Semiparametric estimation of conditional quantiles for time series, with applications in finance

Peter N Mwita
Machakos University, Machakos, Kenya

Semiparametric estimation of conditional quantiles for time series, with applications in finance

Peter Nyamuhanga Mwita

Vom Fachbereich Mathematik
der Universität Kaiserslautern
zur Verleihung des akademischen Grades
Doktor der Naturwissenschaften
(Doctor rerum naturalium, Dr. rer. nat.)
genehmigte Dissertation.

1. Gutachter: Prof. Dr. Jürgen Franke
2. Gutachterin: Prof. Dr. Claudia Klüppelberg

Disputation:11 März 2003

D 386

Abstract

The estimation of conditional quantiles has become an increasingly important issue in insurance and financial risk management. The stylized facts of financial time series data has rendered direct applications of extreme value theory methodologies, in the estimation of extreme conditional quantiles, inappropriate. On the other hand, quantile regression based procedures work well in nonextreme parts of a given data but breaks down in extreme probability levels. In order to solve this problem, we combine nonparametric regressions for time series and extreme value theory approaches in the estimation of extreme conditional quantiles for financial time series. To do so, a class of time series models that is similar to nonparametric AR-(G)ARCH models but which does not depend on distributional and moments assumptions, is introduced. We discuss estimation procedures for the nonextreme levels using the models and consider the estimates obtained by inverting conditional distribution estimators and by direct estimation using Koenker-Basset (1978) version for kernels. Under some regularity conditions, the asymptotic normality and uniform convergence, with rates, of the conditional quantile estimator for α-mixing time series, are established. We study the estimation of scale function in the introduced models using similar procedures and show that under some regularity conditions, the scale estimate is weakly consistent and asymptotically normal. The application of introduced models in the estimation of extreme conditional quantiles is achieved by augmenting them with methods in extreme value theory. It is shown that the overal extreme conditional quantiles estimator is consistent. A Monte Carlo study is carried out to illustrate the good performance of the estimates and real data are used to demonstrate the estimation of Value-at-Risk and conditional expected shortfall in financial risk management and their multiperiod predictions discussed.

Acknowledgement

Best words can not express my sincere gratitude to my supervisor Jürgen Franke, who is also the head of statistics research group, not only for introducing nonparametric time series to me and for all his devotion and relentless support for this thesis but also for allowing me to join his group in the first place. I am very indebted to Andrzej S. Kozek of Macquarie Univeristy, Australia, for having played a pivotal role to my education and for introducing the quantile business to me. I am grateful to Roger Koenker of Illinois-Champaign, Gib Bassett, Jr. of Illinois-Chicago, for their fruitful conversations during various workshops and for continuously making their papers and other resources available to me. I have learned considerably from conversations with Ralf Korn, Olivier Scaillet, Markus Junker, Romanus Odhiambo and Jean-Pierre Stockis. The pleasant research atmosphere at the group is of course also the product of Beate Siegler.

I am grateful to DAAD and the Graduate college of the University of Kaiserslautern for financial support which enabled the completion of the work.

Finally I am most grateful to my family for having provided ideal environment to work and more so, to my beloved mother for having worked so hard to see me through my basic education. To her memory this thesis is dedicated.

Contents

List of Figures

List of Tables

1 Introduction

The increasing awareness in the financial industries (both private and regulators) of the consequences of extreme risks (the possibility of lossing large amount of money) in tradable portfolios has called for effective risk management systems to be put in place for financial institutions, such as banks and investment firms. This has seen the use of quantitative risk measures as essential management alternatives used for internal or external requirements parallel with other models.

Theoretically, one can quantify risk by using measures such as standard deviation, quantile, interquantile range or expected shortfall. The quantile based Value-at-Risk (henceforth VaR) has become a basic tool employed by financial institutions and their regulators to assess riskness of trading activities. It can formally be defined as the maximum potential change in value of a portfolio of financial instruments with a given probability over a certain horizon. Specifically, based on negative returns, VaR is defined so that the probability that a portfolio will lose more than its VaR over a particular time horizon is equal to $1-\varphi$, for the probability level $\varphi \to 1$ prespecified. Its popularity among financial practitioners stems from the fact that it is very simple: It can be used to summerize risk of individual positions or of large financial institutions such as dealer-banks in the OTC derivatives and other portfolios by reducing the (market) risk to just a dollar ammount[1], thereby representing a compromise between the needs of different users. Because of this simplicity, it has been adopted for regularity purposes. In particular, the 1996 market risk ammendments[2] to the Basel Accord stipulates that banks and broker-dealers minimum capital requirements for market risk should be set based on the ten-day 1% VaR for the trading portfolios. Detail analysis and application of this measure to risk management can be found in among others JP Morgan [73], Duffie and Pan [37], Jorion [74], Dowd [33], Stulz [106].

Despite its simplicity, the factor of its accurate measurement at high values of φ (e.g $\varphi > 0.95$) and subsequent monitoring of high risky activities has remained a challenging statistical problem. This is because VaR depends on the joint distribution of all intru-

[1] No matter how complex it is, a single value is provided as a summery.

[2] Which allows ten-day 1% VaR to be measured as a multiple of one-day 1% VaR.

ments in a portfolio whose changes are nonnormal[3] with some hidden information about market movements. The challenge has therefore been to find a suitable model of the extreme conditional time varying statistics for risk measurement that is able to adopt to general returns distribution and simultaneously reflect the latest information. To current, most literature has focused on the VaR from the marginal distribution, see for example Alexander and Leigh [5] and Boudoukh et al. [19].

The extreme quantiles can be estimated by using ideas from Extreme Value Theory (EVT). The use of EVT in financial market calculations is a fairly recent innovation, Embrechts et al. [39] surveys the mathematical theory of EVT and discusses its applications to both financial and insurance risk management. The EVT can be used to characterise the behaviour of the extreme returns or the tail of returns distribution without tying the analysis down to a single parametric family fitted to a whole distribution. However, because of the presence of stochastic volatility[4] in financial data, it is inappropriate to apply such models[5] directly. Furthermore Danielsson and de Vries [31] has shown that this model do not work well in the common low probabilities, such as 0.95. Very few attempts have been made to develop extensions of extreme value statistical methodology to take account of the variable volatility. Among others McNeil and Frey [88] and Barone-Adesi et al. [7] have taken an approach built arround the $GARCH$[6] with heavy tailed innovation estimated by EVT.

A seemingly flexible parametric approach to VaR estimation is being researched in Engle and Manganelli [42], where the estimation of VaR uses regression quantile methodology introduced by Koenker and Basset (1978) to determine the unknown parameters, under the assumption that the quantile process is correctly specified. In nonparametric set up, the estimation of quantiles with application to finance has been observed in Abberger [1]. However, due to the sparsity of data in high risk areas, the nonparametric kernel methods do not guaranree reliable description of the tails.

[3] Empirically, their peaks and tails are higher than normal, see Mendelbrot [83], Fama [43] and in the case of equity returns, the losses have longer tail than the profits.

[4] Changes in portfolio values have the characteristic of being significantly autocorrelated in their squares or absolutes i.e volatilities of market factors tend to cluster.

[5] These models are also nested in a framework of iid variables which is not consistent with the aforemened charactersistics.

[6] To take account of the underlying volatility.

In this thesis a semiparametric approach, to estimating conditional quantiles for time series in both common and extreme levels of $\varphi \in (0,1)$, that has simple structure, robust and tailored for general distributions is developed. It is based on the combination of three pillars: The nonparametric conditional quantile, based directly or indirectly[7] on local version of Koenker-Bassett (1978) methodology, constitute a flexible[8] part of our initial[9] estimator for fitting empirical changes. The second one models the randomly changing volatility as a scale function whose main purpose is to devolatize large observations (or losses) beyond an initial estimator. The third pillar, which is parametric in nature, exploits the results from the EVT and fits the transformed (devolatilized) excesses.

The rest of this chapter gives a general overview of the concepts of methodologies used in the thesis. We state explicitly what we want to estimate and provide a formula for that purpose. We then propose and define a class of time series models which is similar to nonparametric AR-(G)ARCH models but does not depend on the form of the conditional distribution and the finiteness of moments assumptions.

Chapter 2 derives the estimators for various nonparametric functions in the introduced process by inverting the estimates of conditional distributions. We provide pointwise consistency and asymptotic normality as well as uniform convergence of the conditional distribution and respective quantile estimators.

In chaper 3 we present and discuss various forms of regression based approaches for estimating, in particular, the purely heteroscedastic part of the introduced model. The chapter provides the asymptotic properties of the estimators based on direct Koenker and Basset (1978) version for kernels. We then give a standardization procedure for approximating and estimating the true volatility.

Chapter 4 uses results from EVT and nonparametric procedures, based on our model in chapters 2 and 3, to estimate the extreme conditional φ-quantile for time series in (1.2.0.1). Two formulae are derived: The first one is based on a Hill's estimator of shape parameter while the second one, on GPD. Heuristically, it is shown that both overall estimators of the extreme conditional φ-quantile function converges in probability to the true one. Further, a Monte Carlo and backtesting results based on artificial and real data

[7] By inverting conditional distribution function.

[8] In the sense that no strict distributional assumptions and variance specifications are made.

[9] This can be taken as conditional quantile at common probabitlites.

respectively, show that the introduced model argumented with EVT performs better than direct estimators for large levels of φ. We also confirm that the estimate based on GPD is superior than the one based on Hill's, for a wide range of large values of φ. The chapter also discusses the problem of multi-period prediction of VaR and derives a formula[10] based on α-root of time rule.

For completeness, chapter 5 extends the VaR formula based on our model to the case of coherence risk measure. We propose a more general formula for the conditional expected shortfall, for dependent data, that takes simple form in cases of a continuous distribution. Lastly we discuss the estimates of the formulae and their corresponding estimates of multi-period conditional expected shortfall and show heuristically that they converge in probability to the respective true ones.

[10] It is based on the tail of a Pareto distribution whose shape parameter is obtained by using the Hill's estimator.

1.1 Concepts and definitions

1.1.1 Econometric Model

Let $\left\{Y_t\right\}$ be real-valued and $\left\{\mathbf{F}_t, -\infty < t < \infty\right\}$ be an increasing sequence of σ-algebras representing information available at time t. We will assume that Y_t is $\mathbf{F}_t$-measurable. Let $\left\{\mathbf{X}_t\right\}$ be a d-dimensional process such that $\mathbf{X}_t$ is $\mathbf{F}_{t-1}$-measurable. In particular, we have a situation in mind where $\mathbf{X}_t = \left(Y_{t-1}, \ldots, Y_{t-d}\right)$ are the last observed returns up to time $t-1$, or where $\mathbf{X}_t$ consist of $\left(Y_{t-1}, \ldots, Y_{t-\tau}\right)$ and some exogeneous variable S_t which is $\mathbf{F}_{t-1}$-measurable and forms a $(d-\tau)$-dimensional time series. We will assume that the sequence of random variables $\left\{Y_t, \mathbf{X}_t\right\}$ taking values in $\mathbf{R} \times \mathbf{R}^d$ is stationary and that Y_t can be considered as the response variable and $\mathbf{X}_t$, the predictor variable (or the conditioning covariates). Further, we will assume that the underlying process of interest is of the form

$$Y_t = \mu_t + \sigma_t e_t, \quad t = 1, 2 \ldots, \tag{1.1.1.1}$$

where

1. μ_t is the conditional expectation function of Y_t given $\mathbf{F}_{t-1}$,
2. σ_t is the conditional volatility function of Y_t given $\mathbf{F}_{t-1}$,
3. and e_t are variables, independent of $\mathbf{F}_{t-1}$, with mean 0 and variance 1.

