

ESTIMATION OF SMOOTHED CONDITIONAL SCALE
FUNCTION USING QUANTILE AUTOREGRESSIVE PROCESS
WITH CONDITIONAL HETEROSCEDASTIC INNOVATIONS

Lema Logamou Seknewna

MS400-0004/15

A thesis submitted to Pan African University Institute of Basic Sciences,
Technology and innovation in partial fulfillment of the requirement for
the award of the degree of Doctor of Philosophy in Mathematics
(Statistics Option) of Pan African University

2018

Declaration

This thesis is my original work and has not been presented for a degree in any other university

Signature _____

Lema Logamou Seknewna

Date _____

This thesis has been submitted for examination with our approval as University Supervisors

Signature _____

Prof. Peter N. Mwita

Date _____

Signature _____

Dr. Benjamin K. Muema

Date _____

Acknowledgement

I would like to thank my Academic Supervisor Prof. Peter N. Mwita. I express my gratitude for the trust he has placed in me in agreeing to lead this research work. His advice and remarks have allowed me to improve myself on a daily basis.

I would also like to express my deepest gratitude to my Co-Supervisor, Dr. Benjamin K. Muema, who has given everything for this work to come into being.

I would like to thank Dr. Antony G. Waititu, not only for teaching me R Programming but for agreeing to check my codes and giving me a positive comment that filled me with the courage to deep in the research. I will never forget that.

I am thankful to the PAU Institute for Basic Sciences, Technology and Innovation for giving me the opportunity to study in Kenya.

I would also like to thank JICA JKUAT who financed the purchase of some necessity that really helped me at the beginning of my thesis.

Dedication

To my family

Abstract

In this thesis, we carried out the estimation smoothed Conditional Scale Function for an Autoregressive process with conditional heteroscedastic innovations by using the kernel smoothing approach. The estimations were based on the quantile Auregression methodology proposed by Koenker and Bassett. The proof of the asymptotic properties was given. All our estimations were made through inverting conditional distribution functions and we showed that they are weakly consistent under specific assumptions. We performed Monte Carlo studies to show the accuracy of our estimators. This study can use in area requiring conditional quantile estimations can be improve using local polynomial estimation of degree two.

Contents

Declaration	ii
Acknowledgement	iii
Dedication	iv
Abstract	v
Contents	viii
1 Introduction	1
1.1 Background of the study	1
1.2 Problem statement	3
1.3 Justification of the study	4
1.4 Scope and limitations	5
1.5 Objective of the study	5
2 Literature review	7
2.1 Quantile Autoregression (QAR)	7
2.2 Density and distribution functions estimation	8
2.3 Bandwidth selection	9
2.4 Boundary issue	10
2.5 Research Gap	10
3 Methodology	11

3.1	Empirical and inverse CCDF	11
3.2	Unconditional Cumulative Distribution Function	11
3.2.1	Kernel matrix	12
3.2.2	Indicator matrix	14
3.2.3	Empirical CCDF and consistency	16
3.2.4	Inversion	20
3.3	Model specification and simulation	20
3.3.1	Definitions	21
3.3.2	AR(1) process	21
3.3.3	ARCH(1) process	23
3.3.4	GARCH(1,1) process	24
3.3.5	AR(1)-GARCH(1,1)	26
3.3.6	Simulation of AR(1)-ARCH(1) processes	27
3.4	Conclusion	31
4	Estimation of the conditional scale function	32
4.1	Definitions	32
4.1.1	Nadaraya-Watson smoothing method	39
4.1.2	Asymptotic normality of QARCH	43
4.2	Bandwidth selections	47
4.2.1	Optimal bandwidth for density estimations	47
4.2.2	Optimal bandwidth for CCDF	49
4.3	Simulation study	51
4.3.1	Model specification	51
4.3.2	Specifications for AR(1)-GARCH(1,1)	52
4.3.3	Monte Carlo study	54
4.3.4	Boundary correction	55
4.3.5	Accuracy of the estimator	57

4.4	Conclusion	59
5	Estimation of the QAR and QARCH	60
5.1	Conditional Quantile estimations	61
5.1.1	Non-parametric QAR	61
5.1.2	k Nearest Neighbor (k -NN) prediction	62
5.1.3	Non-parametric QARCH	63
5.1.4	Quantile error estimation	64
5.2	Simulation study	65
5.3	Accuracy of the estimates	68
5.4	Conclusion	69

List of Tables

4.2.1 Description of the most used kernel functions	49
4.3.1 MASE for QAR at $\tau = 0.25$	58
4.3.2 MASE for QAR at $\tau = 0.50$	58
4.3.3 MASE for QAR at $\tau = 0.75$	58
4.3.4 MASE for QAR at $\tau = 0.90$	58
5.2.1 Summary of quantile errors	66
5.3.1 MASE for QAR and QARCH at $\tau = 0.25$	68
5.3.2 MASE MASE for QAR and QARCH at $\tau = 0.50$ (median)	69
5.3.3 MASE MASE for QAR and QARCH at $\tau = 0.75$	69
5.3.4 MASE MASE for QAR and QARCH at $\tau = 0.90$	69

List of Figures

2.3.1 Bandwidth influence on the resulting curve	9
3.2.1 Empirical cumulative distribution function	13
3.3.1 AR(1)-ARCH(1) process (1)	28
3.3.2 AR(1)-ARCH(1) process (2)	29
3.3.3 AR(1)-ARCH(1) process (3)	30
3.3.4 AR(1)-ARCH(1) process (4)	31
4.3.1 Plot of the simulated AR(1)-ARCH(1)	54
4.3.2 Conditional scale function estimate at $\tau = 0.75$	55
4.3.3 Scatter plot and outliers detection	55
4.3.4 CSF estimations	56
4.3.5 Smoothed estimate of the CSF	57
5.2.1 Predicted conditional quantile returns	66
5.2.2 Graphical superposition of $\tilde{\alpha}_\tau(z)$ [red points] and $\hat{\alpha}_\tau(z)$ [blue curve] . .	67
5.2.3 Quantile errors for $\tau = 0.75$	68

Chapter 1

Introduction

1.1 Background of the study

In many regression methods, it's usually about finding a linear or curvilinear relationship based on the scatter plot. Most regression methods estimate the average (mean) value of the response variable. Some z - x scatter plots do not obey this dictatorship due to influential points also known as outliers. Financial and insurance data among others have significant variability and are in some cases known as heavy-tailed data (Markovich, 2008). Those data possess isolated points (from the cloud) that distort any attempt to make a simple linear or other average-based regression. This is one of the reasons why many robust methods are being developed in both parametric and non-parametric fashion. Robust because they aim to get rid of being influenced by extreme values. This is the case in methods as LAD (Least Absolute Deviations) which estimate the median or 1/2-quantile value of the response variable Portnoy and Koenker (1997). Conditional quantile regression as developed in (Koenker and Zhao, 1996) is more general and gives a more general description of the response variable at each level in $(0, 1)$. The local polynomial regression method, mostly used for non-parametric estimations, is robust but is still influenced by abnormally far-off points at boundaries. Outliers pull the curve

toward them in places where there are few number of points. (Cowling and Hall, 1996) devised a method to perform the analysis without deleting them by filling the gap between the dense cloud and the very distant points by adding pseudo-points before making the non-parametric estimation of the probability density function. Our approach, in this thesis, gives absolute robustness to these non-parametric methods estimates by solving the problem of outliers, smoothing the estimators and giving the possibility in forecasting. We base our estimations on the (Nadaraya, 1964) - (Watson, 1964) (NW) method which is a particular case of local polynomial regression. The method consists of detecting points likely to change the behavior of the curves towards the borders by using the method of Tukey then making an estimation of the quantile as discussed in (Tukey, 1977) then reintegrating the outliers by predicting their response variable by k -NN algorithm. The latter is a data mining tool with predictive power from observations using distance or similarity. We performed a two step-estimation which consist of estimating the quantile location shift or the QAR (Quantile AutoRegressive). After smoothing it and predicting the response for the outliers (omitted in the first place), the CSF (Conditional scale function) is estimated from the residuals. Specific assumptions, also found in literature, are made to ascertain the consistency of ours estimations. The data generating process is discussed in Chapter 3. The combination of smoothing method and the outliers handling reduce the bias of the estimate compared to the results in (Mwita, 2003). To illustrate that, we simulated identical processes in terms of parameter, then obtained estimates from the processes and computed the quadratic errors. These errors are very small and confirm that our estimates are accurate. An easy-to-program algorithm that allows the empirical estimation of the conditional distribution function and its inverse is discussed in the section 4. Our results can be used in finance in calculating CVaR (Conditional Value-at-Risk), expected shortfall, etc.

1.2 Problem statement

Consider a Quantile Autoregressive model,

$$X_t = \alpha_\tau(Z_t) + u_t, \quad t = 1, 2, \dots \quad (1.2.1)$$

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, (Mwita, 2003). A kernel estimator of $\alpha_\tau(Z_t)$ has been determined and its consistency shown, (Franke and Mwita, 2003). A bootstrap kernel estimator of $\alpha_\tau(Z_t)$ was determined and shown to be consistent, (Mwita and Franke, 2013). This research extends (Mwita and Franke, 2013) by assuming that the innovations follow Quantile Autoregressive Conditional Heteroscedastic process similar to Autoregressive-Quantile Autoregressive Conditional Heteroscedastic process proposed in (Mwita, 2003):

$$X_t = \alpha_\tau(Z_t) + \varpi_\tau(Z_t)\varepsilon_t, \quad t = 1, 2, \dots \quad (1.2.2)$$

where $\alpha_\tau(Z_t)$ is the same as in the model (1.2.1), $\varpi_\tau(Z_t)$ is a conditional scale function at τ -level and ε_t is independent and identically distributed (i.i.d.) error with zero τ -quantile and unit scale. The function $\varpi_\tau(Z_t)$ can be expressed as

$$\varpi_\tau(Z_t) = \lambda\varpi(Z_t) \quad (1.2.3)$$

where $\varpi(Z_t)$ is the so called volatility found in (Bollerslev et al., 1994) and (Shephard, 1996) which are papers of reference on Engle's ARCH models among many others and λ is a positive constant depending on τ (Mwita and Otieno, 2005). An example of this kind of function is Autoregressive - Generalized Autoregressive Conditional Heteroscedastic AR(1)-GARCH(1,1),

$$X_t = \alpha_t + \varpi_t e_t, \quad t = 1, 2, \dots, \quad (1.2.4)$$

where

$$\begin{aligned}\alpha_t &= \mu + \delta X_{t-1} \\ \varpi_t &= \sqrt{w + \alpha X_{t-1}^2 + \beta \varpi_{t-1}^2} \\ e_t &\sim \mathcal{N}(0, 1), \text{ independent of } X_{t-1}\end{aligned}\tag{1.2.5}$$

and $\mu \in (-\infty, \infty)$, $|\delta| < 1$, $\beta > 0$, $\alpha > 0$, $w > 0$, $\alpha + \beta < 1$. Note that α_t may also be an ARMA citepweiss1984arma. The specifications for model (1.2.4) are given in section 3.3.5.

Considering other financial time series models, the model (1.2.1) can be seen as a robust generalization of AR-ARCH- models, introduced in (Weiss, 1984), and their non-parametric generalizations reviewed by (Härdle et al., 1997). For instance, consider a financial time series model of AR(p)-ARCH(p)-type,

$$X_t = \alpha(Z_t) + \varpi(Z_t)e_t, \quad t = 1, 2, \dots\tag{1.2.6}$$

Where $Z_t = (X_{t-1}, X_{t-2}, \dots, X_{t-p})$, $\alpha(\cdot)$ and $\varpi(\cdot)$ arbitrary functions representing, respectively, the conditional mean and conditional variance of the process.

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.

1.3 Justification of the study

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. The asymptotic properties of the scale function

derived through model (1.2.2) have not been found yet. Non-parametric estimation is motivated because of its flexibility to describe relationship between a dependent variable and independent variables.

1.4 Scope and limitations

This study is an extension of (Mwita, 2003) in which an estimate of CSF was made assuming that the QAR part is zero to allow a less complicated estimation. For the case where the QAR is to be estimated non-parametrically, the estimate of the CSF depends on the residuals $X_t - \alpha_\tau(z)$ which gives rise to a dimension problem. We will investigate the methods that will solve this dimension problem. This last estimate must also be smooth in order to reduce the bias. Since data from AR-ARCH processes are very noisy, our estimates may be influenced by outliers that may also increase the bias. Smoothing alone will not be very effective for robustness. We will look at the aspect of the boundary correction of the curves by suitable methods. Smoothing parameters for our estimates will be of great importance as it is a recurring topic in non-parametric estimations. Parametric estimate of the CSF will be omitted in this study.

1.5 Objective of the study

The main objective in this thesis is to derive a smoothed estimator of the CSF from Mwita and Franke (2013) and show its asymptotic properties. For this to be done, we articulated our research on three (3) specific objectives:

1. Derive a smoothed estimator of the conditional scale function when the Quantile Autoregressive part is known (equal to zero) and proof its consistency under specific assumptions;

2. Derive a smoothed estimator of the conditional scale function when the **Quantile Autoregressive** part is unknown and proof its consistency under specific assumptions;
3. Perform Monte Carlo studies to ascertain the consistency of the estimators.

Chapter 2

Literature review

2.1 Quantile Autoregression (QAR)

In recent years, quantile autoregression is recurrent in the literature when it comes to do estimations on financial data that depend on previous value. (Koenker and Bassett, 1978) developed the quantile regression to give full description of the response variable according to each level, say, $\tau \in (0, 1)$. Since the method was created by (Koenker and Bassett, 1978), it is gaining ground and applications are numerous. (Koenker and Zhao, 1996) based their estimation on (Engle, 1982)'s ARCH models to estimate the conditional functions in non-parametric way. An example of modeling today's return on the ones from yesterday are found in (Koenker, 2001). Another application is the calculation of students' scores using quantile regression, see (Koenker and Hallock, 2001). Mwita (2003) proposed alternative non-parametric estimation of the conditional scale function by minimizing a conditional expectation of loss function which lead to the estimation of an estimation conditional cumulative distribution function which is influenced by the outliers. Methods for correcting the boundary where proposed but did not help to ascertain the accuracy of the estimations. The estimation of the CSF assuming the Quantile Autoregression part to be zero was also investigated and the consistency proven under

specific assumptions. The case where the QAR is unknown was theoretically motivated and left for further research. Mwita and Otieno (2005) extended research on the estimation of CSF for unknown QAR which remained theoretical, i.e. without a simulation study to show the accuracy of the estimator. (Franke and Mwita, 2003) discussed bootstrap estimations of the QAR based on the method found in Mwita (2003). The asymptotic properties of the estimates were shown and the estimator was also shown to be more accurate than the previous estimation. Franke et al. (2015) applied method based on the QAR estimation in Mwita (2003) to estimate Value-At-Risk (VaR) for stocks in DAX. Simulation studies were also performed in order to evaluate the performance of the estimates.

2.2 Density and distribution functions estimation

A review on kernel density estimation is found in Zambom and Dias (2012). The most used estimation of the probability density function (pdf) for a sequence of real valued random variable is the kernel density function that require the choice of a kernel function (Min and Lee, 2005), (Baudat and Anouar, 2001) where b is the bandwidth and also called the smoothing parameter. This method of estimation does depend on the smoothing parameter because a very small bandwidth provides a noisy curve and a big one give a flat curve. It's required to use the existing method to get the estimation of b before doing any estimation related to it. It's still a challenge because the estimators for the optimal bandwidth. This kind of estimations have the disadvantage of being influenced by the outliers.

The kernel estimation of the condition distribution function (CDF) requires the probability density function. The CDF also depend on the inverse of the Conditionaribution Function trough inversion. Method for estimating the CDF are numerous and we have the Weighted Nadaraya-Watson (WNW) estimate of the CDF discussed in (Das and Poli-

tis, 2017), (Hall et al., 1999), (Steikert, 2014, p. 3–18) among others.

2.3 Bandwidth selection

Finding the optimal smoothing parameter is a major problem in non-parametric because the shape of the estimated curve depends on. A very small bandwidth underestimate the function of interest and a big bandwidth leads to overestimation (see Table 2.3.1). The bandwidth estimation is not standard because there is no systematic way to do so. The estimation varies from one problem to another. (Avramidis, 2016) proposed cross-validation method to estimate optimal bandwidth for kernel based estimation given a sequence of random variables Z_1, Z_2, \dots, Z_n . A smoothing parameter selection procedure is proposed (Abberger, 1997) for the same objective.

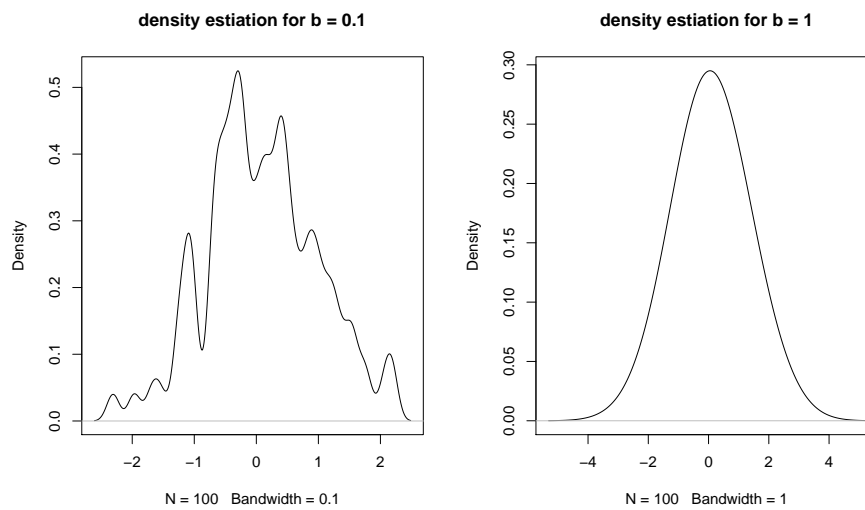


Figure 2.3.1: Bandwidth influence on the resulting curve

2.4 Boundary issue

To correct the boundary effects, we use the method of box-plot fences proposed by (Tukey, 1977) to detect the extreme values that make the estimation too rough at the extremities of the CCDF estimations' curves. 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 because its domain is \mathbb{R} . 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 $[Q1 - 3 \times (Q3 - Q1), Q3 + 3 \times (Q3 - Q1)]$ where $Q1$ and $Q3$ are the first and the third quantiles.

2.5 Research Gap

All the estimations obtained through the inversion of the conditional cumulative distribution function are not smoothed and are influenced by the boundary effects. Performing a Monte Carlo study using those estimators will increase the mean square error between two estimations.

Chapter 3

Methodology

3.1 Empirical and inverse CCDF

The estimation of the conditional cumulative distribution function (CCDF) is highly required because the estimation of the scale function (in chapter 5) is derived from its inverse. In this chapter, the problem of estimating the cdf is addressed for a couple of independent and identically distributed (i.i.d) random variables (Z, X) where X is the daily return and Z the lagged (or simply yesterday's) value of X . Researches based on this estimation are numerous mainly on the estimation of quantiles or conditional quantiles. The estimator is weakly consistent and the approach discussed here is easy to compute. In the following we present the estimation of empirical probability density and conditional distribution functions and their convergences.

3.2 Unconditional Cumulative Distribution Function

Let X_1, X_2, \dots, X_n be a sequence of i.i.d. random variables from the σ -algebra \mathcal{A} such that $X_1 < X_2 < \dots < X_n$. Assuming that $X \sim \mathcal{A}$ possesses a distribution function $F_X(x) = P(X \leq x)$. A good estimator for F is the empirical cumulative distribution

function $F_n(x)$ given by

$$F_n(x) = \frac{\#\{\text{observations less than or equal to } x\}}{n} \quad (3.2.1)$$

$$= \frac{1}{n} \sum_{t=1}^n I(X_t \leq x) \quad (3.2.2)$$

$$= \begin{cases} 0 & \text{if } x < X_1 \\ \frac{k}{n} & \text{if } X_k \leq x < X_{k+1}, \quad k = 1, \dots, n-1 \\ 1 & \text{if } x \geq X_n \end{cases} \quad (3.2.3)$$

Example

For the following sequence of observations of an arbitrary distribution,

$$X = (-1.5, -0.6, -0.4, -0.3, -0.1, 0.3, 0.6, 0.9, 1.7, 2.1),$$

the number of observations less than or equal to -0.3 is 4. This means that $F(-0.3) = 0.4$. Obviously, -0.3 is the 40% quantile of the distribution, i.e.

$$\inf \{X : F_n(X) \geq 0.4\} = -0.3$$

After doing the same computation with all the observations, we obtain the Figure 3.2.1.

The calculation of the quantile using the ecdf its inverse such we have

$$F^{-1}(F(x)) = x = F^{-1}(\tau) \quad (3.2.4)$$

where $\tau \in (0, 1)$ is the accumulated proportion of all the observations that are less or equal to a given observation x . The same idea is used in the following section which consider weights in the calculation to include the influence of the explanatory variable.

3.2.1 Kernel matrix

The estimation of F_n conditionally to the explanatory variable Z will be a weighted version of ECDF in previous section. In our approach, a kernel function $K : \mathbb{R}^d \rightarrow \mathbb{R}$ is required and should satisfy the following conditions over the kernel function.

