Smoothed Conditional Scale Function Estimation in AR(1)-ARCH(1) processes

Logamou S. Lema[1] Benjamin K. Muema[2] Peter N. Mwita[3]

Abstract

This paper is about the estimation of the Smoothed Conditional Scale Function for time series, under Conditional Heteroscedastic Innovations, by imitating the kernel smoothing in non-parametric Autoregressive-Generalized Autoregressive scheme. We based our estimation on the methodology of M-estimators for conditional quantile regression. The proof of the asymptotic properties of the Conditional Scale Function estimator for this type of process is given and its consistency shown.

Keywords & phrases Conditional Cumulative Distribution Function; Conditional Quantile; Convergence; Kernel Estimation; Quantile Autoregression; Heteroscedasticity; Uniform consistency.

1 Introduction

Consider a Quantile Autoregressive model,

$$X_t = \alpha_\tau(Z_t) + u_t, \quad t = 1, 2, \dots \tag{1.1}$$

where $\alpha_\tau(Z_t)$ is the τ^{th} Conditional Quantile Function of X_t given Z_t and the innovation u_t are assumed to be independent and identically distributed with zero τ^{th} quantile and constant scale function, see [19]. A kernel estimator of $\alpha_\tau(Z_t)$ has been determined and its consistency shown, [8]. A bootstrap kernel estimator of $\alpha_\tau(Z_t)$ was determined and shown to be consistent, [22]. This research will extend [22] by assuming that the innovations follow Quantile Autoregressive Conditional Heteroscedastic process similar to Autoregressive-Quantile Autoregressive Conditional Heteroscedastic process proposed in [19]:

$$X_t = \alpha_\tau(Z_t) + \varpi_\tau(Z_t)\varepsilon_t, \quad t = 1, 2, \dots \tag{1.2}$$

where $\alpha_\tau(Z_t)$ is the conditional θ-quantile function of X_t given Z_t; $\varpi_\tau(Z_t)$ is a conditional scale function at τ-level and ε_t is i.i.d. error with zero τ-quantile and unit scale. The function $\varpi_\tau(Z_t)$ can be expressed as

$$\varpi_\tau(Z_t) = \lambda\varpi(Z_t) \tag{1.3}$$

[1]Pan African University Institute for Basic Science, Technology and Innovation

[2]Department of Statistics and Actuarial Sciences, Jomo Kenyatta University of Agriculture and Technology, P.O Box 62000-00200, Nairobi, Kenya.

[3]Deputy Vice Chancellor Research Innovation and Linkages, Machakos University, P.O Box 136 - 90100, Machakos, KENYA.

where $\varpi(Z_t)$ is the so called volatility found in [1] and [26] which are papers of reference on Engle's ARCH models among many others and λ is a positive constant depending on τ [see [21]]. An example of this kind of function is Autoregressive - Generalized Autoregressive Conditional Heteroscedastic AR(1)-GARCH(1,1)),

$$X_t = \alpha_t + \varpi_t e_t,\ t = 1, 2, \ldots, \tag{1.4}$$

where $\alpha_t = \mu + \delta X_{t-1},\ \varpi_t = \sqrt{w + \alpha X_{t-1}^2 + \beta \varpi_{t-1}^2},\ \mu \in (-\infty, \infty), \delta \in (0,1), \beta \geq 0, \alpha > 0, w > 0$ and $e_t \sim$ i.i.d. with 0 mean and variance 1. Note that α_t may also be an ARMA (see [30]).

Considering other financial time series models, the model (1.1) can be seen as a robust generalization of AR-ARCH- models, introduced in [30], and their non-parametric generalizations reviewed by [13]. For instance, consider a financial time series model of AR(p)-ARCH(p)-type,

$$X_t = \alpha(Z_t) + \varpi(Z_t) e_t,\ t = 1, 2, \ldots \tag{1.5}$$

Where $Z_t = (X_{t-1}, X_{t-2}, \cdots, X_{t-p})$, $\alpha(\cdot)$ and $\varpi(\cdot)$ arbitrary functions.

The focus of this paper is to determine a smoothed estimator of the conditional scale function (CSF) and its asymptotic properties. This study is essential since volatility is inherent in many areas, for example, Hydrology, Finance, Weather, etc. The volatility needs to be estimated robustly even when the moments of distribution do not exist.

A partitioned stationary α-mixed time series (X_t, Z_t), where the $X_t \in \mathbb{R}$ and the variate $Z_t \in \mathbb{R}^d$ are respectively $\mathcal{A}_t$-measurable and $\mathcal{A}_{t-1}$-measurable is considered. For some $\tau \in (0,1)$, the conditional τ-quantile of X_t given the past F_{t-1} assumed to be determined by Z_t is estimated. For simplicity, we assume that $Z_t = X_{t-1} \in \mathbb{R}$ throughout the rest of the discussion.

We derive a smoothed non-parametric estimator of $\varpi_\tau(x)$ and show its consistency, asymptotic normality and uniform convergence using standard estimate of Nadaraya (1964)-Watson (1964) type. This estimate is obtained from the estimate of the conditional scale function in [20] which is a type of estimator that has some disadvantages of not being adaptive and having some boundary effects but can be fixed by well-known techniques ([11]). It's though a constrained estimator in (0,1) and a monotonically increasing function. This is very important to our estimation of the conditional distribution function and its inverse.

2 Methods and estimations

Let $f(z)$ and $f(x, z)$, denote the probability density function (pdf) of X_t and the joint pdf of (X_t, Z_t). The dependence between the exogenous X_t and the endogenous variables is described by the following conditional probability density function (CPDF)

$$f(x|z) = \frac{f(x,z)}{f(x)} \tag{2.1}$$

and the conditional cumulative distribution function (CCDF)

$$F(x|z) = \int_{-\infty}^{x} f(s|z)ds = P(X \leq x | Z_t = z) = \mathrm{E}\left[I_{\{X_t \leq x\}} \mid Z_t = z\right] \tag{2.2}$$

The estimation of the conditional scale function is derived through the CCDF. However, the following assumptions and definitions are necessary.

Assumption 2.1. *(i)* $f(x,z)$ *and* $f(z)$ *exist.*

(ii) for fixed (x,z), $0 < F(x|z) < 1$ *and* $f(z) > 0$ *are continuous in the neighborhood of* z *where the estimator is to be estimated.*

(iii) The derivatives $F^{(j)}(x) = \dfrac{d^j F(x|z)}{dz^j}$ *and* $f^{(j)}(z) = \dfrac{d^j f(z)}{dz^j}$, *for* $j = 1, 2$, *exist*

(iv) $F(x|z)$ *is a convex function in* x *for fixed* z.

(v) The conditional density $f(x|z) = \dfrac{dF(x|z)}{dx}$ *exists and is continuous in the neighborhood of* x

(vi) $f\left(\varpi_\tau(z) \mid z\right) > 0$

Assumption 2.2. *The kernel function* $\mathrm{K} : \mathbb{R}^d \longrightarrow \mathbb{R}$ *is:*

- *Symmetrical:* $\mathrm{K}(s) = \mathrm{K}(-s)$ *with* $s \in \mathbb{R}^d$,
- *Nonnegative and bounded: For* $\Gamma < \infty$, $0 < \mathrm{K}(s) \leq \Gamma$, $s \in \mathbb{R}^d$.
- *Lipschitz:* $\exists \lambda > 0, m_k < \infty$ *such that* $\left|\mathrm{K}(s) - \mathrm{K}(t)\right| \leq m_k |s-t|^\lambda$ *for all* $s, t \in \mathbb{R}^d$.
- *a pdf:* $\displaystyle\int \mathrm{K}(s) ds = 1$ *with* $\displaystyle\int_{\mathbb{R}^d} s\,\mathrm{K}(s) = 0$.

Assumption 2.3. *The process* $\left\{(X_t, Z_t)\right\}$ *is strong mixing with* $\alpha(s) = o\left(s^{-2-\delta}\right)$, $\delta > 0$.

Assumption 2.4. *The sequence* $\{b_n\}_{n\in\mathbb{N}}$ *of the smoothing parameters is such that* $b_n \longrightarrow 0$, $nb_n^p \longrightarrow \infty$ *as* $n \to \infty$ *and* $b_n > 0$.

Definition 2.1 (strong mixing). *Let* $X_t = \{\ldots, X_{t-1}, X_t, X_{t+1}, \ldots\}$ *be a stationary time series endowed with* σ*-algebras* $\mathcal{A}_t = \{X_j, -\infty < j \leq t\}$ *and* $\mathcal{A}^t = \{X_j, t \leq j < \infty\}$. *Define* $\alpha(s)$ *as*

$$\alpha(s) = \sup_{A \in \mathcal{A}_t,\, B \in \mathcal{A}^{t+s}} \left\{\left|\mathrm{P}(A \cap B) - \mathrm{P}(A)\,\mathrm{P}(B)\right|\right\}$$

If $\alpha(s) \longrightarrow 0$ *as* $s \longrightarrow \infty$, *then the process is strong mixing.*

The results in this section are about the case when the Autoregressive part of the model (1.4), i.e, $\alpha_{t,\tau} = \alpha_\tau(z) = 0$ for any $\tau \in (0,1)$. We therefore consider the model

$$X_t = \varpi_\tau(Z_t)\varepsilon_t, \quad t = 1, 2, \ldots \tag{2.3}$$

Define the check-function as

$$\gamma_\tau(X, \mu) = \gamma_\tau(X - \mu) = \left(\tau - I_{\{X-\mu\leq 0\}}\right)(X - \mu) \tag{2.4}$$

Here, $I_{\{\}}$ is the indicator function. Therefore, γ_τ is a piece-wise monotone increasing function. $\gamma_\tau(\cdot,\cdot)$ is a function of any real random variable X with distribution function $F_X(x) = P(X \leq x) = \mathrm{E}\, I_{\{X\leq x\}}$, and a real value $\mu \in \mathbb{R}$, is the asymmetric absolute value function whose amount of asymmetry depends on τ , see [18]. In case where X_t is symmetric and $\tau = 1/2$, then we have $2\gamma_\tau(X_t, \mu)$ is an absolute value function and $\varpi_{0.5}(Z_t)$ is the conditional median absolute deviation (CMAD) of X_t . When α became 0 in model (1.5), we have a purely heteroscedastic ARCH model introduced in [7] and $\alpha_\tau(Z_t)$ for $\tau > 0.5$, in this particular case, can be seen as a conditional scale function at τ-level.