The conditional θ-quantile of (1.1.1.1) given $\mathbf{F}_{t-1}$ is then given by

$$\mu_{t,\theta} = \mu_t + \sigma_t F_e^{-1}\left(\theta\right) \tag{1.1.1.2}$$

where $F_e^{-1}\left(\theta\right)$ is the θ-quantile of e_t and $\theta \in (0,1)$. For instance, let e_t be independent and identically distributed (iid) standard normal random variable (rv), then

$$Y_t \sim N\left(\mu_t, \sigma_t^2\right), \quad t = 1, 2 \ldots, \tag{1.1.1.3}$$

If the time series of μ_t and σ_t^2 are known, the conditional θ-quantile is then given by

$$\mu_{t,\theta} = \mu_t + \sigma_t \Phi^{-1}\Big(\theta\Big)$$

where $F_e^{-1}\Big(\theta\Big) = \Phi^{-1}\Big(\theta\Big)$ is the inverse of the standard normal distribution at θ. This holds analogously for other known distributions. In practice, however, the misspecification of time series structural functions and strict distributional assumptions imposed on the standardized residuals can lead to serious under or overestimation of conditional quantiles.

In order to avoid such specification and distributional assumptions, a variety of nonparametric approaches for quantile estimation are available: The Historical simulation (HS) and the Hybride methods. In HS a random sample of size n, say from (1.1.1.1), is split up into a number of equally long (overlapping) subsamples of length,k say, called a rolling window (or window size). Then $n - k + 1$ subsamples are constructed such that for any two subsamples, there is all but one datum in common. Next, the θ^{th}-percentile of each subsample is picked[11] as the θ-quantile. A major[12] problem of HS is the rareness of extreme observations. In the interior the sampling observations are very close to each other and therefore the empirical quantile function is reasonably smooth, i.e does not show jumps. But as one goes to the extreme, the distance between adjacent observations becomes large. This results in quantile estimates that are highly variable and therefore either underpredict or overpredict. The blue curves in figure (1) illustrate this feature. The returns data is represented in black, while the continuous and dotted blue curves represents the 0.75 and 0.99-quantile esimates respectively.

The second approach is a variation of HS proposed in Boudoukh et al. [19] that combines RiskMetrics approach with HS methodologies by applying exponentially declining weights to the past observations. First, each of the most recent k observations, $y_t, y_{t-1}, \ldots, y_{t-k+1}$ is associated a weight $\frac{1-\lambda}{1-\lambda^k}, \Big(\frac{1-\lambda}{1-\lambda^k}\Big)\lambda, \ldots, \Big(\frac{1-\lambda}{1-\lambda^k}\Big)\lambda^{k-1}$ respectively[13] and then the observations are ordered in ascending order. The corresponding weights are then accumulated starting with the smallest observation until the $100\theta\%$ is reached. The θ-quantile of

[11] For intermediate percentile, a linear interpolation is performed.

[12] Others include the iid assumption on the observations and equal weight given for all of them within a window. The problem for determining the size of the window is still open for debate.

[13] As there is no statistical method available to estimate λ, it is usually taken between 0.97 and 0.99 and the role of $\frac{1-\lambda}{1-\lambda^k}$ is to ensure that the weights sum to 1.

the random variable Y_t corresponds to the last weight used in the previous sum, i.e the θ-quantile is given as

$$\sum_{j=t-k+1}^{t} y_j \mathbf{I}\Big(\sum_{i=1}^{k} w_i\Big(\lambda, k\Big)\mathbf{I}\Big(y_{t+1-i} \leq y_j\Big) = \theta\Big)$$

where $w_i\Big(\lambda, k\Big)$ are the weights associated with observation y_i and $\mathbf{I}\Big(.\Big)$ is the indicator function. This approach prevails a significant improvement over the HS as it removes most of the drawbacks in HS. However, the hybride method is not efficient in allowing for the volatility, σ_t in model (1.1.1.1), see Hull and White [69] and they both share the same sparseness problem.

The following section introduces a method[14] that automatically takes account of the conditional volatility. Note that if $\{e_t\}$ satisfies only martingale difference condition, then still μ_t becomes the conditional expectation and σ_t^2, the conditional variance[15]. However, for the sake of simplicity, we stick to the strong assumption in (1.1.1.1(3)).

1.1.2 Quantile autoregression(QAR)

The regression quantile models were introduced by Koenker and Basset(1978). They represent a substantially more general and informative method of regression analysis than the conventional mean-variance regression, since the former fully[16] describes the conditional distribution of a response variable Y_t given a covariate $\mathbf{X}_t$ in $\mathbf{F}_{t-1}$, without imposing any rigid distributional assumptions. Let $\mu_\theta \in \mathbf{R}^d \to \mathbf{R}$ be an unknown smooth function and define, analogously to Koenker-Bassett, the loss function M_θ as

$$M_\theta\Big(y, \mu\Big) = \theta\Big|y-\mu\Big|^+ + (1-\theta)\Big|y-\mu\Big|^-$$

where $\Big|y-\mu\Big|^-$ and $\Big|y-\mu\Big|^+$ stands for the absolute of negative and positive values respectively. This equation can be rewritten in terms of indicator function,

[14]Where both hybride and HS are just but special cases.

[15] In various forms, see [40] and [13], [70],[26],[104].

[16] Quantile regression method can be used to measure the effect of covariates anywhere in unknown distribution.

$$M_\theta\Big(y,\mu\Big) = \Big(y-\mu\Big)\Big(\theta - \mathbf{I}_{\{y-\mu\le 0\}}\Big) \tag{1.1.2.1}$$

The conditional θ-quantile function is then obtained as

$$\mu_\theta\Big(\mathbf{x}\Big) = \arg\min_{\mu} E\Big[M_\theta\Big(Y_t,\mu\Big)\Big|\mathbf{X}_t = \mathbf{x}\Big].$$

In the following, we use the abreviation $\mu_{t,\theta} = \mu_\theta\Big(\mathbf{X}_t\Big)$ for the conditional θ-quantile of Y_t given $\mathbf{X}_t$. Let $\{Y_t\}$ be conditionally distributed according to

$$P\Big(Y_t \le y\Big|\mathbf{X}_t = \mathbf{x}\Big) = F_{\mathbf{x}}\Big(y\Big), \quad t = 1,2,\ldots,. \tag{1.1.2.2}$$

We assume that $F_{\mathbf{x}}$ has the density $f_{\mathbf{x}}$, and then $f_{\mathbf{X}_t}$ is the conditional density of Y_t given $\mathbf{X}_t$. The conditional θ-quantile $\mu_{t,\theta}$, of Y_t given $\mathbf{X}_t$ satisfying (1.1.2.2) can then be written as

$$\theta = \int_{-\infty}^{\mu_{t,\theta}} f_{\mathbf{X}_t}\Big(y\Big)dy \tag{1.1.2.3}$$

or in the usual regression convention as

$$Y_t = \mu_{t,\theta} + \widetilde{\mu}_{t,\theta}$$

where $\widetilde{\mu}_{t,\theta}$ is a r.v with conditional θ-quantile 0, i.e

$$P\Big(\widetilde{\mu}_{t,\theta} \le 0\Big|\mathbf{X}_t\Big) = \theta$$

The term *Quantile Autoregression (QAR)* was introduced in literature in Abbegger [1] to mean the conditional quantile regression of a response, Y_t, given its past observations as a covariate. We will adopt this terminology. For instance, if $\mathbf{X}_t = \Big(Y_{t-1},\ldots,Y_{t-d}\Big)$, $\mu_{t,0.5}$ is the 0.5^{th} QAR, $\mu_{t,0.95}$ is the 0.95^{th} QAR and for $\theta \to 1$ or $\theta \to 0$, $\mu_{t,\theta}$ will be called here θ^{th} extreme QAR. If the conditional returns distribution function, $F_{X_t}\Big(y\Big)$, of Y_t conditional on $\mathbf{X}_t$, is continuous and strictly monotone, then $\mu_{t,\theta}$ is its inverse and hence unique. More generally, the QAR of Y_t will be given by

$$\mu_{t,\theta} = F^{-1}_{\mathbf{X}_t}\Big(\theta\Big) \tag{1.1.2.4}$$

where $F_{\mathbf{X}_t}^{-1}\Big(\theta\Big)$ is the general inverse of $F_{\mathbf{X}_t}\Big(y\Big)$ at a fixed θ. We use the term Quantile Autoregression in the following slightly more general sense by allowing $\mathbf{X}_t$ to contain not only the past observations of Y_t but also exogeneous $\mathbf{F}_{t-1}$-measurable random variables.

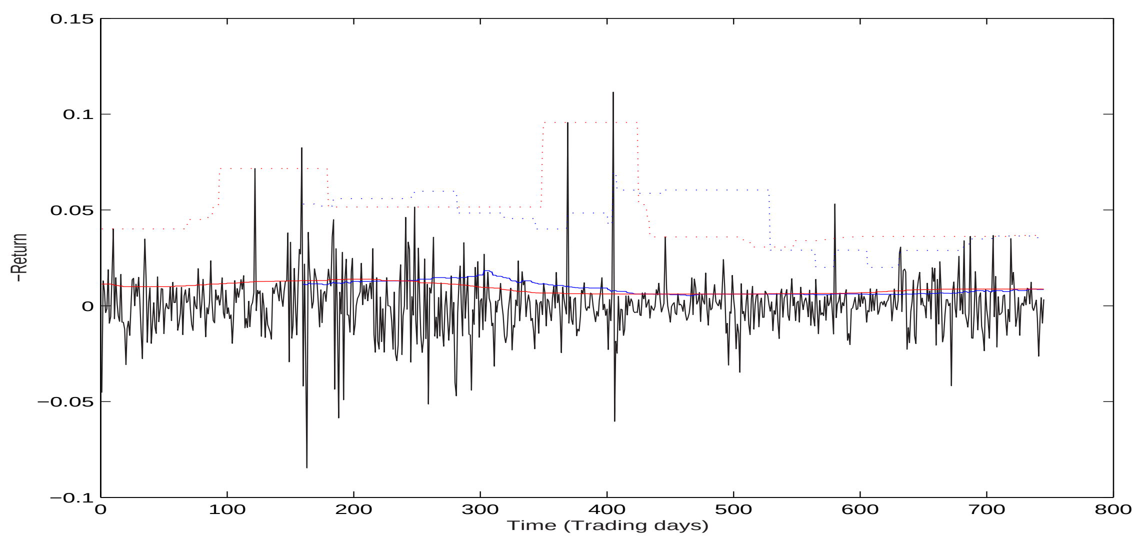

Figure 1: Returns on BASF:1/90-12/92. The difference between the quantile estimates in the interior and for the level close to 1. The blue and red curves are HS and QAR estimates respectively. The solid and dotted curves depict estimates at $\theta = 0.75$ and 0.99 respectively.

