{aligned}
Var(A) &:= Var\Bigg[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{b}K\Big(\frac{X_i-x}{b}\Big)Z_i\Bigg]\\
&= \frac{1}{(nb)^2}\sum_{i=1}^{n}Var\Bigg[K\Big(\frac{X_i-x}{b}\Big)Z_i\Bigg]\\
&= \frac{1}{nb^2}Var\Bigg[K\Big(\frac{X_i-x}{b}\Big)Z_i\Bigg]\\
&= \frac{1}{nb^2}\Bigg\{\mathbb{E}\Bigg[K^2\Big(\frac{X_i-x}{b}\Big)Z_i^2\Bigg]-\Bigg(\mathbb{E}\Bigg[K\Big(\frac{X_i-x}{b}\Big)\Bigg]\Bigg)^2\Bigg\}\\
&\approx \frac{1}{nb}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}v^2K^2(t)f(v|x-bt)f(x-bt)dtdv-o(n^{-1})\\
&= \frac{1}{nb}\int_{-\infty}^{\infty}K^2(t)f(x-bt)\Bigg[\int_{-\infty}^{\infty}v^2f(v|x-bt)dv\Bigg]dt-o(n^{-1})\\
&= \frac{1}{nb}R(K)f(x)[\sigma_z^2+F_z^{(2)}(x)]
\end{aligned}
\tag{35}
$$

The Covariance:

$$
\begin{aligned}
Cov(A,B) &= Cov\Bigg[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{b}K\Big(\frac{X_i-x}{b}\Big)Z_i,\ \frac{1}{n}\sum_{i=1}^{n}\frac{1}{b}K\Big(\frac{X_i-x}{b}\Big)\Bigg]\\
&= \frac{1}{nb^2}Cov\Bigg[K\Big(\frac{X_i-x}{b}\Big)Z_i,\ K\Big(\frac{X_i-x}{b}\Big)\Bigg]\\
&= \frac{1}{nb^2}\mathbb{E}\Bigg[K^2\Big(\frac{X_i-x}{b}\Big)Z_i\Bigg]-\mathbb{E}\Bigg[K\Big(\frac{X_i-x}{b}\Big)Z_i\Bigg]\mathbb{E}\Bigg[K\Big(\frac{X_i-x}{b}\Big)\Bigg]\\
&\approx \frac{1}{nb}R(K)f(x)F_z(x)
\end{aligned}
\tag{36}
$$

The above results are based on first order Taylor's expansion.
Substitution into (34) yields,

$$
Var\big(\tilde{F}_{n,z}(x)\big)=\frac{R(K)\sigma_z^2}{nbf(x)}
\tag{37}
$$

Hence, by central limit theorem;

$$
\sqrt{nb}\big(\tilde{F}_{n,z}(x)-F_z(x)-Bias\big(\tilde{F}_{n,z}(x)\big)\big)\xrightarrow{d}\mathbf{N}\Bigg(0,\frac{R(K)\sigma_z^2}{nbf(x)}\Bigg)
\tag{38}
$$

The Asymptotic Mean Square Error (AMSE)
A very good way of checking the performance of $\tilde{F}_{n,z}(x)$ is by its AMSE;

$$\begin{aligned} AMSE(\tilde{F}_{n,z}(x)) &= \mathbb{E}\Big[\big\{\tilde{F}_{n,z}(x) - F_z(x)\big\}^2\Big] \\ &= \mathbb{E}\Big[\big\{\tilde{F}_{n,z}(x) - \mathbb{E}[\tilde{F}_{n,z}(x)]\big\}^2\Big] + \big\{\mathbb{E}[\tilde{F}_{n,z}(x)] - F_z(x)\big\}^2 \\ &= Var\Big(\tilde{F}_{n,z}(x)\Big) + \big\{Bias[\tilde{F}_{n,z}(x)]\big\}^2 \\ &= \frac{R(K)\sigma_z^2}{nbf(x)} + \frac{b^4}{4}\mu_2^2(K)\big(F_z^{(2)}(x)\big)^2 \end{aligned} \tag{39}$$