The check-function in (2.4) is Lipschitz continuous by the following theorem.

Theorem 2.1. *Let γ_τ be defined as in (2.4) and $(x, \sigma) \in \mathbb{R}^2$. Then, γ_τ satisfies the Lipschitz continuity condition:*

$$\left|\gamma_\tau(x, \sigma) - \gamma_\tau(x, \sigma')\right| \le M\left|\sigma - \sigma'\right|$$

with the Lipschitz constant $M = 1$ and for all σ, σ'.

***Proof of** Theorem 2.2.* See the proof of Lemma 3.1 in [19, p .74-75] □

By the next theorem we show clearly why the errors $\{\varepsilon_t\}$ in model (1.2) are assumed to be zero τ-quantile and unit scale

Theorem 2.2. *Consider the model (1.5) and the so-called check function in (2.4), then for $\varpi_\tau(Z_t) \in \mathbb{R}_+^*$,*

$$\varepsilon_t = \frac{X_t - \alpha_\tau(Z_t)}{\varpi_\tau(Z_t)} \tag{2.5}$$

is zero τ-quantile and unit scale. And the following equations are verifiable

$$\mathrm{P}\left(X_t \le \alpha_\tau(Z_t)) \mid Z_t\right) = \tau \quad \textit{and} \tag{2.6}$$

$$\mathrm{P}\left(\gamma_\tau\left(X_t, \alpha_\tau(Z_t)\right) \le \varpi_\tau(Z_t) \mid Z_t\right) = \tau \tag{2.7}$$

***Proof of** Theorem 2.2.* The τ^{th}-quantile operator is

$$Q_\tau(Y_t) = \inf\left\{\mu \in \mathbb{R} : \mathrm{P}(Y_t \le \mu \mid Z_t) \ge \tau\right\} \tag{2.8}$$

with well-defined properties in [19, p .9-10]. From the model (1.5), the conditional τ-quantile of X_t is

$$q_\tau(Z_t) = Q_\tau(X_t) = \alpha(Z_t) + \varpi(Z_t)q_\tau^e \tag{2.9}$$

Where q_τ^e is the τ-quantiles of e_t. Then, using model (1.5) and the equation (2.9), we get

$$X_t - q_\tau(Z_t) = \varpi(Z_t)\left(e_t - q_\tau^e\right) \tag{2.10}$$

and

$$\gamma_\tau\left(X_t, q_\tau(Z_t)\right) = \varpi(Z_t)\gamma_\tau\left(e_t, q_\tau^e\right). \tag{2.11}$$

and the τ^{th}-quantile of (2.11) is

$$Q_\tau\left(\gamma_\tau\left(X_t, q_\tau(Z_t)\right)\right) = \varpi(Z_t)Q_\tau\left(\gamma_\tau\left(e_t, q_\tau^e\right)\right) = \varpi(Z_t)Q_\tau^e \tag{2.12}$$

where Q_τ^e is the τ-quantile of $\gamma_\tau\left(e_t, q_\tau^e\right)$. Note that from (2.10), $Q_\tau\left(X_t - q_\tau(Z_t)\right) = 0$. The quotient

$$\frac{X_t - \alpha_\tau(Z_t)}{Q_\tau\left(\gamma_\tau\left(X_t, \alpha_\tau(Z_t)\right)\right)} = \frac{e_t - q_\tau^e}{Q_\tau^e} \tag{2.13}$$

is zero τ-quantile and unit scale and can be seen as model (1.2) if $\varepsilon_t = (e_t - q_\tau^e)/Q_\tau^e$, $\alpha_\tau(Z_t) = q_\tau(Z_t)$ and $\varpi_\tau(Z_t) = Q_\tau\left(\gamma_\tau\left(X_t, \alpha_\tau(Z_t)\right)\right)$.

Now, assuming that ε_t (independent of Z_t) in model (1.2) is zero τ-quantile, this is equivalent to write

$$\Pr\left(\varepsilon_t \le 0\right) = \Pr\left(\varepsilon_t \le 0 | Z_t\right) = \tau$$

$$\Rightarrow \Pr\left(\frac{X_t - \alpha_\tau(Z_t)}{\varpi_\tau(Z_t)} \le 0 \mid Z_t\right) = \tau$$

This prove (2.6) for $\varpi_\tau(z) > 0$. Also, ε_t is unit-scale, means

$$\Pr\Big(\gamma_\tau(\varepsilon_t) \leq 1\Big) = \tau \Rightarrow \Pr\left(\gamma_\tau\left(\frac{X_t - \alpha_\tau(Z_t)}{\varpi_\tau(Z_t)}\right) \leq 1 \mid Z_t\right) = \tau$$

$$\Rightarrow \Pr\Big(\gamma_\tau\left(X_t - \alpha_\tau(Z_t)\right) \leq \varpi_\tau(Z_t) \mid Z_t\Big) = \tau$$

□

Assuming $\alpha_\tau(Z_t) = 0$, the estimator, $\hat{\varpi}_\tau(Z_t)$ of the conditional scale function $\varpi_\tau(Z_t)$, is obtained through the minimization of the objective function

$$\varphi(z, \varpi) = \mathrm{E}\left[\gamma_\tau(\gamma_\tau(X_t), \varpi) \mid Z_t = z\right] \tag{2.14}$$

Thus, the conditional scale function may be obtained by minimizing $\varphi(z, \varpi)$ with respect to ϖ, i.e,

$$\varpi_\tau(z) = \underset{\varpi \in \mathbb{R}_+}{\arg\min}\, \varphi(z, \varpi) \tag{2.15}$$

and

$$\hat{\varpi}_\tau(z) = \inf\left\{\mu \in R_+^* : F(\mu|z) \geq \tau\right\} \equiv F^{-1}(\tau \mid z) \tag{2.16}$$

The kernel estimator of (2.15) at $Z_t = z$ is given by

$$\hat{\varpi}_\tau(z) = \underset{\varpi \in \mathbb{R}_+}{\arg\min}\, \hat{\varphi}_n(z, \varpi) \tag{2.17}$$

We can express the estimate of $\varphi(z, \varpi)$ in the random design as it was developed in [14]. Let $Y_t^* = \gamma_\tau(\gamma_\tau(X_t), \varpi)$ be a non-negative function of X_t and $Y^* = (Y_1^*, Y_2^*, \ldots, Y_n^*)$ a random vector in $\mathbb{R}_+^* = (0, \infty)$, $t = 1, 2, \ldots, n$. In the random design, the conditional expectation (2.14) can be rewritten as follow

$$\varphi(z, \varpi) = \mathrm{E}\left[Y^* \mid Z_t = z\right] = \int y^* f(y^* \mid z) dy^* = \int y^* \frac{f(y^*, z)}{f(z)} dy^* \tag{2.18}$$

Where $f(y^* \mid z)$ represents for the conditional pdf of $Y_t^* = y^*$ given $Z_t = z$, $f(y^*, z)$ is the joint pdf of the two random variables Y^* and Z and $f(z)$ the pdf of $Z_t = z$. Using the [23] and [29] with $\mathrm{K}_b(u) = b^{-1}\,\mathrm{K}\left(ub^{-1}\right)$, a 1-dimensional rescaled kernel with bandwidth $b > 0$, we have the following estimates of $f(y^*, z)$ and $f(z)$.

$$\begin{aligned} \hat{f}(y^*, z) &= \frac{1}{n}\sum_{t=1}^{n} \mathrm{K}_{b_z}(Z_t - z)\,\mathrm{K}_{b_{y^*}}(y^* - Y_t^*) \\ \hat{f}(z) &= \frac{1}{n}\sum_{t=1}^{n} \mathrm{K}_{b_z}(Z_t - z) \end{aligned} \tag{2.19}$$