The QAR enjoys the robustness property against the effect of the outlying events as the effect on the θ^{th}-QAR is bounded so long as the number of outlying events is lower than $n\min\{\theta, 1-\theta\}$. The following equivariance properties are exhibited by the QAR. Analoguous version can be found in Koenker and Basset(1978).
For a $\mathbf{F}_t$-measurable r.v η_t let $Q_{t,\theta}\Big(\eta_t\Big) = \inf\Big\{\mu \in \mathbf{R}\Big|P\Big(\eta_t \le \mu\Big|\mathbf{X}_t\Big) \ge \theta\Big\}$ denote the conditional θ-quantile of η_t given $\mathbf{X}_t$. If, in particular, $\eta_t = Y_t$, we have $Q_{t,\theta}\Big(Y_t\Big) = \mu_{t,\theta}$. Let a_t be a r.v which is a measurable function of $\mathbf{X}_t$, then

1. $Q_{t,\theta}$ is translation -equivariant, that is

$$Q_{t,\theta}\Big(Y_t + a_t\Big) = Q_{t,\theta}\Big(Y_t\Big) + a_t$$

2. $Q_{t,\theta}$ is positively homogeneous or scale equivariant, that is

$$(i) \qquad Q_{t,\theta}\Big(a_t Y_t\Big) = a_t Q_{t,\theta}\Big(Y_t\Big)$$

3. $Q_{t,\theta}$ is invariant to monotonic transformation: That is, for any random variable Y_t and nondecreasing function μ on $\mathbf{R}$, then

$$Q_{t,\theta}\Big(\mu\Big(Y_t\Big)\Big) = \mu\Big(Q_{t,\theta}\Big(Y_t\Big)\Big),$$

i.e the QAR of the transformed random variable $\mu\Big(Y_t\Big)$ are the transformed QAR of the original variable Y_t.

These properties have immediate application in the estimation of scale function in our model. They constitute part of the definition of a risk measure.
It is noteworthy to mention that, just as in the case of both HS and Hybride, the QAR is based on the conditional ordering and extrapolation of observations and quantiles respectively and, therefore, the empirical distribution is a step function for data not so close to each other. In particular, the estimates of the true function far out in the tails can cause biased results due to sparseness of the data. See the red curves in figure (1) for an illustration. The solid and dotted red curves correspond to 0.75 and 0.99-quantile functions respectively. It can be seen that as $\theta \to 1$, the estimate becomes a step function. This unreliability of the estimate is a very undesirable feature in the prediction of extreme risks.

1.1.3 The extreme sample quantiles

The drawback for QAR can be remedied by fitting a smooth function through the tail of the distribution. We propose to use Extreme Value Theory (EVT). EVT concerns the asymptotic behaviour of extreme order statistics, such as minimum and the maximum.

Let $e_t \in \mathbf{R}$ be iid random variables with distribution F. The conditional excess distribution of e_t given that it exceeds a threshold value u is

$$F_u\Big(z\Big) = P\Big(e_t - u \leq z \Big| e_t > u\Big)$$

for $z \geq 0$. We assume that the threshold is somehow marking the beginning of the right hand tail. The main principle behind EVT is that for any general distribution, G, such that

$$\lim_{u \to e_F} \sup_{0<z<e_F-u} \left| \bar{F}_u\Big(z\Big) - \bar{G}\Big(z\Big) \right| = 0 \qquad (1.1.3.1)$$

where e_F is the endpoint of the distribution function F, the tail $\bar{F} = 1 - F$ of F can be estimated by means of the tail of G. The extreme unconditional quantiles, z_φ, are then derived from $G\Big(z\Big)$ for any values of $z > 0$ or alternatively, for any $\varphi \geq F\Big(u\Big)$. The behaviour of the tail forms is an essential part of EVT. The analysis of the extreme statistics started in 1920's and still continues. This includes the fundamental results on the distribution of extremes, obtained by Fisher, Frechet, Tippet, Gnedenko, von Mises, Galambos, de Haan among others-see Embrechts et al. [39] for an excellent review. Several monographs and lecture notes focusing on the extreme statistics includes Galambos [54], Leadbetter et al. [80], Huesler and Reiss [68].

1.1.4 Scale function

The application of EVT in the estimation of extreme quantiles requires that the series be independent. In the estimation of extreme QAR, the clustering or evolving volatility in financial time series is found to be significant at high (also applies to low) levels of θ and cannot be disregarded. Adjusting a series of its QAR at such levels, we still find the excesses to exhibit some siginificant dependence[17]. This dependence can be reduced by extracting the volatility.

The autoregressive conditional heteroscedasticity (ARCH) model, in Engle [40] and its variants, were introduced to allow the conditional variance of time series model to depend on the past information (conditional heteroscedasticity). Because of the well established empirical facts about financial (returns) data (see Mandelbrot [83] and Fama [43] among others), Bollerslev [14] and Nelson [91] among others have taken the approach of likelihood based on the student's t-distribution to estimate the volatility. However, the misspecification of the form of such conditional distribution used to define the likelihood can create

[17] The investigation of this feature is carried out in chapter 4 using real data.

problems in parameter estimation. The theoretical work on adaptive estimator of ARCH-type models which provides an alternative approach to the problem has been observed in Linton [81] and Drost et al. [34]. For recent review of the ARCH literature see among others Bollerslev et al. [15] and Bollerslev et al. [16]. The clustering of volatility and heavy tailedness in financial(returns) data has seen the generalization of ARCH models and its variants to models such as generalized ARCH(GARCH), Integrated GARCH (IGARCH), exponential GARCH(EGARCH) and threshold GARCH(TGARCH). In all these models, the hidden variable volatility depends on lagged values of the process and lagged values of the volatility. A detailed review of these models and their many variants can be found in among others Bollerslev et al. [15] and Shephard [102]. There estimation is usually based on symmetric distribution of the error or a robust quasi-maximum likelihood method, see Bollerslev and Wooldridge [17]. The former suffers the same consequence as in ARCH. In mean-variance[18], the latter suffers from the fact that it depends on the properties of the estimated mean, which is sensitive to model misspecification. Furthermore, Hall and Yao (2002) have shown that for heteroscedastic data with heavy tailed errors, the method suffers from the complex limit distributions and slow convergence rates. These problems have motivated us to look for more flexible methods that are not based on symmetry assumptions and are in general less sensitive to model misspecification. Thus, in order to reduce the dependence we have incoporated a scale function of the form

$$\sigma_{t,\theta} = b\sigma_t,$$

with b being a positive constant at time t but depends on θ. The function σ_t is the conditional volatility defined in (1.1.1.1).

The estimation of such a scale function is not novel in literature, see variation in Welsh et al. [111] in the case of heteroscedastic regression with independent variables. Because quantiles are readily interpretable in location-scale models and are robustly estimable than moments, Koenker and Zhao [76] has exploited quantile regression ideas of Koenker and Basset(1978) to ARCH setting. Instead of modeling conditional variance, it focuses on ARCH models for conditional scale, where the standardized error is assumed to be iid random variable with zero mean-finite variance. Similar models based on conditional scale (standard deviation), but restricted to Gaussian context, have been observed in Tayler

[18] Examples includes AR-(G)ARCH models

[107], Shwert [101] and Nelson and Foster [92]. The scale function provide a more natural concept of dispersion than variance, see Bickel and Lehmann [11], and offers advantages from robustness viewpoint, see Bickel [12] and Carrol and Ruppert [22] among others. The nonparametric approach to ARCH/GARCH estimation, see Franke [51], Franke et al. [52] and Buehlmann and McNeil [18] among others, enjoys the advantage of being less sensitive to model misspecification. However, the reasons given in Bickel and Lehmann [11], Bickel [12] and Carrol and Ruppert [22] in the case of parametric models, still infers. Instead of assuming that the errors are iid with zero mean -finite variance, as in Koenker and Zhao [76], we are introducing a new nonparametric model which only assumes that the standardized residuals are zero quantile-unit scale. We believe the model can also be used in the case of infinite variance distribution. Further, because it is a combination of nonparametric regression methodology (see Stone [105], Robinson [96], Haerdle [58], Franke [51]) with quantile regression methodologies, it is less sensitive to model misspecification. In the application part, the more closely related to our spirit is the work in Turner and Weigel [109] which analysed the volatility of returns of S&P 500 and Dow Jones indices using the interquartile range and other measures of volatility.

1.2 What we want to estimate

Suppose we are interested in the QAR corresponding to the level of probability φ. Apart from the usual threshold problem in EVT, we are also faced with the decision on whether to really incorporate the EVT or not.

If $\varphi \approx 1$, the sparseness of data at the extreme right end of the sample makes it hard to directly estimate the QAR, $\mu_{t,\varphi}$, reliably. In that case, a similar approach as the peak-over-threshold (POT)-method of quantile estimation may be useful. The basic idea is to estimate the QAR, $\mu_{t,\theta}$, for some smaller level $\theta < \varphi$ nonparametrically and correct it in the following form

$$\widehat{\mu}_{t,\theta,\varphi} = \widehat{\mu}_{t,\theta} + \widehat{\sigma}_{t,\theta}\widehat{z}_{\varphi} \tag{1.2.0.1}$$

with $\widehat{sigma}_{t,\theta}$ and $\widehat{Z}_{\varphi}$ being appropriate estimates for the scale function based on conditional data and the extreme quantile based on iid assumption respectively. This idea

has motivated us to introduce a form of nonparametric quantile autoregression conditional heteroscedastic function described in section (1.1.4), where the heteroscedastic part is in the form of a scale function depending on θ. Clearly, all functions in (1.2.0.1), have the advantage of being robust estimators than the ones in contrast to (1.1.1.1), even in cases where moments do not exist. Secondly, the problem of sparseness of data faced by direct estimation of extreme quantiles is reduced, since EVT does not necessarily require very large samples and comprise of various smooth functions.
Let the probability density function (*pdf*) of $\mathbf{X}_t$ and the joint *pdf* of $\left(Y_t, \mathbf{X}_t\right)$ denoted by $g\left(\mathbf{x}\right)$ and $f\left(y, \mathbf{x}\right)$. The joint cumulative distribution function *cdf* of $\left(Y_t, \mathbf{X}_t\right)$ is given by

$$F\left(y, \mathbf{x}\right) = \int_{-\infty}^{y} \int_{-\infty}^{\mathbf{x}} f\left(u_1, \mathbf{u}_2\right) du_1 d\mathbf{u}_2$$

The dependence structure between Y_t and $\mathbf{X}_t$ is described by the conditional *pdf* of Y_t given $\mathbf{X}_t$, defined as

$$f_{\mathbf{x}}\left(y\right) = \frac{f\left(y, \mathbf{x}\right)}{g\left(\mathbf{x}\right)}$$

and its conditional *cdf*,

$$F_{\mathbf{x}}\left(y\right) = \int_{-\infty}^{y} f_{\mathbf{x}}\left(u_1\right) du_1.$$

We can estimate $\mu_{t,\theta,\varphi}$ or $\mu_{t,\theta}$ via the conditional distribution of Y_t on its past or directly as will be seen in chapters 2 and 3 respectively. Definition 1.2.1 gives a general definition of the conditional Value-at-Risk (VaR).

Definition 1.2.1 *The Value-at-Risk, $VaR_{t,\varphi}$, of negative returns or losses Y_t at time t given is past information of $\mathbf{X}_t$, is*

$$VaR_{t,\varphi} = \inf\left\{y \in \mathbf{R} \middle| F_{\mathbf{X}_t}\left(y\right) \geq \varphi\right\}$$

for $\varphi \in \left(0, 1\right)$, i.e $VaR_{t,\varphi}$ is just the conditional φ-quantile of Y_t given $\mathbf{X}_t$. $1 - \varphi$ is the probability of extreme losses greater than the VaR usually taking values 5% or 1% corresponding to one or ten day trading[19] periods respectively.

[19] May refer, for example to stock returns, holdings in banks etc.