Optimal Bandwidth Selection

The smoothing parameter (bandwidth) that minimizes (39) is optimal for estimating (22), hence;

$$b_{opt} = \arg minAMSE(\tilde{F}_{n,z}(x)), \quad b > 0 \tag{40}$$

Therefore,

$$\frac{d}{db}AMSE(\tilde{F}_{n,z}(x)) = 0$$

this yields;

$$b_{opt} = \left\{\frac{R(K)F_z(x)[1 - F_z(x)]}{\mu_2^2(K)f(x)\big(F_z^{(2)}(x)\big)^2}\right\}^{\frac{1}{5}} * n^{-\frac{1}{5}} \tag{41}$$

4. Simulation Study

To examine the performance of our estimators, we conducted a simulation study considering the following data generating location-scale model

$$Y_t = m(Y_{t-1}) + h(t)^{1/2}\epsilon_t, \quad t = 1, 2, ..., n \tag{42}$$

where

$m(Y_{t-1}) = sin(0.5Y_{t-1}), \quad \epsilon_t \sim t(\nu = 3)$

$h(t) = h_i(Y_{t-1}) + \theta h(t-1), \quad i = 1, 2$

and

$h_1(Y_{t-1}) = 1 + 0.01Y_{t-1}^2 + 0.5sin(Y_{t-1})$

$h_1(Y_{t-1}) = 1 - 0.9exp\big(-2Y_{t-1}^2\big)$

Y_t and $h(t)$ are set to zero (0) initially, then Y_t is generated recursively from (42) above. To reduce the effect of the choice of our initial values on the samples, the first 1000 observations are discarded, the above data generating process was also considered by [9].

In this section, using the mean square error (MSE) we compared the two estimators of the innovation; the unconditional cumulative distribution (UCD) estimator (17) and the conditional cumulative distribution (CCD) estimator (22). We found that the

MSE of UCDE was 0.0 and that of CCDE was 0.07853961. The mean and variance of the error term was also verified to be approximately 0 and 1, confirming the assumption made. A sample of 10000 observations was considered in this study, with 20 replications.