From the estimations above, $\hat{\varphi}(z, \varpi)$ the estimate of $\varphi(z, \varpi)$, is

$$\begin{aligned} \hat{\varphi}_n(z, \varpi) &= \int \frac{y^* \sum_{t=1}^{n} \mathrm{K}_{b_z}(Z_t - z)\,\mathrm{K}_{b_{y^*}}(y^* - Y_t^*)}{\sum_{t=1}^{n} \mathrm{K}_{b_z}(Z_t - z)} dy^* \\ &= \frac{\sum_{t=1}^{n} \mathrm{K}_{b_z}(Z_t - z) \int y^*\,\mathrm{K}_{b_{y^*}}(y^* - Y_t^*) dy^*}{\sum_{t=1}^{n} \mathrm{K}_{b_z}(Z_t - z)} \\ &= \frac{\sum_{t=1}^{n} \mathrm{K}_{b_z}(Z_t - z) \int \left[(y^* - Y_t^*) + Y_t^*\right] \mathrm{K}_{b_{y^*}}(y^* - Y_t^*) dy^*}{\sum_{t=1}^{n} \mathrm{K}_{b_z}(Z_t - z)} \end{aligned} \tag{2.20}$$

and considering the regularity conditions of K_b in Assumption 2.2 and also the fact that $d(y^* - Y_t^*) = dy^*$, $Y_t^* \in \mathbb{R}_+$, we have

$$\hat{\varphi}_n(z, \varpi) = \frac{\sum_{t=1}^n \mathrm{K}_{b_z}(Z_t - z)Y_t^*}{\sum_{t=1}^n \mathrm{K}_{b_z}(Z_t - z)} = n^{-1}\sum_{t=1}^n \mathrm{K}_{b_z}(Z_t - z)Y_t^*/\hat{f}(z) \tag{2.21}$$

where $\hat{g}(z)$ is the estimate of the marginal pdf of Z_t at point z and Y^* can be rewritten as

$$Y_t^* = \left[X_t\left(\tau - I_{\{X_t \le 0\}}\right) - \varpi\right]\left(\tau - I_{\{X_t\left(\tau - I_{\{X_t\le 0\}}\right)\le\varpi\}}\right) \tag{2.22}$$

and the derivative of $\hat{\varphi}_n(z, \varpi)$ w.r.t ϖ is

$$\frac{d\hat{\varphi}_n(z,\varpi)}{d\varpi} = \left(n\hat{f}(z)\right)^{-1}\sum_{t=1}^n \mathrm{K}_{b_z}(Z_t - z)\left(I_{\{X_t\left(\tau - I_{\{X_t\le 0\}}\right)\le\varpi\}} - \tau\right) \tag{2.23}$$

The minimizer of (2.21) is obtained from $\frac{d\hat{\varphi}_n(z,\varpi)}{d\varpi} = 0$. This leads to the following equation

$$\left(n\hat{f}(z)\right)^{-1}\sum_{t=1}^n \mathrm{K}_{b_z}(Z_t - z)\left(I_{\{X_t^* \le \varpi\}}\right) = \tau \tag{2.24}$$

where

$$X_t^* = X_t(\tau - I_{\{X_t \le 0\}}) \in \mathbb{R}_+^*, \tag{2.25}$$

for all $X_t \in \mathbb{R}$, $t = 1, 2, \ldots$ Note that $Y_t^* = I_{\{X_t^* \le \varpi\}}$ in (2.18). The left part of the equation (2.24) is a (unsmoothed) conditional cumulative distribution function (CCDF),

$$\hat{F}(x^* \mid z) = \left(n\hat{f}(z)\right)^{-1}\sum_{t=1}^n \mathrm{K}_{b_z}(Z_t - z)\left(I_{\{X_t^* \le x^*\}}\right), \tag{2.26}$$

that needs to be estimated and our estimator is therefore

$$\hat{\varpi}_\tau(z) = \inf\left\{x^* \in \mathbb{R}_+ : \hat{F}(x^* \mid z) \ge \tau\right\} \equiv \hat{F}^{-1}(\tau \mid z) \tag{2.27}$$

which is equivalent to $\hat{F}\left(\hat{\varpi}(z) \mid z\right) = \tau$.

An algorithm to estimating $\hat{F}(x^* \mid z)$ is proposed in the following section. This estimator suffers from the problem of boundary effects as we can see it on figure 4.2 due to outliers. We obtain unsmoothed curves of the CCDF because of the smoothness is only in the Z direction. A method proposed by [12] to smooth in the y direction is adopted here. The form of Smoothed Conditional Distribution Estimator is

$$\tilde{F}(x^* \mid z) = \left(n\hat{f}(z)\right)^{-1}\sum_{t=1}^n \mathrm{K}_h(z - Z_t)G\left(\frac{x^* - X_t^*}{h_0}\right) \tag{2.28}$$

where $G(\cdot)$ is an integrated kernel with the smoothing parameter h_0 in the X^* direction. This estimate is smooth rather than the NW which is a jump function in y. To deal with boundary effects, one may think of the Weighted Nadaraya-Watson (WNW) estimate of the CDF discussed in [5], [11], Steikert [28, p. 3–18] among others. The WNW estimator's expression is

$$\tilde{F}_{WNW}(x^* \mid z) = \frac{\sum_{t=1}^n p_t(z,\lambda)\,\mathrm{K}_{b_z}(Z_t - z)I_{\{X_t^* \le x^*\}}}{\sum_{t=1}^n p_t(z,\lambda)\,\mathrm{K}_{b_z}(Z_t - z)} \tag{2.29}$$

with conditions $\sum_{t=1}^n p_t(z,\lambda) = 1$ and . Lambda is determined using the Newton-Raphson iteration. Smoothing the CDF does not smooth the estimator in (2.27).

2.1 Algorithm to estimate the CCDF

The denominator is easy to compute as the estimator of the probability density function of Z at point z. Below, we give an algorithm to determine $\hat{F}_n(x^* \mid z)$,

1. Obtain $X_t^*, t = 1, 2, \ldots,$ by passing X_t through the check-function defined in (2.4)
2. Check if each x_{t+1}^* is less than or equal to each observation of the whole sequence $x^* = (x_1^*, x_2^*, \ldots) \in \mathbb{R}^{n+1}$. The result determines $I_{\{x_t^* \leq x\}}$ which can be expressed in $(0,1)$-matrix of order $n \times N$. N is the number of bins.
3. Construct $z_1^* = \min(Z) < z_2^* < \cdots < z_N^* = \max(Z)$ from the sequence of i.i.d random variables $Z_1, Z_2, \ldots, Z_n$ with observations $z_1,\ z_2,\ \ldots,\ z_n$. N is the number of z_i^* from which the probability density function (pdf) of Z_t is to be estimated.
4. Determine the matrix of kernels K which is

$$K = \begin{pmatrix} K_b(z_1^* - Z_1) & K_b(z_1^* - Z_2) & \cdots & K_b(z_1^* - Z_n) \\ K_b(z_2^* - Z_1) & K_b(z_2^* - Z_2) & \cdots & K_b(z_2^* - Z_n) \\ \vdots & \vdots & \vdots & \vdots \\ K_b(z_N^* - Z_1) & K_b(z_N^* - Z_2) & \cdots & K_b(z_N^* - Z_n) \end{pmatrix}$$

   The row sums of K over n, give the estimator of the pdf of Z_t at z_i^*, $\hat{g}(z_i^*)$, $i = 1, 2, \ldots, N$. We obtain the matrix of weights W by multiplying K by the inverse of $K\mathbb{I}_n$, where $\mathbb{I}_n$ is a vector of n ones. Note that the row sums of W are 1.

   Let M be the $(0,1)$-matrix from 2. The estimator of the Conditional Cumulative Distribution Function (CCDF) is

$$\hat{F}(x^* \mid z^*) = WM = KM\,(K\mathbb{I}_n)^{-1}$$

5. For each row of $\hat{F}(\cdot \mid \cdot)$, find the smallest x^* such that $\hat{F}(x^* \mid z^*) \geq \tau$, $\tau \in (0,1)$.
6. The quantiles $\hat{\varpi}_\tau(z)$ are the x^*'s which satisfy (2.27). This gives an unsmoothed estimator curve with bad shape at boundaries (see figure 4.2)

2.2 Nadaraya-Watson smoothing method

We can make $\hat{\varpi}_\tau(z)$ smooth by using NW regression[1]. This will provide a smoothed curve at each level $\tau \in (0,1)$. We can write the regression equation as

$$Y_t = \varpi_{\tau,s}(Z_t) + \eta_t \tag{2.30}$$

with $Y_t = \varpi_\tau(Z_t)$, $\varpi_{\tau,s}(x) = \mathrm{E}[\varpi_\tau(z)|Z_t = z]$ and the errors $\{\eta_i\}$ satisfy $\mathrm{E}[\eta_i] = 0$, $\mathrm{V}(\eta_i) = \sigma_\eta^2$ and $\mathrm{Cov}(\eta_i) = 0$ for $i \neq j$. Note that $\varpi_{\tau,s}(x)$ can be derived using joint pdf $f(y,z)$ as

$$\varpi_{\tau,s}(z) = \mathrm{E}[Y|Z = z] = \int y \frac{f(y,z)}{f(z)} dy \tag{2.31}$$

where $f(y,z)$ and $f(z)$ can be estimated as in (2.19).

[1] One can also use LOWESS (LOcally WEighted Scatter-plot Smoother) regression introduced by [4] to smooth the estimator in (2.27) and which solves the problem of boundary effects.