The risk measure $VaR_{t,\varphi}$ has the following properties[20]. For any two random variables $Y^{(1)}$ and $Y^{(2)}$

1. $VaR_{t,\varphi}$ is monotonic with respect to stochastic dominance of order one (SD(1))[21]. That is

$$VaR_{t,\varphi}\Big(Y^{(1)}\Big) \leq VaR_{t,\varphi}\Big(Y^{(2)}\Big)$$

2. $VaR_{t,\varphi}$ is comonotone additive. That is, if $Y^{(1)}$ and $Y^{(2)}$ are commonotone, then

$$VaR_{t,\varphi}\Big(Y^{(1)} + Y^{(2)}\Big) = VaR_{t,\varphi}\Big(Y^{(1)}\Big) + VaR_{t,\varphi}\Big(Y^{(2)}\Big)$$

Recently, there has been an intense discussion on good measures of risk, see Artzner et al. [6], which provides the requirements for a coherent risk measure. These requirements have ruled out measures that are based on second moments, including the standard deviation as well as quantile based measures, like VaR. A measure that has gained preference in the wake of these findings is the expected shortfall. In relation to $VaR_{t,\varphi}$, and by some appropriate moment condition, we formally define it as

$$\Im_{Y_t}\Big(VaR_{t,\varphi}\Big) = E\Big[Y_t \Big| Y_t > VaR_{t,\varphi}; \mathbf{X}_t\Big] \tag{1.2.0.2}$$

which is just the conditional expection of those losses exceeding the VaR. In iid and univariate case, expected shortfall was first proposed in Acerbi and Tasche [4]. Its variants have been suggested in different names; Conditional VaR in Rockafellar and Uryasev [98] and conditional expection in Artzner et al. [6].

1.3 Nonparametric Method

Later on, we consider the nonparametric estimation of $\mu_{t,\theta}$ and $\sigma_{t,\theta}$. Their properties will be studied under α-mixing conditions. For convenience, the following definition is refered.

[20] Properties formulated in terms of preference structures induced by dominance relations, see Fishburn [50] and Georg [95].

[21] The relationship $Y^{(1)} \prec_{SD(1)} Y^{(2)}$ hold if and only if $E\Big[\mu\Big(Y^{(1)}\Big)\Big] \leq E\Big[\mu\Big(Y^{(2)}\Big)\Big]$ for all (integrable) monotonic function μ.

Definition 1.3.1 (Strong mixing) *Let* $\left\{\eta_t\right\}$ *be a stationary time series and let* $\mathbf{F}_t$ *and* $\mathbf{F}^t$ *denote* σ*-fields generated, respectively, by* $\eta_i, -\infty < i \leq t$ *and* $\eta_i, t \leq i < \infty$*. Given a positive number* s*, then*

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

is the strong mixing coefficient. If

$$\lim_{s\to\infty} \alpha\left(s\right) = 0$$

the process is called strongly mixing or α*-mixing.*

The mixing conditions indicate the maximum dependence between two time events which are a least s-steps apart. For instance if a stationary sequence is m-dependent, namely Y_t depends only on previous m observations, then the mixing coefficient is zero for $s > m$. There are a number of mixing conditions in literature, among them α-mixing is reasonably weak and known to be fulfilled for many time series models. For instance, under the conditions derived in Gorodetskii [55] and Withers [112], a linear process is α-mixing. Chen and Tsay [25] has shown that the functional autoregressive process is geometrically ergodic under certain conditions. Franke et al. [53] provide sufficient conditions for general markov chain process to be geometrically ergodic, with coefficient which depend on some explicit constants. Futhermore Masry and Tjostheim ([84],[85]) have demonstrated that under some conditions both autoregressive conditional heteroscedastic $(ARCH)$ process and nonlinear additive autoregressive models with exogeneous variables, which are popular in finance and econometrics, are stationary and α-mixing.
Let $\jmath_1\left(.\right)$ and $\jmath_2\left(.\right)$ be real valued, measurable functions. Set $J_1 = \jmath_1\left(\eta_t\right), \quad J_2 = \jmath_2\left(\eta_{t'}\right)$.
The proof of the following lemma can be found on page 10 of Doukhan [32].

Lemma 1.1 (Covariance inequality) *Suppose that* J_1 *and* J_2 *are bounded random variables with respect to* $\mathbf{F}_t$ *and* $\mathbf{F}^{t'}$ *respectively, then*

$$\left|cov\left(J_1, J_2\right)\right| \leq c\alpha\left(t' - t\right)\left|\left|J_1\right|\right|_\infty\left|\left|J_2\right|\right|_\infty$$

where c is a generic constant and

$$\Big|\Big|V\Big|\Big|_{\infty} = ess.sup|V| = \inf\Big\{c \in \mathbf{R}\Big|P\Big(|V| > c\Big) = 0\Big\}$$

1.3.1 Kernel method

We assume, for $d \in \mathbf{N}$, our data consist of a realization of $\Big(Y_t, \mathbf{X}_t\Big), t = 1, \ldots, n$, which may correspond to the observed returns information Y_t and $\mathbf{X}_t \in \mathbf{R}^d$ variables as in (1.1.1.1) at several dates. We assume that the conditional distribution function $F_{\mathbf{x}_i}\Big(y\Big)$, of Y_t given $X_t = \mathbf{x}_i \in \mathbf{R}^d, \quad i = 1, 2, \ldots, n$ is such that the equation $F_{\mathbf{x}_i}\Big(y\Big) = \theta$ admits a unique solution, $\mu_\theta\Big(\mathbf{x}_i\Big)$, for each $\mathbf{x}_i$. Let $k_{i,j} : \mathbf{R} \to \mathbf{R}$ be bounded and symmetric functions such that

$$\int k_{i,j}\Big(u\Big)du = 1, \qquad i = 1, \ldots, n, \quad j = 1, \ldots, d.$$

The kernel functions $k_{i,j}$ assign weights to the observation $X_{t,j} \in \mathbf{R}$ which decreases with the distance between the point of estimation $x_{i,j} \in \mathbf{R}$ and $X_{t,j}$. The various forms of kernel functions include Uniform,Triangle, Epanechnikov, Bisquare, Triweight, Gaussian among others for univariate. The latter one has an infinite support while the rest are bounded in $[-1, 1]$. The Bisquare kernel has the form

$$k_{i,j}\Big(u\Big) = \frac{15}{16}\Big(1 - u^2\Big)^2 \mathbf{I}_{[-1,1]}\Big(u\Big)$$

and that of normal takes the form

$$k_{i,j}\Big(u\Big) = \frac{1}{2\pi}\exp\Big(-\frac{1}{2}u^2\Big), \quad \text{for} \quad -\infty < u < \infty.$$

We evaluate it at the point $x_{i,j}$ for observation $X_{t,j}$ for

$$u = \frac{x_{i,j} - X_{t,j}}{h_{i,j}}$$

The bandwidth $h_{i,j}$ plays an important role in determining the number of data in a local neighborhood of the estimation point, $x_{i,j}$. Hence a very small bandwidth will lead to a wiggly curve of the estimated quantile function, while at the same time, a large bandwidth gives a smooth curve but with a possibility of obscuring the interesting

structures. Detailed information on this subject can be found in Haerdle [58].
For $\mathbf{x}_i \in \mathbf{R}^d$, a multivariate kernel $\mathbf{K} : \mathbf{R}^d \to \mathbf{R}$, is used. In this case we may choose a norm kernel,

$$\mathbf{K}\Big(\mathbf{u}\Big) = k_i\Big(\Big|\Big|\mathbf{u}\Big|\Big|\Big), \quad \text{where} \quad \mathbf{u} \in \mathbf{R}^d$$

for some univariate kernel k_i. For further details on the norm kernel, see Michels [89], page 16. In connection with time series application, a frequently used multivariate kernel alternative, is the product kernel. We define the product kernel as

$$\mathbf{K}\Big(\mathbf{u}\Big) = \prod_{j=1}^{d} k_{i,j}\Big(u_j\Big), \quad i = 1, \ldots, n$$

where the multivariate bandwidth $\mathbf{h}^{(i)}$ is a diagonal matrix with elements $\Big(h_{i,j}\Big)$ whose determinant is $\Big|\mathbf{h}^{(i)}\Big| = \prod_j^d h_{i,j}$. In terms of the random variable $\Big(Y_t, \mathbf{X}_t\Big)$,

$$\begin{aligned}\mathbf{K}\Big(\mathbf{x}_i - \mathbf{X}_t; \mathbf{h}^{(i)}\Big) &= \mathbf{K}\left(\frac{x_{i1} - X_{t1}}{h_{i1}}, \frac{x_{i2} - X_{t2}}{h_{i2}}, \ldots, \frac{x_{id} - X_{td}}{h_{id}}\right) \\ &= \prod_{j=1}^{d} k\left(\frac{x_{ij} - X_{tj}}{h_{ij}}\right), \quad \text{for product kernel} \end{aligned} \qquad (1.3.1.1)$$

where k is a fixed univariate kernel.
The following conditions are imposed on the kernel, densities, bandwidth and the process. Conditions (B1)-(B5), (C1)-(C2), (D1) and (E1) are standard regularity conditions. Additional conditions (B6), (C3)-(C5) and (D2) are used for deriving the asymptotic properties of the conditional distribution estimator. The asymptotic properties of QAR are studied under further conditions, (C6)-(C7). The letter c and $c_i, 1, 2 \ldots$ will denote a generic constant which might take a different value at different places.

Conditions 1.3.1
(B1) $\int_{\mathbf{R}^d} \mathbf{K}\Big(\mathbf{u}\Big) d\mathbf{u} = 1,$
(B2) $\mathbf{K}\Big(\mathbf{u}\Big) \leq \overline{\mathbf{K}} < \infty, \quad \mathbf{u} \in \mathbf{R}^d$
(B3) $\mathbf{K}$ *is a compactly supported density*
(B4) $\mathbf{K}\Big(\mathbf{u}\Big) \geq 0, \quad \forall \mathbf{u} \in \mathbf{R}^d$

(B5) $\mathbf{K}$, *is symmetric i.e* $\mathbf{K}\Big(\mathbf{u}\Big) = \mathbf{K}\Big(-\mathbf{u}\Big), \quad \mathbf{u} \in \mathbf{R}^d$
(B6) $\mathbf{K}$ *satisfies Lipschitz condition,* $\Big|\mathbf{K}\Big(\mathbf{u}\Big) - \mathbf{K}\Big(\mathbf{v}\Big)\Big| \le c_k\Big|\mathbf{u} - \mathbf{v}\Big|, \forall \mathbf{u}, \mathbf{v} \in \mathbf{R}^d, c_k > 0.$

Conditions 1.3.2
(C1) $\Big(Y_t, \mathbf{X}_t\Big)$ *has a joint density* $f\Big(\mathbf{x}, y\Big)$. *Then, the density* $g\Big(\mathbf{x}_i\Big)$ *of* $\mathbf{X}_t$, *exists too.*

(C2) For fixed $\Big(y, \mathbf{x}\Big), \quad F_{\mathbf{x}_i}\Big(y\Big) \in (0,1)$ *and* $g\Big(\mathbf{x}_i\Big) > 0$ *are continuous in a neighborhood of* $\mathbf{x}_i$, *where we want to estimate the quantile function. Then, the conditional density* $f_{\mathbf{x}_i}\Big(y\Big)$ *exist at* $\mathbf{x}_i$