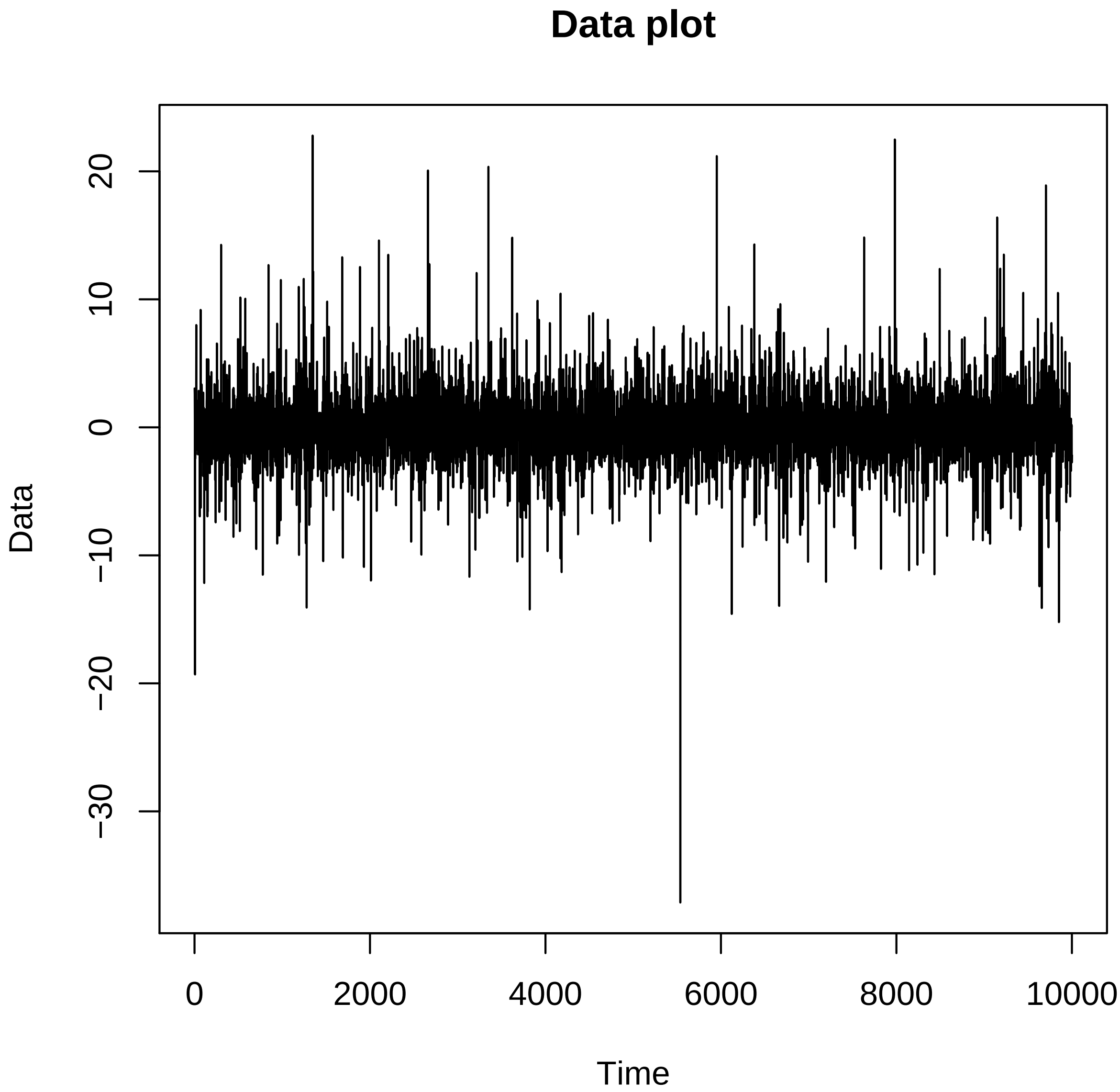

Figure 1: Plot of the simulated data showing its evolution.