We can perform some transformations on (2.31) in order to show that it's actually better that the unsmoothed one. By assumption 2.1 (iv) and the fact that $F\left(\varpi_\tau(z) \mid z\right) = \tau$, we have

$$\begin{aligned} F\left(\varpi_{\tau,s}(Z_t) \mid z\right) &= F\left(\mathrm{E}[\varpi_\tau(z)|Z_t = z] \mid z\right) \\ &\le \mathrm{E}\left[F\left(\varpi_\tau(z) \mid z\right)|Z_t = z\right] \\ &= F\left(\varpi_\tau(Z_t) \mid z\right) \\ &= \tau \end{aligned}$$

We've have used the Jensen's theorem for conditional expectation found in [3]. $\varpi_{\tau,s}(Z_t)$ is also element of the set in which the unsmoothed estimator belongs. This means that $F\left(\varpi_{\tau,s}(Z_t) \mid z\right) \ge \tau$. The estimator is therefore, given by

$$\hat{\varpi}_{\tau,s}(z) = \frac{\sum_{t=1}^n K_b(Z_t - z)y_t}{\sum_{t=1}^n K_b(Z_t - z)} = \frac{\sum_{t=1}^n K_b(Z_t - z)\varpi_\tau(Z_t)}{\sum_{t=1}^n K_b(Z_t - z)} \tag{2.32}$$

2.2.1 Asymptotic properties

To show the asymptotic properties of our estimator, we compute its expectation and variance. Assuming the data (Y_t, Z_t) is i.i.d, the expectation of the numerator is given by

$$\begin{aligned} \mathrm{E}\left[K_b(Z_t - z)Y_t\right] &= \int\int \frac{v}{b}K\left(\frac{u-z}{b}\right) f(u,v)dudv \\ &= \int\int vK(s)f(v|z+sb)f(z+sb)dsdv \\ &= \int K(s)f(z+sb)\left(\int vf(v|z=sb)dv\right)ds \\ &= \int K(s)f(z+sb)\varpi_{\tau,s}(z+sh)ds \end{aligned}$$

We assume that the first and the second derivatives of $\varpi_{\tau,s}(z)$ at point $Z_t = z$ exist. That is, by the Taylor's expansion of $f(z+sb)$ and $\varpi_{\tau,s}(z+sh)$, we get

$$\begin{aligned} &\mathrm{E}\left[K_b(Z_t - z)Y_t\right] \\ &= \varpi_{\tau,s}(z)f(z) + \frac{1}{2}b^2\mu_2(K)\left(f(z)\varpi_{\tau,s}^{(2)}(z) + f^{(1)}(z)\varpi_{\tau,s}^{(1)}(z) + f^{(2)}(z)\varpi_{\tau,s}(z)\right) + o(h^3) \end{aligned} \tag{2.33}$$

Similarly, the expectation of the numerator is

$$\mathrm{E}\left[K_b(Z_t - z)\right] = f(z) + \frac{1}{2}b^2\mu_2(K)f^{(2)}(z) + O(h^2). \tag{2.34}$$

For b^2 small enough, $\left(1 + \frac{1}{2}b^2\mu_2(K)\frac{f^{(2)}(z)}{f(z)}\right)^{-1} \approx 1 - \frac{1}{2}b^2\mu_2(K)\frac{f^{(2)}(z)}{f(z)}$. Thus,

$$\mathrm{E}\left[\hat{\varpi}_{\tau,s}(z)\right] \approx \varpi_{\tau,s}(z) + \frac{1}{2}b^2\mu_2(K)\left(\varpi_{\tau,s}^{(2)}(z) + 2\frac{f^{(1)}(z)}{f(z)}\varpi_{\tau,s}^{(1)}(z)\right) \tag{2.35}$$

The variance of the numerator, say $V(N)$, is

$$
\begin{aligned}
\mathrm{V}\left(\frac{1}{n}\sum_{t=1}^{n} K_b(Z_t - z)Y_t\right) &= \frac{1}{nb^2}\,\mathrm{V}\left(K\left(\frac{Z_t - z}{b}\right)Y_t\right) \\
&= \frac{1}{nb^2}\left(\mathrm{E}\left[K^2\left(\frac{Z_t - z}{b}\right)y_t^2\right] - \left(\mathrm{E}\left[K\left(\frac{Z_t - z}{b}\right)Y_t\right]\right)^2\right) \\
&\approx \frac{1}{nb}\iint v^2 K^2(s) f(v|z+sb) f(z+sb)\, ds\, dv - o\left(\frac{1}{n}\right) \\
&= \frac{1}{nb}\int K^2(s) f(z+sb)\left(\int v^2 f(v|z+sb)\, dv\right) ds - o\left(\frac{1}{n}\right) \\
&\approx \frac{1}{nb} R(K) f(z)\left[\sigma_\eta^2 + \varpi_{\tau,s}^2(z)\right]
\end{aligned}
$$

Note that $\int v^2 f(v|z+sb)ds \approx \mathrm{E}[Y_t^2|Z_t = z]$. Similarly, the variance of the denominator, $\mathrm{V}(D)$, is $\mathrm{V}\left(\frac{1}{n}\sum_{t=1}^{n} K_b(Z_t - z)\right) \approx \frac{1}{nb} f(z) R(K)$.

The covariance of the numerator and the denominator of the estimator in (2.32) is given by

$$
\begin{aligned}
\mathrm{Cov}(N, D) &= \mathrm{Cov}\left(\frac{1}{nb}\sum_{t=1}^{n} K\left(\frac{Z_t - z}{b}\right)Y_t, \frac{1}{nb}\sum_{t=1}^{n} K\left(\frac{Z_t - z}{b}\right)\right) \\
&= \frac{1}{nb^2}\,\mathrm{Cov}\left(K\left(\frac{Z_t - z}{b}\right)Y_t, K\left(\frac{Z_t - z}{b}\right)\right) \\
&= \frac{1}{nb^2}\left(\mathrm{E}\left[K^2\left(\frac{Z_t - z}{b}\right)Y_t\right] - \mathrm{E}\left[K\left(\frac{Z_t - z}{b}\right)Y_t\right]\mathrm{E}\left[K\left(\frac{Z_t - z}{b}\right)\right]\right) \\
&\approx \frac{1}{nb} R(K) f(z) \varpi_{\tau,s}(z) - o\left(\frac{1}{n}\right)
\end{aligned}
$$

The variance of the estimator in (2.32) is the variance of a ratio of correlated variables that can be calculated using the approximation found in [25]

$$
\mathrm{V}\left(\frac{N}{D}\right) \approx \left(\frac{\mathrm{E}[N]}{\mathrm{E}[D]}\right)^2\left[\frac{\mathrm{V}(N)}{(\mathrm{E}[N])^2} + \frac{\mathrm{V}(D)}{(\mathrm{E}[D])^2} - \frac{2\,\mathrm{Cov}(N, D)}{\mathrm{E}[N]\,\mathrm{E}[D]}\right] \tag{2.36}
$$

$$
= \frac{R(K)\sigma_\eta^2}{nbf(z)} \tag{2.37}
$$

The Central Limit Theorem (CLT) yields

$$
\sqrt{nb}\left(\hat{\varpi}_{\tau,s}(z) - \varpi_{\tau,s}(z) - \mathrm{Bias}\left(\hat{\varpi}_{\tau,s}(z)\right)\right) \xrightarrow{D} \mathcal{N}\left(0, \frac{R(K)\sigma_\eta^2}{f(z)}\right) \tag{2.38}
$$

2.3 Asymptotic normality of QARCH

The CCDF in (2.26) can be written in the form of an arithmetic mean of a random variable L:

$$\hat{F}(x^*|z) = \frac{1}{n}\sum_{t=1}^{n} L_t \quad \text{with} \quad L_t = \frac{K_{b_z}(Z_t - z) I_{\{X_t^* \leq x^*\}}}{\frac{1}{n}\sum_{t=1}^{n} K_{b_z}(Z_t - z)} \tag{2.39}$$

with and the approximation of the expectation of L is

$$\mathrm{E}\left[L_t\right] \approx \frac{\mathrm{E}\left[K_{b_z}(Z_t - z) I_{\{X_t^* \leq x^*\}}\right]}{\mathrm{E}\left[\frac{1}{n}\sum_{t=1}^{n} K_{b_z}(Z_t - z)\right]} = \frac{\mathrm{E}[N]}{\mathrm{E}[D]} \tag{2.40}$$

[see [25]]. Using the i.i.d assumption over the data, the numerator is

$$\begin{aligned}
\mathrm{E}[N] &= \frac{1}{b_z}\,\mathrm{E}\left[K\left(\frac{Z_t - z}{b_z}\right) I_{\{X_t^* \leq x^*\}}\right] \\
&= \frac{1}{b_z}\int\int_{-\infty}^{x^*} K\left(\frac{u-z}{b_z}\right) f(u,v)\,du\,dv \\
&= \int F(x^* \mid z + sh) K(s) f(z+sh)\,ds
\end{aligned} \tag{2.41}$$

We have used the change of variables $s = (u-z)/b_z$, the definition of the conditional density function turned into $f(z + sb_z, v) = f(v \mid z + sh) f(z + sb_z)$ and Fubuni's theorem for multiple integrals. Taylor series expansions of $F(v \mid z + sh)$ and $f(z + sh)$[2], yield

$$\begin{aligned}
\mathrm{E}[N] =& f(z)F(x^* \mid z) + b_z^2\mu_2(K)\Big[f^{(1)}(z)F^{(1)}(x^* \mid z) + \frac{1}{2}f^{(2)}(z)F(x^* \mid z) + \\
&\frac{1}{2}f(z)F^{(2)}(x^* \mid z) + o(b_z^2)\Big]
\end{aligned} \tag{2.42}$$

and for the denominator, we have

$$\mathrm{E}\left[D\right] = f(z) + \frac{1}{2}b_z^2\mu_2(K)f^{(2)}(z) + o(b_z^2) \tag{2.43}$$

Thus,

$$\begin{aligned}
&\mathrm{E}\left[L_t\right] \\
&\approx \frac{f(z)\left[F(x^* \mid z) + b_z^2\mu_2(K)\left(\frac{f^{(1)}(z)}{f(z)}F^{(1)}(x^* \mid z) + \frac{1}{2}\frac{f^{(2)}(z)}{f(z)}F(x^* \mid z) + \frac{1}{2}F^{(2)}(x^* \mid z)\right)\right]}{f(z)\left(1 + \frac{1}{2}b_z^2\mu_2(K)\frac{f^{(2)}(z)}{f(z)}\right)} \\
&= F(x^* \mid z) + \frac{1}{2}b_z^2\mu_2(K)\left(2\frac{f^{(1)}(z)}{f(z)}F^{(1)}(x^* \mid z) + F^{(2)}(x^* \mid z)\right) + o(b_z^4)
\end{aligned}$$

[2] For instance, $f(z+sh) = f(z) + \frac{f^{(1)}(z)}{1!}sb_z + \frac{f^{(2)}(z)}{2!}(sb_z)^2 + o(b_z^2)$.