The following derivatives exist for $\mathbf{x} = \mathbf{x}_i$
(C3) $\nabla^2 F_{\mathbf{x}}\Big(y\Big) = F''_{\mathbf{x}}\Big(y\Big) = \frac{\partial^2 F_{\mathbf{x}}\Big(y\Big)}{\partial \mathbf{x}^2}$ *where* ∇^2 *is the Hessian w.r.t* $\mathbf{x}$ *evaluated at* $\mathbf{x}$ *for fixed* y
(C4) $\nabla g\Big(\mathbf{x}\Big) = \frac{\partial g\Big(\mathbf{x}\Big)}{\partial \mathbf{x}}$ *and* ∇ *is the gradient w.r.t* $\mathbf{x}$
(C5) $\nabla^2 g\Big(\mathbf{x}\Big) = \frac{\partial^2 g\Big(\mathbf{x}\Big)}{\partial \mathbf{x}^2}$
(C6)The conditional density $f_{\mathbf{x}_i}\Big(y\Big)$ *is continuous in a neighborhood of* $\mathbf{x}_i$
(C7) $f_{\mathbf{x}_i}\Big(\mu_\theta\Big(\mathbf{x}_i\Big)\Big) > 0.$

Conditions 1.3.3
With $\Big|\mathbf{h}^{(i)}\Big|$ *and* $\Big|\Big|\mathbf{h}^{(i)}\Big|\Big|$ *denoting the determinant, respectively the Euchlidean norm, of the diagonal bandwidth matrix* $\mathbf{h}^{(i)} = diag\Big(h_{i,j}\Big), j = 1, \dots, d$:
(D1) $h_{i,j} > 0, \quad h_{i,j} \to 0$ *and* $n\Big|\mathbf{h}^{(i)}\Big| \to \infty$ *for* $n \to \infty$
(D2) $\Big(\sqrt{n}\Big|\mathbf{h}^{(i)}\Big|\Big)^{-2} \to \infty$, *as* $n \to \infty$.

Conditions 1.3.4
(E1) The process $\Big\{\Big(Y_t, \mathbf{X}_t\Big)\Big\}$ *is* α*- mixing with coefficient satisfying* $\alpha\Big(s\Big) = o\Big(s^{-(2+\delta)}\Big)$, *for some* $\delta > 0$

If the kernel function $\mathbf{K}$ has support $\Big[-1, 1\Big]^d$, then we expect the relevant estimator, for instance of the conditional distribution $F_{\mathbf{x}_i}\Big(y\Big)$, to use the observations in the intervals $\Big(\mathbf{x}_i - \mathbf{h}^{(i)}\bar{\mathbf{I}}, \mathbf{x}_i + \mathbf{h}^{(i)}\bar{\mathbf{I}}\Big)$, where $\bar{\mathbf{I}}$ is a unit vector of dimension d. In situation where the dependent observations are used, the local estimator is affected only by the dependence of observations in a small window and not by the whole data. The rate on which $\alpha(s)$ in (E1) goes to zero plays an important role in showing the asymptotic behaviour of the estimators. In general we use the following lemma, due to Volkonskii and Rozanov (1959), to shows that dependent random variables can be approximated by a sequence of independent random variables having the same marginal distributions.

Lemma 1.2 *Let $V_1, \dots, V_L$, be random variables measurable with respect to the σ-algebras $\mathbf{F}_{i1}^{j1}, \dots, \mathbf{F}_{iL}^{jL}$ respectively with $i_{l+1} - j_l \geq w \geq 1$ and $|V_j| \leq 1$ for $j = 1, \dots, L$. Then*

$$\Big| E \prod_{j=1}^{L} V_j - \prod_{j=1}^{L} E\Big(V_j\Big) \Big| \leq 16\Big(L-1\Big)\alpha\Big(w\Big)$$

where $V_j = \exp\Big(it_j X_j\Big)$, is the characteristic function of the random variable X.

1.4 Model definition

Let $\Big\{V_t, t \in \mathbf{Z}\Big\}$ be a real-valued financial time series on a complete probability space $\Big(\Omega, \mathbf{F}, P\Big)$, where P is such that either

1. $\Big\{V_t\Big\}$ is an iid process or
2. $\Big\{V_t\Big\}$ is a stationary α-mixing process such that condition (E1) holds.

We assume that V_t can be partitioned as $\Big(Y_t, \mathbf{X}_t\Big)$, where $Y_t \in \mathbf{R}$ is $\mathbf{F}_t$-measurable and $\mathbf{X}_t \in \mathbf{R}^d$ is $\mathbf{F}_{t-1}$-measurable, and that V_t may have representation (1.1.1.1).

In the time series case 2., we consider the quantile autoregression-heteroscedastic process of the form

$$Y_t = \mu_{t,\theta} + \sigma_{t,\theta} Z_t, \quad t = 1, 2, \dots, \tag{1.4.0.2}$$

where $\mu_{t,\theta} = \mu_\theta\Big(\mathbf{X}_t\Big)$ is the conditional θ-quantile of Y_t given $\mathbf{X}_t$ and $\sigma_{t,\theta} = \sigma_\theta\Big(\mathbf{X}_t\Big)$ is the conditional scale function of Y_t given $\mathbf{X}_t$. The residuals Z_t are iid with zero θ-quantile and unit scale, i.e they satisfy the following conditions,

Conditions 1.4.1 *Z_t and $M_\theta\Big(Z_t, F_{Z_t}^{-1}\big(\theta\big)\Big) - 1$ have a continuous positive density in a neighborhood of 0 and*

$$P\Big(Z_t \le 0\Big) = \theta \tag{1.4.0.3}$$

and

$$P\Big(M_\theta\Big(Z_t, F_{Z_t}^{-1}\big(\theta\big)\Big) \le 1\Big) = \theta \tag{1.4.0.4}$$

Condition (1.4.1) ensures that both $E\Big[M_\theta\Big(Z_t, \mu\Big) - M_\theta\Big(Z_t, 0\Big)\Big]$ and $E\Big[M_\theta\Big(M_\theta\Big(Z_t\Big) - 1, \sigma\Big) - M_\theta\Big(M_\theta\Big(Z_t\Big) - 1, 0\Big)\Big]$ are nonnegative and have a unique minimum at 0 with respect to μ and σ. The second terms in the brackets of both expectations ensure the respective first moments are finite, see Huber [67] and Kozek [77] for similar expressions and arguments.

In the following we use the notation

$$M_\theta\Big(Z\Big) = M_\theta\Big(Z, F_Z^{-1}\big(\theta\big)\Big)$$

for any real random variable Z with distribution function F_Z, i.e we evaluate the distance function $M_\theta\Big(y, \mu\Big)$ at the random point $y = Z$ and at its θ-quantile $\mu = F_Z^{-1}\Big(\theta\Big)$.

If we take the residuals in (1.1.1.1) and define

$$Z_t = \frac{e_t - q_\theta^e}{M_\theta^e}$$

where q_θ^e is the θ-quantile of e_t and M_θ^e the θ-quantile of $M_\theta\Big(e_t\Big)$, then the resulting Z_t satisfies (1.4.0.3) and (1.4.0.4) by the following lemma:

Lemma 1.3 *Let U be a real random variable with absolutely continuous distribution F_U and θ-quantile $q_\theta = F_U^{-1}\Big(\theta\Big)$.*
(a) $P\Big(M_\theta\Big(U\Big) \le \mu\Big) = P\Big(q_\theta - \frac{\mu}{1-\theta} \le U \le q_\theta + \frac{\mu}{\theta}\Big)$, for all $0 \le \mu < \infty$.
(b) Let M_θ be the θ-quantile of $M_\theta\Big(U\Big)$. Then $W = \frac{U - q_\theta}{M_\theta}$ has zero θ-quantile and unit scale, i.e it satisfies (1.4.0.3) and (1.4.0.4).

Proof of lemma 1.3

(a) By definition of $M_\theta\Big(U\Big) = M_\theta\Big(U, q_\theta^U\Big)$:

$$\begin{aligned} P\Big(M_\theta\Big(U\Big) \leq \mu\Big) &= P\Big(U > q_\theta, \theta\Big(U - q_\theta\Big) \leq \mu\Big) + P\Big(U \leq q_\theta, \Big(1-\theta\Big)\Big(q_\theta - U\Big) \leq \mu\Big) \\ &= P\Big(q_\theta < U \leq q_\theta + \frac{\mu}{\theta}\Big) + P\Big(q_\theta - \frac{\mu}{1-\theta} \leq U \leq q_\theta\Big) \\ &= P\Big(q_\theta - \frac{\mu}{1-\theta} \leq U \leq q_\theta + \frac{\mu}{\theta}\Big). \end{aligned}$$

(b) $P\Big(W \leq 0\Big) = P\Big(U - q_\theta \leq 0\Big) = \theta$, ie the θ-quantile of W is 0 and, therefore, using (a)

$$\begin{aligned} P\Big(M_\theta\Big(W\Big) \leq 1\Big) &= P\Big(-\frac{1}{\theta} \leq W \leq \frac{1}{\theta}\Big) \\ &= P\Big(q_\theta - \frac{M_\theta}{1-\theta} \leq U \leq q_\theta + \frac{M_\theta}{\theta}\Big) \\ &= P\Big(M_\theta\Big(U\Big) \leq M_\theta\Big) = \theta, \end{aligned}$$

where, again we have used (a)

□

If we choose $\mathbf{X}_t = \Big(Y_{t-1}, \ldots, Y_{t-\tau}, \mu_{t-1,\theta}, \ldots, \mu_{t-\tau,\theta}, S_t\Big)$ to consist of a finite past of the observed process Y_t, the corresponding conditional θ-quantiles and an exogeneous series S_t, then in (1.4.0.2) we assume that

$$\begin{cases} \mu_{t,\theta} = \mu\Big(Y_{t-1}, \ldots, Y_{t-d}, S_t\Big) & : \\ \sigma_{t,\theta} = \sigma_\theta\Big(Y_{t-1} - \mu_{t-1,\theta}, \ldots, Y_{t-q} - \mu_{t-q,\theta}, S_t\Big) & : \end{cases} \tag{1.4.0.5}$$

i.e we let the volatility be a function of the past residuals instead of the data themselves. Instead, we could as well take the past information as $\Big(M_\theta\Big(Y_{t-1}\Big), \ldots, M_\theta\Big(Y_{t-q}\Big)\Big)$. That is a special case of (1.4.0.5) as $\sigma_{t,\theta}$ is an arbitrary function and $Y_{t-k} - \mu_{t-k,\theta}$ determines $M_\theta\Big(Y_{t-k}\Big)$ uniquely. In this case, regressing $M_\theta\Big(Y_t\Big)$ against $\Big(M_\theta\Big(Y_{t-1}\Big), \ldots, M_\theta\Big(Y_{t-q}\Big)\Big)$[22] produces a linear surface which could easily be fitted by linear parametric, if desired, as

[22] As in Engle [40], in the case of square.

well as nonparametric methods. This is illustrated in Figures (2) and (3) where the underlying regression is based on the process[23]

$$Y_t = \Big(0.075 + 0.45 M_{0.75}\Big(Y_{t-1}\Big) + 0.50 M_{0.75}\Big(Y_{t-2}\Big)\Big) Z_t, \quad t = 1, 2, \ldots,$$

with Y_t being variables in the interval $\Big(-4, 4\Big)$, and Z_t comprising of student's-t distributed error with 4 d.f. The former surface results by regressing $M_{0.75}\Big(Y_t\Big)$ on $\Big(M_{0.75}\Big(Y_{t-2}\Big), M_{0.75}\Big(Y_{t-1}\Big)\Big)$ with $M_{0.75}\Big(Y_{t-j}\Big) = Stheta\Big(Y_{t-j}\Big)$ for $j = 1, 2$ and the latter one by regressing $M_{0.75}\Big(Y_t\Big)$ on $Y_{t-j}, j = 1, 2$.