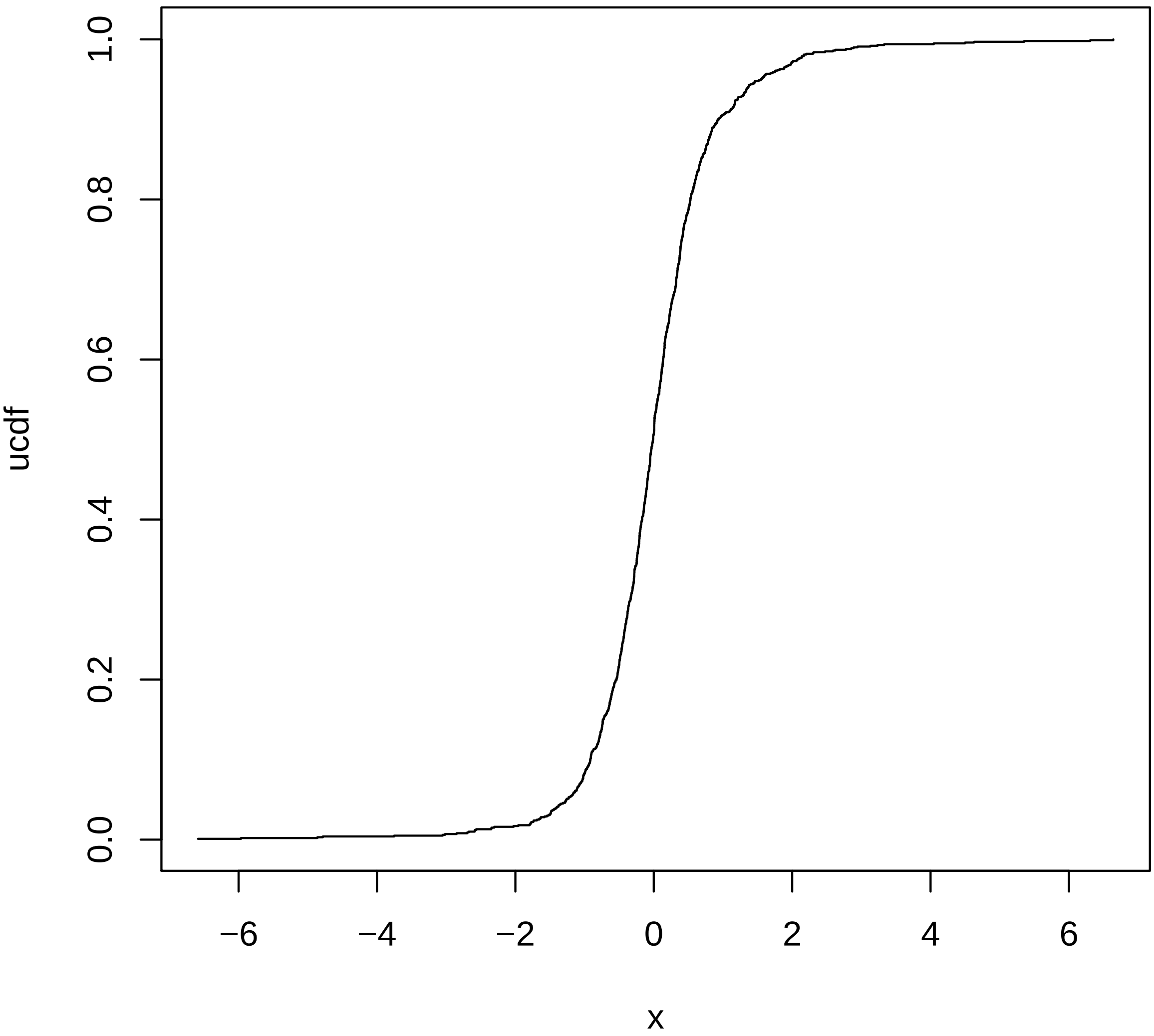

Figure 2: The unconditional cumulative distribution function.