From the assumption that $b_z \longrightarrow 0$, the denominator is approximated to $1 - b_z^2 \mu_2(K) \dfrac{f^{(2)}(z)}{2f(z)}$. Hence,

$$\operatorname{Bias}\left(\hat{F}(x^*|z)\right) \approx \frac{1}{2} b_z^2 \mu_2(K) \left(2 \frac{f^{(1)}(z)}{f(z)} F^{(1)}(x^* \mid z) + F^{(2)}(x^* \mid z)\right) \tag{2.44}$$

Some authors assumed that, in this case, the first derivative of the true pdf of Z at point z can be zero [[12]] as the one for the fixed design and therefore, the base can be

$$\operatorname{Bias}\left(\hat{F}(x^*|z)\right) \approx \frac{1}{2} b_z^2 \mu_2(K) \left(F^{(2)}(x^* \mid z)\right) \tag{2.45}$$

We have

$$\begin{aligned}
\mathrm{V}(N) &= \mathrm{V}\left(\frac{1}{b_z} K\left(\frac{Z_t - z}{b_z}\right) I_{\{X_t^* \leq x^*\}}\right) = \frac{1}{b_z^2} \mathrm{V}\left(K\left(\frac{Z_t - z}{b_z}\right) I_{\{X_t^* \leq x^*\}}\right) \\
&= \frac{1}{b_z^2}\left(\mathrm{E}\left[K^2\left(\frac{Z_t - z}{b_z}\right) I_{\{X_t^* \leq x^*\}}\right] - \left(\mathrm{E}\left[K\left(\frac{Z_t - z}{b_z}\right) I_{\{X_t^* \leq x^*\}}\right]\right)^2\right) \\
&\approx \frac{F(x^*|z) f(z) R(K)}{b_z} - o(1),
\end{aligned}$$

$$\begin{aligned}
\mathrm{V}(D) &= \mathrm{V}\left(\frac{1}{n}\sum_{t=1}^{n} K_{b_z}(Z_t - z)\right) = \frac{1}{n b_z^2} \mathrm{V}\left(K\left(\frac{Z_t - z}{b_z}\right)\right) \\
&= \frac{1}{n b_z^2}\left(\mathrm{E}\left[K^2\left(\frac{Z_t - z}{b_z}\right)\right] - \left(\mathrm{E}\left[K\left(\frac{Z_t - z}{b_z}\right)\right]\right)^2\right) \\
&\approx \frac{f(z) R(K)}{n b_z} - o\left(\frac{1}{n}\right),
\end{aligned}$$

$$\begin{aligned}
\operatorname{Cov}(N, D) &= \frac{1}{n b_z^2} \operatorname{Cov}\left(K\left(\frac{Z_t - z}{b_z}\right) I_{\{X_t^* \leq x^*\}}, K\left(\frac{Z_t - z}{b_z}\right)\right) \\
&\approx \frac{1}{n b_z^2} \mathrm{E}\left[K^2\left(\frac{Z_t - z}{b_z}\right) I_{\{X_t^* \leq x^*\}}\right] - o\left(\frac{1}{n}\right) \\
&\approx \frac{1}{n b_z} F(x^*|z) f(z) R(K)
\end{aligned}$$

Using the same approximation in (2.36), the variance of $\hat{F}(x^*|z)$ is

$$\mathrm{V}\left(L_t\right) \approx F(x^*|z)\left[\frac{R(K)\left(1 - F(x^*|z)\right)}{b_z f(z)}\right] \tag{2.46}$$

and by the Central Limit Theorem

$$\sqrt{n}\left(\hat{F}(x^*|z) - F(x^*|z) - \operatorname{Bias}\left(F(x^*|z)\right)\right) \xrightarrow{D} \mathcal{N}\left(0, \mathrm{V}\left(L_t\right)\right) \tag{2.47}$$

Notice that the expectation of $\hat{F}(x^*|z)$ is the same as the one of L and the variance is $\mathrm{V}(L_t)/n$. To show the asymptotic normality of $\hat{\varpi}_\tau(z)$, we use the following theorem.

Theorem 2.3 (**Delta Method**). *Suppose $\hat{F}(x^*|z)$ has the asymptotic normal distribution as in (2.47). Suppose $g(\cdot)$ is a continuous function that has a derivative $g^{(1)}(\cdot)$ at $\mu = \mathrm{E}\left[\hat{F}(x^*|z)\right]$. Then*

$$\sqrt{nb_z}\left(g\left(\hat{F}(x^*|z)\right) - g(\mu)\right) \xrightarrow{D} \mathcal{N}\left(0, \left[g^{(1)}(\mu)\right]^2 \frac{R(K)\left(1-F(x^*|z)\right)}{f(z)}\right) \tag{2.48}$$

Proof of *Theorem 2.3.* The first-order Taylor expansion of $g(\cdot)$ about the point μ, and evaluated at the random variable $\hat{F}(x^*|z)$ is

$$g\left(\hat{F}(x^*|z)\right) \approx g(\mu) + g^{(1)}(\mu)\left(\hat{F}(x^*|z) - \mu\right)$$

and subtracting $g(\mu)$ from both sides and multiplying by $\sqrt{nb}$, we get

$$\sqrt{nb}\left(g\left(\hat{F}(x^*|z)\right) - g(\mu)\right) \approx \sqrt{nb}g^{(1)}(\mu)\left(\hat{F}(x^*|z) - \mu\right)$$

which tends to $\mathcal{N}\left(0, \left[g^{(1)}(\mu)\right]^2 \frac{R(K)\left(1-F(x^*|z)\right)}{f(z)}\right)$ in distribution. □

For $g(\mu) = F^{-1}(\mu|z)$, thus, $g^{(1)}(\mu) = \frac{1}{f\left(F^{-1}(\mu|z)|z\right)}$. In the next section, it's shown that the AMSE (Asymptotic Mean Squared Error) of $\hat{F}(x^*|z)$ is equal to $o\left(b^4\right) + o\left(1/(nb)\right)$ which tends to 0 as $n \longrightarrow \infty$ and $b \longrightarrow 0$. This shows the consistency of the CCDF estimate, i.e, $\hat{F}(x^*|z) \longrightarrow^p F(x^*|z)$ and we have

$$\frac{1}{f\left(F^{-1}(\mu|z) \mid z\right)} \xrightarrow{p} \frac{1}{f\left(F^{-1}(\tau \mid z) \mid z\right)} = \frac{1}{f\left(\varpi_\tau(z) \mid z\right)}$$

at points x^*'s that satisfy (2.27). We also have

$$\begin{aligned} g(\mu) &= g\left(F(x^*|z) + \mathrm{Bias}\left(\hat{F}(x^*|z)\right)\right) \\ &\approx g\left(F(x^*|z)\right) + \mathrm{Bias}\left(\hat{F}(x^*|z)\right) \times g^{(1)}\left(F(x^*|z)\right) \\ &= x^* + \frac{\mathrm{Bias}\left(\hat{F}(x^*|z)\right)}{f(x^*|z)} \end{aligned} \tag{2.49}$$

for x^*'s satisfying (2.27) and replacing $\hat{F}(\varpi_\tau(z)|z)$ by $F(\hat{\varpi}_\tau(z)|z)$ using the uniqueness assumption of $\varpi_\tau(z)$, (2.48) becomes

$$\sqrt{nb}\left(\hat{\varpi}_\tau(z) - \varpi_\tau(z) - \mathrm{Bias}\left(\hat{\varpi}_\tau(z)\right)\right) \xrightarrow{D} \mathcal{N}\left(0, \frac{R(K)\tau(1-\tau)}{f(x)\left[f\left(\varpi_\tau(z) \mid z\right)\right]^2}\right) \tag{2.50}$$

with $\mathrm{Bias}\left(\hat{\varpi}_\tau(z)\right) = \dfrac{\mathrm{Bias}\left(\hat{F}(\varpi_\tau(z)|z)\right)}{f(\varpi_\tau(z)|z)} \approx \dfrac{1}{2f(\varpi_\tau(z)|z)} b_z^2 \mu_2(K)\left(F^{(2)}(\varpi_\tau(z) \mid z)\right)$

This result can be used to calculate the optimal bandwidth to compute the good estimation of the CSF.