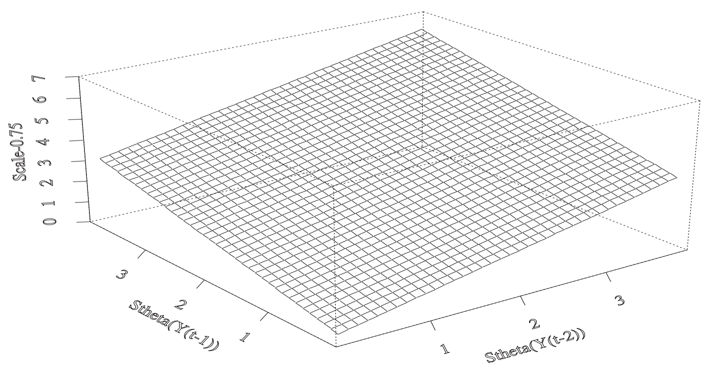

Figure 2: Example of linear regression surface

We will call processes of the form (1.4.0.2) satisfying (1.4.0.3) and (1.4.0.4) *Quantile Autoregressive-Quantile Autoregressive Conditional Heteroscedastic* of orders d and q (QAR(d)-QARCH(q)), where again we do not explicitly denote the presence of the exogeneous variable S_t which allows for a considerable degree of flexibility. If, for example, we

[23] With $\mu_{t,0.75} = 0$ and because our interest here is to give an example of a heteroscedastic part, we have set M^e_θ to be equal to 1, so that Z_t is just $\Big(e_t - F_e^{-1}\Big(0.75\Big)\Big)$.

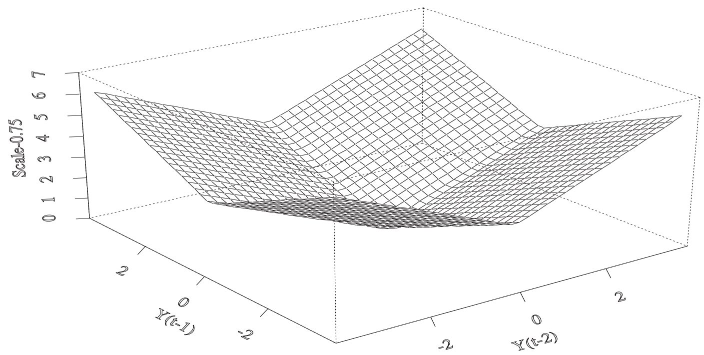

Figure 3: Example of nonlinear regression surface

choose $S_t = \Big(\sigma_{t-1,\theta}, \ldots, \sigma_{t-\tau,\theta}\Big)$ in (1.4.0.5) then we get a nonparametric quantile analogue to an AR-GARCH-process which we call *Quantile Autoregressive-Generalized Quantile Autoregressive Conditional Heteroscedastic* (QAR(d)-GQARCH(p,q)), where d, p and q are the respective orders. The QAR of Y_t in (1.4.0.2) under condition (1.4.1) given the information, $\mathbf{F}_{t-1}$, are seen as

$$Q_{t,\theta}\Big(Y_t\Big) = \mu_{t,\theta}$$

and the scale function at $\theta^{th} \quad QAR$

$$Q_{t,\theta}\Big(M_\theta\Big(Y_t\Big)\Big) = \sigma_{t,\theta}$$

where $Q_{t,\theta}$ is the quantile operator defined in section (1.1.2) and, for a time series $Y_t, \quad M_\theta\Big(Y_t\Big) = M_\theta\Big(Y_t, \mu_{t,\theta}\Big)$ denotes the distance function evaluated at the conditional θ-quantile.

1.4.1 Parametric examples

1. For a parametric $QAR(d) - QARCH(q)$, the components in (1.4.0.2) would take the form

$$\mu_{t,\theta} = \omega_0 + \sum_{i=1}^{d} \omega_i Y_{t-i} + \sum_{j=1}^{q} \upsilon_j M_\theta \Big(Y_{t-j} \Big) \tag{1.4.1.1}$$

and

$$\sigma_{t,\theta} = \alpha_0 + \sum_{j=1}^{q} \alpha_j M_\theta \Big(Y_{t-j} \Big)$$

where $\omega_o > 0, \alpha_o > 0 \quad \upsilon_j, \omega_j > 0, \forall i = 1, \ldots, d, \quad \alpha_j \geq 0, \quad \forall j = 1, 2, \ldots, q$. Note that in (1.1.1.1), if e_t is zero-mean random variable with symmetric distribution, F_e, and $\theta = 0.5$, we obtain the QAR which is in exact form as the one represented in Koenker and Zhao [76]. The paper considers a stochastic process $\Big\{ Y_t \Big\}$ generated by the autoregressive process of the form

$$Y_t = \omega_0 + \sum_{i=1}^{d} \omega_i Y_{t-i} + \left(\alpha_0 + \sum_{j=1}^{q} \alpha_j \Big| Y_{t-j} - \mu_{t-j} \Big| \right) e_t \tag{1.4.1.2}$$

where $\alpha_0 > 0 \quad \Big(\alpha_1, \ldots, \alpha_q \Big)' \in \mathbf{R}_+^q$ are the parameters and e_t are iid random variables with zero mean-unit variance.

Both the parametric form (1.4.1.1) and model (1.4.1.2) could be viewed as a general type introduced by Engle (1982). However, the difference comes in via the way heteroscedastic structures and the innovations are defined. In Engle's model the heteroscedastic term is given by

$$\sigma_t = \left(\alpha_0 + \sum_{j=1}^{q} \alpha_j \Big(Y_{t-j} - \mu_{t-j} \Big)^2 \right)^{\frac{1}{2}} \tag{1.4.1.3}$$

where μ_t is the conditional expectation and $\Big\{ e_t \Big\}$ are iid $N\Big(0, 1\Big)$. Because of moment conditions in (1.4.1.2) and (1.4.1.3), it is apparent that symmetry in the distribution of e_t plays an important role in the asymptotic behaviour of the corresponding estimators. In the case of homoscedasticity, the parametric form of (1.4.0.2) and (1.4.1.2) have the same conditional quantile.

2. For a paramerec $QAR(d) - GQARCH(q, p)$ which includes stochastic volatility process of order (q, p) that allows a more complicated dependence of the present scale function on the past volatility, it can be written as

$$Y_t = \mu_{t,\theta} + \Big(\alpha_0 + \sum_{j=1}^{q} \alpha_j M_\theta \Big(Y_{t-j}\Big) + \sum_{k=1}^{p} \beta_k \sigma_{t-k}\Big) Z_t \tag{1.4.1.4}$$

   where $\sigma_{t-k} = c\sigma_{t-k,\theta}$ for some rescaling constant $c > 0$. The constants $\alpha_0, \alpha_j^{'}s,$ and $\beta_k^{'}s$ are non-negative parameters with Z_t satisfying condition (1.4.1). Other derivatives of $GARCH$, like $TGARCH$ could be reformulated in a similar manner, as will be seen later on.

In chapter 2, we start by giving the asymptotic properties of the estimators of $\mu_{t,\theta}$ and $\sigma_{t,\theta}$ when they are obtained through the inverse of conditional distribution. In chapter (3), we give the asymptotic properties for the estimators obtained by direct minimization involving QAR-QARCH in detail and its extensions to QAR-GQARCH.

1.5 Conclusion

This chapter has stated the problem at hand and discussed the estimation methodologies. We have proposed a class of time series models which is similar to nonparametric AR-(G)ARCH models but which is more suitable for estimating quantiles even in cases of infinite variance.

2 Estimation via conditional distribution

In this chapter, we assume $\left\{Y_t, \mathbf{X}_t\right\}$ follows the process in (1.1.1.1) and estimate the QAR and the scale function in (1.4.0.2). We present the consistency and asymptotic normality results of the conditional θ-quantile functional estimator for QAR. These results are then subsequently used to derive further consistency and asymptotic normality results on the the scale functional estimator of $\sigma_\theta\left(\mathbf{x}_i\right)$, both under known and unknown QAR cases.

2.1 The kernel estimator for QAR

The following definitions of the estimators will be used. The pdf g of $\mathbf{X}_t$ at $\mathbf{x}_i$ will be estimated by

$$\widehat{g}\left(\mathbf{x}_i\right) = \left(n\left|\mathbf{h}^{(i)}\right|\right)^{-1} \sum_{t=1}^{n} \mathbf{K}\left(\mathbf{x}_i - \mathbf{X}_t; \mathbf{h}^{(i)}\right), \tag{2.1.0.5}$$

see Parzen [94], Rosenblatt [97]. The joint pdf $f\left(y, \mathbf{x}\right)$ of $\left(Y_t, \mathbf{X}_t\right)$ at $\left(y_j, \mathbf{x}_i\right)$ will be estimated by

$$\widehat{f}\left(y_j, \mathbf{x}_i\right) = \left(n\left|\mathbf{h}^{(i)}\right|h_j\right)^{-1} \sum_{t=1}^{n} \ell_j\left(\frac{y_j - Y_t}{h_j}\right)\mathbf{K}\left(\mathbf{x}_i - \mathbf{X}_t; \mathbf{h}^{(i)}\right), \tag{2.1.0.6}$$

where the functions $\ell_j : \mathbf{R} \to \mathbf{R}$, and $h_j \in \mathbf{R}$ are the kernel and bandwidth respectively for Y_t at point $\left\{y_j\right\}$[24]. The conditional *pdf* $f_{\mathbf{x}_i}\left(y\right)$ of Y_t given that $\mathbf{X}_t = \mathbf{x}_i$ will be estimated by

$$\widehat{f}_{\mathbf{x}_i}\left(y\right) = \frac{\widehat{f}\left(y, \mathbf{x}_i\right)}{\widehat{g}\left(\mathbf{x}_i\right)} \tag{2.1.0.7}$$

The conditional *cdf* $F_{\mathbf{x}_i}\left(y\right)$ of Y_t given $\mathbf{X}_t = \mathbf{x}_i$ at distinct points y can be obtained by integrating (2.1.0.7)

$$\begin{aligned}
\widehat{F}_{\mathbf{x}_i}\left(y\right) &= \int_{-\infty}^{y} \frac{\widehat{f}\left(u, \mathbf{x}_i\right)du}{\widehat{g}\left(\mathbf{x}_i\right)} \\
&= \frac{1}{h}\frac{\sum_{t=1}^{n} \mathbf{K}\left(\mathbf{x}_i - \mathbf{X}_t; \mathbf{h}^{(i)}\right) \int_{-\infty}^{y} \ell\left(\frac{u - Y_t}{h}\right)du}{\sum_{t=1}^{n} \mathbf{K}_i\left(\mathbf{x}_i - \mathbf{X}_t; \mathbf{h}^{(i)}\right)}
\end{aligned}$$

[24] For convenience, the subscript is henceforth dropped.