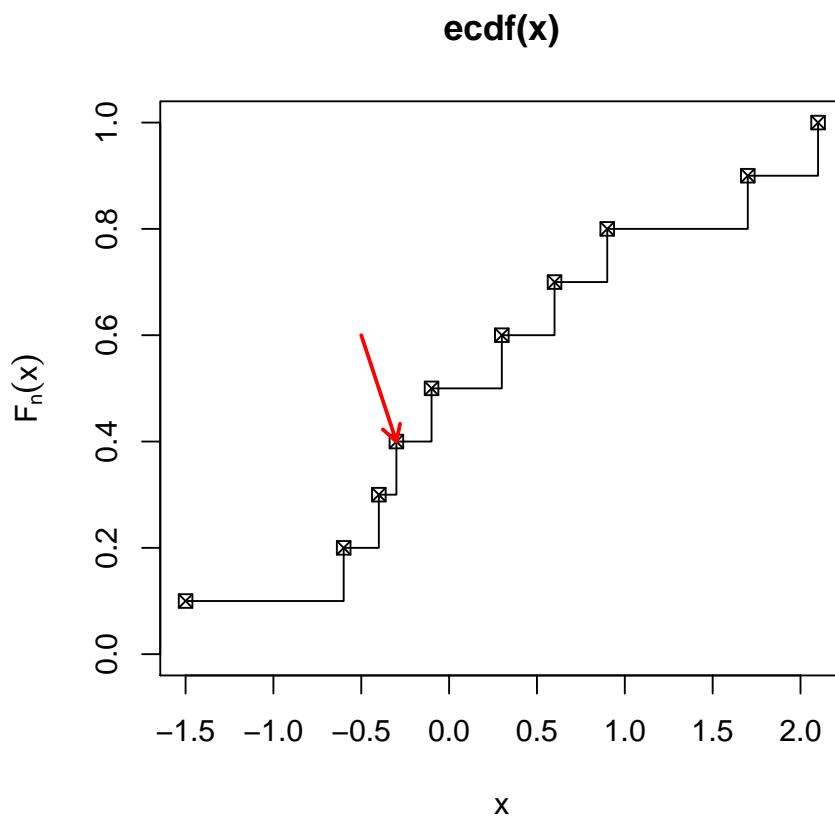


Figure 3.2.1: Empirical cumulative distribution function

Assumption 1.

(i) *Symmetrical:* $K(s) = K(-s)$ with $s \in \mathbb{R}^d$,

(ii) *Nonnegative and bounded:* For $\Gamma < \infty$, $0 < K(s) \leq \Gamma$, $s \in \mathbb{R}^d$.

(iii) *Lipschitz:* $\exists \lambda > 0, m_k < \infty$ such that $|K(s) - K(t)| \leq m_k |s - t|^\lambda$ for all $s, t \in \mathbb{R}^d$.

(iv) *a pdf:* $\int K(s)ds = 1$ with $\int_{\mathbb{R}^d} sK(s) = 0$.

We have the notations $\mu_2(K) = \int s^2 K(s)ds$ and $R(K) = \int K^2(s)ds$. Now, given the kernel function, we can estimate the kernel matrix from the following steps:

1. Divide a span of our data into N non-overlapping bins of the same size, z_1^*, \dots, z_N^* , such that $z_1^* = \min(Z) < z_2^* < \dots < z_N^* = \max(Z)$.
2. Determine the matrix

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

where $K_b(u) = b^{-1}K(ub^{-1})$, a 1-dimensional rescaled kernel and $b > 0$, the smoothing parameter or bandwidth.

3.2.2 Indicator matrix

Each column of the indicator matrix is the calculation of $I(X_t \leq x)$ for fixed x and $t = 1, 2, \dots, n$. The product of the kernel matrix K and the matrix M contains all the summations (also seen as joint probability density function at $X_t = x$ and $Z_t = z^*$).

$$\hat{f}(x, z^*) = \frac{1}{n} \sum_{t=1}^n K_b(z^* - Z_t) I(X_t \leq x) \quad (3.2.5)$$

for all fixed couple $(z^*, x) \in \mathbb{R}^2$.

$$M = \begin{pmatrix} I(x_1 \leq x_1) & I(x_1 \leq x_2) & \dots & I(x_1 \leq x_n) \\ I(x_2 \leq x_1) & I(x_2 \leq x_2) & \dots & I(x_2 \leq x_n) \\ \vdots & & & \\ I(x_n \leq x_1) & I(x_n \leq x_2) & \dots & I(x_n \leq x_n) \end{pmatrix} \quad (3.2.6)$$

$$= \begin{pmatrix} 1 & I(x_1 \leq x_2) & \dots & I(x_1 \leq x_n) \\ I(x_2 \leq x_1) & 1 & \dots & I(x_2 \leq x_n) \\ \vdots & & & \\ I(x_n \leq x_1) & I(x_n \leq x_2) & \dots & 1 \end{pmatrix} \quad (3.2.7)$$

The elements of M are 1 where the inequalities are true and 0 otherwise. Note that the unconditional empirical distribution function from the previous section is equal to

$$F_n(x) = \frac{1}{n} M^T \mathbf{1}_n, \quad \mathbf{1}_n = (1, 1, \dots, 1)^T \in \mathbb{R}^n \quad (3.2.8)$$

$$= \frac{1}{n} \begin{pmatrix} 1 & I(x_1 \leq x_2) & \dots & I(x_1 \leq x_n) \\ I(x_2 \leq x_1) & 1 & \dots & I(x_2 \leq x_n) \\ \vdots & & & \\ I(x_n \leq x_1) & I(x_n \leq x_2) & \dots & 1 \end{pmatrix}^T \times \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \quad (3.2.9)$$

$$= \frac{1}{n} \begin{pmatrix} 1 + I(x_2 \leq x_1) + \dots + I(x_n \leq x_1) \\ I(x_1 \leq x_2) + 1 + \dots + I(x_n \leq x_2) \\ \vdots \\ I(x_1 \leq x_n) + I(x_2 \leq x_n) + \dots + 1 \end{pmatrix} \quad (3.2.10)$$

The proceeding is the same for the CCDF by multiplying the kernel matrix K and the indicator matrix M and diving the product by $K\mathbf{1}_n$ (point-wise division). Note that the empirical probability density function of Z_t is given by

$$\hat{f}(z^*) = \frac{1}{n} K \mathbf{1}_n \quad (3.2.11)$$

$$= \frac{1}{n} \begin{pmatrix} K_b(z_1^* - Z_1) & K_b(z_1^* - Z_2) & \cdots & K_b(z_1^* - Z_n) \\ K_b(z_2^* - Z_1) & K_b(z_2^* - Z_2) & \cdots & K_b(z_2^* - Z_n) \\ \vdots & \vdots & \vdots & \vdots \\ K_b(z_N^* - Z_1) & K_b(z_N^* - Z_2) & \cdots & K_b(z_N^* - Z_n) \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \quad (3.2.12)$$

$$= \frac{1}{n} \begin{pmatrix} K_b(z_1^* - Z_1) + K_b(z_1^* - Z_2) + \cdots + K_b(z_1^* - Z_n) \\ K_b(z_2^* - Z_1) + K_b(z_2^* - Z_2) + \cdots + K_b(z_2^* - Z_n) \\ \vdots \\ K_b(z_N^* - Z_1) + K_b(z_N^* - Z_2) + \cdots + K_b(z_N^* - Z_n) \end{pmatrix} \quad (3.2.13)$$

a vector of N elements.

The CCDF estimation is therefore given by

$$F_n(x | z^*) = KM / (K \mathbf{1}_n) \quad (3.2.14)$$

$$= \left[n \hat{f}(z^*) \right]^{-1} KM \quad (3.2.15)$$

is a matrix of order $N \times n$

3.2.3 Empirical CCDF and consistency

In this section, we consider a pair of random variables $(Z_t, X_t) \in \mathbb{R}^{n \times n}$, $t = 1, \dots, n$, with Z and X the exogenous variable (the yesterday's returns) and the endogenous variable (today's return) respectively. The CCDF of X conditionally¹ to Z is given by

$$F_n(x | z^*) = \sum_{t=1}^n w_t I(X_t \leq x) \quad (3.2.16)$$

¹Weights from the exogenous variable Z_t

where $w_t = K_b(z^* - Z_t) / \sum_{t=1}^n K_b(z^* - Z_t)$ are the weights that verify $\sum_{t=1}^n w_t = 1$. The following theorem shows the weak asymptotic representation (Welsh, 1986) of the estimation in 3.2.16 by use of the following assumptions that are also found in (Mwita and Franke, 2013)

Assumption 2.

- (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 derived.
- (iii) The derivatives $F^{(j)}(x) = \frac{d^j F(x|z)}{dz^j}$ and $f^{(j)}(z) = \frac{d^j f(z)}{dz^j}$, for $j = 1, 2$, exist
- (iv) The conditional density $f(x|z) = \frac{dF(x|z)}{dx}$ exists and is continuous in the neighborhood of x

Theorem 3.2.1. Suppose that the assumptions 1, 2, 4, 5 hold. Then,

$$F_n(x | z^*) - F(x | z^*) \xrightarrow{D} 0, \quad \text{as } n \rightarrow \infty \quad (3.2.17)$$

Proof. The CCDF in (3.2.16) can be written in the form of an arithmetic mean of a random variable L :

$$F_n(x|z^*) = \frac{1}{n} \sum_{t=1}^n L_t \quad \text{with} \quad L_t = \frac{K_{b_z}(Z_t - z^*) I_{\{X_t \leq x\}}}{\frac{1}{n} \sum_{t=1}^n K_{b_z}(Z_t - z^*)} \quad (3.2.18)$$

and the approximation of the expectation of L is

$$\mathbb{E}[L_t] \approx \frac{\mathbb{E} [K_{b_z}(Z_t - z^*) I_{\{X_t \leq x\}}]}{\mathbb{E} [\frac{1}{n} \sum_{t=1}^n K_{b_z}(Z_t - z^*)]} = \frac{\mathbb{E}[N]}{\mathbb{E}[D]} \quad (3.2.19)$$

see Seltman (2012). Using the i.i.d assumption over the data, the numerator is

$$\begin{aligned} \mathbb{E}[N] &= \frac{1}{b_z} \mathbb{E} \left[K \left(\frac{Z_t - z^*}{b_z} \right) I_{\{X_t \leq x\}} \right] \\ &= \frac{1}{b_z} \iint_{-\infty}^x K \left(\frac{u - z^*}{b_z} \right) f(u, v) dudv \\ &= \int F(x | z^* + sh) K(s) f(z^* + sh) ds \end{aligned} \quad (3.2.20)$$

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

$$\begin{aligned} E[N] = & f(z)F(x | z^*) + b_z^2\mu_2(K) \left[f^{(1)}(z)F^{(1)}(x | z^*) + \frac{1}{2}f^{(2)}(z)F(x | z^*) + \right. \\ & \left. \frac{1}{2}f(z^*)F^{(2)}(x | z^*) + o(b_z^2) \right] \end{aligned} \quad (3.2.21)$$

and for the denominator, we have

$$E[D] = f(z^*) + \frac{1}{2}b_z^2\mu_2(K)f^{(2)}(z^*) + o(b_z^2) \quad (3.2.22)$$

Thus,

$$\begin{aligned} E[L_t] & \approx \\ & \frac{\left[F(x | z^*) + \frac{1}{2}b_z^2\mu_2(K) \left(2\frac{f^{(1)}(z)}{f(z^*)}F^{(1)}(x | z^*) + \frac{f^{(2)}(z)}{f(z)}F(x | z^*) + F^{(2)}(x | z^*) \right) \right]}{\left(1 + \frac{1}{2}b_z^2\mu_2(K)\frac{f^{(2)}(z^*)}{f(z^*)} \right)} \\ & = F(x | z^*) + \frac{1}{2}b_z^2\mu_2(K) \left(2\frac{f^{(1)}(z^*)}{f(z^*)}F^{(1)}(x | z^*) + F^{(2)}(x | z^*) \right) + o(b_z^4) \end{aligned} \quad (3.2.23)$$

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

$$\text{Bias}(F_n(x|z^*)) \approx \frac{1}{2}b_z^2\mu_2(K) \left(2\frac{f^{(1)}(z^*)}{f(z^*)}F^{(1)}(x | z^*) + F^{(2)}(x | z^*) \right) \quad (3.2.24)$$

Some authors assumed that, in this case, the first derivative of the true pdf of Z at point z can be zero (Hansen, 2004) as the one for the fixed design and therefore, the bias can be given by

$$\text{Bias}(F_n(x|z^*)) \approx \frac{1}{2}b_z^2\mu_2(K) \left(F^{(2)}(x | z^*) \right) = O(b^2/2). \quad (3.2.25)$$

We have

$$\begin{aligned}
\mathbf{V}(N) &= \mathbf{V} \left(\frac{1}{b_z} K \left(\frac{Z_t - z}{b_z} \right) I_{\{X_t^* \leq x^*\}} \right) = \frac{1}{b_z^2} \mathbf{V} \left(K \left(\frac{Z_t - z^*}{b_z} \right) I_{\{X_t \leq x\}} \right) \\
&= \frac{1}{b_z^2} \left(\mathbf{E} \left[K^2 \left(\frac{Z_t - z^*}{b_z} \right) I_{\{X_t \leq x\}} \right] - \left(\mathbf{E} \left[K \left(\frac{Z_t - z^*}{b_z} \right) I_{\{X_t \leq x\}} \right] \right)^2 \right) \\
&\approx \frac{F(x|z^*)f(z^*)R(K)}{b_z} - o(1),
\end{aligned} \tag{3.2.26}$$

$$\begin{aligned}
\mathbf{V}(D) &= \mathbf{V} \left(\frac{1}{n} \sum_{t=1}^n K_{b_z}(Z_t - z^*) \right) = \frac{1}{nb_z^2} \mathbf{V} \left(K \left(\frac{Z_t - z^*}{b_z} \right) \right) \\
&= \frac{1}{nb_z^2} \left(\mathbf{E} \left[K^2 \left(\frac{Z_t - z^*}{b_z} \right) \right] - \left(\mathbf{E} \left[K \left(\frac{Z_t - z^*}{b_z} \right) \right] \right)^2 \right) \\
&\approx \frac{f(z)R(K)}{nb_z} - o\left(\frac{1}{n}\right),
\end{aligned} \tag{3.2.27}$$

$$\begin{aligned}
\text{Cov}(N, D) &= \frac{1}{nb_z^2} \text{Cov} \left(K \left(\frac{Z_t - z^*}{b_z} \right) I_{\{X_t \leq x\}}, K \left(\frac{Z_t - z^*}{b_z} \right) \right) \\
&\approx \frac{1}{nb_z^2} \mathbf{E} \left[K^2 \left(\frac{Z_t - z^*}{b_z} \right) I_{\{X_t \leq x\}} \right] - o\left(\frac{1}{n}\right) \\
&\approx \frac{1}{nb_z} F(x|z^*)f(z^*)R(K)
\end{aligned} \tag{3.2.28}$$

Using the same approximation in (4.1.45), the variance of $F_n(x|z^*)$ is

$$\mathbf{V}(L_t) \approx F(x|z^*) \left[\frac{R(K)(1 - F(x|z^*))}{b_z f(z^*)} \right] \tag{3.2.29}$$

and by the Central Limit Theorem, using assumption 4 for $\{(X_t^*, Z_t), t = 1, 2, \dots\}$

$$\sqrt{n} \left(F_n(x|z^*) - F(x|z^*) - \text{Bias}(F(x|z^*)) \right) \xrightarrow{D} \mathcal{N}(0, \mathbf{V}(L_t)) \tag{3.2.30}$$

Notice that the expectation of $F_n(x|z^*)$ is the same as the one of L and the variance is $\mathbf{V}(L_t)/n$. □

3.2.4 Inversion

There is no systematic expression of the inverse CCDF, F_n^{-1} , unless we know the true conditional distribution of $X_t = x$ given $Z_t = z^*$. Given the level τ , a technique allowing the calculation of the τ^{th} conditional quantile of $X_t = x$ given $Z_t = z^*$ follows the following steps:

1. For each row of $F_n(x | z^*)$, find the smallest x such that $F_n(x | z^*) \geq \tau$
2. The ICCDF or τ^{th} conditional quantile is

$$Q_\tau(X_t | Z_t) = \inf \{x \in \mathbb{R} : F_n(x | z^*) \geq \tau\} \quad (3.2.31)$$

This gives a vector of N elements. The consistency of the ICCDF is proven in chapter 4. We have motivated the estimation of both the CDF unconditionally and conditionally using matrices. We also showed theoretically that the CCDF is consistent. These estimations and their features will help on the estimation of the quantile autoregressive function and conditional scale function.

3.3 Model specification and simulation

In this section, we introduce the process the estimations will be obtained from. Given that the process AR(1)-ARCH(1) is a combination of two process, AR(1) and ARCH(1), is stationary as summation of two stationary processes. Stationary processes verify the conditions in the following definition.

3.3.1 Definitions

Definition 3.3.1. *A process is said to be weakly stationary, if its first and second moments exist and are time invariant. Meaning that*

$$E[X_t] = E[X_{t-1}] = \lambda < \infty, \quad \forall t \quad (3.3.1)$$

$$V(X_t) = \rho_0 < \infty, \quad \forall t \text{ and} \quad (3.3.2)$$

$$\text{Cov}(X_t, X_{t-k}) = \rho_k, \quad \forall t, \forall k. \quad (3.3.3)$$

The third property only depends on the difference $t - (t - k)$.

In this chapter, we discuss the properties of the model AR(1)-ARCH(1) that will be simulated for the application of our findings.

3.3.2 AR(1) process

Recall that the process of application or to be simulated is a combination of two processes. The first is the AR(1) represented by

$$X_t = \mu + \delta X_{t-1} + e_t \quad (3.3.4)$$

where $\mu \in \mathbb{R}$ is a constant, $|\delta| < 1$ is the parameter of the model and e_t is white noise with mean 0, constant variance σ_e^2 and is independent of the lagged value X_{t-1} . This model represents some outputs, in financial time series for instance, that depend on their own previous values and an innovation term (stochastic term). Using the definition 3.3.1, we specify the parameter that yield the stationarity of the AR(1) process.

$$\begin{aligned} E[X_t] &= \mu + \delta E[X_{t-1}] + 0 \\ \lambda &= \mu + \delta \lambda \\ &= \frac{\mu}{1 - \delta} \end{aligned} \quad (3.3.5)$$

and

$$\begin{aligned}
V(X_t) &= 0 + V(\delta X_{t-1} + e_t) \\
\rho_0 &= \delta^2 V(X_{t-1}) + V(e_t) + 2 \underbrace{\text{Cov}(X_{t-1}, e_t)}_{=0} \\
\rho_0 &= \delta^2 \rho_0 + \sigma_e^2 \\
\rho_0 &= \frac{\sigma_e^2}{1 - \delta^2}
\end{aligned} \tag{3.3.6}$$

We calculate the covariance, for $k = 1$, as

$$\begin{aligned}
\text{Cov}(X_t, X_{t-1}) &= E[X_t X_{t-1}] - E[X_t] E[X_{t-1}] \\
\rho_1 &= E[\mu X_{t-1} + \delta X_{t-1}^2 + e_t X_{t-1}] - \frac{\mu^2}{(1 - \delta)^2} \\
&= \frac{\mu^2}{1 - \delta} + \delta E[X_t^2] - \frac{\mu^2}{(1 - \delta)^2} \\
&= \frac{-\mu^2 \delta}{(1 - \delta)^2} + \delta \left(V(X_t) + (E[X_t])^2 \right) \\
&= \frac{-\mu^2 \delta}{(1 - \delta)^2} + \delta \left(\frac{\sigma_e^2}{1 - \delta^2} + \frac{\mu^2}{(1 - \delta)^2} \right) \\
&= \delta \frac{\sigma_e^2}{1 - \delta^2}
\end{aligned} \tag{3.3.7}$$

Now, for $k = 2$ and using the properties of the Covariance, we have

$$\begin{aligned}
\text{Cov}(X_t, X_{t-2}) &= \text{Cov}(\mu + \delta X_{t-1} + e_t, X_{t-2}) \\
\rho_2 &= \text{Cov}(\mu, X_{t-2}) + \delta \text{Cov}(X_{t-1}, X_{t-2}) + \text{Cov}(e_t, X_{t-2}) \\
&= 0 + \delta \rho_1 + 0 \\
&= \delta^2 \frac{\sigma_e^2}{1 - \delta^2}
\end{aligned} \tag{3.3.8}$$

We conclude that

$$\text{Cov}(X_t, X_{t-k}) = \rho_k = \delta^k \frac{\sigma_e^2}{1 - \delta^2} \tag{3.3.9}$$

3.3.3 ARCH(1) process

As the AR(1) models the outputs from the previous ones, the ARCH(1) is the modelization of the actual innovation as function of the previous ones too. ARCH-based process are being utilized in most of the current time series analysis in finance, economics, etc because they model the volatility. An ARCH(1) is depicted by

$$\begin{aligned}\varepsilon_t &= \varpi e_t, \\ \varpi &= (\omega + \alpha \varepsilon_{t-1}^2)^{\frac{1}{2}}, \quad t = 1, 2, \dots\end{aligned}\tag{3.3.10}$$

with the conditions $\omega > 0$, $\alpha < 1$ and e_t i.i.d with zero mean and variance 1 and independent to ε_{t-1} . These conditions allow the data generation process to be stationary.

To show it, we calculate the following statistics:

$$\begin{aligned}\mathbb{E}[\varepsilon_t] &= \mathbb{E} \left[(\omega + \alpha \varepsilon_{t-1}^2)^{\frac{1}{2}} e_t \right] \\ &= \mathbb{E} \left[(\omega + \alpha \varepsilon_{t-1}^2)^{\frac{1}{2}} \right] \times \underbrace{\mathbb{E}[e_t]}_{=0} \\ &= 0.\end{aligned}\tag{3.3.11}$$

Let's also introduce the conditional statistics that will enable the calculation the variance of the process.