3 Bandwidth selections

3.1 Optimal bandwidth for density estimations

In non-parametric, specially in kernel density estimations, computing a curve of an arbitrary function from the data without guessing the shape in advance, requires an adequate choice of the smoothing parameter. The most used method is the "plug-in" method which consist of assigning a pilot bandwidth in order to estimate the derivatives of $\hat{f}(z)$. We choose the bandwidth that minimizes the AMISE (Asymptotic Mean Integrated Squared Error) below.

$$
\begin{aligned}
\text{AMISE}\left(\hat{f}(z)\right) &= \int \mathrm{E}\left[\left(\hat{f}(z) - f(z)\right)^2\right] dz \\
&= \int \mathrm{E}\left[\left(\hat{f}(z) - \mathrm{E}\left[\hat{f}(z)\right] + \text{Bias}\left(\hat{f}(z)\right)\right)^2\right] dz \\
&= \int \left\{\mathrm{E}\left[\left(\hat{f}(z) - \mathrm{E}\left[\hat{f}(z)\right]\right)^2\right] + \text{Bias}^2\left(\hat{f}(z)\right)\right\} dz \\
&= \int \left\{\mathrm{V}\left(\hat{f}(z)\right) + \text{Bias}^2\left(\hat{f}(z)\right)\right\} dz \\
&= \int \left\{\frac{R(K)f(z)}{nb} + \frac{1}{4}b^4\mu_2^2(K)\left[f^{(2)}(z)\right]^2\right\} dz \\
&= \frac{R(K)}{nb} + \frac{1}{4}b^4\mu_2^2(K)R\left(f^{(2)}(z)\right)
\end{aligned}
\tag{3.1}
$$

The general form of the r^{th} derivatives of the AMISE w.r.t b where studied in [24], considering that the unknown functions in (3.1) are also functions of the smoothing parameter [citations].

$$
\frac{d}{dz^r}\text{AMISE}\left(\hat{f}(z)\right) = \frac{R(K^{(r)})}{nb^{2r+1}} + \frac{1}{4}b^4\mu_2^2(K)R\left(f^{(2+r)}(z)\right) \tag{3.2}
$$

The optimal smoothing parameter minimizing (3.2) is

$$
b^* = \left[\frac{(2r+1)R(K^{(r)})}{\mu_2^2(K)R\left(f^{(2+r)}(z)\right)}\right]^{1/(2r+5)} \times n^{-1/(2r+5)} \tag{3.3}
$$

Using this result, we came up with the optimal version of optimal bandwidth for CCDF. The aim of derivation the AMISE in (3.1) is to get the optimal bandwidth for each $f^{(r)}$ directly. As an example, we consider the Epanechnikov Kernel function in order to compute $R(K)$, $\mu_2(K)$ and the efficiency of the kernel function given by $\sqrt{\mu_2(K)}R(K)$. The Epanechnikov's kernel function is

$$
K(u) = \frac{3}{4}(1-u^2)I_{\{|u|\leq 1\}} \Rightarrow R(K) = \frac{3}{4}\int_{-1}^{1}\left(1 - 2u^2 + u^4\right) du = \frac{3}{5},
$$

$$
\mu_2(K) = \int_{-1}^{1} u^2 K(u) du = \int_{-1}^{1}\left(u^2 - u^4\right) du = \frac{1}{5}
$$

and it's efficiency is measured by

$$
\text{Eff}(K) = R(K)\sqrt{\mu_2(K)} = \frac{3}{4}\sqrt{\frac{1}{5}} = 0.268
$$

which is the smallest of all the other kernel functions.

Table 3.1: Description of the most used kernel functions

Kernel functions	**Expressions** $K(u)$	r	$R(K)$	$\mu_2(K)$	Eff(K)
Gaussian	$\frac{1}{\sqrt{2}}\exp\left(\frac{-u^2}{2}\right) I_{\mathbb{R}}$	∞	$1/2\sqrt{2}$	1	0.2821
Epanechnikov	$\frac{3}{4}(1-u^2)I_{\{\|u\|\leq 1\}}$	2	3/5	1/5	0.2683
Uniform	$\frac{1}{2}I_{\{\|u\|\leq 1\}}$	0	1/2	1/3	0.2887
Triangular	$(1-\|u\|)I_{\{\|u\|\leq 1\}}$	1			
Triweight	$\frac{35}{32}(1-u^2)^3 I_{\{\|u\|\leq 1\}}$	6	2/3	1/6	0.2722
Tricube	$\frac{70}{81}(1-\|u\|^3)^3 I_{\{\|u\|\leq 1\}}$	9	175/247	35/243	0.2689
Biweight	$\frac{15}{16}(1-u^2)^2 I_{\{\|u\|\leq 1\}}$	4	5/7	1/7	0.2700
Cosine	$\frac{\pi}{4}\cos\left(\frac{\pi}{2}u\right)$	∞	$\frac{\pi^2}{16}$	$\frac{-8+\pi^2}{\pi^2}$	0.2685

3.2 Optimal bandwidth for CCDF

The optimal bandwidth for the CCDF estimate is the one that minimizes the AMSE. It is shown below that the AMSE is actually the summation of the variance and the bias of the CCDF estimate. This is useful because when the two are linked. When the variance is big, the bias also is big and when the variance is small, the bias is small.

$$\begin{aligned}
\text{AMSE}\left(\hat{F}(x^*|z)\right) &= \text{E}\left[\left(\hat{F}(x^*|z) - F(x^*|z)\right)^2\right] \\
&= \text{E}\left[\left(\hat{F}(x^*|z) - \text{E}\left[\hat{F}(x^*|z)\right] + \text{Bias}\left(\hat{F}(x^*|z)\right)\right)^2\right] \\
&= \text{E}\left[\left(\hat{F}(x^*|z) - \text{E}\left[\hat{F}(x^*|z)\right]\right)^2\right] + \text{Bias}\left(\hat{F}(x^*|z)\right) \\
&\times \text{E}\left[\hat{F}(x^*|z) - \text{E}\left[\hat{F}(x^*|z)\right]\right] + \text{Bias}^2\left(\hat{F}(x^*|z)\right) \\
&= \text{V}\left(\hat{F}(x^*|z)\right) + \text{Bias}^2\left(\hat{F}(x^*|z)\right) \\
&= \frac{R(K)}{nb_z f(z)} F(x^*|z)\left(1 - F(x^*|z)\right) + \frac{b^4}{4}\mu_2^2(K)\left(F^{(2)}(x^* \mid z)\right)^2
\end{aligned} \tag{3.4}$$

which is given by (2.46) and (2.45). Therefore,

$$b^* = \underset{b>0}{\arg\min}\, \text{AMSE}\left(\hat{F}(x^*|z)\right) \tag{3.5}$$

and $\frac{d}{db}\text{AMSE}\left(\hat{F}(x^*|z)\right) = 0$ leads to

$$b^* = \left\{\frac{R(K)F(x^*|z)\left(1 - F(x^*|z)\right)}{\mu_2^2(K)f(z)\left(F^{(2)}(x^*|z)\right)^2}\right\}^{\frac{1}{5}} \times n^{-\frac{1}{5}} \tag{3.6}$$

This result is practically possible by estimating the unknown functions which are dependent of the smoothing parameter. $\hat{F}^{(2)}$ is the second derivative of the CCDF from (2.26) at point $Z_t = z$. The estimator of the r^{th} derivatives of (2.26) is:

$$\hat{F}^{(r)}(x^*|z) = \frac{d^r}{dz^r}\sum_{t=1}^{n} W_t(z) X_{\{X_t^* \le x^*\}} = \sum_{t=1}^{n} W_t^{(r)}(z) X_{\{X_t^* \le x^*\}} \tag{3.7}$$

with

$$W_t(z) = \frac{K\left(\frac{Z_t - z}{b}\right)}{\sum_{t=1}^{n} K\left(\frac{Z_t - z}{b}\right)} = \frac{K\left(\frac{Z_t - z}{b}\right)}{nb\hat{f}(z)} \tag{3.8}$$

The function of weights. Thus, the first derivative is given by

$$W_t^{(1)}(z) = \frac{1}{nb^2}\frac{K^{(1)}\left(\frac{Z_t - z}{b}\right)\hat{f}(z) - bK\left(\frac{Z_t - z}{b}\right)\hat{f}^{(1)}(z)}{\left[f^{(1)}(z)\right]^2} = \frac{1}{nb^2}\frac{A}{B} \tag{3.9}$$

and the second derivative is also

$$W_t^{(2)}(z) = \frac{1}{nb^2}\frac{A^{(1)}B - B^{(1)}A}{B^2} \tag{3.10}$$

with $A^{(1)} = \frac{1}{b}K^{(2)}\left(\frac{Z_t - z}{b}\right)\hat{f}(z) - bK\left(\frac{Z_t - z}{b}\right)\hat{f}^{(2)}(z)$ and $B^{(1)} = 2\hat{f}^{(1)}(z)\hat{f}(z)$. Note that the estimation of the CCDF is function of the estimation of the empirical pdf of z. An optimal bandwidth that minimizes the AMISE of $\hat{f}(z)$ can also be the one that is optimal for the estimation of the CCDF.

Recent findings on the estimation of an optimal bandwidth for KDE (Kernel Density Estimation) are numerous [[2], [10], [24]] but the estimation of an optimal smoothing parameter remains irksome due to computation issue and time consuming routines. To do so, we adopt what had been done by [10] to estimate the r^{th} derivatives of the pdf of Z_t with respect to z. We extend the idea to estimate the first and the second derivative of the CCDF with respect to z.

4 Simulation study

4.1 Model specification

The ARCH(q) models introduced by [7] is widely used in financial applications. An AR(1)-ARCH(1) is a mixed model from an AR(d) and GARCH(p,q) for $d = 1$, $p = 1$ and $q = 0$. In time series, an observation at one time can be correlated with the observations in the previous time. That is:

(∗) **Autoregressive process of order** $p = 1, 2, \ldots$

$$AR(p): \quad X_t = \mu + \delta_1 X_{t-1} + \delta_2 X_{t-2} + \cdots + \delta_p X_{t-p} + e_t, \quad \text{with } \varepsilon_t \text{ i.i.d.}$$

(∗) **Autoregressive (p)- General Autoregressive Conditional Heteroscedastic process of order** $(d = 1, 2, \ldots;\ p = 1, 2, \ldots;\ q = 1, 2, \ldots)$

$$AR(d) - GARCH(p,q): \quad X_t = \sum_{i=1}^{p} a_i X_{t-i} + \varpi_t e_t,$$

with e_t i.i.d. and $\varpi_t = \left(w + \sum_{i=1}^{p} \alpha_i u_{i-1}^2 + \sum_{i=1}^{q} \beta_i \varpi_{i-1}^2 \right)^{1/2}$.