An alternative estimator for $F_{\mathbf{x}_i}\Big(y\Big)$ proposed in Colomb [27] is the empirical estimator;

$$\begin{aligned}\widehat{F}_{\mathbf{x}_i}\Big(y\Big) &= \frac{\sum_{t=1}^{n}\mathbf{K}\Big(\mathbf{x}_i-\mathbf{X}_t;\mathbf{h}^{(i)}\Big)\mathbf{I}_{\{Y_t\le y\}}}{\sum_{t=1}^{n}\mathbf{K}\Big(\mathbf{x}_i-\mathbf{X}_t;\mathbf{h}^{(i)}\Big)}\\ &= \frac{\widehat{r}_{\mathbf{x}_i}\Big(y\Big)}{\widehat{g}\Big(\mathbf{x}_i\Big)} \qquad (2.1.0.8)\end{aligned}$$

where the indicator function $\mathbf{I}_{\{\eta_t\le .\}} = \mathbf{I}_{\{\eta_t\le .\}}\Big(\eta_t - .\Big)$, will be used throughout. We will base our estimator of the conditional distribution function on (2.1.0.8). The conditional QAR estimator is then obtained by inverting (2.1.0.8) at θ

$$\widehat{\mu}_\theta\Big(\mathbf{x}_i\Big) = \inf_{y\in\mathbf{R}}\Big\{y : \widehat{F}_{\mathbf{x}_i}\Big(y\Big) \ge \theta\Big\}$$

Because $0 \le F_{\mathbf{x}_i}\Big(y\Big) \le 1$ and is strictly monotone in y, $\mu_\theta\Big(\mathbf{x}_i\Big)$ exist and is unique.

2.1.1 The asymptotic properties of the conditional distribution estimator

Let $W = \Big(n\Big|\mathbf{h}^{(i)}\Big|\Big)^{-1}\sum_{t=1}^{n}\eta_t$ where $\eta_t = \mathbf{K}\Big(\mathbf{x}_i-\mathbf{X}_t;\mathbf{h}^{(i)}\Big)\Big(\mathbf{I}_{\{Y_t\le y\}} - F_{\mathbf{X}_t}\Big(y\Big)\Big)$. In order to establish the order of the bias and the variability of the estimator, the following lemma is necessary.

Lemma 2.1 *Under regularity conditions (B1)-(B5), (C1)-(C5), (D1)-(D2) and (E1),*

(1)

$$var\Big[W\Big] = \Big(n\Big|\mathbf{h}^{(i)}\Big|\Big)^{-1}\mathbf{V}^2\Big(y\Big)g^2\Big(\mathbf{x}_i\Big) + o\Big(\Big(n\Big|\mathbf{h}^{(i)}\Big|\Big)^{-1}\Big), \qquad (2.1.1.1)$$

where $\mathbf{V}^2\Big(y\Big) = \frac{1}{g\Big(\mathbf{x}_i\Big)}\Big(F_{\mathbf{x}_i}\Big(y\Big) - F^2_{\mathbf{x}_i}\Big(y\Big)\Big)\int\mathbf{K}^2\Big(\mathbf{u}\Big)d\mathbf{u}$

(2)

$$\widehat{g}\Big(\mathbf{x}_i\Big) \to^p g\Big(\mathbf{x}_i\Big) \qquad (2.1.1.2)$$

(3)

$$\widehat{r}_{\mathbf{x}_i}\Big(y\Big) \to^p r_{\mathbf{x}_i}\Big(y\Big) \qquad (2.1.1.3)$$

Proof of lemma 2.1(1) :

Observe that $E\Big[W\Big] = 0$ since $E\Big[\eta_t\Big|\mathbf{X}_t\Big] = 0$ and

$$\begin{aligned} var\Big[W\Big] &= \Big(n\Big|\mathbf{h}^{(i)}\Big|\Big)^{-2} var\Big(\sum_{t=1}^{n}\eta_t\Big) \\ &= \Big(n\Big|\mathbf{h}^{(i)}\Big|\Big)^{-2}\Big\{\sum_{t=1}^{n} var\Big[\eta_t\Big] + \sum_{t\neq t'}^{n} cov\Big[\eta_t, \eta_{t'}\Big]\Big\} \\ &= \Big(n\Big|\mathbf{h}^{(i)}\Big|\Big)^{-2}\Big\{nE\Big[\eta_1^2\Big] + 2\sum_{t=2}^{n}\Big(n-t+1\Big)cov\Big[\eta_1, \eta_t\Big]\Big\} \\ &= \Big(\sqrt{n}\Big|\mathbf{h}^{(i)}\Big|\Big)^{-2} E\Big[\eta_1^2\Big] + 2\Big(\sqrt{n}\Big|\mathbf{h}^{(i)}\Big|\Big)^{-2}\sum_{t=2}^{n}\Big(1-\frac{t-1}{n}\Big)cov\Big(\eta_1, \eta_t\Big)\Big\} \end{aligned} \tag{2.1.1.4}$$

by stationarity. Now

$$E\Big[\eta_1^2\Big] = E\Big[\mathbf{K}^2\Big(\mathbf{x}_i - \mathbf{X}_t; \mathbf{h}^{(i)}\Big)\Big(F_{\mathbf{X}_t}\Big(y\Big) - F^2_{\mathbf{X}_t}\Big(y\Big)\Big)\Big] \tag{2.1.1.5}$$

Using conditions (C3)-(C5), the taylor expansion of $F_{\mathbf{X}_t}\Big(y\Big)$ about $F_{\mathbf{x}_i}\Big(y\Big)$ and the resulting terms involving the density about $g\Big(\mathbf{x}_i\Big)$, we get

$$E\Big[\eta_t^2\Big] = \Big(\Big|\mathbf{h}^{(i)}\Big|\Big)\mathbf{V}^2\Big(y\Big)g^2\Big(\mathbf{x}_i\Big) + o\Big(\Big|\mathbf{h}^{(i)}\Big|\Big). \tag{2.1.1.6}$$

Next, we show that the second term on the right hand side of (2.1.1.4) is of negligible magnitude,

$$\begin{aligned} &\leq 2\Big|\Big(\sqrt{n}\Big|\mathbf{h}^{(i)}\Big|\Big)^{-2}\sum_{t=2}^{n}\Big(1-\frac{t-1}{n}\Big)cov\Big(\eta_1, \eta_t\Big)\Big| \\ &\leq 2\Big(\sqrt{n}\Big|\mathbf{h}^{(i)}\Big|\Big)^{-2}\sum_{t=2}^{n}\Big|cov\Big(\eta_1, \eta_t\Big)\Big| \end{aligned} \tag{2.1.1.7}$$

By condition (B2) and lemma 1.1, (2.1.1.7) becomes

$$\begin{aligned} &\leq \Big(\sqrt{n}\Big|\mathbf{h}^{(i)}\Big|\Big)^{-2}\sum_{t=2}^{n} c\alpha\Big(t-1\Big)||\eta_1||_\infty||\eta_t||_\infty \\ &\leq \Big(\sqrt{n}\Big|\mathbf{h}^{(i)}\Big|\Big)^{-2}\sum_{t=2}^{n} c\alpha\Big(t-1\Big) \to 0, \end{aligned}$$

Where in the last part we have used condition (D2) and (E1). This together with (2.1.1.6) establishes (2.1.1.1).

To show (2.1.1.2) and (2.1.1.3) is straightforward, see also for example, in Robinson [96].

We therefore only give the leading terms in the necessary steps. The resulting covariance terms are approached in precisely the same way as above. Under conditions (B1-B5), (C1-C5) and (E1), the bias of the density estimator is obtained as

$$E\left[\widehat{g}\left(\mathbf{x}_i\right)\right] - g\left(\mathbf{x}_i\right) = \frac{\left|\left|\mathbf{h}^{(i)}\right|\right|^2}{2} \int \mathbf{u}^T \nabla^2 g\left(\mathbf{x}_i\right) \mathbf{u} \mathbf{K}\left(\mathbf{u}\right) d\mathbf{u} + o\left(\left|\left|\mathbf{h}^{(i)}\right|\right|^3\right) \qquad (2.1.1.8)$$

and the variance,

$$var\left[\widehat{g}\left(\mathbf{x}_i\right)\right] = \left(n\left|\mathbf{h}^{(i)}\right|\right)^{-1} g\left(\mathbf{x}_i\right) \int \mathbf{K}^2\left(\mathbf{u}\right) d\mathbf{u} + o\left(\left(n\left|\mathbf{h}^{(i)}\right|\right)^{-1}\right. \qquad (2.1.1.9)$$

The mean squared error of $\widehat{g}\left(\mathbf{x}_i\right)$ then becomes of the following order

$$MSE\left(\widehat{g}\left(\mathbf{x}_i\right)\right) = O\left(\left|\left|\mathbf{h}^{(i)}\right|\right|^4\right) + O\left(\left(n\left|\mathbf{h}^{(i)}\right|\right)^{-1}\right)$$

which goes to zero as n increases. Hence $\widehat{g}\left(\mathbf{x}_i\right) \to^p g\left(\mathbf{x}_i\right)$. In (2.1.1.3), we use similar lines as the preceeding steps to obtain the bias,

$$\begin{aligned} E\left[\widehat{r}_{\mathbf{x}_i}\left(y\right)\right] &- F_{\mathbf{x}_i}\left(y\right) g\left(\mathbf{x}_i\right) = \frac{1}{2}\left|\left|\mathbf{h}^{(i)}\right|\right|^2 F_{\mathbf{x}_i}\left(y\right) \int \mathbf{u}^T \nabla^2 g\left(\mathbf{x}_i\right) \mathbf{u} \mathbf{K}\left(\mathbf{u}\right) d\mathbf{u} \\ &+ \left|\left|\mathbf{h}^{(i)}\right|\right|^2 \nabla F_{\mathbf{x}_i}\left(y\right) \int \mathbf{u} \nabla g\left(\mathbf{x}_i\right)^T \mathbf{u} \mathbf{K}\left(\mathbf{u}\right) d\mathbf{u} \\ &+ \frac{1}{2} g\left(\mathbf{x}_i\right) \left|\left|\mathbf{h}^{(i)}\right|\right|^2 \int \mathbf{u}^T \nabla^2 F_{\mathbf{x}_i}\left(y\right) \mathbf{u} K\left(\mathbf{u}\right) d\mathbf{u} + o\left(\left|\left|\mathbf{h}^{(i)}\right|\right|^3\right) \end{aligned} \qquad (2.1.1.10)$$

and the same arguments as in the proof of (2.1.1.1) to get the variance,

$$var\left[\widehat{r}_{\mathbf{x}_i}\left(y\right)\right] = \left(n\left|\mathbf{h}^{(i)}\right|\right)^{-1} F_{\mathbf{x}_i}\left(y\right) g\left(\mathbf{x}_i\right) \int \mathbf{K}^2\left(\mathbf{u}\right) d\mathbf{u} + o\left(\left(n\left|\mathbf{h}^{(i)}\right|\right)^{-1}\right). \qquad (2.1.1.11)$$

The mean squared error for $\widehat{r}_{\mathbf{x}_i}\left(y\right)$ is of order $O\left(\left|\left|\mathbf{h}^{(i)}\right|\right|^4 + \left(n\left|\mathbf{h}^{(i)}\right|\right)^{-1}\right)$ and by condition (D1), it goes to zero with n and hence $\widehat{r}_{\mathbf{x}_i}\left(y\right) \to^p r_{\mathbf{x}_i}\left(y\right)$. □

Considering (2.1.0.8), both the numerator and denominator are consistent and hence by Slutsky's theorem, this is true also of $\frac{\widehat{r}_{\mathbf{x}_i}\left(y\right)}{\widehat{g}\left(\mathbf{x}_i\right)} = \widehat{F}_{\mathbf{x}_i}\left(y\right)$ for $\left|\left|\mathbf{h}^{(i)}\right|\right| + \left(n\left|\mathbf{h}^{(i)}\right|\right)^{-1} \to 0$, as $n \to \infty$. Hence, $\widehat{F}_{\mathbf{x}_i}\left(y\right)$ is a consistent estimator for $F_{\mathbf{x}_i}\left(y\right)$, i.e $\widehat{F}_{\mathbf{x}_i}\left(y\right) \to^p F_{\mathbf{x}_i}\left(y\right)$. This result will be used to obtain the bias for $\widehat{F}_{\mathbf{x}_i}\left(y\right)$ for subsequent use throughout this chapter.