Conditional expectation

The conditional expectation of the ARCH(1) process is

$$\begin{aligned}\mathbb{E}[\varepsilon_t | \varepsilon_{t-1}] &= \mathbb{E} \left[(\omega + \alpha \varepsilon_{t-1}^2)^{\frac{1}{2}} e_t | \varepsilon_{t-1} \right] \\ &= (\omega + \alpha \varepsilon_{t-1}^2)^{\frac{1}{2}} \mathbb{E}[e_t | \varepsilon_{t-1}] \\ &= (\omega + \alpha \varepsilon_{t-1}^2)^{\frac{1}{2}} \mathbb{E}[e_t] \\ &= 0.\end{aligned}\tag{3.3.12}$$

Conditional variance

$$\begin{aligned} \mathbf{V} [\varepsilon_t | \varepsilon_{t-1}] &= \mathbf{V} \left[(\omega + \alpha \varepsilon_{t-1}^2)^{\frac{1}{2}} e_t | \varepsilon_{t-1} \right] \\ &= \mathbf{E} \left[(\omega + \alpha \varepsilon_{t-1}^2) e_t^2 | \varepsilon_{t-1} \right] \\ &= (\omega + \alpha \varepsilon_{t-1}^2) \mathbf{E}[e_t^2] \\ &= \omega + \alpha \varepsilon_{t-1}^2. \end{aligned} \tag{3.3.13}$$

The variance of the process is therefore given by the law of total variance

$$\begin{aligned} \mathbf{V}(\varepsilon_t) &= \mathbf{E} [\mathbf{V}(\varepsilon_t | \varepsilon_{t-1})] + \mathbf{V} (\mathbf{E}[\varepsilon_t | \varepsilon_{t-1}]) \\ &= \mathbf{E} [\omega + \alpha \varepsilon_{t-1}^2] \\ &= \omega + \alpha \mathbf{E}[\varepsilon_{t-1}^2] \\ &= \omega + \alpha \left(\mathbf{V}(\varepsilon_t) + (\mathbf{E}[\varepsilon_t])^2 \right) \\ &= \omega + \alpha \mathbf{V}(\varepsilon_t) \\ \mathbf{V}(\varepsilon_t) &= \frac{\omega}{1 - \alpha}. \end{aligned} \tag{3.3.14}$$

For this process, the covariance

$$\text{Cov}(\varepsilon_t, \varepsilon_{t-k}) = 0 \quad \forall k > 0. \tag{3.3.15}$$

3.3.4 GARCH(1,1) process

This process depends on both the previous innovation and the previous conditional variance. It's defined as

$$\begin{aligned} \varepsilon_t &= \varpi e_t, \\ \varpi &= (\omega + \alpha \varepsilon_{t-1}^2 + \beta \varpi_{t-1}^2)^{\frac{1}{2}}, \\ e_t &\sim \mathcal{N}(0, 1), \text{ independent of } \varepsilon_{t-1} \text{ and } \varpi_{t-1}, \quad t = 1, 2, \dots \end{aligned} \tag{3.3.16}$$

Using the definition 3.3.1, we can show the specifications of the GARCH(1,1). We calculate, as in the previous section, the statistics

$$\begin{aligned}
\mathbf{E}[\varepsilon_t] &= \mathbf{E} \left[(\omega + \alpha\varepsilon_{t-1}^2 + \beta\varpi_{t-1}^2)^{\frac{1}{2}} e_t \right] \\
&= \mathbf{E} \left[(\omega + \alpha\varepsilon_{t-1}^2 + \beta\varpi_{t-1}^2)^{\frac{1}{2}} \right] \mathbf{E}[e_t] \\
&= 0.
\end{aligned} \tag{3.3.17}$$

The conditional expectation of the GARCH(1,1) process is given by

$$\begin{aligned}
\mathbf{E}[\varepsilon_t \mid \varepsilon_{t-1}] &= \mathbf{E} \left[(\omega + \alpha\varepsilon_{t-1}^2 + \beta\varpi_{t-1}^2)^{\frac{1}{2}} e_t \mid \varepsilon_{t-1} \right] \\
&= \mathbf{E} \left[(\omega + \alpha\varepsilon_{t-1}^2 + \beta\varpi_{t-1}^2)^{\frac{1}{2}} \right] \mathbf{E}[e_t \mid \varepsilon_{t-1}] \\
&= 0,
\end{aligned} \tag{3.3.18}$$

and the conditional variance

$$\begin{aligned}
\mathbf{V}(\varepsilon_t \mid \varepsilon_{t-1}) &= \mathbf{E}[\varepsilon_t^2 \mid \varepsilon_{t-1}] \\
&= \mathbf{E} \left[(\omega + \alpha\varepsilon_{t-1}^2 + \beta\varpi_{t-1}^2) e_t^2 \mid \varepsilon_{t-1} \right] \\
&= \mathbf{E} \left[(\omega + \alpha\varepsilon_{t-1}^2 + \beta\varpi_{t-1}^2) \mid \varepsilon_{t-1} \right] \mathbf{E}[e_t^2 \mid \varepsilon_{t-1}] \\
&= \omega + \alpha\varepsilon_{t-1}^2 + \beta\varpi_{t-1}^2.
\end{aligned} \tag{3.3.19}$$

The law of total variance yields

$$\begin{aligned}
\mathbf{V}(\varepsilon_t) &= \mathbf{E}[\varpi^2] + \mathbf{V}(0) \\
&= \mathbf{E}[\omega + \alpha\varepsilon_{t-1}^2 + \beta\varpi_{t-1}^2] \\
&= \omega + \alpha \mathbf{E}[\varepsilon_{t-1}^2] + \beta \mathbf{E}[\varpi_{t-1}^2] \\
&= \omega + \alpha \mathbf{V}(\varepsilon_t) + \beta \mathbf{V}(\varepsilon_t) \\
\mathbf{V}(\varepsilon_t) &= \frac{\omega}{1 - \alpha - \beta}.
\end{aligned} \tag{3.3.20}$$

This variance is positive and finite for $\omega > 0$ and $\alpha + \beta < 1$.

3.3.5 AR(1)-GARCH(1,1)

A financial time series can be of this form which is function of the previous return and the previous volatility or innovation. It's represented by

$$\begin{aligned}
 X_t &= \alpha_t + u_t \\
 \alpha_t &= \mu + \delta X_{t-1} \\
 u_t &= \varpi e_t \\
 \varpi_t &= (\omega + \alpha u_{t-1}^2 + \beta \varpi_{t-1}^2)^{\frac{1}{2}} \\
 e_t &\sim \mathcal{N}(0, 1), \text{ independent of } X_{t-1}.
 \end{aligned} \tag{3.3.21}$$

Here, we also calculate the statistics using the definition 3.3.1 in order to show the conditions over the coefficients that ascertain the stationarity of the process. The first moment is given by

$$\begin{aligned}
 E[X_t] &= E \left[\mu + \delta X_{t-1} + (\omega + \alpha u_{t-1}^2 + \beta \varpi_{t-1}^2)^{\frac{1}{2}} e_t \right] \\
 &= \mu + \delta E[X_{t-1}] + E \left[(\omega + \alpha u_{t-1}^2 + \beta \varpi_{t-1}^2)^{\frac{1}{2}} \right] E[e_t] \\
 &= \mu + \delta E[X_t] \\
 E[X_t] &= \frac{\mu}{1 - \delta}.
 \end{aligned} \tag{3.3.22}$$

Conditional expectation

$$\begin{aligned}
 E[X_t | X_{t-1}] &= \mu + \delta X_{t-1} + E \left[(\omega + \alpha u_{t-1}^2 + \beta \varpi_{t-1}^2)^{\frac{1}{2}} e_t | X_{t-1} \right] \\
 &= \mu + \delta X_{t-1}.
 \end{aligned} \tag{3.3.23}$$

Conditional variance

$$\begin{aligned}
 V(X_t | X_{t-1}) &= E[X_t^2 | X_{t-1}] - (\mu + \delta X_{t-1})^2 \\
 &= E \left[(\omega + (\alpha e_{t-1}^2 + \beta) \varpi_{t-1}^2) | X_{t-1} \right] \times E[e_t^2 | X_{t-1}] \\
 &= \omega + (\alpha + \beta) E[\varpi_{t-1}^2 | X_{t-1}]
 \end{aligned} \tag{3.3.24}$$

Law of total variance

$$\begin{aligned} V(X_t) &= E [V(X_t | X_{t-1})] + V (E[X_t | X_{t-1}]) \\ &= E \left[\omega + (\alpha + \beta) E [\varpi_{t-1}^2 | X_{t-1}] \right] + V (\mu + \delta X_{t-1}) \\ &= \omega + (\alpha + \beta) E [\varpi_{t-1}^2] + \delta^2 V(X_t) \\ (1 - \delta^2) V[X_t] &= \omega + (\alpha + \beta) E [\varpi_{t-1}^2] \end{aligned} \tag{3.3.25}$$

We have

$$E [\varpi_t^2] = \omega + (\alpha + \beta) E [\varpi_{t-1}^2] \tag{3.3.26}$$

and for stationary, we'll assume the moments to be time-independent. That is,

$$E [\varpi_t^2] = \frac{\omega}{1 - \alpha - \beta} \tag{3.3.27}$$

Finally,

$$V[X] = \frac{\omega}{(1 - \delta^2)(1 - \alpha - \beta)} \tag{3.3.28}$$

which is positive and finite for $\omega > 0, |\delta| < 1$ and $\alpha + \beta < 1$.

3.3.6 Simulation of AR(1)-ARCH(1) processes

All our estimations will take into account a data generated from an AR(1)-ARCH(1), a process as in the section 3.3.5 where the GARCH term $\beta = 0$. In order to graphically show how the curves behave in view of the variation of the coefficients satisfying the conditions and which do not (See Figure 3.3.1, 3.3.2, 3.3.3 and 3.3.4). The Figure 3.3.3 and Figure 3.3.4 show non-stationary process because the parameter do not satisfy the conditions discussed in the previous section.

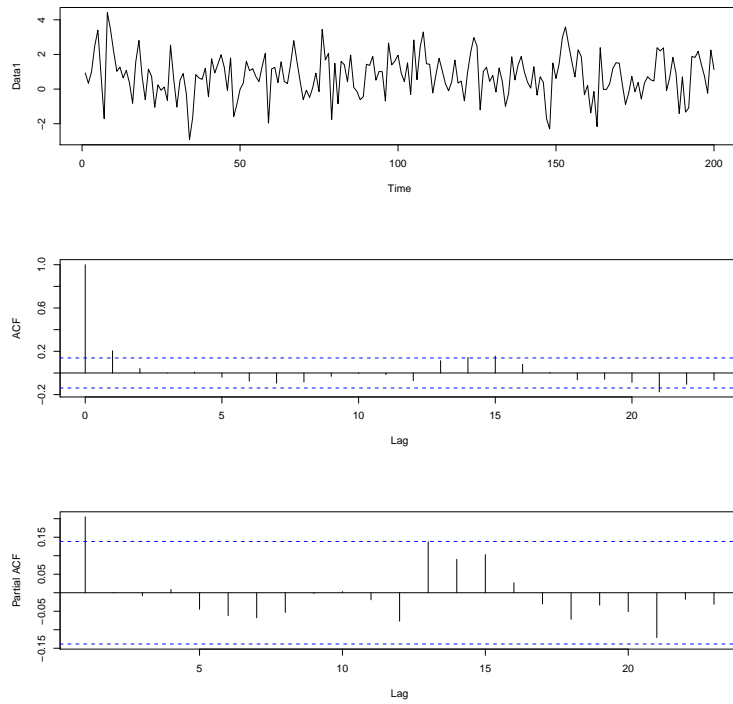


Figure 3.3.1: AR(1)-ARCH(1) process for $\mu = 0.5, \delta = 0.25, \omega = 1, \alpha = 0.35$

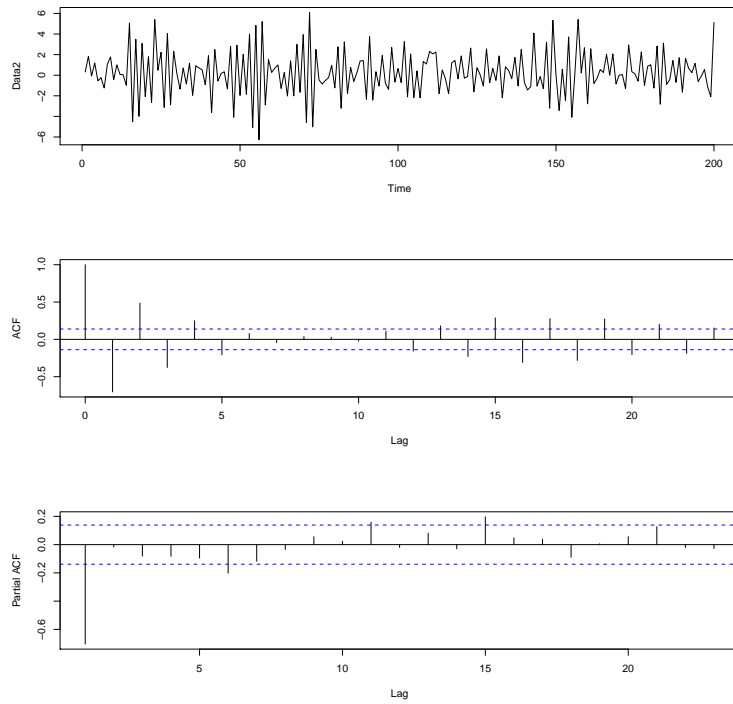


Figure 3.3.2: AR(1)-ARCH(1) process for $\mu = 0.5$, $\delta = -0.75$, $\omega = 1$, $\alpha = 0.5$

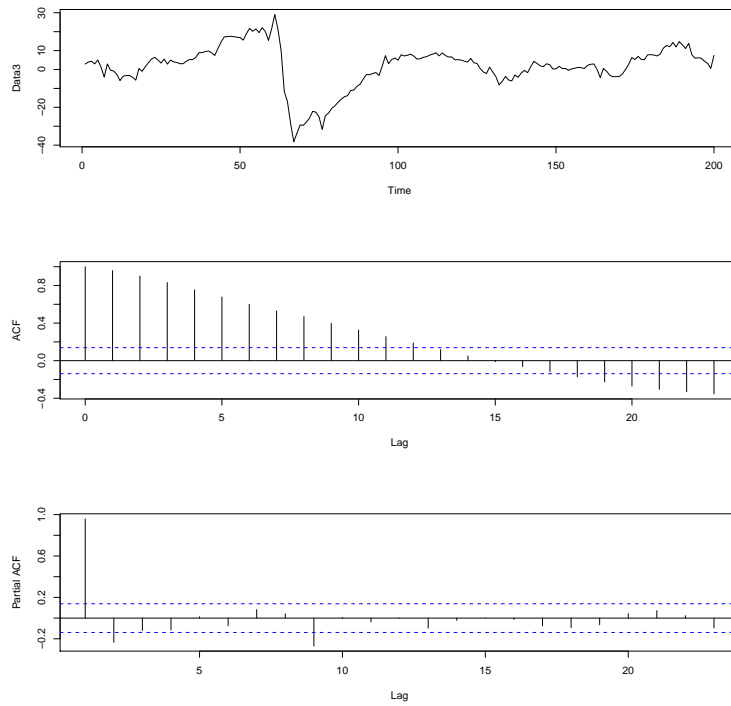


Figure 3.3.3: AR(1)-ARCH(1) process for $\mu = 0.5, \delta = 0.95, \omega = 1, \alpha = 1.2$

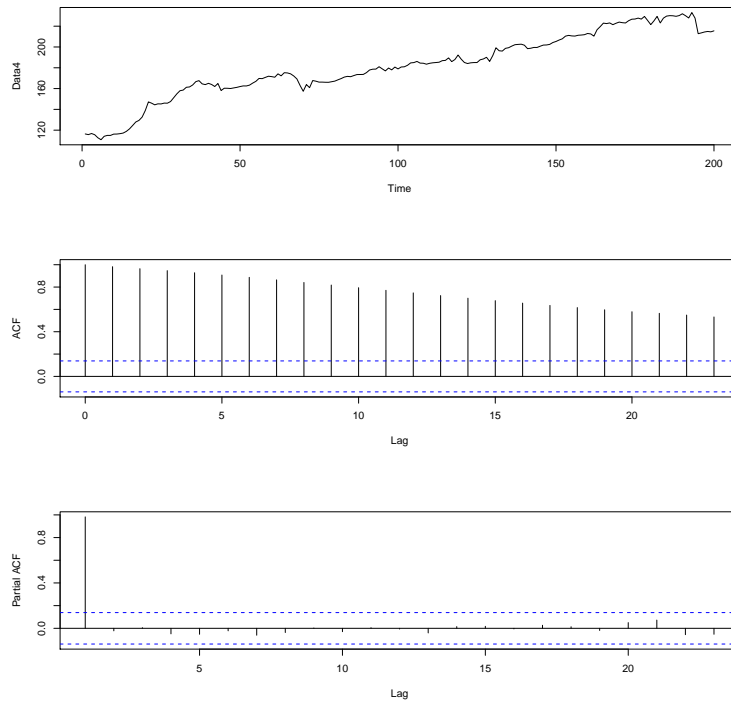


Figure 3.3.4: AR(1)-ARCH(1) process for $\mu = 0.5, \delta = 1, \omega = 1, \alpha = 1$

The calculation of Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) help to determine the order for AR and Moving Average (MA) processes that are known to be stationary. For a non-stationary as shown by figure 3.3.3 and 3.3.4, it is not possible to get the order of the AR and MA.

3.4 Conclusion

Having a clear information of the parameters that will come into play, we can simulate a stationary (Figure 3.3.1) AR(1)-ARCH(1) process in order to apply our estimations. In order to have accurate estimations, we'll perform Monte Carlo studies bases on the stationary processes. This means the data generation process should be from the choice of adequate coefficients.

Chapter 4

Estimation of the conditional scale function

In this chapter, the estimation of the Smoothed Conditional Scale Function for time series is carried out under the conditional heteroscedastic innovations by imitating the kernel smoothing in nonparametric QAR-QARCH scheme. The estimation was carried out based on the quantile regression methodology proposed by Koenker and Bassett. And the proof of the asymptotic properties of the Conditional Scale Function estimator for this type of process was given and its consistency was shown by Franke et al. (2015).

4.1 Definitions

Let $f_{Z_t}(z)$ and $f(x, z)$, denote the probability density function (pdf) of X_t and the joint pdf of (X_t, Z_t) . If $f_{Z_t}(z) > 0$ (assumption 2.(i)), the dependence between the exogenous X_t and the endogenous variables is described by the following conditional probability density function (CPDF)

$$f(x|z) = \frac{f(x, z)}{f(z)} \quad (4.1.1)$$

and the conditional cumulative distribution function (CCDF)

$$F(x|z) = \int_{-\infty}^x f(s|z)ds = P(X \leq x|Z_t = z) \quad (4.1.2)$$

$$= E [I_{\{X_t \leq x\}} | Z_t = z] \quad (4.1.3)$$

The estimation of the conditional scale function is derived through the CCDF. However, the following assumptions and definitions (these assumptions are commonly used for kernel density estimation (KDE), bias reduction (Mynbaev and Martins-Filho, 2010), asymptotic properties, and normality proof) are necessary (see Table 4.3.1). We add the following assumptions to Assumption 2 in Chapter 3.1.

Assumption 3.

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

(ii) $f(\varpi_\tau(z) | z) > 0$

Assumption 4. The process $\{(X_t, Z_t), t = 1, 2, \dots\}$ is strong mixing with $\alpha(s) = o(s^{-2-\delta})$, $\delta > 0$, see (Bosq, 2012, Theorem 1.7).

Assumption 5. The smoothing parameter $b > 0$ of the smoothing parameters is such that $b \rightarrow 0$, $nb^p \rightarrow \infty$ as $n \rightarrow \infty$.

Definition 4.1.1 (strong mixing). Let $X_t = \{\dots, X_{t-1}, X_t, X_{t+1}, \dots\}$ 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

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

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

Assuming that the Autoregressive part in the model (1.2.4) is equal to zero, i.e., $\alpha_{t,\tau} = \alpha_\tau(z) = 0$ for any $\tau \in (0, 1)$, we consider the model

$$X_t = \varpi_\tau(Z_t)\varepsilon_t, \quad t = 1, 2, \dots \quad (4.1.4)$$

Define the check-function as

$$\gamma_\tau(X, \mu) = \gamma_\tau(X - \mu) = (\tau - I_{\{X - \mu \leq 0\}})(X - \mu) \quad (4.1.5)$$

Here, $I(\cdot)$ 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 $F_X(x) = P(X \leq x) = E I_{\{X \leq x\}}$, and a real value $\mu \in \mathbb{R}$, is the asymmetric absolute value function whose amount of asymmetry depends on τ (Koenker and Bassett, 1978). 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 is assumed to be 0 in model (1.2.6), we have a purely heteroscedastic ARCH model introduced in (Engle, 1982) 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 (4.1.5) is Lipschitz continuous by the following theorem.