The data to be simulated is given by $X_t = \mu + \delta X_{t-1} + \left(w + \alpha X_{t-1}^2\right)^{1/2} e_t$, $t = 1, 2, \ldots$

4.2 Model simulation

We simulated the data from (1.1) with $\mu = 0.5, \delta = .3$, for the AR(1) part, $w = 0.1, \alpha = .35$, for the ARCH(1) and $e_t \sim$ i.i.d. $\mathcal{N}(0,1)$. The data plot is represented by figure 4.2.

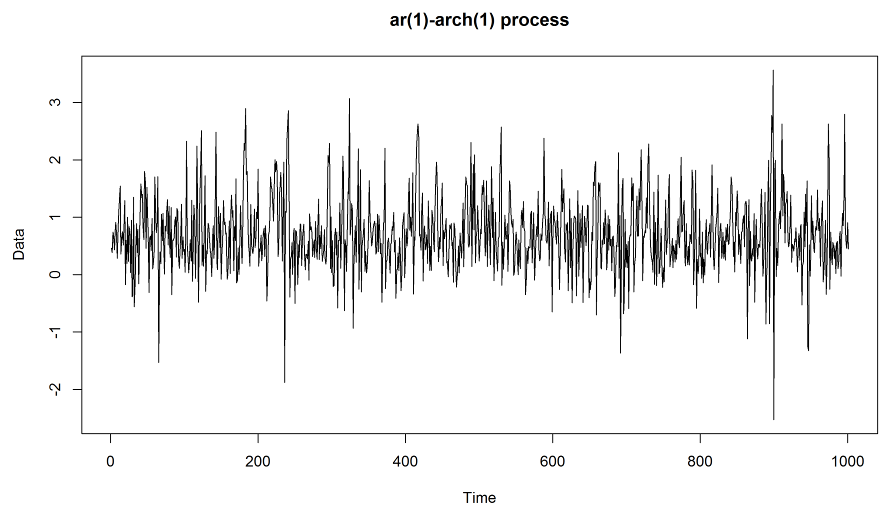

Figure 4.1: Plot of the simulated AR(1)-ARCH(1)

Our algorithm gives the estimation of the conditional scale function which suffers from boundary effects as it's seen on figure 4.2. This issue is recurrent while performing Kernel Density Estimations. The reason is that at the boundaries, $g(z)$ is underestimated because of the minimal number of points (see [17]). The consistency of our estimator is dependent on this problem of big variations at the boundaries. This increases the Average Squared Error (see the next section) between two different estimation from a same model.

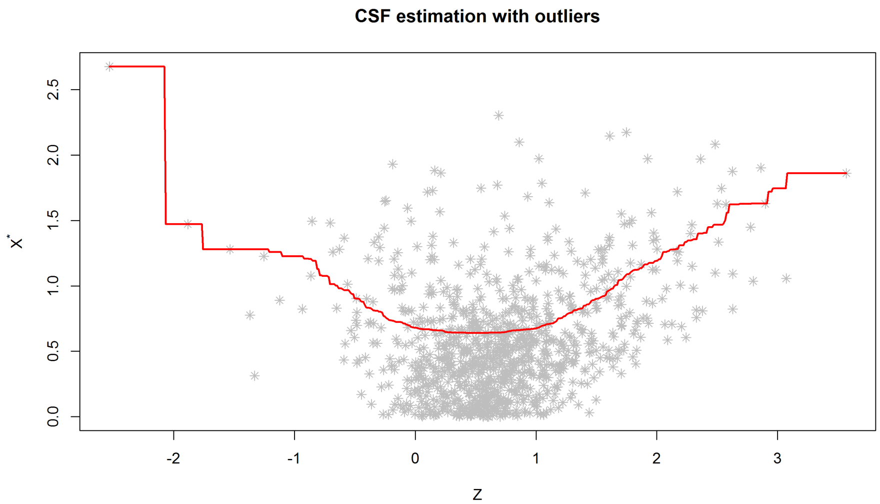

Figure 4.2: Conditional scale function estimate at level 0.75

4.3 Boundary correction

To correct the boundary effects, we use the method of box-plot fences to detect the extreme values that make the estimation too rough at the ends of the curves of the CCDF. Our estimator, being the inverse of the CCDF, is naturally rough at extremities. Among the Kernel functions, only the Gaussian can handle the sparseness of points at boundaries. The other kernel functions can bring zero at extremities and make the estimation of the CCDF wrong. What we do is to omit the points that are extremely far from the others by the box-plot fences method. The method consist of determining the first and the third quantiles from the Z_t's. Outliers are the points that are located outside the interval

$$[Q1 - 3 \times (Q3 - Q1),\ Q3 + 3 \times (Q3 - Q1)] \tag{4.1}$$

where $Q1$ and $Q3$ are the first and the third quantiles. The following figure 4.3 is the representation of Z_t and the transformed response variable X_t^* defined in (2.25) at level $\tau = 0.75$.

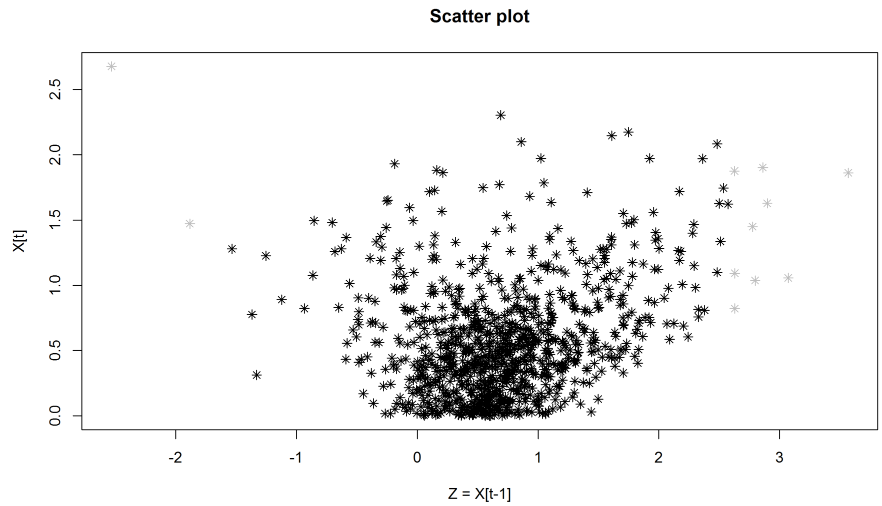

Figure 4.3: Scatter plot and outliers detection

The gray points are outliers from (4.1). We loose some information by deleting them but we get the possibility to perform the estimation a continuous curve of the CSF. The next figure is the estimations of the CSF at levels 0.25, 0.5 (median), 0.75 and 0.9. As we can see on the graphic, despite the optimal bandwidth for the empirical pdf of Z_t at point z, we get unsmoothed curves at high level $\tau > 0.5$.

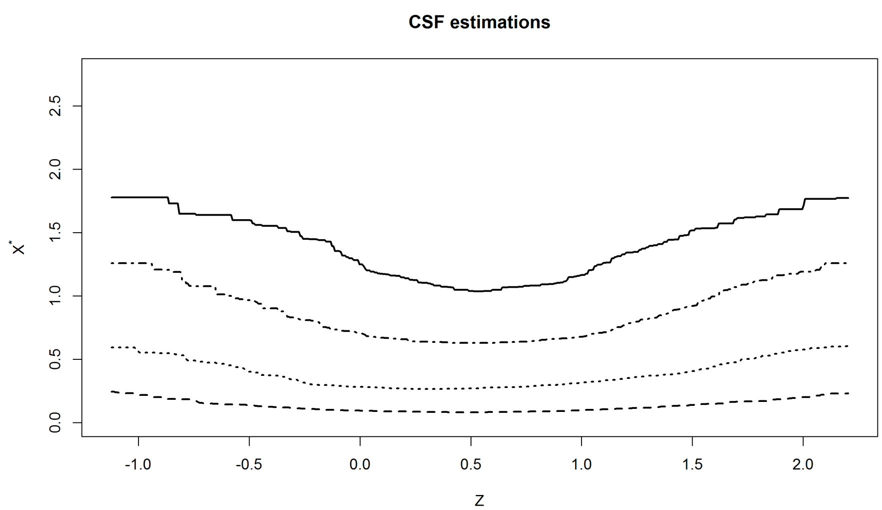

Figure 4.4: CSF estimations

The curves represent the estimations of the CSF at τ = 0.9, 0.75, 0.50, 0.25 from up to down. As it's seen on figure 4.3, the curve are not smooth that why the NW method discussed in section 2.2 which requires that unsmoothed estimator and the bins $z_1^*, z_2^*, \ldots, z_N^*$. We obtain the following graphic which combines the two estimations.