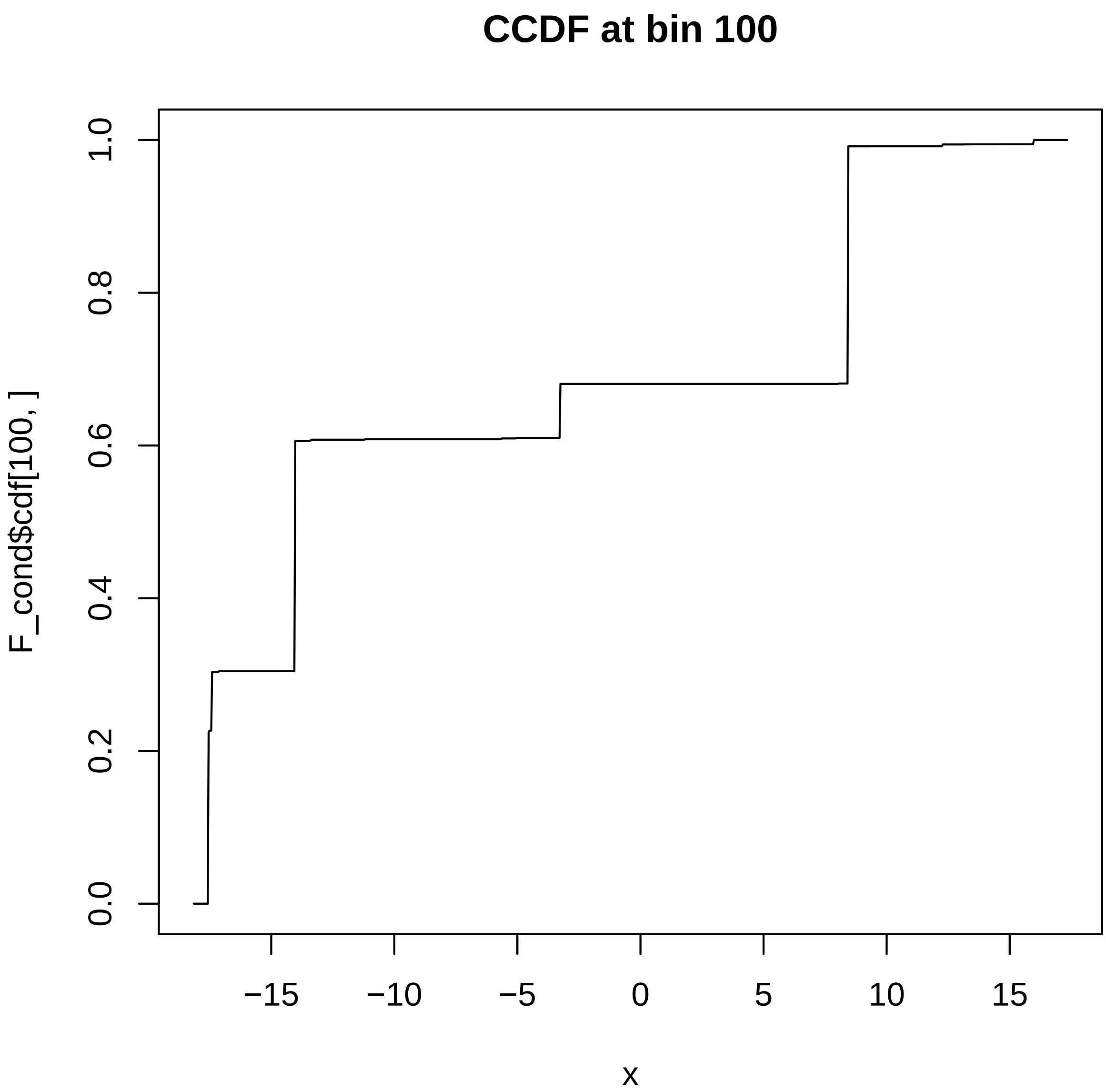

Figure 3: The conditional cumulative distribution function.

5. Conclusion

Assuming the distribution function of the innovation is unknown in a location-scale model, we proposed two estimators.The first one was unconditional and the second was conditioned. We compared the two estimators using their MSEs, the unconditional one performed better.

Acknowledgement(s)

The first author wishes to thank African Union and Pan African University, Instutute for Basic Sciences Technology and Innovation, Kenya for their supporting this research.

Disclosure statement

The authors declare that there is no conflict of interest regarding the publication of this paper.

References

[1] Akritas, M. G. and Van Keilegom, I. (2001). Non-parametric estimation of the residual distribution. *Scandinavian Journal of Statistics*, 28(3):549–567.
[2] Buckley, M., Eagleson, G., and Silverman, B. (1988). The estimation of residual variance in nonparametric regression. *Biometrika*, 75(2):189–199.
[3] Durbin, J. (1973). Weak convergence of the sample distribution function when parameters are estimated. *The Annals of Statistics*, pages 279–290.
[4] Fan, J. (1992). Design-adaptive nonparametric regression. *Journal of the American statistical Association*, 87(420):998–1004.
[5] Fan, J. and Yao, Q. (1998). Efficient estimation of conditional variance functions in stochastic regression. *Biometrika*, 85(3):645–660.
[6] Hansen, B. E. (2009). Lecture notes on nonparametrics. *Lecture notes.*
[7] Koul, H. L. and Ling, S. (2006). Fitting an error distribution in some heteroscedastic time series models. *The Annals of Statistics*, pages 994–1012.
[8] Loynes, R. (1980). The empirical distribution function of residuals from generalised regression. *The Annals of Statistics*, pages 285–298.
[9] Martins-Filho, C., Yao, F., and Torero, M. (2016). Nonparametric estimation of conditional value-at-risk and expected shortfall based on extreme value theory. *Econometric Theory*, pages 1–45.
[10] Seifert, B. and Gasser, T. (1993). Nonparametric estimation of residual variance revisited. *Biometrika*, 80(2):373–383.
[11] Seknewna, L. L., Mwita, P. N., and Muema, B. K. (2018). Smoothed conditional scale function estimation in ar (1)-arch (1) processes. *Journal of Probability and Statistics*, 2018.
[12] Seltman, H. (2012). Approximations for mean and variance of a ratio. *unpublished note.*