Theorem 4.1.1. *Let γ_τ be defined as in (4.1.5) and $(x, \sigma) \in \mathbb{R}^2$. Then, γ_τ satisfies the Lipschitz continuity condition:*

$$|\gamma_\tau(x, \sigma) - \gamma_\tau(x, \sigma')| \leq M|\sigma - \sigma'|$$

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

Proof of Theorem 4.1.2. See the proof of Lemma 3.1 in (Mwita, 2003, p .74-75) \square

By the next theorem we show clearly why the errors $\{\varepsilon_t\}$ in model (1.2.2) are assumed to be zero τ -quantile and unit scale

Theorem 4.1.2. *Consider the model (1.2.6) and the check function in (4.1.5), then for $\varpi_\tau(Z_t) \in \mathbb{R}_+^*$,*

$$\varepsilon_t = \frac{X_t - \alpha_\tau(Z_t)}{\varpi_\tau(Z_t)} \quad (4.1.6)$$

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

$$\mathbf{P} (X_t \leq \alpha_\tau(Z_t)) | Z_t) = \tau \quad \text{and} \quad (4.1.7)$$

$$\mathbf{P} \left(\gamma_\tau (X_t, \alpha_\tau(Z_t)) \leq \varpi_\tau(Z_t) | Z_t) = \tau \quad (4.1.8)$$

Proof of Theorem 4.1.2. The τ^{th} -quantile operator is

$$Q_\tau(Y_t) = \inf \{ \mu \in \mathbb{R} : \mathbf{P}(Y_t \leq \mu | Z_t) \geq \tau \} \quad (4.1.9)$$

with well-defined properties in (Mwita, 2003, p .9-10). From the model (1.2.6), the conditional τ -quantile of X_t is

$$q_\tau(Z_t) = Q_\tau(X_t) = \alpha(Z_t) + \varpi(Z_t)q_\tau^e \quad (4.1.10)$$

Where q_τ^e is the τ -quantiles of e_t . Then, using model (1.2.6) and the equation (4.1.10), we get

$$X_t - q_\tau(Z_t) = \varpi(Z_t) (e_t - q_\tau^e) \quad (4.1.11)$$

and

$$\gamma_\tau (X_t, q_\tau(Z_t)) = \varpi(Z_t)\gamma_\tau (e_t, q_\tau^e). \quad (4.1.12)$$

and the τ^{th} -quantile of (4.1.12) is

$$Q_\tau \left(\gamma_\tau (X_t, q_\tau(Z_t)) \right) = \varpi(Z_t)Q_\tau (\gamma_\tau (e_t, q_\tau^e)) = \varpi(Z_t)Q_\tau^e \quad (4.1.13)$$

where Q_τ^e is the τ -quantile of $\gamma_\tau (e_t, q_\tau^e)$. Note that from (4.1.11),

$$Q_\tau (X_t - q_\tau(Z_t)) = 0. \quad (4.1.14)$$

The quotient

$$\frac{X_t - \alpha_\tau(Z_t)}{Q_\tau \left(\gamma_\tau (X_t, \alpha_\tau(Z_t)) \right)} = \frac{e_t - q_\tau^e}{Q_\tau^e} \quad (4.1.15)$$

is zero τ -quantile and unit scale and can be seen as model (1.2.2) if $\varepsilon_t = (e_t - q_\tau^e)/Q_\tau^e$, $\alpha_\tau(Z_t) = q_\tau(Z_t)$ and $\varpi_\tau(Z_t) = Q_\tau \left(\gamma_\tau (X_t, \alpha_\tau(Z_t)) \right)$.

Now, assuming that ε_t (independent of Z_t) in model (1.2.2) is zero τ -quantile, this is equivalent to write

$$\begin{aligned} \Pr(\varepsilon_t \leq 0) &= \Pr(\varepsilon_t \leq 0 | Z_t) = \tau \\ \Rightarrow \Pr\left(\frac{X_t - \alpha_\tau(Z_t)}{\varpi_\tau(Z_t)} \leq 0 \mid Z_t\right) &= \tau \end{aligned}$$

This prove (4.1.7) for $\varpi_\tau(z) > 0$. Also, ε_t is unit-scale, means

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

□

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

$$\varphi(z, \varpi) = \mathbb{E} [\gamma_\tau(\gamma_\tau(X_t), \varpi) \mid Z_t = z] \quad (4.1.16)$$

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

$$\varpi_\tau(z) = \arg \min_{\varpi \in \mathbb{R}_+} \varphi(z, \varpi) \quad (4.1.17)$$

and

$$\varpi_\tau(z) = \arg \min_{\varpi \in \mathbb{R}_+} \varphi(z, \varpi) \quad (4.1.18)$$

and

$$\varpi_\tau(z) = \inf \{ \mu \in \mathbb{R}_+^* : F(\mu | z) \geq \tau \} = F^{-1}(\tau | z) \quad (4.1.19)$$

The kernel estimator of (4.1.18) at $Z_t = z$ is given by

$$\hat{\varpi}_\tau(z) = \arg \min_{\varpi \in \mathbb{R}_+} \hat{\varphi}_n(z, \varpi) \quad (4.1.20)$$

We can express the estimate of $\varphi(z, \varpi)$ in the random design as it was developed in (Härdle et al., 2004). Let $Y_t^* = \gamma_\tau(\gamma_\tau(X_t), \varpi)$ be a non-negative function of X_t and $Y^* = (Y_1^*, Y_2^*, \dots, Y_n^*)$ a random vector in $\mathbb{R}_+^* = (0, \infty)$, $t = 1, 2, \dots, n$. In the random design, the conditional expectation (4.1.16) can be rewritten as follow

$$\varphi(z, \varpi) = \mathbb{E} [Y^* | Z_t = z] = \int y^* f(y^* | z) dy^* = \int y^* \frac{f(y^*, z)}{f(z)} dy^* \quad (4.1.21)$$

Where $f(y^* | z)$ 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 (Nadaraya, 1964) and (Watson, 1964) with $K_b(u) = b^{-1}K(ub^{-1})$, a 1-dimensional rescaled kernel with bandwidth $b > 0$, we have the following estimates of $f(y^*, z)$ and $f(z)$ (Silverman, 1986).

$$\begin{aligned} \hat{f}(y^*, z) &= \frac{1}{n} \sum_{t=1}^n K_{b_z}(Z_t - z) K_{b_{y^*}}(y^* - Y_t^*) \\ \hat{f}(z) &= \frac{1}{n} \sum_{t=1}^n K_{b_z}(Z_t - z) \end{aligned} \quad (4.1.22)$$

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

$$\begin{aligned} \hat{\varphi}_n(z, \varpi) &= \int \frac{y^* \sum_{t=1}^n K_{b_z}(Z_t - z) K_{b_{y^*}}(y^* - Y_t^*)}{\sum_{t=1}^n K_{b_z}(Z_t - z)} dy^* \\ &= \frac{\sum_{t=1}^n K_{b_z}(Z_t - z) \int y^* K_{b_{y^*}}(y^* - Y_t^*) dy^*}{\sum_{t=1}^n K_{b_z}(Z_t - z)} \\ &= \frac{\sum_{t=1}^n K_{b_z}(Z_t - z) \int [(y^* - Y_t^*) + Y_t^*] K_{b_{y^*}}(y^* - Y_t^*) dy^*}{\sum_{t=1}^n K_{b_z}(Z_t - z)} \end{aligned} \quad (4.1.23)$$

and considering the regularity conditions of K_b in Assumption 1 and also the fact that $d(y^* - Y_t^*) = dy^*$, $Y_t^* \in \mathbb{R}_+$, we have

$$\hat{\varphi}_n(z, \varpi) = \frac{\sum_{t=1}^n K_{b_z}(Z_t - z) Y_t^*}{\sum_{t=1}^n K_{b_z}(Z_t - z)} = n^{-1} \sum_{t=1}^n K_{b_z}(Z_t - z) Y_t^* / \hat{f}(z) \quad (4.1.24)$$

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

$$Y_t^* = \left[X_t (\tau - I_{\{X_t \leq 0\}}) - \varpi \right] \left(\tau - I_{\{X_t (\tau - I_{\{X_t \leq 0\}}) \leq \varpi\}} \right) \quad (4.1.25)$$

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

$$\frac{d\hat{\varphi}_n(z, \varpi)}{d\varpi} = \left(n\hat{f}(z)\right)^{-1} \sum_{t=1}^n K_{b_z}(Z_t - z) \left(I_{\{X_t(\tau - I_{\{X_t \leq 0\}}) \leq \varpi\}} - \tau \right) \quad (4.1.26)$$

The minimizer of (4.1.24) is obtained from $\frac{d\hat{\varphi}_n(z, \varpi)}{d\varpi} = 0$. This leads to the following equation

$$\left(n\hat{f}(z)\right)^{-1} \sum_{t=1}^n K_{b_z}(Z_t - z) \left(I_{\{X_t^* \leq \varpi\}} \right) = \tau \quad (4.1.27)$$

where

$$X_t^* = X_t(\tau - I_{\{X_t \leq 0\}}) \in \mathbb{R}_+, \quad (4.1.28)$$

for all $X_t \in \mathbb{R}$, $t = 1, 2, \dots$. Note that $Y_t^* = I_{\{X_t^* \leq \varpi\}}$ in (4.1.21). The left part of the equation (4.1.27) is a (unsmoothed) conditional cumulative distribution function (CCDF),

$$\hat{F}(x^* | z) = \left(n\hat{f}(z)\right)^{-1} \sum_{t=1}^n K_{b_z}(Z_t - z) \left(I_{\{X_t^* \leq x^*\}} \right), \quad (4.1.29)$$

that needs to be estimated and our estimator is therefore

$$\hat{\omega}_\tau(z) = \inf \left\{ x^* \in \mathbb{R}_+ : \hat{F}(x^* | z) \geq \tau \right\} \equiv \hat{F}^{-1}(\tau | z) \quad (4.1.30)$$

which is equivalent to $\hat{F}(\hat{\omega}(z) | z) = \tau$.

An algorithm to estimating $\hat{F}(x^* | z)$ is proposed in the following section. This estimator suffers from the problem of boundary effects as we can see it on figure 4.3.2 due to outliers. We obtain unsmoothed curves of the CCDF because the smoothness is only in the Z direction. A method is proposed by (Hansen, 2004) to smooth in the y . The form of Smoothed Conditional Distribution Estimator is

$$\tilde{F}(x^* | z) = \left(n\hat{f}(z)\right)^{-1} \sum_{t=1}^n K_h(z - Z_t) G\left(\frac{x^* - X_t^*}{h_0}\right) \quad (4.1.31)$$

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 (Das and Politis, 2017), (Hall et al., 1999), (Steikert, 2014, p. 3–18) among others. The WNW estimator's expression is

$$\tilde{F}_{WNW}(x^* | z) = \frac{\sum_{t=1}^n p_t(z, \lambda) K_{b_z}(Z_t - z) I_{\{X_t^* \leq x^*\}}}{\sum_{t=1}^n p_t(z, \lambda) K_{b_z}(Z_t - z)} \quad (4.1.32)$$

with conditions $\sum_{t=1}^n p_t(z, \lambda) = 1$ and λ is determined using the Newton-Raphson iteration. Smoothing the CDF does not smooth the estimator in (4.1.30).

4.1.1 Nadaraya-Watson smoothing method

We can make $\hat{\varpi}_\tau(z)$ smooth by using NW regression¹. This will provide a smoothed curve at each level $\tau \in (0, 1)$. We write the regression equation as

$$Y_t = \varpi_{\tau,s}(Z_t) + \eta_t \quad (4.1.33)$$

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

$$\varpi_{\tau,s}(z) = E[Y | Z = z] = \int y \frac{f(y, z)}{f(z)} dy \quad (4.1.34)$$

where $f(y, z)$ and $f(z)$ are estimated as in (4.1.22).

We can perform some transformations on (4.1.34) in order to show that it's actually better than the unsmoothed one. By assumption 2 (i) and the fact that $F(\varpi_\tau(z) | z) = \tau$, we have

$$\begin{aligned} F(\varpi_{\tau,s}(Z_t) | z) &= F(E[\varpi_\tau(z) | Z_t = z] | z) \\ &\leq E[F(\varpi_\tau(z) | z) | Z_t = z] \\ &= F(\varpi_\tau(Z_t) | z) \\ &= \tau \end{aligned}$$

¹One can also use LOWESS (LOcally WEighted Scatter-plot Smoother) regression introduced by (Cleveland, 1981) to smooth the estimator in (4.1.30) and which solves the problem of boundary effects.

We've have used the Jensen's theorem for conditional expectation found in (Chen et al., 2003) and stated as follows:

Theorem 4.1.3 (Jensen's inequality). *For any convex function l ,*

$$E [l(X)] \geq l (E[X]) \quad (4.1.35)$$

Proof of Theorem 4.1.35. Suppose that l is differentiable. The function l is convex if

$$l(x) \geq l(y) + (x - y)l'(y), \quad \text{for any } x, y. \quad (4.1.36)$$

Let $x = X$ and $y = E[X]$. The inequality $l(X) \geq l(E[X]) + (X - E[X])l'(E[X])$ is true for all X and taking its expectation on both sides prove the theorem. \square

This inequality is applicable when f is a conditional convex function and when $E[\cdot]$ is a conditional expectation. The estimator $\varpi_{\tau,s}(Z_t)$ is also element of the set in which the unsmoothed estimator belongs. This means that $F(\varpi_{\tau,s}(Z_t) | z) \geq \tau$. The estimator is empirically given by

$$\hat{\varpi}_{\tau,s}(z) = \frac{\sum_{t=1}^n K_b(Z_t - z)y_t}{\sum_{t=1}^n K_b(Z_t - z)} = \frac{\sum_{t=1}^n K_b(Z_t - z)\varpi_{\tau}(Z_t)}{\sum_{t=1}^n K_b(Z_t - z)} \quad (4.1.37)$$

Asymptotic properties

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

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

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 $\varpi_{\tau,s}(z + sh)$ given by

$$f(z + sh) = f(z) + \frac{f^{(1)}(z)}{1!}sb_z + \frac{f^{(2)}(z)}{2!}(sb_z)^2 + o(b_z^2) \quad (4.1.38)$$

$$\varpi_{\tau,s}(z + sh) = \varpi_{\tau,s}(z) + \frac{\varpi_{\tau,s}^{(1)}(z)}{1!}sb_z + \frac{\varpi_{\tau,s}^{(2)}(z)}{2!}(sb_z)^2 + o(b_z^2). \quad (4.1.39)$$

We get

$$\begin{aligned} & \mathbb{E} [K_b(Z_t - z)Y_t] \\ &= \varpi_{\tau,s}(z)f(z) + \frac{1}{2}b^2\mu_2(K) \left(f(z)\varpi_{\tau,s}^{(2)}(z) + f^{(1)}(z)\varpi_{\tau,s}^{(1)}(z) + f^{(2)}(z)\varpi_{\tau,s}(z) \right) + o(h^3) \end{aligned} \quad (4.1.40)$$

Similarly, the expectation of the numerator is

$$\mathbb{E} [K_b(Z_t - z)] = f(z) + \frac{1}{2}b^2\mu_2(K)f^{(2)}(z) + o(h^2). \quad (4.1.41)$$

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,

$$\mathbb{E} [\hat{\varpi}_{\tau,s}(z)] \approx \varpi_{\tau,s}(z) + \frac{1}{2}b^2\mu_2(K) \left(\varpi_{\tau,s}^{(2)}(z) + 2\frac{f^{(1)}(z)}{f(z)}\varpi_{\tau,s}^{(1)}(z) \right) \quad (4.1.42)$$

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

$$\begin{aligned} \mathbb{V} \left(\frac{1}{n} \sum_{t=1}^n K_b(Z_t - z)Y_t \right) &= \frac{1}{nb^2} \mathbb{V} \left(K \left(\frac{Z_t - z}{b} \right) Y_t \right) \\ &= \frac{1}{nb^2} \left(\mathbb{E} \left[K^2 \left(\frac{Z_t - z}{b} \right) y_t^2 \right] - \left(\mathbb{E} \left[K \left(\frac{Z_t - z}{b} \right) Y_t \right] \right)^2 \right) \\ &\approx \frac{1}{nb} \iint v^2 K^2(s) f(v|z + sb) f(z + sb) ds dv - o\left(\frac{1}{n}\right) \\ &= \frac{1}{nb} \int K^2(s) f(z + sb) \left(\int v^2 f(v|z + sb) dv \right) ds - o\left(\frac{1}{n}\right) \\ &\approx \frac{1}{nb} R(K) f(z) \left[\sigma_\eta^2 + \varpi_{\tau,s}^2(z) \right] \end{aligned} \quad (4.1.43)$$

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

The covariance of the numerator and the denominator of the estimator in (4.1.37) is given by

$$\begin{aligned}
\text{Cov}(N, D) &= \text{Cov}\left(\frac{1}{nb} \sum_{t=1}^n K\left(\frac{Z_t - z}{b}\right) Y_t, \frac{1}{nb} \sum_{t=1}^n K\left(\frac{Z_t - z}{b}\right)\right) \\
&= \frac{1}{nb^2} \text{Cov}\left(K\left(\frac{Z_t - z}{b}\right) Y_t, K\left(\frac{Z_t - z}{b}\right)\right) \\
&= \frac{1}{nb^2} \left(\mathbb{E}\left[K^2\left(\frac{Z_t - z}{b}\right) Y_t\right] - \mathbb{E}\left[K\left(\frac{Z_t - z}{b}\right) Y_t\right] \mathbb{E}\left[K\left(\frac{Z_t - z}{b}\right)\right] \right) \\
&\approx \frac{1}{nb} R(K) f(z) \varpi_{\tau,s}(z) - o\left(\frac{1}{n}\right)
\end{aligned} \tag{4.1.44}$$

The variance of the estimator in (4.1.37) is the variance of a ratio of correlated variables that can be calculated using the approximation found in (Seltman, 2012)

$$V\left(\frac{N}{D}\right) \approx \left(\frac{\mathbb{E}[N]}{\mathbb{E}[D]}\right)^2 \left[\frac{V(N)}{(\mathbb{E}[N])^2} + \frac{V(D)}{(\mathbb{E}[D])^2} - \frac{2 \text{Cov}(N, D)}{\mathbb{E}[N] \mathbb{E}[D]} \right] \tag{4.1.45}$$

$$= \frac{R(K) \sigma_\eta^2}{nb f(z)} \tag{4.1.46}$$

If the assumption 4 for strong mixing processes holds, then from the Central Limit Theorem (CLT) we have

$$\sqrt{nb} \left(\hat{\varpi}_{\tau,s}(z) - \varpi_{\tau,s}(z) - \text{Bias}(\hat{\varpi}_{\tau,s}(z)) \right) \xrightarrow{D} \mathcal{N}\left(0, \frac{R(K) \sigma_\eta^2}{f(z)}\right) \tag{4.1.47}$$

4.1.2 Asymptotic normality of QARCH

The CCDF in (4.1.29) can be written in the form of an arithmetic mean of a random variable L :

$$\hat{F}(x^*|z) = \frac{1}{n} \sum_{t=1}^n L_t \quad \text{with} \quad L_t = \frac{K_{b_z}(Z_t - z) I_{\{X_t^* \leq x^*\}}}{\frac{1}{n} \sum_{t=1}^n K_{b_z}(Z_t - z)} \quad (4.1.48)$$

and the approximation of the expectation of L is

$$E[L_t] \approx \frac{E \left[K_{b_z}(Z_t - z) I_{\{X_t^* \leq x^*\}} \right]}{E \left[\frac{1}{n} \sum_{t=1}^n K_{b_z}(Z_t - z) \right]} = \frac{E[N]}{E[D]} \quad (4.1.49)$$

see (Seltman, 2012). Using the i.i.d assumption over the data, the numerator is

$$\begin{aligned} E[N] &= \frac{1}{b_z} E \left[K \left(\frac{Z_t - z}{b_z} \right) I_{\{X_t^* \leq x^*\}} \right] \\ &= \frac{1}{b_z} \iint_{-\infty}^{x^*} K \left(\frac{u - z}{b_z} \right) f(u, v) du dv \\ &= \int F(x^* | z + sh) K(s) f(z + sh) ds \end{aligned} \quad (4.1.50)$$

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

$$\begin{aligned} E[N] &= f(z)F(x^* | z) + b_z^2 \mu_2(K) \left[f^{(1)}(z)F^{(1)}(x^* | z) + \frac{1}{2}f^{(2)}(z)F(x^* | z) + \right. \\ &\quad \left. \frac{1}{2}f(z)F^{(2)}(x^* | z) + o(b_z^2) \right] \end{aligned} \quad (4.1.51)$$

and for the denominator, we have

$$E[D] = f(z) + \frac{1}{2}b_z^2 \mu_2(K) f^{(2)}(z) + o(b_z^2) \quad (4.1.52)$$

Thus,

$$\begin{aligned}
& \mathbb{E}[L_t] \\
& \approx \frac{f(z) \left[F(x^* | z) + b_z^2 \mu_2(K) \left(\frac{f^{(1)}(z)}{f(z)} F^{(1)}(x^* | z) + \frac{1}{2} \frac{f^{(2)}(z)}{f(z)} F(x^* | z) + \frac{1}{2} F^{(2)}(x^* | z) \right) \right]}{f(z) \left(1 + \frac{1}{2} b_z^2 \mu_2(K) \frac{f^{(2)}(z)}{f(z)} \right)} \\
& = F(x^* | z) + \frac{1}{2} b_z^2 \mu_2(K) \left(2 \frac{f^{(1)}(z)}{f(z)} F^{(1)}(x^* | z) + F^{(2)}(x^* | z) \right) + o(b_z^4)
\end{aligned} \tag{4.1.53}$$

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

Hence,

$$\text{Bias} \left(\hat{F}(x^* | z) \right) \approx \frac{1}{2} b_z^2 \mu_2(K) \left(2 \frac{f^{(1)}(z)}{f(z)} F^{(1)}(x^* | z) + F^{(2)}(x^* | z) \right) \tag{4.1.54}$$

Some authors assumed that, in this case, the first derivative of the true pdf of Z at point z can be zero (Hansen, 2004) as the one for the fixed design and therefore, the bias can be given by

$$\text{Bias} \left(\hat{F}(x^* | z) \right) \approx \frac{1}{2} b_z^2 \mu_2(K) \left(F^{(2)}(x^* | z) \right) \tag{4.1.55}$$