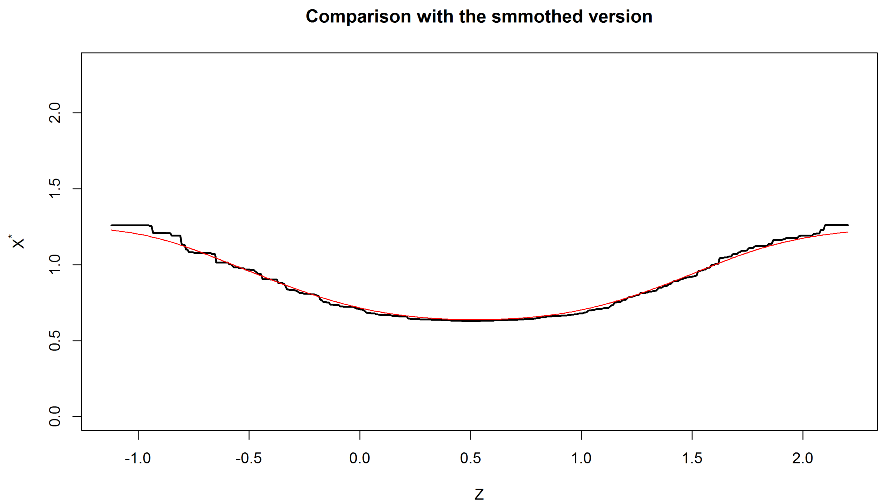

Figure 4.5: Smoothed estimate of the CSF

The next section discusses how precised is our estimation with the optimal bandwidth selection with the calculation of the MASE(Mean Average Squared Errors).

4.4 Consistency

The consistency of the estimator can be shown with the calculation of the Mean Average Squared Error providing the quantitative assessment of the accuracy of our estimator. This is a kind bootstrap method to calculate the average gap between m estimated CSFs. The formula is

$$\text{MASE}\left(\hat{\varpi}_\tau(z)\right) = \frac{1}{n}\sum_{j=1}^{n}\left[\frac{1}{m}\sum_{i=1}^{m}\left(\hat{\varpi}_{\tau,1}(z_i) - \hat{\varpi}_{\tau,j}(z_i)\right)^2\right] \tag{4.2}$$

Table (4.1) shows that the estimator of the CSF is more precised at level $\tau \leq 0.55$ for both the smoothed and the LOWESS versions.

5 Conclusion

We've derived an estimator for the conditional scale function in an AR(1)-GARCH(1) despite the heavy-tail of the data, we could deal with the boundary effect and were able to show the consistency of the estimator through a Monte Carlo study. We assumed that the QAR(1) is known and is zero and along with the regularity assumptions, we derived the estimator which can be improved in some next papers. The very next paper will focus on the estimation when the QAR(1) is unknown.

Acknowledgment

The authors thank the Pan-African University of basic science, Technology and Innovation (PAUISTI) for funding this research.

Table 4.1: Mean Average Squared Errors (MASE)

		mase(.25)						mase(.50)					
Kern. func.	n	$m = 10$		$m = 50$		$m = 100$		$m = 10$		$m = 50$		$m = 100$	
Gaussian	250	0.0017	0.0013	0.0010	0.0008	0.0010	0.0009	0.0054	0.0045	0.0063	0.0055	0.0051	0.0043
	500	0.0011	0.0008	0.0008	0.0007	0.0007	0.0006	0.0036	0.0029	0.0048	0.0042	0.0047	0.0038
	1,000	0.0006	0.0004	0.0005	0.0004	0.0005	0.0004	0.0028	0.0022	0.0025	0.0021	0.0028	0.0023
Epanech	200	0.0012	0.0009	0.0013	0.0011	0.0011	0.0009	0.0067	0.0057	0.0105	0.0091	0.0071	0.0060
	500	0.0007	0.0006	0.0010	0.0008	0.0011	0.0009	0.0045	0.0036	0.0057	0.0046	0.0042	0.0034
	1,000	0.0006	0.0004	0.0007	0.0005	0.0006	0.0005	0.0031	0.0026	0.0041	0.0033	0.0029	0.0023
Triweight	200	0.0005	0.0005	0.0006	0.0005	0.0006	0.0005	0.0008	0.0007	0.0030	0.0026	0.0039	0.0034
	500	0.0006	0.0005	0.0005	0.0004	0.0005	0.0004	0.0023	0.0020	0.0021	0.0018	0.0025	0.0021
	1,000	0.0003	0.0002	0.0003	0.0002	0.0004	0.0003	0.0019	0.0016	0.0016	0.0013	0.0016	0.0013

		mase(.75)						mase(.90)					
Kern. func.	n	$m = 10$		$m = 50$		$m = 100$		$m = 10$		$m = 50$		$m = 100$	
Gaussian	250	0.0234	0.0183	0.0237	0.0197	0.0294	0.0253	0.0880	0.0692	0.1180	0.0893	0.0971	0.0770
	500	0.0227	0.0178	0.0223	0.0178	0.0171	0.0132	0.0468	0.0377	0.0890	0.0644	0.0742	0.0525
	1,000	0.0099	0.0079	0.0138	0.0110	0.0125	0.0095	0.0932	0.0690	0.0491	0.0367	0.0510	0.0365
Epanech	200	0.0156	0.0123	0.0266	0.0223	0.0274	0.0227	0.0664	0.0577	0.1074	0.0866	0.1050	0.0844
	500	0.0184	0.0152	0.0235	0.0189	0.0181	0.0147	0.0816	0.0515	0.0827	0.0625	0.0879	0.0617
	1,000	0.0162	0.0130	0.0102	0.0074	0.0136	0.0106	0.0740	0.0510	0.0449	0.0315	0.0373	0.0274
Triweight	200	0.0190	0.0176	0.0145	0.0127	0.0167	0.0150	0.0510	0.0434	0.0452	0.0382	0.0467	0.0402
	500	0.0112	0.0099	0.0131	0.0113	0.0097	0.0081	0.0453	0.0337	0.0390	0.0333	0.0391	0.0328
	1,000	0.0075	0.0064	0.0073	0.0058	0.0069	0.0056	0.0172	0.0133	0.0268	0.0205	0.0267	0.0209

References

[1] Bollerslev, T., Engle, R. F., and Nelson, D. B. (1994). Arch models. *Handbook of econometrics*, 4:2959–3038.

[2] Chen, S. (2015). Optimal bandwidth selection for kernel density functionals estimation. *Journal of Probability and Statistics*, 2015.

[3] Chen, Z., Kulperger, R., and Jiang, L. (2003). Jensen's inequality for g-expectation: part 1. *Comptes Rendus Mathematique*, 337(11):725–730.

[4] Cleveland, W. S. (1981). Lowess: A program for smoothing scatterplots by robust locally weighted regression. *The American Statistician*, 35(1):54–54.

[5] Das, S. and Politis, D. N. (2017). Nonparametric estimation of the conditional distribution at regression boundary points. *arXiv preprint arXiv:1704.00674*.

[6] Doukhan, P. (1994). Mixing. In *Mixing*, pages 15–23. Springer.

[7] Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society*, pages 987–1007.

[8] Franke, J. and Mwita, P. (2003). Nonparametric estimates for conditional quantiles of time series.

[9] Franke, J., Mwita, P., and Wang, W. (2015). Nonparametric estimates for conditional quantiles of time series. *AStA Advances in Statistical Analysis*, 99(1):107–130.

[10] Guidoum, A. C. (2013). Kernel estimator and bandwidth selection for density and its derivatives.

[11] Hall, P., Wolff, R. C., and Yao, Q. (1999). Methods for estimating a conditional distribution function. *Journal of the American Statistical Association*, 94(445):154–163.

[12] Hansen, B. E. (2004). Nonparametric estimation of smooth conditional distributions. *Unpublished paper: Department of Economics, University of Wisconsin*.

[13] Härdle, W., Lütkepohl, H., and Chen, R. (1997). A review of nonparametric time series analysis. *International Statistical Review*, 65(1):49–72.

[14] Härdle, W., Werwatz, A., Müller, M., and Sperlich, S. (2004). Nonparametric density estimation. *Nonparametric and Semiparametric Models*, pages 39–83.

[15] Harvey, A. C. and Shephard, N. (1996). Estimation of an asymmetric stochastic volatility model for asset returns. *Journal of Business & Economic Statistics*, 14(4):429–434.

[16] Huber, P. J. (1981). Robust statistics. 1981.

[17] Karunamuni, R. J. and Alberts, T. (2005). On boundary correction in kernel density estimation. *Statistical Methodology*, 2(3):191–212.

[18] Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica: journal of the Econometric Society*, pages 33–50.

[19] Mwita, P. (2003). *Semiparametric estimation of conditional quantiles for time series with applications in finance*. PhD thesis, University of Kaiserslautern.

[20] Mwita, P. (2004). On conditional scale function: Estimate and asymptotic properties.

[21] Mwita, P. and Otieno, R. (2005). Conditional scale function estimate in the presence of unknown conditional quantile function. *African Journal of Science and Technology*, 6(1).

[22] Mwita, P. N. and Franke, J. (2013). Bootstrap of kernel smoothing in quantile autoregression process. *Journal of Statistical and Econometric Methods*, 2(3):175–196.

[23] Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability & Its Applications*, 9(1):141–142.

[24] Raykar, V. C. and Duraiswami, R. (2006). Fast optimal bandwidth selection for kernel density estimation. In *Proceedings of the 2006 SIAM International Conference on Data Mining*, pages 524–528. SIAM.

[25] Seltman, H. (2012). Approximations for mean and variance of a ratio. *unpublished note*.

[26] Shephard, N. (1996). Statistical aspects of arch and stochastic volatility. *Monographs on Statistics and Applied Probability*, 65:1–68.

[27] Silverman, B. W. (1986). *Density estimation for statistics and data analysis*, volume 26. CRC press.

[28] Steikert, K. U. (2014). *The Weighted Nadaraya-Watson Estimator: Strong Consistency to the Faculty of Economics, Business Administration and Information Technology of the University of Zurich*. PhD thesis.

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

[30] Weiss, A. A. (1984). Arma models with arch errors. *Journal of time series analysis*, 5(2):129–143.