We have

$$\begin{aligned}
\mathbb{V}(N) &= \mathbb{V} \left(\frac{1}{b_z} K \left(\frac{Z_t - z}{b_z} \right) I_{\{X_t^* \leq x^*\}} \right) = \frac{1}{b_z^2} \mathbb{V} \left(K \left(\frac{Z_t - z}{b_z} \right) I_{\{X_t^* \leq x^*\}} \right) \\
&= \frac{1}{b_z^2} \left(\mathbb{E} \left[K^2 \left(\frac{Z_t - z}{b_z} \right) I_{\{X_t^* \leq x^*\}} \right] - \left(\mathbb{E} \left[K \left(\frac{Z_t - z}{b_z} \right) I_{\{X_t^* \leq x^*\}} \right] \right)^2 \right) \\
&\approx \frac{F(x^* | z) f(z) R(K)}{b_z} - o(1),
\end{aligned} \tag{4.1.56}$$

$$\begin{aligned}
\mathbf{V}(D) &= \mathbf{V} \left(\frac{1}{n} \sum_{t=1}^n K_{b_z}(Z_t - z) \right) = \frac{1}{nb_z^2} \mathbf{V} \left(K \left(\frac{Z_t - z}{b_z} \right) \right) \\
&= \frac{1}{nb_z^2} \left(\mathbf{E} \left[K^2 \left(\frac{Z_t - z}{b_z} \right) \right] - \left(\mathbf{E} \left[K \left(\frac{Z_t - z}{b_z} \right) \right] \right)^2 \right) \\
&\approx \frac{f(z)R(K)}{nb_z} - o\left(\frac{1}{n}\right),
\end{aligned} \tag{4.1.57}$$

$$\begin{aligned}
\text{Cov}(N, D) &= \frac{1}{nb_z^2} \text{Cov} \left(K \left(\frac{Z_t - z}{b_z} \right) I_{\{X_t^* \leq x^*\}}, K \left(\frac{Z_t - z}{b_z} \right) \right) \\
&\approx \frac{1}{nb_z^2} \mathbf{E} \left[K^2 \left(\frac{Z_t - z}{b_z} \right) I_{\{X_t^* \leq x^*\}} \right] - o\left(\frac{1}{n}\right) \\
&\approx \frac{1}{nb_z} F(x^*|z) f(z) R(K)
\end{aligned} \tag{4.1.58}$$

Using the same approximation in (4.1.45), the variance of $\hat{F}(x^*|z)$ is

$$\mathbf{V}(L_t) \approx F(x^*|z) \left[\frac{R(K) (1 - F(x^*|z))}{b_z f(z)} \right] \tag{4.1.59}$$

and by the Central Limit Theorem, using assumption 4 for $\{(X_t^*, Z_t), t = 1, 2, \dots\}$

$$\sqrt{n} \left(\hat{F}(x^*|z) - F(x^*|z) - \text{Bias}(F(x^*|z)) \right) \xrightarrow{D} \mathcal{N}(0, \mathbf{V}(L_t)) \tag{4.1.60}$$

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

Theorem 4.1.4 (Delta Method). *Suppose $\hat{F}(x^*|z)$ has the asymptotic normal distribution as in (??). Suppose $g(\cdot)$ is a continuous function that has a derivative $g^{(1)}(\cdot)$ at $\mu = \mathbf{E}[\hat{F}(x^*|z)]$. Then*

$$\sqrt{nb_z} \left(g \left(\hat{F}(x^*|z) \right) - g(\mu) \right) \xrightarrow{D} \mathcal{N} \left(0, \left[g^{(1)}(\mu) \right]^2 \frac{R(K) (1 - F(x^*|z))}{f(z)} \right) \tag{4.1.61}$$

Proof of Theorem 4.1.4. 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 $\mathcal{N}\left(0, \left[g^{(1)}(\mu)\right]^2 \frac{R(K)(1-F(x^*|z))}{f(z)}\right)$ in distribution. \square

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

$$\frac{1}{f(F^{-1}(\mu|z)|z)} \xrightarrow{p} \frac{1}{f(F^{-1}(\tau|z)|z)} = \frac{1}{f(\varpi_\tau(z)|z)} \quad (4.1.62)$$

at points x^* 's that satisfy (4.1.30). Using again the first-order Taylor expansion, we also have

$$\begin{aligned} g(\mu) &= g\left(F(x^*|z) + \text{Bias}\left(\hat{F}(x^*|z)\right)\right) \\ &\approx g(F(x^*|z)) + \text{Bias}\left(\hat{F}(x^*|z)\right) \times g^{(1)}(F(x^*|z)) \\ &= x^* + \frac{\text{Bias}\left(\hat{F}(x^*|z)\right)}{f(x^*|z)} \end{aligned} \quad (4.1.63)$$

for x^* 's satisfying (4.1.30) and replacing $\hat{F}(\varpi_\tau(z)|z)$ by $F(\hat{\varpi}_\tau(z)|z)$ using the uniqueness assumption of $\varpi_\tau(z)$, (4.1.61) becomes

$$\sqrt{nb}\left(\hat{\varpi}_\tau(z) - \varpi_\tau(z) - \text{Bias}\left(\hat{\varpi}_\tau(z)\right)\right) \xrightarrow{D} \mathcal{N}\left(0, \frac{R(K)\tau(1-\tau)}{f(z)\left[f(\varpi_\tau(z)|z)\right]^2}\right) \quad (4.1.64)$$

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

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

4.2 Bandwidth selections

4.2.1 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.

$$\begin{aligned} \text{AMISE} \left(\hat{f}(z) \right) &= \int \text{E} \left[\left(\hat{f}(z) - f(z) \right)^2 \right] dz \\ &= \int \text{E} \left[\left(\hat{f}(z) - \text{E} \left[\hat{f}(z) \right] + \text{Bias} \left(\hat{f}(z) \right) \right)^2 \right] dz \\ &= \int \left\{ \text{E} \left[\left(\hat{f}(z) - \text{E} \left[\hat{f}(z) \right] \right)^2 \right] + \text{Bias}^2 \left(\hat{f}(z) \right) \right\} dz \quad (4.2.1) \\ &= \int \left\{ \text{V} \left(\hat{f}(z) \right) + \text{Bias}^2 \left(\hat{f}(z) \right) \right\} dz \\ &= \int \left\{ \frac{R(K)f(z)}{nb} + \frac{1}{4} b^4 \mu_2^2(K) \left[f^{(2)}(z) \right]^2 \right\} dz \\ &= \frac{R(K)}{nb} + \frac{1}{4} b^4 \mu_2^2(K) R \left(f^{(2)}(z) \right) \end{aligned}$$

The general form of the r^{th} derivatives of the AMISE w.r.t b was studied in (Raykar and Duraiswami, 2006), considering that the unknown functions in (4.2.1) are also functions

of the smoothing parameter.

$$\frac{d}{dz^r} \text{AMISE} \left(\hat{f}(z) \right) = \frac{R(K^{(r)})}{nb^{2r+1}} + \frac{1}{4}b^4\mu_2^2(K)R \left(f^{(2+r)}(z) \right) \quad (4.2.2)$$

The optimal smoothing parameter minimizing (4.2.2) is

$$b^* = \left[\frac{(2r+1)R(K^{(r)})}{\mu_2^2(K)R \left(f^{(2+r)}(z) \right)} \right]^{1/(2r+5)} \times n^{-1/(2r+5)} \quad (4.2.3)$$

Using this result, we came up with the optimal version of optimal bandwidth for CCDF.

The aim of derivation the AMISE in (4.2.1) 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

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

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

and its 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 in terms of efficiency (see Table 4.2.1)

Table 4.2.1: Description of the most used kernel functions

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

4.2.2 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 the two are linked. When the variance is big, the bias also is big and when the variance is small, the bias is small.

$$\begin{aligned}
 \text{AMSE} \left(\hat{F}(x^*|z) \right) &= \text{E} \left[\left(\hat{F}(x^*|z) - F(x^*|z) \right)^2 \right] \\
 &= \text{E} \left[\left(\hat{F}(x^*|z) - \text{E} \left[\hat{F}(x^*|z) \right] + \text{Bias} \left(\hat{F}(x^*|z) \right) \right)^2 \right] \\
 &= \text{E} \left[\left(\hat{F}(x^*|z) - \text{E} \left[\hat{F}(x^*|z) \right] \right)^2 \right] + \text{Bias}^2 \left(\hat{F}(x^*|z) \right) \\
 &\times \text{E} \left[\hat{F}(x^*|z) - \text{E} \left[\hat{F}(x^*|z) \right] \right] + \text{Bias}^2 \left(\hat{F}(x^*|z) \right) \\
 &= \text{V} \left(\hat{F}(x^*|z) \right) + \text{Bias}^2 \left(\hat{F}(x^*|z) \right) \\
 &= \frac{R(K)}{nb_z f(z)} F(x^*|z) (1 - F(x^*|z)) + \frac{b^4}{4} \mu_2^2(K) \left(F^{(2)}(x^*|z) \right)^2
 \end{aligned} \tag{4.2.4}$$

which is given by (3.2.29) and (4.1.55). Therefore,

$$b^* = \arg \min_{b>0} \text{AMSE} \left(\hat{F}(x^*|z) \right) \quad (4.2.5)$$

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

$$b^* = \left\{ \frac{R(K)F(x^*|z) (1 - F(x^*|z))}{\mu_2^2(K)f(z) (F^{(2)}(x^*|z))^2} \right\}^{\frac{1}{5}} \times n^{-\frac{1}{5}} \quad (4.2.6)$$

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 (4.1.29) at point $Z_t = z$. The estimator of the r^{th} derivatives of (4.1.29) is:

$$\hat{F}^{(r)}(x^*|z) = \frac{d^r}{dz^r} \sum_{t=1}^n W_t(z) X_{\{X_t^* \leq x^*\}} = \sum_{t=1}^n W_t^{(r)}(z) X_{\{X_t^* \leq x^*\}} \quad (4.2.7)$$

with

$$W_t(z) = \frac{K \left(\frac{Z_t - z}{b} \right)}{\sum_{t=1}^n K \left(\frac{Z_t - z}{b} \right)} = \frac{K \left(\frac{Z_t - z}{b} \right)}{nb \hat{f}(z)} \quad (4.2.8)$$

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

$$W_t^{(1)}(z) = \frac{1}{nb^2} \frac{K^{(1)} \left(\frac{Z_t - z}{b} \right) \hat{f}(z) - bK \left(\frac{Z_t - z}{b} \right) \hat{f}^{(1)}(z)}{[f^{(1)}(z)]^2} = \frac{1}{nb^2} \frac{A}{B} \quad (4.2.9)$$

and the second derivative is also

$$W_t^{(2)}(z) = \frac{1}{nb^2} \frac{A^{(1)}B - B^{(1)}A}{B^2} \quad (4.2.10)$$

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 (Chen, 2015), (Guidoum, 2013), (Raykar and Duraiswami,

2006). 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 (Guidoum, 2013) 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 .

4.3 Simulation study

4.3.1 Model specification

The ARCH(q) models introduced by (Engle, 1982) 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 $p = 1, 2, \dots$**

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

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

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

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

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

4.3.2 Specifications for AR(1)-GARCH(1,1)

Unconditional expectation

The unconditional expectation is

$$E[X_t] = \mu + \delta E[X_{t-1}] + E[\varpi_t e_t] = \mu + \delta E[X_t] + E[\varpi_t] E[e_t] \quad (4.3.1)$$

Note that $E[X_t] = E[X_{t-1}]$ is used to ensure the stationarity of the process. That is, the expectation is therefore given by

$$E[X_t] = \frac{\mu}{1 - \delta} \quad (4.3.2)$$

Unconditional variance

The unconditional variance of the model is given by the law of total variance

$$V(X_t) = E[V(X_t | X_{t-1})] + V(E[X_t | X_{t-1}]) \quad (4.3.3)$$

$$= E[\varpi_t^2] + V[\alpha_t] \quad (4.3.4)$$

We have

$$E[\varpi_t^2] = \omega + \alpha E[X_{t-1}^2] + \beta E[\varpi_{t-1}^2] \quad (4.3.5)$$

Using the i.i.d. assumption on the sequence of random variables X_1, X_2, \dots, X_n , the expected value of X_t^2 can be calculated as follow

$$E[X_t^2] = E[\mu X_t + \delta X_{t-1} X_t + \varpi_t e_t X_t] \quad (4.3.6)$$

$$= \mu E[X_t] + \delta (E[X_t])^2 \quad (4.3.7)$$

$$= \frac{\mu^2}{1 - \delta} + \frac{\delta \mu^2}{(1 - \delta)^2} \quad (4.3.8)$$

$$= \frac{\mu^2}{(1 - \delta)^2} \quad (4.3.9)$$

Which is independent of time. In another way,

$$\mathbb{E}[X_t^2] = \mathbb{E}[\alpha_t^2 + 2\alpha\varpi_t e_t + \varpi_t^2 e_t^2] \quad (4.3.10)$$

$$= \mathbb{E}[\alpha_t^2] + \mathbb{E}[\varpi_t^2] \quad (4.3.11)$$

The equation (4.3.5) becomes

$$\mathbb{E}[\varpi_t^2] = \omega + \alpha (\mathbb{E}[\alpha_t^2] + \mathbb{E}[\varpi_t^2]) + \beta \mathbb{E}[\varpi_{t-1}^2] \quad (4.3.12)$$

$$= \omega + \alpha \mathbb{E}[\alpha_t^2] + (\alpha + \beta) \mathbb{E}[\varpi_t^2] \quad (\text{stationarity}) \quad (4.3.13)$$

We obtain

$$\mathbb{E}[\varpi_t^2] = \frac{\omega + \alpha \mathbb{E}[\alpha_t^2]}{1 - \alpha - \beta} \quad (4.3.14)$$

The expectation of α_t^2 is given by

$$\mathbb{E}[\alpha_t^2] = \mathbb{E}[(\mu + \delta X_{t-1})^2] \quad (4.3.15)$$

$$= \mu^2 + 2\mu\delta \mathbb{E}[X_t] + \delta^2 \mathbb{E}[X_t^2] \quad (4.3.16)$$

$$= \mu^2 + 2\frac{\delta\mu^2}{1-\delta} + \frac{\delta^2\mu^2}{(1-\delta)^2} \quad (4.3.17)$$

$$= \frac{\mu^2}{(1-\delta)^2} \quad (4.3.18)$$

It follows that

$$\mathbb{E}[\varpi_t^2] = \frac{\omega(1-\delta)^2 + \alpha\mu^2}{(1-\alpha-\beta)(1-\delta)^2} \quad (4.3.19)$$

and the variance in (4.3.4) becomes

$$\mathbb{V}(X_t) = \frac{\omega(1-\delta)^2 + \alpha\mu^2}{(1-\alpha-\beta)(1-\delta)^2} + \mathbb{V}(\mu + \delta X_{t-1}) \quad (4.3.20)$$

$$= \frac{\omega(1-\delta)^2 + \alpha\mu^2}{(1-\alpha-\beta)(1-\delta)^2} + \delta^2 \mathbb{V}(X_t) \quad (4.3.21)$$

$$= \frac{\omega(1-\delta)^2 + \alpha\mu^2}{(1-\alpha-\beta)(1-\delta^2)(1-\delta)^2} \quad (4.3.22)$$

This variance is positive and finite for $\mu \in \mathbb{R}$, $|\delta| < 1$, $\omega > 0$, $\alpha > 0$, $\beta > 0$ and $\alpha + \beta < 1$.

4.3.3 Monte Carlo study

We simulated the data from (1.2.1) with $\mu = 0.5, \delta = 0.3$, for the AR(1) part, $w = 0.1, \alpha = 0.35$, for the ARCH(1) and $e_t \sim \text{i.i.d. } \mathcal{N}(0, 1)$. The data plot is represented by figure 4.3.1.

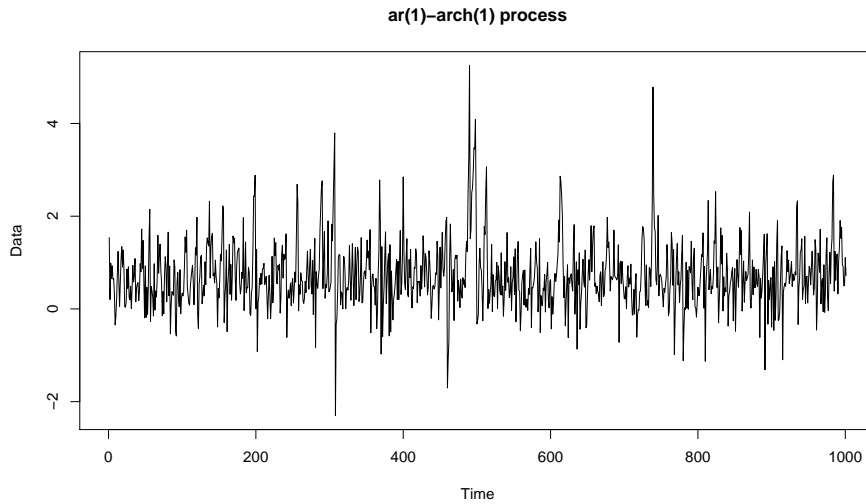


Figure 4.3.1: Plot of the simulated AR(1)-ARCH(1)

Our algorithm gives the estimation of the conditional scale function which suffers from boundary effects as it's seen on figure 4.3.2. 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 (Karunamuni and Alberts, 2005). The consistency of our estimator is dependent on this problem of big variations at the boundaries. This increases the Average Squared Error between two different estimation from a same model.

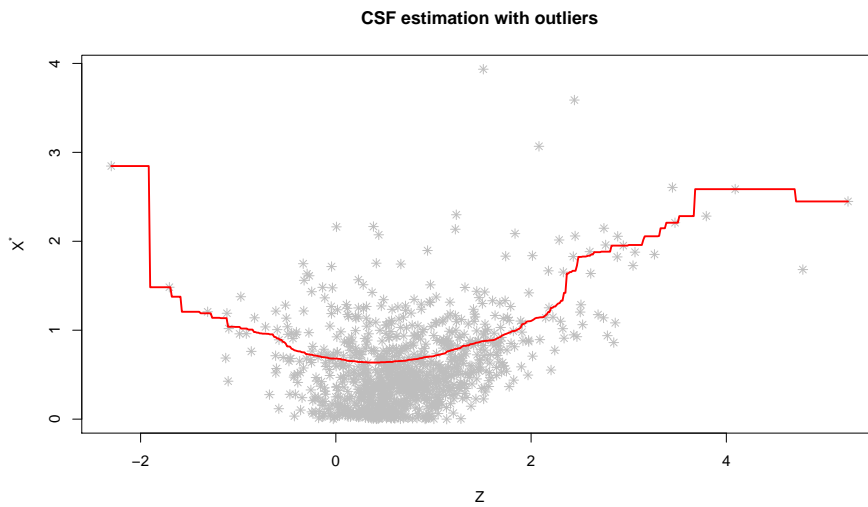


Figure 4.3.2: Conditional scale function estimate at $\tau = 0.75$

4.3.4 Boundary correction

The following figure 4.3.3 is the representation of Z_t and the transformed response variable X_t^* defined in (4.1.28) at level $\tau = 0.75$.

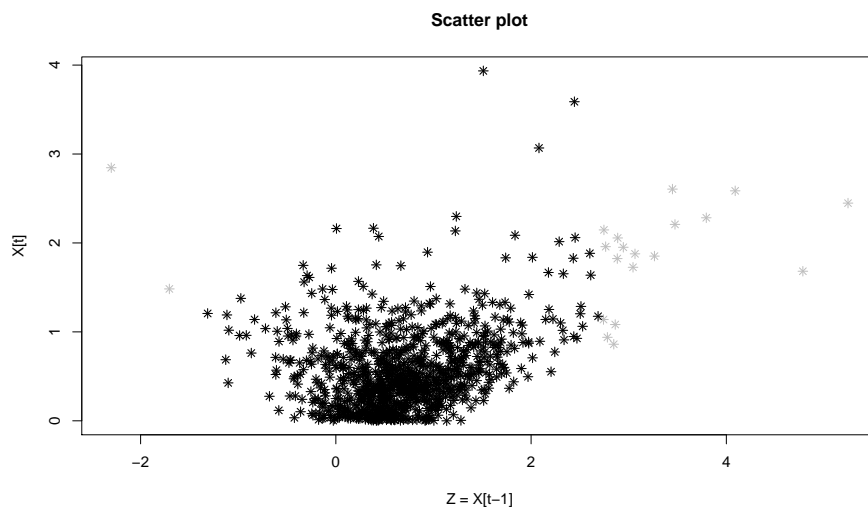


Figure 4.3.3: Scatter plot and outliers detection

The gray points are outliers from (4.3.3). We lose 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 $\tau > 0.5$.

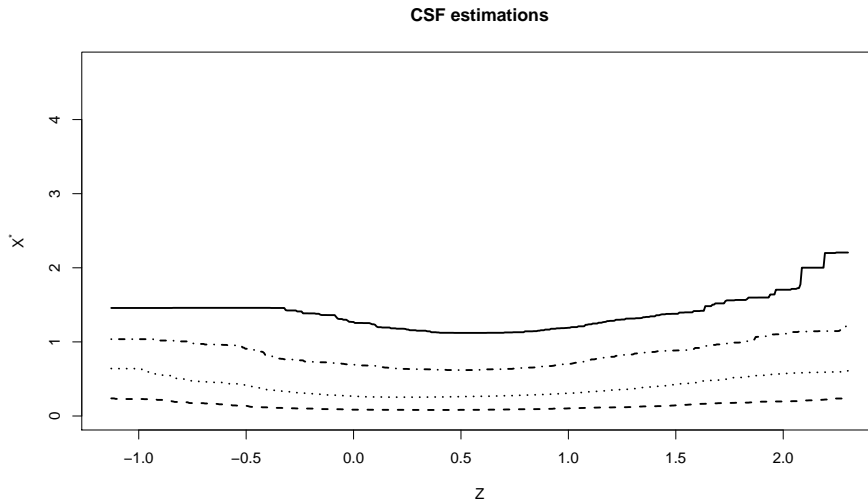


Figure 4.3.4: CSF estimations

The curves represent the estimations of the CSF at $\tau = 0.9, 0.75, 0.50, 0.25$ from up to down. As it's seen on figure 4.3.4, the curve are not smooth that why the NW method discussed in section 4.1.1 which requires that unsmoothed estimator and the bins $z_1^*, z_2^*, \dots, z_N^*$. We obtain the following graphic which combines the two estimations.

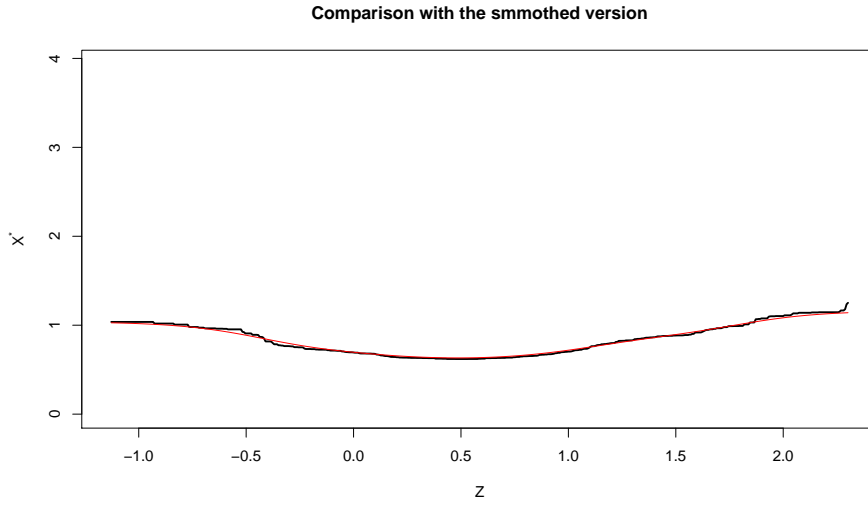


Figure 4.3.5: Smoothed estimate of the CSF

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

4.3.5 Accuracy of the estimator

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 of bootstrap method to calculate the average gap between m estimated CSFs. The formula is

$$\text{MASE}(\hat{w}_\tau(z)) = \frac{1}{n} \sum_{j=1}^n \left[\frac{1}{m} \sum_{i=1}^m (\hat{w}_{\tau,1}(z_i) - \hat{w}_{\tau,j}(z_i))^2 \right] \quad (4.3.23)$$

The following tables show the accuracy of the smoothed CSF compared to the rough (nonsmoothed) one. We can see that our estimation is more accurate for all proportions.

Table 4.3.1: MASE for QAR at $\tau = 0.25$

n	rough $\hat{\omega}_{0.25}$	smooth $\hat{\omega}_{0.25}$
250	0.04557	0.00082
500	0.09579	0.00036
1000	0.19245	0.00063

Table 4.3.2: MASE for QAR at $\tau = 0.50$

n	rough $\hat{\omega}_{0.50}$	smooth $\hat{\omega}_{0.50}$
250	0.24196	0.00567
500	0.23448	0.00367
1000	0.29932	0.00392

Table 4.3.3: MASE for QAR at $\tau = 0.75$

n	rough $\hat{\omega}_{0.75}$	smooth $\hat{\omega}_{0.75}$
250	0.54523	0.01936
500	0.49205	0.01658
1000	1.2227	0.01191

Table 4.3.4: MASE for QAR at $\tau = 0.90$

n	rough $\hat{\omega}_{0.90}$	smooth $\hat{\omega}_{0.90}$
250	2.60788	0.10582
500	2.11358	0.05296
1000	2.13707	0.04107

4.4 Conclusion

We've derived an estimator for the conditional scale function in an AR(1)-GARCH(1) despite the high variability of the simulated data, we were able to deal with the boundary effect and were able to show the asymptotic behavior 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. In the chapter, the estimation when the QAR(1) is unknown is carried out.

Chapter 5

Estimation of the QAR and QARCH

In the previous chapter, the quantile autoregressive (QAR) was assumed to be zero (0) and we derived the estimator of the conditional scale function QARCH and showed that it is asymptotically normal with bias $O(b^2/2)$ and variance $O(1/nb)$ where n and b are respectively the data size and bandwidth. In this chapter, the estimations of the QAR and the QARCH functions are carried out considering the model (1.2.2) with the same assumptions made for consistency proof. The estimation of the QARCH follows the same proceeding and depends on the estimation of the QAR function. This can be written theoretically but can have a dimension issue when it comes to computation because the QAR's estimate is a set of points that may not be of the same length as the response variable due to the fact that, in non-parametric regression using (Nadaraya, 1964)-(Watson, 1964) methods, the unknown functions are estimated using bins. It's feasible if the number of bins are chosen to be equal to the length of X_t . To take care of the dimension issue, a similar work was done in (Hall et al., 2002) on the prediction based on non-parametric estimation of the conditional median using local least absolute regression. Our approach will take care of all given quantile $\tau \in (0, 1)$. The estimation of the QARCH is based on residuals (Koenker and Bassett, 1978), (Koenker and Zhao, 1996), (Koenker, 2005) and (Givord and Dhaultfoeuille, 2013) by removing the effect

of the QAR effect from X_t at each level τ . The kernel function used for this application is the Gaussian kernel function.

5.1 Conditional Quantile estimations

To obtain the QAR-QARCH model from (1.2.1), we simply take its τ^{th} conditional quantile and we obtain:

$$Q_\tau(X_t | X_{t-1}) = \alpha_\tau(X_{t-1}) = \alpha(X_{t-1}) + \varpi(X_{t-1})q_\tau^e \quad (5.1.1)$$

where $q_\tau^e = F_e^{-1}(\tau)$ is the τ^{th} quantile of $\{e_t\}$. To make the reading less difficult, X_{t-1} is changed to Z_t . Note that (5.1.1) is the estimation of the CVaR (Conditional Value-at-Risk) discussed in . Now, centering the response variable in (1.2.1) at its τ^{th} -quantile in (5.1.1), we get:

$$X_t - \alpha_\tau(Z_t) = \varpi(Z_t) (e_t - q_\tau^e) \quad (5.1.2)$$

which is equivalent to the quantile autoregressive model:

$$X_t = \alpha_\tau(Z_t) + \varepsilon_\tau, \quad (5.1.3)$$

where $\varepsilon_\tau = \varpi(Z_t) (e_t - q_\tau^e)$ is 0 τ -quantile, i.e, $Q_\tau(\varepsilon_\tau) = 0$.

5.1.1 Non-parametric QAR

Consider the model (5.1.1) and the assumption made on the innovation ε_τ . By definition, ε_τ is zero τ -quantile meaning

$$\Pr(\varepsilon_\tau \leq 0) = \Pr(\varepsilon_\tau \leq 0 | Z_t) = \tau \quad (5.1.4)$$

and using (5.1.4), we have

$$\Pr(X_t \leq \alpha_\tau(Z) | Z_t) = E \left[I (X_t \leq \alpha_\tau(Z_t)) | Z_t \right] = \tau \quad (5.1.5)$$

which is equivalent to $F(\alpha_\tau(Z_t) | Z_t) = \tau$. The conditional quantile function α_τ minimizes the objective function $E[\gamma_\tau(X_t, \alpha_\tau) | Z_t]$, i.e.

$$\alpha_\tau(z) = \arg \min_{\alpha_\tau} E[\gamma_\tau(X, \alpha_\tau) | Z_t = z] \quad (5.1.6)$$

and is empirically given by

$$\hat{\alpha}_\tau(z) = \arg \min_{\alpha_\tau} \frac{1}{n} \sum_{t=1}^n K_b(Z_t - z) \gamma_\tau(X_t, \alpha_\tau) \quad (5.1.7)$$

Let's denote $\hat{\varphi}_{n,\tau} = \frac{1}{n} \sum_{t=1}^n K_b(Z_t - z) \gamma_\tau(X_t, \alpha_\tau)$. The zero of the equation $\frac{d}{d\alpha_\tau} \varphi_{n,\tau} = 0$ is

$$\hat{\alpha}_\tau(z) = \inf \{ \mu : F(\mu | z) \geq \tau \} \equiv \hat{F}^{-1}(\tau | z) \quad (5.1.8)$$

where

$$\hat{F}(x | z) = \left[n \hat{f}(z) \right]^{-1} \sum_{t=1}^n K_b(Z_t - z) I(X_t \leq x) \quad (5.1.9)$$

Theorem 5.1.1. *Under the assumptions 1, 2 and 5,*

$$\hat{F}(x | z) \xrightarrow{d} F(x | z) \quad (5.1.10)$$

if the assumption 4 holds, we have in addition

$$\sqrt{n} \left(\hat{F}(x^* | z) - F(x^* | z) - \text{Bias}(F(x^* | z)) \right) \xrightarrow{D} \mathcal{N}(0, V(L_t)) \quad (5.1.11)$$

It follows that

$$\hat{\alpha}_\tau(z) \xrightarrow{D} \alpha_\tau(z). \quad (5.1.12)$$

Proof. The proof is similar to the one given in the chapter 4 using CLT and by the use of the Delta Method we showed the consistency of $\hat{\alpha}_\tau(z)$ \square

5.1.2 k Nearest Neighbor (k -NN) prediction

The prediction $\tilde{\alpha}_\tau(z)$ of a future value or any value $Z_{n+1} = z$ is easy in parametric regression once we have the estimated coefficients of a model. But in non-parametric

regression, this prediction is somehow impossible. Recent research on this problem suggests methods more or less feasible for our type of estimation. There is the k NN (k Nearest Neighbor)(Caires and Ferreira, 2005) method which consists of finding the k values, z_1^*, \dots, z_k^* close to z . The requirement of this method is that the estimator α_τ is to be smooth (Caires and Ferreira, 2005). Unfortunately, the estimation of the QAR in (5.1.7) is not smooth and suffers from boundary issues. Having estimated $\hat{\alpha}_\tau(z_i^*)$ and the bin points $z_i^*, i = 1, \dots, N$, thus, $\tilde{\alpha}_\tau(z)$ will be the average of $\hat{\alpha}_\tau(z_1^*), \dots, \hat{\alpha}_\tau(z_k^*)$. In other words,

$$\tilde{\alpha}_\tau(z) = \frac{1}{k} \sum_{i=1}^k \hat{\alpha}_\tau(z_i^*) \quad (5.1.13)$$

This approach is used to predict the values $\tilde{\alpha}_\tau(Z_t)$ which is a sequence of n points. The figure 5.2.1 represents the prediction for the entire data (for instance, the daily returns) at $\tau = 0.25, 0.50, 0.75, 0.9$. In order to see if the prediction is accurate, the following error is calculated (the mean squared error of the difference between $\hat{\alpha}_\tau(z_i^*)$ and $\tilde{\alpha}_\tau(z_i^*)$ for bins z_1^*, \dots, z_N^*)

$$\tilde{e}_p = \frac{1}{N} \sum_{i=1}^N (\hat{\alpha}_\tau(z_i^*) - \tilde{\alpha}_\tau(z_i^*))^2 \quad (5.1.14)$$

The same prediction applies when we have the non-parametric estimation of the conditional scale function $\hat{\omega}_\tau$.

5.1.3 Non-parametric QARCH

Considering that the QAR is already estimated, we have

$$Q_\tau \left[\gamma_\tau (X_t - \alpha_\tau(Z_t)) \right] = \varpi(Z_t) Q_\tau \left[\gamma_\tau (e_t - q_\tau^e) \right] \quad (5.1.15)$$

The ratio of $X_t - \alpha_\tau$ in (5.1.2) and the left part in (5.1.15) gives

$$\frac{X_t - \alpha_\tau(Z_t)}{Q_\tau \left[\gamma_\tau (X_t - \alpha_\tau(Z_t)) \right]} = \frac{e_t - q_\tau^e}{Q_\tau \left[\gamma_\tau (e_t - q_\tau^e) \right]} \quad (5.1.16)$$

This transformation leads to the QAR-QARCH model

$$X_t = \alpha_\tau(Z_t) + \varpi_\tau(Z_t)\eta_\tau \quad (5.1.17)$$

where $\varpi_\tau(Z_t) = Q_\tau \left[\gamma_\tau (X_t - \alpha_\tau(Z_t)) \right]$ and $\eta_\tau = \frac{e_t - q_\tau^e}{Q_\tau \left[\gamma_\tau (e_t - q_\tau^e) \right]}$ is zero τ -quantile with unit scale. This property leads to the expression

$$\Pr (\gamma_\tau(\eta_\tau) \leq 1) = \Pr (\gamma_\tau(\eta_\tau) \leq 1 \mid Z) = \tau \quad (5.1.18)$$

This is identifiable to (5.1.5), if X_t and $\alpha_\tau(Z_t)$ are replaced by $\gamma_\tau (X_t - \alpha_\tau(Z_t))$ and $\varpi_\tau(Z_t)$ respectively. Thus, $\varpi_\tau(Z_t)$ minimizes $E \left[\gamma_\tau \left(\gamma_\tau (X_t, \alpha_\tau(Z_t)), \varpi_\tau(Z_t) \right) \mid Z_t \right]$, i.e.

$$\varpi_\tau(Z_t) = \arg \min_{\varpi_\tau} E \left[\gamma_\tau (X_t^*, \varpi_\tau) \mid Z_t \right] \quad (5.1.19)$$

or is empirically given by

$$\hat{\varpi}_\tau(Z_t) = \arg \min_{\varpi_\tau} \frac{1}{n} \sum_{t=1}^n K_b(Z_t - z) \gamma_\tau(X_t^*, \varpi_\tau) \quad (5.1.20)$$

where $X_t^* = \gamma_\tau (X_t, \alpha_\tau(Z_t))$. Again, if we denote $\hat{\varphi}_{n,\tau} = \frac{1}{n} \sum_{t=1}^n K_b(Z_t - z) \gamma_\tau (X_t^*, \varpi_\tau)$, then $\frac{d\hat{\varphi}_{n,\tau}}{d\varpi_\tau} = 0$ has as solution

$$\hat{\varpi}_\tau(z) = \inf \left\{ x^* \in \mathbb{R}_*^+ : \hat{F} (x^* \mid z) \geq \tau \right\} \equiv \hat{F}^{-1}(\tau \mid z) \quad (5.1.21)$$

with

$$\hat{F} (x^* \mid z) = \left[n \hat{f}(z) \right]^{-1} \sum_{t=1}^n K_b(Z_t - z) I(X_t^* \leq x^*) \quad (5.1.22)$$

Where $I(\cdot)$ is the indicator function which is 1 if the condition $X_t^* \leq x^*$ holds and 0 otherwise. The estimation of (5.1.22) and it's inverse are the same as in chapter 3.1.

5.1.4 Quantile error estimation

In chapter 4, we showed the asymptotic properties of the conditional scale function estimate through inversion of the conditional CCDF as in (5.1.22) with the assumption

that the quantile location shift α_τ is zero. The properties for the QAR estimate are the same given that the two CCDFs in (5.1.9) and (5.1.22) differ respectively in indicator parts $I(X_t \leq x)$ and $I(X_t^* \leq x^*)$ only. Thus, assuming we have estimated the two components using the prediction method, the quantile error η_τ can be estimate as

$$\hat{\eta}_\tau = \frac{X_t - \tilde{\alpha}_\tau(Z_t)}{\tilde{\omega}_\tau(Z_t)} \quad (5.1.23)$$

and should verify the conditions (5.1.4) and (5.1.18). Moreover, if the conditions hold, then the estimators are accurate. From our simulation, the estimations seem to be accurate for quantile $\tau \geq 0.75$ (see Table 5.2.1).

5.2 Simulation study

The data of size $n = 1000$ was simulated from the model $X_t = 0.5 + 0.3X_{t-1} + \sqrt{1 + 0.35X_{t-1}^2}e_t$ with e_t generated from a student t -distribution with 4 degrees of freedom. The figure 5.2.1 represents the superposition of the process and the estimated $\tilde{\alpha}_\tau(z)$ using the k NN prediction method. In fact, the non-parametric estimation of $\hat{\alpha}_\tau(z)$ was first carried out using the smoothed estimator along with the outliers detection using box-plot fences in order to correct the boundary issue (see Chapter 5). The comparison between $\hat{\alpha}_\tau(z)$ and the predicted $\tilde{\alpha}_\tau(z)$ for bins z is represented by Figure 5.2.2. Note that the prediction error in (5.1.14) was evaluated to 10^{-6} and the Figure 5.2.2 illustrates it as well. The outliers detection technique and prediction give less weight to extreme points that are not considered in the first estimation, then are re-involved in the prediction. This made our estimations less sensitive to the boundaries (see Figure 5.2.2).

Table 5.2.1: Summary of quantile errors

τ	Min.	1st Qu.	Med	Mean	3rd Qu.	Max.	$\Pr(\eta_\tau \leq 0)$	$\Pr(\eta_\tau^* \leq 1)$
0.25	-47.18	0.03	3.00	3.53	6.48	61.04	0.25	0.42
0.50	-16.78	-1.46	0.02	0.17	1.61	22.99	0.50	0.62
0.75	-16.61	-2.86	-1.49	-1.32	0.06	21.53	0.74	0.74
0.90	-16.62	-2.79	-1.95	-1.92	-1.09	12.96	0.89	0.96

$$\eta_\tau^* = \gamma_\tau(\eta_\tau).$$

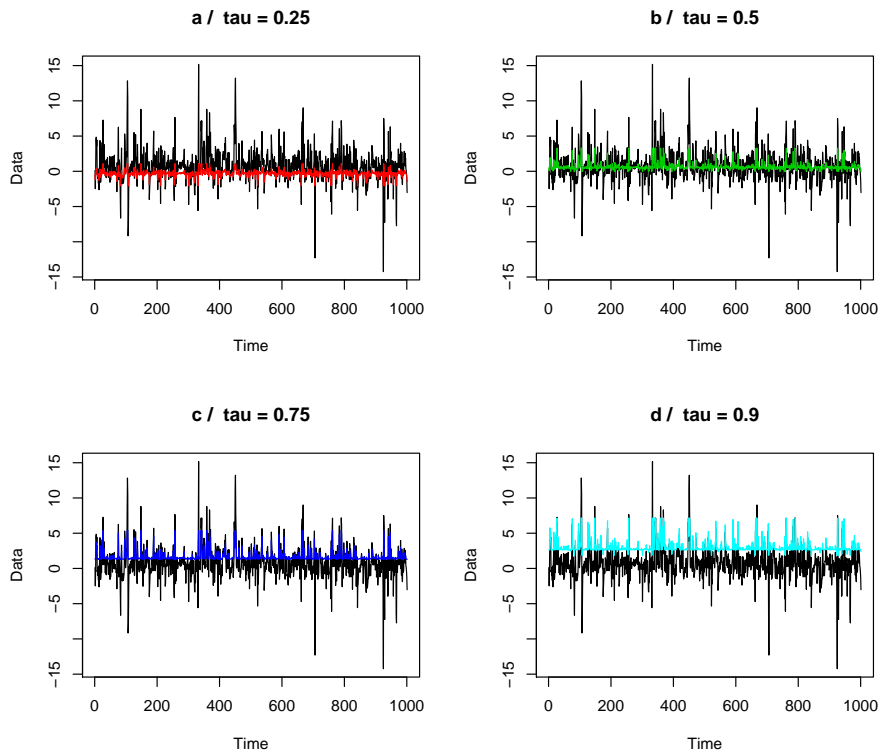


Figure 5.2.1: Predicted conditional quantile returns

On the graphs above, we see the smoothed estimation of the QAR for a given data set from an AR(1)-ARCH(1) process. The fact that we are able to estimate the daily conditional returns from previous values, help here to calculate the smoothed QAR for every point

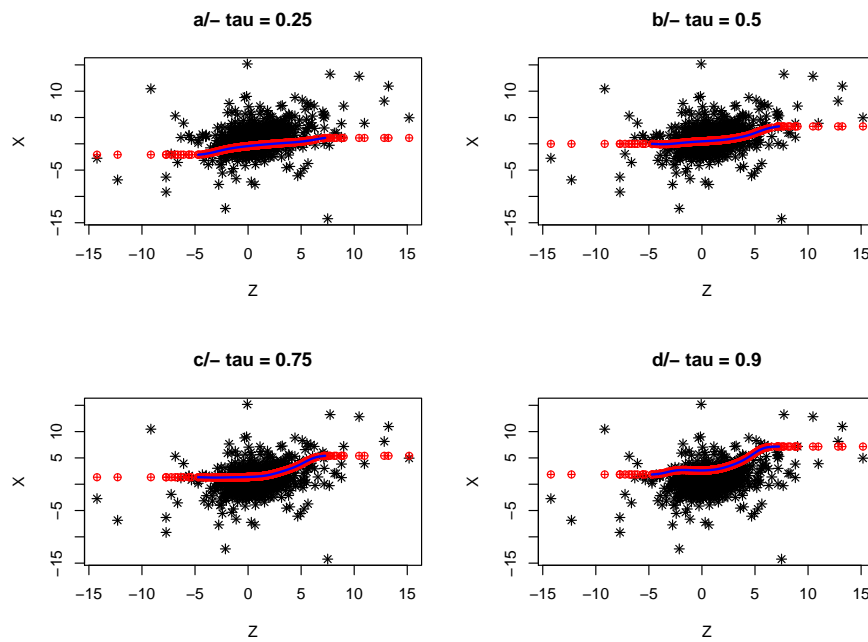


Figure 5.2.2: Graphical superposition of $\tilde{\alpha}_\tau(z)$ [red points] and $\hat{\alpha}_\tau(z)$ [blue curve]

The figures above represent the estimations of smoothed QAR for proportions 0.25, 0.50, 0.75 and 0.90. Given an AR(1)-ARCH(1) time series, we are able to calculate today's return based on the yesterday's at proportion $\tau \in (0, 1)$. This feature describe the response variable but not totally because we need to know the conditional variation at the given proportion τ . That's why the smoothed estimation of the CSF is necessary.

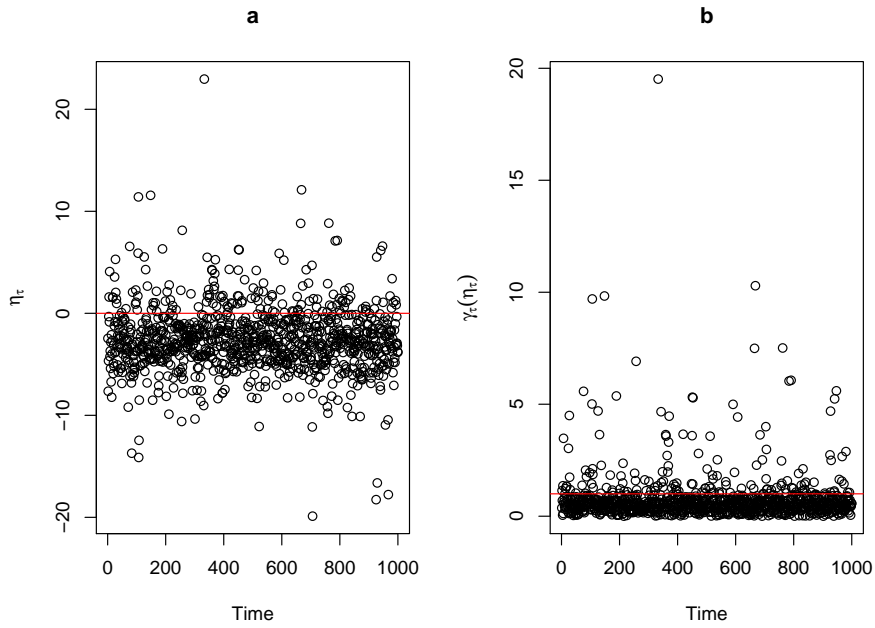


Figure 5.2.3: Quantile errors for $\tau = 0.75$

5.3 Accuracy of the estimates

The calculation of the MASE as in (4.3.23) for QAR and QARCH are recapitulated in the following table

Table 5.3.1: MASE for QAR and QARCH at $\tau = 0.25$

n	rough $\hat{\alpha}_{0.25}$	smooth $\hat{\alpha}_{0.25}$	rough $\hat{\omega}_{0.25}$	smooth $\hat{\omega}_{0.25}$
250	1.13482	0.03078	0.03457	0.00075
500	0.94149	0.04128	0.04916	0.00075
1000	1.22881	0.00671	0.15645	0.00115

Table 5.3.2: MASE MASE for QAR and QARCH at $\tau = 0.50$ (median)

n	rough $\hat{\alpha}_{0.50}$	smooth $\hat{\alpha}_{0.50}$	rough $\hat{\omega}_{0.50}$	smooth $\hat{\omega}_{0.50}$
250	0.6184	0.01963	0.08938	0.00401
500	1.21301	0.00873	0.3448	0.00526
1000	1.54507	0.0091	0.3595	0.00816

Table 5.3.3: MASE MASE for QAR and QARCH at $\tau = 0.75$

n	rough $\hat{\alpha}_{0.75}$	smooth $\hat{\alpha}_{0.75}$	rough $\hat{\omega}_{0.75}$	smooth $\hat{\omega}_{0.75}$
250	1.88628	0.03351	1.36976	0.0126
500	0.39451	0.02664	0.69214	0.02616
1000	1.28018	0.01356	1.21384	0.02367

Table 5.3.4: MASE MASE for QAR and QARCH at $\tau = 0.90$

n	rough $\hat{\alpha}_{0.90}$	smooth $\hat{\alpha}_{0.90}$	rough $\hat{\omega}_{0.90}$	smooth $\hat{\omega}_{0.90}$
250	0.66136	0.222	0.62655	0.17793
500	0.99836	0.12574	1.09674	0.27295
1000	1.76097	0.07349	1.47431	0.1794

From the tables we see that the MASE for smooth estimations are the smallest. This means the smoothed estimations are more accurate than the "rough" ones. This results is very significant for further researches that involve conditional quantile estimations.

5.4 Conclusion

In this chapter, the problem of estimating the conditional scale function when the autoregressive part is not zero was addressed by first estimating the smoothed estimate of the QAR, then, we used the k NN prediction to adjust the dimension in order to calculate

the residuals from which the smoothed CSF was estimated. The showed that the soothed estimates are accurate compared to the rough ones.

Conclusion and recommendations

In this thesis, was derived an estimator of the Conditional Scale Function in two different cases. The first is the one assuming the QAR function is zero and the second is the one that requires the estimation of the QAR in the first place. This was possible by inverting the CDF. We estimated the matrix of the Conditional Cumulative Distribution Function and then its inverse. This approach is a first in the field of conditional quantile estimation. The consistency of our estimates has been proven according to specific assumptions. Monte Carlo studies were performed in order to show the performance of our estimators. An application on real data could have been a very good application for our theories but since it already works on simulations, we hope that these results will be of great use. The method of Nadaraya - Watson is according to the literature very influenced by very distant data whereas more researches could be based on very robust methods like local polynomial of degree greater than equal to two (2) which is more robust than the former. This may avoid using the outliers detection method. The estimations can be improved using bootstrapping methods.

References

- Abberger, K. (1997). Quantile smoothing in financial time series. *Statistical Papers*, 38(2):125–148.
- Avramidis, P. (2016). Adaptive likelihood estimator of conditional variance function. *Journal of Non-parametric Statistics*, 28(1):132–151.
- Baudat, G. and Anouar, F. (2001). Kernel-based methods and function approximation. In *Neural Networks, 2001. Proceedings. IJCNN'01. International Joint Conference on*, volume 2, pages 1244–1249. IEEE.
- Bollerslev, T., Engle, R. F., and Nelson, D. B. (1994). Arch models. *Handbook of econometrics*, 4:2959–3038.
- Bosq, D. (2012). *Nonparametric statistics for stochastic processes: estimation and prediction*, volume 110. Springer Science & Business Media.
- Caires, S. and Ferreira, J. A. (2005). On the non-parametric prediction of conditionally stationary sequences. *Statistical inference for stochastic processes*, 8(2):151–184.
- Chen, S. (2015). Optimal bandwidth selection for kernel density functionals estimation. *Journal of Probability and Statistics*, 2015.
- Chen, Z., Kulperger, R., and Jiang, L. (2003). Jensen’s inequality for g-expectation: part 1. *Comptes Rendus Mathématique*, 337(11):725–730.
- Cleveland, W. S. (1981). Lowess: A program for smoothing scatterplots by robust locally weighted regression. *The American Statistician*, 35(1):54–54.
- Cowling, A. and Hall, P. (1996). On pseudodata methods for removing boundary effects in kernel density estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 551–563.

- Das, S. and Politis, D. N. (2017). Nonparametric estimation of the conditional distribution at regression boundary points. *arXiv preprint arXiv:1704.00674*.
- 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.
- Franke, J. and Mwita, P. N. (2003). Nonparametric estimates for conditional quantiles of time series.
- Franke, J., Mwita, P. N., and Wang, W. (2015). Nonparametric estimates for conditional quantiles of time series. *AStA Advances in Statistical Analysis*, 99(1):107–130.
- Givord, P. and Dhaultfoeuille, X. (2013). La régression quantile en pratique. Technical report, Institut National de la Statistique et des Etudes Economiques.
- Guidoum, A. C. (2013). Kernel estimator and bandwidth selection for density and its derivatives.
- Hall, P., Peng, L., and Yao, Q. (2002). Prediction and nonparametric estimation for time series with heavy tails. *Journal of Time Series Analysis*, 23(3):313–331.
- 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.
- Hansen, B. E. (2004). Nonparametric estimation of smooth conditional distributions. *Unpublished paper: Department of Economics, University of Wisconsin*.
- Härdle, W., Lütkepohl, H., and Chen, R. (1997). A review of nonparametric time series analysis. *International Statistical Review*, 65(1):49–72.
- Härdle, W., Werwatz, A., Müller, M., and Sperlich, S. (2004). Nonparametric density estimation. *Nonparametric and Semiparametric Models*, pages 39–83.
- Karunamuni, R. J. and Alberts, T. (2005). On boundary correction in kernel density estimation. *Statistical Methodology*, 2(3):191–212.
- Koenker, R. (2001). Quantile regression. *J. of Economics Perspectives*.
- Koenker, R. (2005). *Quantile regression*. Number 38. Cambridge university press.
- Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica: journal of the Econometric Society*, pages 33–50.

- Koenker, R. and Hallock, K. F. (2001). Quantile regression. *Journal of economic perspectives*, 15(4):143–156.
- Koenker, R. and Zhao, Q. (1996). Conditional quantile estimation and inference for arch models. *Econometric Theory*, 12(5):793–813.
- Markovich, N. (2008). *Nonparametric analysis of univariate heavy-tailed data: research and practice*, volume 753. John Wiley & Sons.
- Min, J. H. and Lee, Y.-C. (2005). Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. *Expert systems with applications*, 28(4):603–614.
- Mwita, P. N. (2003). *Semiparametric estimation of conditional quantiles for time series with applications in finance*. PhD thesis, University of Kaiserslautern.
- Mwita, P. N. (2004). On conditional scale function: Estimate and asymptotic properties.
- 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.
- Mwita, P. N. and Otieno, R. O. (2005). Conditional scale function estimate in the presence of unknown conditional quantile function. *African Journal of Science and Technology*, 6(1).
- Mynbaev, K. and Martins-Filho, C. (2010). Bias reduction in kernel density estimation via lipschitz condition. *Journal of Nonparametric Statistics*, 22(2):219–235.
- Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability & Its Applications*, 9(1):141–142.
- Portnoy, S. and Koenker, R. (1997). The gaussian hare and the laplacian tortoise: computability of squared-error versus absolute-error estimators. *Statistical Science*, 12(4):279–300.
- 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.
- Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, pages 832–837.
- Seltman, H. (2012). Approximations for mean and variance of a ratio. *unpublished note*.

- Shephard, N. (1996). Statistical aspects of arch and stochastic volatility. *Monographs on Statistics and Applied Probability*, 65:1–68.
- Silverman, B. W. (1986). *Density estimation for statistics and data analysis*, volume 26. CRC press.
- 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.
- Tukey, J. W. (1977). Exploratory data analysis.
- Watson, G. S. (1964). Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 359–372.
- Weiss, A. A. (1984). Arma models with arch errors. *Journal of time series analysis*, 5(2):129–143.
- Welsh, A. (1986). Bahadur representations for robust scale estimators based on regression residuals. *The Annals of Statistics*, pages 1246–1251.
- Zambom, A. Z. and Dias, R. (2012). A review of kernel density estimation with applications to econometrics. *arXiv preprint arXiv:1212.2812*.

Appendix

R codes

Data Generating process

```
#-----  
# Simulation of processes with heteroscedasticity  
#-----  
# size: (output) data length  
#-----  
# (ar.coef = c(.01, .3), arch.coef = c(.1, .1, .6), size = 1001)  
  
if(!require(kedd)) install.packages("kedd")  
  
ArGarch <- function(ar.coef = c(.5, .3), arch.coef = c(.1, .1, .75),  
                   size = 1001){  
  if (sum(arch.coef[2:3]) >= 1){  
    stop(  
      "The process may not be stationary because the sum of coefficients ",  
      arch.coef[2], " and ", arch.coef[3],  
      " for GARCH(1,1) is greater than or equal to 1"  
    )  
  }  
  m <- size*3  
  varpi <- runif(1)
```

```

x <- 1
for(i in 2:m){
  varpi[i] <- sqrt(arch.coef[1] + arch.coef[2]*x[i-1]^2 +
                 arch.coef[3]*varpi[i-1]^2)
  x[i] <- ar.coef[1] + ar.coef[2]*x[i-1] + varpi[i]*rt(1, df = 3)
}
Data <- tail(x, size)
return(Data)
}

# Simulating an AR(1)-ARCH(1) process
ar1arch1 <- function(ar.coef = c(.5, .3), arch.coef = c(.1, .35),
                    size = 1001){
  m <- size*2
  e <- rnorm(m)
  varpi <- abs(runif(1))
  x <- runif(1)
  for(i in 2:m){
    varpi[i] <- sqrt(arch.coef[1] + arch.coef[2]*x[i-1]^2)
    x[i] <- ar.coef[1] + ar.coef[2]*x[i-1] + varpi[i]*e[i-1]
  }
  Data <- tail(x, size)
  return(Data)
}

# Testing the parameters
# set.seed(1)
size <- 200
Data <- list(
  Data1 = ar1arch1(c(.5, 0.25), arch.coef = c(1, 0.35), size),
  Data2 = ar1arch1(c(.5, -0.75), arch.coef = c(1, 0.5), size),
  Data3 = ar1arch1(c(.5, 0.95), arch.coef = c(1, 1.20), size),
  Data4 = ar1arch1(c(.5, 1), arch.coef = c(1, 1), size)
)

```



```

)

for (i in 1:length(Data)) {
  pdf(paste(names(Data[i]),"pdf", sep = "."), width = 8,
      height = 8)
  par(mfrow = c(3,1))
  ts.plot(Data[[i]], main = NULL, ylab = names(Data[i]))
  acf(Data[[i]], main = "")
  pacf(Data[[i]], main = "")
  dev.off()
}

## Other way of simulating a AR(1)-ARCH(1) process
# m function: conditional mean
m <- function(x) sin(0.5*x)

# h function: conditional variance
h <- function(x){
  # Definition of h1
  h1 <- 1+0.01*x^2+0.5*sin(x)
  # h2 <- 1-0.9*exp(-2*x^2)
  h2 <- 0

  return(h1+h2)
}

# Generating data and deleting the first 1000 points
processGen <- function(n, mfun = m, hfun = h, delete = 1000){
  n <- n+delete
  e <- rt(n, df=3)

  # seed
  y = 0

```

```

for (t in 2:n) {
  y[t] <- m(y[t-1]) + sqrt(h(y[t-1]))*e[t-1]
}
y <- y[-(1:delete)]
return(y)
}

```

Check-function γ_T

```

## Check-function
checkfun <- function(x, tau) return(x*(tau - 1*(x <= 0)))

```

Outliers detection

```

# Code for outliers detection. The code returns the indices of
# outliers in a sequence
extremes <- function(x, range = 2.5){
  nx <- length(x)

  q1 <- quantile(x, 0.25)
  q3 <- quantile(x, 0.75)
  iqr <- q3 - q1

  # Extrem outlier are points outside
  lower <- (q1 - range)*iqr
  upper <- (q3 + range)*iqr

  # Indices
  id <- which(x < lower | x > upper);

  return(id)
}

```

```

rmextrems <- function(x, which.vect = 1, range = 3){

  if(!is.data.frame(x)) stop(
    "The input should be a data frame or a matrix of order (n x 2)"
  )

  id <- extremes(x[,which.vect], range)
  rest <- x[-id, ]

  freq.rm <- length(id)*100/nrow(x)
  cat("Note: ", round(freq.rm,1), "% of your data is deleted \n")
  return(rest)
}

```

Nadaray-Watson smoother

```

#-----
# Smoothing the rough estimator from the inverse of the CCDF
#-----
# x @matrix of order (N, 2).
#       Y (exogenous)
qsmooth <- function(x, n = 512L, bw = h.amise(x[,1])$h){
  if(!(is.data.frame(x))) stop(
    "The input should be a data.frame or a matrix
    of order (n x 2)"
  )
  N <- nrow(x)

  # Calculate weights
  FUN1 <- function(x, z) {
    Mat <- t(sapply(x, FUN = function(m) m - z))

```

```

    return(Mat)
}
z <- x[,1]
Knorm <- function(u) (2*pi)^(-.5) * exp(-0.5 * u^2)

M <- Knorm(FUN1(z, x[,1])/bw)/bw
a <- as.numeric(M%*%x[,2])
D <- rowMeans(M)

estim <- cbind(x = z , y = a/(N*D))

return(estim)
}

```

Conditional quantile function calculator

```

# -----
# THE CDF CALCULATION and CSF estimation
# -----

np.qarch2 <- function(x, n = 512, tau,
                      Kernel = "gaussian",
                      rm.outliers = TRUE,
                      bw = h.amise(x[,1],
                                   kernel = Kernel)$h){

# Backup
original.data <- x

# Check if x is a data.frame of order (n x 2)
if(!is.data.frame(x)) stop(
  "The input should be a data.frame or a matrix of
  order (n x 2)"
)

```

```

)

# Check if the level entered is between 0 and 1
if (tau < 0 | tau > 1) stop("Choose the level tau such
                             that : 0 < tau < 1")

if(rm.outliers == TRUE){
  # Outliers detection and removal
  x <- x[-extrems(x[,1], 2), ]
}

x[,2] <- checkfun(x[,2], tau = tau)

## Different Kernels functions
### - Gaussian kernel
Knorm <- function(u) (2*pi)^(-.5) * exp(-0.5 * u^2)
### - Epanechnikov kernel
Kepan <- function(u) 0.75 * (1 - u^2) * (abs(u) <= 1)
### - Uniform kernel
Kunif <- function(u) (abs(u) < 1)*0.5
### - Triangle kernel
Ktria <- function(u) (1-abs(u))*(abs(u) < 1)
### - Biweight
Kbiwe <- function(u) 0.9375*((1 - u^2)^2*(abs(u) < 1))
### - Cosine
Kcosi <- function(u) pi*cos(pi*u/2)*(abs(u) < 1)/4

if (Kernel == "gaussian"){K <- Knorm ;
} else if (Kernel == "epanechnikov"){K <- Kepan;
} else if (Kernel == "uniform"){K <- Kunif;
} else if (Kernel == "triangle"){K <- Ktria;
} else if (Kernel == "biweight"){K <- Kbiwe;
} else if (Kernel == "cosine"){K <- Kcosi;

```

```

} else {K <- Kepan}

# xx <- seq(min(Z) - bw, max(Z) + bw, le = n)
xx <- seq(from = min(x[,1]), to = max(x[,1]), le = n)

# y <- seq(min(X) - 3*bw, max(X) + 3*bw, le = n)
y <- x[,2]

## substractions
FUN1 <- function(x, z) {
  Mat <- t(sapply(x, FUN = function(m) m - z))
  return(Mat)
}

## Test function
FUN2 <- function(x, test){
  Mat2 <- sapply(test, FUN = function(m) 1*(x <= m))
  return(Mat2)
}

## Matrix of xi - Zt
M <- FUN1(xx, x[,1])
MatKern <- K(M/bw)/bw

## The pdf estimate of Z
ngz <- rowSums(MatKern)
gz <- rowMeans(MatKern)

## Calculation of the numerator of F(Y|X)
MatCond <- FUN2(y, y)

Fxy <- (MatKern%*%MatCond)/ngz

```

```

# Finding the y for which Fxy >= tau
csf <- apply(Fxy, 1, FUN = function(m){
  min(x[,2][which(m >= tau)])
})

estim <- data.frame(x = xx, csf = csf);

## Smoothed version (mine)
csfs <- qsmooth(estim)

# Percentage of deleted data
freq.outliers <- round(100 - nrow(x)/nrow(original.data)*100,2)

# Message to be displayed
cat("Percentage of outliers (deleted):\t\t\t",
    freq.outliers, "%\n")
cat("Optimal bandwidth for ", Kernel ,
    " density estimation:\t", bw, "\n")

List <- list(
  "Transformed" = data.frame(Z = original.data[,1],
                            X = checkfun(original.data[,2],
                                          tau = tau)),
  "Trimed" = x,
  "csf" = estim,
  "smooth.csf" = csfs
)
return(List)
}

```

Predicting and correcting dimension: k-NN method

```
#-----  
# k-NN version CQR  
#-----  
# new @ is the futur value (can be a vector)  
# result @ is the nx2 data frame from quantiilestim2.R  
  
# Function that calculate the distance between new  
# and existing values:  
  
abs.dist <- function(x,y) sum(abs(x-y))  
distL2 <- function(x,y) mean((x-y)^2)  
  
knn.cqr <- function(k, New, result, distance = distL2){  
  n.New <- length(New)  
  pred <- numeric()  
  
  for (i in 1:n.New) {  
    # Obtaining the KNN of new  
    d <- sapply(result[,1], function(a) distance(New[i], a))  
  
    # Selecting the k nearest in distance  
    ind.knn <- order(d)[1:k]  
  
    # Prediction for new  
    pred[i] <- mean(result[ind.knn,2])  
  }  
  
  return(pred)  
}
```


Mean Average Squared Error

```
## MASE to show the accuracy of the estimator

MASE <- function(ss = c(250, 500, 1000), n.iter = 10,
                 tau = tau, Kernel = "gaussian"){

  errors <- matrix(NA, nrow = n.iter, ncol = 2)
  colnames(errors) <- c("csf", "smooth.csf")

  Errors <- matrix(NA, nrow = length(ss), ncol = 2)
  rownames(Errors) <- paste("n = ", ss, sep = '')

  for(k in 1:length(ss)){
    for(i in 1:n.iter){
      a = np.qarch2(transf.data(ar1arch1(size = ss[k])),
                   tau = tau, Kernel = Kernel,
                   rm.outliers = T)
      b = np.qarch2(transf.data(ar1arch1(size = ss[k])),
                   tau = tau, Kernel = Kernel)
      errors[i,] <- colMeans(data.frame(csf = a$csf[,2],
                                       smooth.csf =
                                         a$smooth.csf[,2]) -
                           data.frame(csf = b$csf[,2],
                                       smooth.csf =
                                         b$smooth.csf[,2])
                           )
    }

    Errors[k,] <- round(colMeans(errors),5)
  }
  colnames(Errors) <- c("csf", "smooth.csf")
  write.csv(Errors, paste("mase", Kernel, tau, n.iter,
```

```

        ".csv", sep = '-')
    return(Errors)
}

```

Estimation of the CSF assuming $QAR = 0$

```

# Level tau in (0,1)
tau = 0.75
set.seed(1111)
Data <- ar1arch1()

w = 8; he = 6 # for pdf(): save graphs in pdf

pdf("process.pdf", width = w, height = he)
# par(mfrow = c(2,2))
ts.plot(Data, main = "ar(1)-arch(1) process")
# acf(Data); pacf(Data); qqnorm(Data); qqline(Data)
dev.off()

#1.3 Data splitting
transf.data <- function(x){
  x <- data.frame("Z" = x[2:length(x)], "X" = x[1:(length(x)-1)])
  return(x)
}

# Estimation without outliers detection
pdf("curvewithouts.pdf", width = w, height = he)
aa1 <- np.qarch2(transf.data(Data), tau = tau, rm.outliers = F,
  Kernel = "epanechnikov")
plot(aa1$Transformed, main = "CSF estimation with outliers",
  xlab = 'Z', ylab = expression(X^'*'), pch = 8, col = 'grey') #
lines(aa1$csf, col = 'red', lwd = 2)
dev.off()

```

```

# Boundary correction
aa1 <- np.qarch2(transf.data(Data), tau = tau,
                rm.outliers = TRUE, Kernel = "epanechnikov")
pdf("curvewithouts1.pdf", width = w, height = he)
plot(aa1$Transformed, pch = 8, col = 'grey')
points(aa1$Trimmed, main = "CSF estimation with outliers", xlab = 'Z',
       ylab = expression(X^'*'), pch = 8) #
lines(aa1$smooth.csf, col = 'red', lwd = 2)
dev.off()

# Try to plot at different levels on the same graph

Levels <- c(.25, 0.5, 0.75)

pdf("curves.pdf", width = w, height = he)
B <- np.qarch2(transf.data(Data), tau = 0.9)
plot(B$csf, main = "CSF estimations", typ = 'l', xlab = 'Z',
     xlim = c(min(B$Trimmed$Z), max(B$Trimmed$Z)),
     ylim = c(min(B$Trimmed$X), max(B$Trimmed$X)),
     ylab = expression(X^'*'), col = 1, lty = 1, lwd = 2) #
for(i in 1:length(Levels)){
  C <- np.qarch2(transf.data(Data), tau = Levels[i])
  lines(C$csf, pch = 8, lty = i+1, lwd = 2)
}
dev.off()

# Application of our smoothing method
pdf("comparisonofcurves.pdf", width = w, height = he)
C <- np.qarch2(transf.data(Data), tau = tau)
plot(C$csf, main = "Comparison with the smmothed version",
     typ = 'l', xlab = 'Z', xlim = c(min(C$Trimmed$Z),
                                     max(C$Trimmed$Z)),

```

```

ylim = c(min(C$Trimed$X), max(C$Trimed$X)),
ylab = expression(X^'*'), col = 1, lty = 1, lwd = 2) #
lines(C$smooth.csf, pch = 8, col = "red")
dev.off()

```

```
## Application
```

```

n.iter = 10
# MASE(tau = tau, Kernel = "gaussian", n.iter = n.iter)
# MASE(tau = tau, Kernel = "epanechnikov", n.iter = n.iter)
# MASE(tau = tau, Kernel = "triweight", n.iter = n.iter)

```

Estimation of the CSF when the QAR is unknown

```
## Function that calculate QAR
```

```
#----- THE CDF CALCULATION and CSF estimation -----
```

```

np.qar <- function(x, n = 512, tau, Kernel = "gaussian",
                  rm.outliers = TRUE, range = 2){

# Backup
original.data <- x

# Check if x is a data.frame of order (n x 2)
if(!is.data.frame(x)) stop("The input should be a data.frame
                           or a matrix of order (n x 2)")

# Check if the level entered is between 0 and 1
if (tau < 0 | tau > 1) stop("Choose the level tau such that :
                           0 < tau < 1")

if(rm.outliers == TRUE){

```

```

# Outliers detection and removal
x <- x[-extrems(x[,1], range), ]
}

## Different Kernels functions
### - Gaussian kernel
Knorm <- function(u) (2*pi)^(-.5) * exp(-0.5 * u^2)
### - Epanechnikov kernel
Kegan <- function(u) 0.75 * (1 - u^2) * (abs(u) <= 1)
### - Uniform kernel
Kunif <- function(u) (abs(u) < 1)*0.5
### - Triangle kernel
Ktria <- function(u) (1-abs(u))*(abs(u) < 1)
### - Biweight
Kbiwe <- function(u) 0.9375*((1 - u^2)^2*(abs(u) < 1))
### - Cosine
Kcosi <- function(u) pi*cos(pi*u/2)*(abs(u) < 1)/4

if (Kernel == "gaussian"){K <- Knorm ;
} else if (Kernel == "epanechnikov"){K <- Kegan;
} else if (Kernel == "uniform"){K <- Kunif;
} else if (Kernel == "triangle"){K <- Ktria;
} else if (Kernel == "biweight"){K <- Kbiwe;
} else if (Kernel == "cosine"){K <- Kcosi;
} else {K <- Kegan}

# Smoothing parameter estimation
bw = h.amise(x[,1], kernel = Kernel)$h

# xx <- seq(min(Z) - bw, max(Z) + bw, le = n)
xx <- seq(from = min(x[,1]), to = max(x[,1]), le = n)

# y <- seq(min(X) - 3*bw, max(X) + 3*bw, le = n)

```

```

y <- x[,2]

## substractions (x - Z_t), t = 1,2, ..., n
FUN1 <- function(x, z) {
  Mat <- t(sapply(x, FUN = function(m) m - z))
  return(Mat)
}

## Test function
FUN2 <- function(x, test){
  Mat2 <- sapply(test, FUN = function(m) 1*(x <= m))
  return(Mat2)
}

## Matrix of xi - Zt
M <- FUN1(xx, x[,1])
MatKern <- K(M/bw)/bw

## The pdf estimate of Z
ngz <- rowSums(MatKern)
gz <- rowMeans(MatKern)

## Calculation of the numerator of F(Y|X)
MatCond <- FUN2(y, y)

Fxy <- (MatKern%*%MatCond)/ngz

## Finding the y for which Fxy >= tau
csf <- apply(Fxy, 1, FUN = function(m){
  min(x[,2][which(m >= tau)])
})

```

```

estim <- data.frame(x = xx, csf = csf);

## Smmothed version (mine)
csfs <- qsmooth(estim)

# Percentage of deleted data
freq.outliers <- round(100 - nrow(x)/nrow(original.data)*100,2)

# Message to be displayed
cat("Percentage of outliers (deleted):\t\t\t",
    freq.outliers, "%\n")
cat("Optimal bandwidth for ", Kernel ,
    " density estimation:\t", bw, "\n")

List <- data.frame('x' = xx, 'qar' = estim[,2],
                  "smooth.qar" = csfs[,2])
return(List)
}

#-----
# APPLICATION FOR np.qar.R
#-----

# set.seed(1111)
Data <- ar1arch1(arch.coef = c(1, 0.35), size = 1001)
# ts.plot(Data)

#1# Compute QAR using bins (NW method)
#2# Use the estimated points to predict the QAR for the
# whole data
tau.vect <- c(0.25, 0.50, 0.75, 0.9)

pdf("predictions.pdf", width = w, height = 7)

```

```

par(mfrow = c(2,2))
for (k in 1:length(tau.vect)) {
  result <- np.qar(transf.data(Data), tau = tau.vect[k])

  predict.data <- knn.cqr(10, New = Data,
                        result = result[,c(1,3)])

  ts.plot(Data, main = letters[k])
  lines(predict.data, col = (k+1))
}
dev.off()

#3# Using the results above to estimate the conditional variance
Z <- transf.data(Data)[,1]
X <- transf.data(Data)[,2]

result <- np.qar(transf.data(Data), tau = tau, rm.outliers = F,
                range = 2)
qAR.predicted <- knn.cqr(3, New = Z, result = result[,c(1,3)])
pred.bins <- knn.cqr(3, New = result$x, result = result[,c(1,3)])

# Precision

ep <- mean((result$smooth.qar-pred.bins)^2)

Res <- X-qAR.predicted
pdf("compredic.pdf", width = w, height = he)
plot(Z, X, pch = 8)
points(Z, qAR.predicted, col = 'red', pch = 10)
lines(result[, c(1,3)], col = 'blue', lwd = 2)
dev.off()

```



```

# # Data simulation
# set.seed(1125)
# Data <- ar1arch1(arch.coef = c(1, 0.35))
# tau.vect <- c(0.25, 0.50, 0.75, 0.9)
# ts.plot(Data)

Z <- transf.data(Data)[,1]
X <- transf.data(Data)[,2]

pdf("compredic1.pdf", width = w, height = he)
par(mfrow = c(2,2))
for (i in 1:length(tau.vect)) {
  result <- np.qar(transf.data(Data), tau = tau.vect[i],
                  rm.outliers = T, range = 2)
  qAR.predicted <- knn.cqr(3, New = Z, result = result[,c(1,3)])
  plot(transf.data(Data), main = paste(letters[i], "/- tau = ",
                                     tau.vect[i], sep = '),
       pch = 8)
  points(Z, qAR.predicted, col = 'red', pch = 10)
  lines(result[, c(1,3)], col = 'blue', lwd = 2)
}
dev.off()

# Use the result to
new.x <- data.frame(x = Z, y = Res)

# QARCH

result2 <- np.qarch2(new.x, tau = tau, rm.outliers = T)
qarch.predicted <- knn.cqr(10, New = Z,
                          result = result2$smooth.csf)

```

```

# Estimation of epsilon

eta <- (X - qAR.predicted)/qarch.predicted

# Calculate the quantile of eps at tau
# Calculate the probability that eps <= 0

mean(eta <= 0)
summary(eta)

mean(checkfun(eta, tau) <= 1)

pdf("error.pdf", width = w, height = he)
par(mfrow = c(1,2))

plot(qAR.predicted, eta, main = "a",
     xlab = expression(hat(alpha)[tau](z)),
     ylab = expression(eta[tau])); abline(h = 0, col = 'red')

plot(qAR.predicted, checkfun(eta, tau), main = "b",
     xlab = expression(hat(alpha)[tau](z)),
     ylab = expression(gamma[tau](eta[tau])))
abline(h = 1, col = 'red')
dev.off()

# Table for eta_\tau
eta <- c()
pvalue0 <- c()
pvalue1 <- c()
pvalue2 <- c()
for (tau in tau.vect){
  # Nonparametric estimation of the QAR

```

```

qar.hat <- np.qar(transf.data(Data), tau = tau,
                 rm.outliers = T, range = 2)

# Prediction for QAR
qar.tilde <- knn.cqr(10, New = Z, result = qar.hat[,c(1,3)])

# Nonparametric estimation of the QARCH
qarch.hat <- np.qarch2(new.x, tau = tau, rm.outliers = T)

# Prediction for QARCH
qarch.tilde <- knn.cqr(10, New = Z,
                     result = qarch.hat$smooth.csf)

# The quantile error at each level tau
new.eta <- (X - qar.tilde)/qarch.tilde
eta <- cbind(eta, new.eta)
# print(head((X - qar.tilde)/qarch.tilde))
pvalue0 <- c(pvalue0, mean(new.eta <= 0))
pvalue1 <- cbind(pvalue1, mean(checkfun(new.eta, tau = tau) <= 1))
pvalue2 <- c(pvalue2, mean(checkfun(new.eta, tau = tau) <= 1))
}

eta.summary <- t(apply(eta, 2, summary))
rownames(eta.summary) <- paste("\tau = ", tau.vect)

df.eta <- round(cbind(tau.vect, eta.summary, pvalue0, pvalue2),2)
write.csv(df.eta, "table_eta.csv", row.names = F)
print(df.eta)

```

Publications

Published paper

Seknewna, L. L., Mwita, P. N., & Muema, B. K. (2018). Smoothed Conditional Scale Function Estimation in AR (1)-ARCH (1) Processes. *Journal of Probability and Statistics*, 2018.

Research Article

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

Lema Logamou Seknewna ¹, Peter Mwita Nyamuhanga,² and Benjamin Kyalo Muema³

¹Department of Mathematics, Pan African University Institute for Basic Sciences, Technology and Innovation, P.O. Box 62000, Nairobi 00200, Kenya

²Department of Mathematics, Machakos University, P.O. Box 136, Machakos 90100, Kenya

³Department of Statistics and Actuarial Sciences, Jomo Kenyatta University of Agriculture and Technology, P.O. Box 62000, Nairobi 00200, Kenya

Correspondence should be addressed to Lema Logamou Seknewna; logamou.lema@students.jkuat.ac.ke

Received 27 November 2017; Revised 11 January 2018; Accepted 24 January 2018; Published 12 March 2018

Academic Editor: Aera Thavaneswaran

Copyright © 2018 Lema Logamou Seknewna et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The estimation of the Smoothed Conditional Scale Function for time series was taken out under the conditional heteroscedastic innovations by imitating the kernel smoothing in nonparametric QAR-QARCH scheme. The estimation was taken out based on the quantile regression methodology proposed by Koenker and Bassett. And the proof of the asymptotic properties of the Conditional Scale Function estimator for this type of process was given and its consistency was shown.

1. Introduction

Consider a Quantile Autoregressive model,

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

where $\alpha_\tau(Z_t)$ is the τ^{th} Conditional Quantile Function of X_t given Z_t and the innovation u_t is assumed to be independent and identically distributed with zero τ^{th} quantile and constant scale function; see [1]. A kernel estimator of $\alpha_\tau(Z_t)$ has been determined and its consistency is shown [2]. A bootstrap kernel estimator of $\alpha_\tau(Z_t)$ was determined and shown to be consistent [3]. This research will extend [3] by assuming that the innovations follow Quantile Autoregressive Conditional Heteroscedastic process similar to Autoregressive-Quantile Autoregressive Conditional Heteroscedastic process proposed in [1]:

$$X_t = \alpha_\tau(Z_t) + \hat{\omega}_\tau(Z_t) \varepsilon_t, \quad t = 1, 2, \dots, \quad (2)$$

where $\alpha_\tau(Z_t)$ is the conditional θ -quantile function of X_t given Z_t ; $\hat{\omega}_\tau(Z_t)$ is a conditional scale function at τ -level, and

ε_t is independent and identically distributed (i.i.d.) error with zero τ -quantile and unit scale. The function $\hat{\omega}_\tau(Z_t)$ can be expressed as

$$\hat{\omega}_\tau(Z_t) = \lambda \hat{\omega}(Z_t), \quad (3)$$

where $\hat{\omega}(Z_t)$ is the so-called volatility found in [4, 5] which are papers of reference on Engle's ARCH models among many others and λ is a positive constant depending on τ [see [6]]. An example of this kind of function is Autoregressive-Generalized Autoregressive Conditional Heteroscedastic AR(1)-GARCH(1,1),

$$X_t = \alpha_t + \hat{\omega}_t e_t, \quad t = 1, 2, \dots, \quad (4)$$

where $\alpha_t = \mu + \delta X_{t-1}$, $\hat{\omega}_t = \sqrt{w + \alpha X_{t-1}^2 + \beta \hat{\omega}_{t-1}^2}$, $\mu \in (-\infty, \infty)$, $|\delta| < 1$, $\beta > 0$, $\alpha > 0$, $w > 0$, $\alpha + \beta < 1$, and $e_t \sim$ i.i.d. with 0 mean and variance 1. Note that α_t may also be an ARMA (see [7]). The specifications for model (4) are given in Section 4.2.

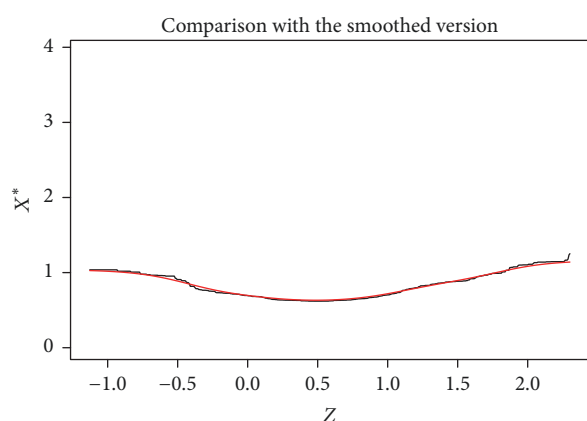


FIGURE 5: Smoothed estimate of the CSE.

Acknowledgments

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

References

- [1] P. Mwita, *Semiparametric estimation of conditional quantiles for time series with applications in finance [Ph.D. thesis]*, University of Kaiserslautern, 2003.
- [2] J. Franke, P. Mwita, and W. Wang, "Nonparametric estimates for conditional quantiles of time series," *AStA Advances in Statistical Analysis*, vol. 99, no. 1, pp. 107–130, 2014.
- [3] P. N. Mwita and J. Franke, "Bootstrap of kernel smoothing in quantile autoregression process," *Journal of Statistical and Econometric Methods*, vol. 2, no. 3, pp. 175–196, 2013.
- [4] T. Bollerslev, R. F. Engle, and D. B. Nelson, "Chapter 49 Arch models," *Handbook of Econometrics*, vol. 4, pp. 2959–3038, 1994.
- [5] N. Shephard, "Statistical aspects of arch and stochastic volatility," in *Monographs on Statistics and Applied Probability*, vol. 65, 1996.
- [6] P. Mwita and R. Otieno, "Conditional scale function estimate in the presence of unknown conditional quantile function," *African Journal of Science and Technology*, vol. 6, no. 1, 2005.
- [7] A. A. Weiss, "ARMA MODELS WITH ARCH ERRORS," *Journal of Time Series Analysis*, vol. 5, no. 2, pp. 129–143, 1984.
- [8] W. Härdle, H. Lütkepohl, and R. Chen, "A review of nonparametric time series analysis," *International Statistical Review*, vol. 65, no. 1, pp. 49–72, 1997.
- [9] E. A. Nadaraya, "Some new estimates for distribution function," *Theory of Probability and Its Applications*, vol. 9, pp. 497–500, 1964.
- [10] G. S. Watson, "Smooth regression analysis," *The Indian Journal of Statistics. Series A*, vol. 26, pp. 359–372, 1964.
- [11] P. Mwita, *On conditional scale function: Estimate and asymptotic properties*, 2004.
- [12] P. Hall, R. C. Wolff, and Q. Yao, "Methods for estimating a conditional distribution function," *Journal of the American Statistical Association*, vol. 94, no. 445, pp. 154–163, 1999.
- [13] K. Mynbaev and C. Martins-Filho, "Bias reduction in kernel density estimation via Lipschitz condition," *Journal of Nonparametric Statistics*, vol. 22, no. 1-2, pp. 219–235, 2010.
- [14] D. Bosq, *Nonparametric statistics for stochastic processes*, vol. 110, Springer Science & Business Media, New York, NY, USA, 1996.
- [15] R. Koenker and G. Bassett Jr., "Regression quantiles," *Econometrica*, vol. 46, no. 1, pp. 33–50, 1978.
- [16] R. F. Engle, "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," *Econometrica*, vol. 50, no. 4, pp. 987–1007, 1982.
- [17] W. Härdle, M. Müller, S. Sperlich, and A. Werwatz, *Nonparametric and Semiparametric Models*, Springer Series in Statistics, Springer-Verlag, New York, NY, USA, 2004.
- [18] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*, Chapman & Hall, London, UK, 1986.
- [19] B. E. Hansen, *Nonparametric estimation of smooth conditional distributions*, Department of Economics, University of Wisconsin, 2004, Unpublished paper.
- [20] S. Das and D. N. Politis, *Nonparametric estimation of the conditional distribution at regression boundary points*, 2017.
- [21] K. U. Steikert, *The Weighted Nadaraya-Watson Estimator: Strong Consistency to the Faculty [Ph.D. thesis]*, Business Administration and Information Technology of the University of Zurich, 2014.
- [22] W. S. Cleveland, "Lowess: A program for smoothing scatterplots by robust locally weighted regression," *The American Statistician*, vol. 35, no. 1, pp. 54–55, 1981.
- [23] Z. Chen, R. Kulperger, and L. Jiang, "Jensen's inequality for g-expectation: Part 1," *Comptes Rendus Mathématique*, vol. 337, no. 11, pp. 725–730, 2003.
- [24] H. Seltman, "Approximations for mean and variance of a ratio," *unpublished note*, 2012.
- [25] V. C. Raykar and R. Duraiswami, "Fast optimal bandwidth selection for kernel density estimation," in *Proceedings of the SIAM International Conference on Data Mining, '06*, pp. 524–528, SIAM, Philadelphia, PA.
- [26] S. Chen, "Optimal Bandwidth Selection for Kernel Density Functionals Estimation," *Journal of Probability and Statistics*, vol. 2015, Article ID 242683, 2015.
- [27] A. C. Guidoum, *Kernel estimator and bandwidth selection for density and its derivatives*, 2013, Kernel estimator and bandwidth selection for density and its derivatives.
- [28] R. J. Karunamuni and T. Alberts, "On boundary correction in kernel density estimation," *Statistical Methodology*, vol. 2, no. 3, pp. 191–212, 2005.
- [29] F. N. David and J. W. Tukey, "Exploratory Data Analysis," *Biometrics*, vol. 33, no. 4, p. 768, 1977.

Accepted manuscript

Title: Non-parametric estimation of conditional quantile functions
for AR(1)-ARCH(1) processes (1241107)

Journal: Open Journal of Statistics (OJS)