The definition of $\mathbf{V}^2\left(y\right)$ will be used throughout this chapter.

The following lemma gives bias and variance for $\widehat{F}_{\mathbf{x}_i}\left(y\right)$.

Lemma 2.2 *Suppose the conditions in lemma 2.1 hold. Then*

$$E\left(\widehat{F}_{\mathbf{x}_i}\left(y\right) - F_{\mathbf{x}_i}\left(y\right)\right) = B_n\left(y\right) + o\left(\left|\left|\mathbf{h}^{(i)}\right|\right|^3\right) \tag{2.1.1.12}$$

and the variance

$$var\left[\widehat{F}_{\mathbf{x}_i}\left(y\right)\right] = \left(n\left|\mathbf{h}^{(i)}\right|\right)^{-1}\mathbf{V}^2\left(y\right) + o\left(\left(n\left|\mathbf{h}^{(i)}\right|\right)^{-1}\right) \tag{2.1.1.13}$$

where

$$\begin{aligned} B_n\left(y\right) &= \frac{\left|\left|\mathbf{h}^{(i)}\right|\right|^2}{g\left(\mathbf{x}_i\right)}\Big\{\nabla F_{\mathbf{x}_i}\left(y\right)^T \int \mathbf{u}\nabla g\left(\mathbf{x}_i\right)^T \mathbf{u}K\left(\mathbf{u}\right)d\mathbf{u} \\ &+ \frac{1}{2}g\left(\mathbf{x}_i\right)\int \mathbf{u}^T\nabla^2 F_{\mathbf{x}_i}\left(y\right)\mathbf{u}K\left(\mathbf{u}\right)d\mathbf{u}\Big\} \end{aligned}$$

Proof of lemma 2.2 :

Because the numerator and denomenator of (2.1.0.8) are stochastic, we proceed by linearizing the estimator;

$$\widehat{F}_{\mathbf{x}_i}\left(y\right) = F_{\mathbf{x}_i}\left(y\right) + \frac{\widehat{r}_{\mathbf{x}_i}\left(y\right) - F_{\mathbf{x}_i}\left(y\right)\widehat{g}\left(\mathbf{x}_i\right)}{g\left(\mathbf{x}_i\right)} + \frac{1}{g\left(\mathbf{x}_i\right)}\left(\widehat{F}_{\mathbf{x}_i}\left(y\right) - F_{\mathbf{x}_i}\left(y\right)\right)\left(g\left(\mathbf{x}_i\right) - \widehat{g}\left(\mathbf{x}_i\right)\right) \tag{2.1.1.14}$$

From lemma 2.1, the consistency of $\widehat{r}_{\mathbf{x}_i}\left(y\right)$ and $\widehat{g}\left(\mathbf{x}_i\right)$ implies that for large n and $\left|\mathbf{h}^{(i)}\right| \to 0$, we have $\widehat{F}_{\mathbf{x}_i}\left(y\right) - F_{\mathbf{x}_i}\left(y\right) = o_p\left(1\right)$. And because $g\left(\mathbf{x}_i\right) - \widehat{g}\left(\mathbf{x}_i\right) = O_p\left(\left|\left|\mathbf{h}^{(i)}\right|\right|^2\right)$, the product of the two quantities is of smaller order in probability. Hence by using (2.1.1.8) and (2.1.1.10) in

$$E\left[\widehat{F}_{\mathbf{x}_i}\left(y\right)\right] - F_{\mathbf{x}_i}\left(y\right) = \frac{E\left[\widehat{r}_{\mathbf{x}_i}\left(y\right)\right] - F_{\mathbf{x}_i}\left(y\right)E\left[\widehat{g}\left(\mathbf{x}_i\right)\right]}{g\left(\mathbf{x}_i\right)} + o_p\left(\left|\left|\mathbf{h}^{(i)}\right|\right|^2\right) \tag{2.1.1.15}$$

results in (2.2). And by using (2.1.1.9), (2.1.1.11) and (2.1.1.14) in

$$var\left[\widehat{F}_{\mathbf{x}_i}\left(y\right)\right] \approx \frac{1}{g^2\left(\mathbf{x}_i\right)}\left(var\left[\widehat{r}_{\mathbf{x}_i}\left(y\right)\right] + F^2_{\mathbf{x}_i}\left(y\right)var\left[\widehat{g}\left(\mathbf{x}_i\right)\right] - 2F_{\mathbf{x}_i}\left(y\right)cov\left[\widehat{r}_{\mathbf{x}_i}\left(y\right), \widehat{g}\left(\mathbf{x}_i\right)\right]\right) \tag{2.1.1.16}$$

where

$$\begin{aligned} cov\Big[\widehat{r}_{\mathbf{x}_i}\Big(y\Big),\widehat{g}\Big(\mathbf{x}_i\Big)\Big] &= \Big(n\Big|\mathbf{h}^{(i)}\Big|\Big)^{-2}\sum_{t=1}^{n} cov\Big[\mathbf{K}\Big(\mathbf{x}_i-\mathbf{X}_t;\mathbf{h}^{(i)}\Big)\mathbf{I}_{\{Y_t\le y\}},\mathbf{K}\Big(\mathbf{x}_i-\mathbf{X}_t;\mathbf{h}^{(i)}\Big)\Big] \\ &+ \Big(n\Big|\mathbf{h}^{(i)}\Big|\Big)^{-2}\sum_{t\neq t'}^{n} cov\Big[\mathbf{K}\Big(\mathbf{x}_i-\mathbf{X}_t;\mathbf{h}^{(i)}\Big)\mathbf{I}_{\{Y_t\le y\}},\mathbf{K}\Big(\mathbf{x}_i-\mathbf{X}_{t'};\mathbf{h}^{(i)}\Big)\Big] \\ &\approx \Big(\sqrt{n}\Big|\mathbf{h}^{(i)}\Big|\Big)^{-2}\Big\{E\Big[\mathbf{K}^2\Big(\mathbf{x}_i-\mathbf{X}_t;\mathbf{h}^{(i)}\Big)\mathbf{I}_{\{Y_t\le y\}}\Big] \\ &- E\Big[\mathbf{K}\Big(\mathbf{x}_i-\mathbf{X}_t;\mathbf{h}^{(i)}\Big)\mathbf{I}_{\{Y_t\le y\}}\Big]E\Big[\mathbf{K}^2\Big(\mathbf{x}_i-\mathbf{X}_t;\mathbf{h}^{(i)}\Big)\Big]\Big\}+A\Big(\mathbf{x}_i\Big) \\ &\approx \Big(n\Big|\mathbf{h}^{(i)}\Big|\Big)^{-1}F_{\mathbf{x}_i}\Big(y\Big)g\Big(\mathbf{x}_i\Big)\int\mathbf{K}^2\Big(\mathbf{u}\Big)d\mathbf{u}+o\Big(\Big(n\Big|\mathbf{h}^{(i)}\Big|\Big)^{-1}\Big) \end{aligned} \tag{2.1.1.17}$$

with $A\Big(\mathbf{x}_i\Big)$ evaluated as in lemma 2.1, we obtain (2.1.1.13).

□

From lemma 2.2, the order of the mean squared error for $\widehat{F}_{\mathbf{x}_i}\Big(y\Big)$ is again $O\Big(\Big|\Big|\mathbf{h}^{(i)}\Big|\Big|^4+(n\Big|\mathbf{h}^{(i)}\Big|\Big)^{-1}\Big)$, and so for $n\to\infty$ we have $MSE\Big(\widehat{F}_{\mathbf{x}_i}\Big(y\Big)\Big)\to 0$. Hence $\widehat{F}_{\mathbf{x}_i}\Big(y\Big)\to^p F_{\mathbf{x}_i}\Big(y\Big)$ with a rate implying consistency, as in lemma 2.1.

Note that the bias is quadratic in $\Big|\Big|\mathbf{h}^{(i)}\Big|\Big|$ and therefore the sequence $\Big|\Big|\mathbf{h}^{(i)}\Big|\Big|$ have to be small to reduce it. On the other hand the variance is proportional to $\Big(n\Big|\mathbf{h}^{(i)}\Big|\Big)^{-1}$ and large bandwidth would be prefered. Hence to get a compromise of both effects, we choose $\Big|\Big|\mathbf{h}^{(i)}\Big|\Big|$ such that $n\Big|\mathbf{h}^{(i)}\Big|\Big|\Big|\mathbf{h}^{(i)}\Big|\Big|^4=c$, where c is a positive constant. In particular, for equal bandwidths, we have $\Big|\Big|\mathbf{h}^{(i)}\Big|\Big|=ch$ and h has to be choosen so that $h=cn^{-\frac{1}{4+d}}$.

The definition of $B_n\Big(y\Big)$ will be used throughout this chapter.

Using lemma 2.2, we now present the asymptotic normality of the conditional distribution estimator results in theorem 2.1 below.

Theorem 2.1 (Asymptotic normality) *Under the conditions (B1)-(B5), (C1)-(C5), (D1)-(D2) and (E1),*

$$\Big(n\Big|\mathbf{h}^{(i)}\Big|\Big)^{\frac{1}{2}}\Big[\widehat{F}_{\mathbf{x}_i}\Big(y\Big)-F_{\mathbf{x}_i}\Big(y\Big)-B_n\Big(y\Big)+o_p\Big(\Big|\Big|\mathbf{h}^{(i)}\Big|\Big|^2\Big)\Big]\to^D N\Big(0,\mathbf{V}^2\Big(y\Big)\Big) \tag{2.1.1.18}$$

where $\mathbf{V}^2\Big(y\Big)$ *and* $B_n\Big(y\Big)$ *are the variance and bias defined in lemmas 2.1 and 2.2 respectively.*

Proof of theorem 2.1 :

We observe that the remainder after adjusting for the bias is

$$\left(n\left|\mathbf{h}^{(i)}\right|\right)^{\frac{1}{2}}\left[\widehat{F}_{\mathbf{x}_i}\left(y\right)-F_{\mathbf{x}_i}\left(y\right)-B_n\left(y\right)+o_p\left(\left|\left|\mathbf{h}^{(i)}\right|\right|^2\right)\right]=\left(n\left|\mathbf{h}^{(i)}\right|\right)^{\frac{1}{2}}\frac{W}{g\left(\mathbf{x}_i\right)}+o_p\left(1\right) \tag{2.1.1.19}$$

Let $J_1=\left(n\left|\mathbf{h}^{(i)}\right|\right)^{\frac{1}{2}}W=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\overline{\eta}_t$, where $\overline{\eta}_t=\left(\left|\mathbf{h}^{(i)}\right|\right)^{-\frac{1}{2}}\eta_t$. It suffice to establish the asymptotic normality of J_1 by (2.1.1.19). We employ Doob's small-block and large-block techniques (see Ibragimov and Linnik, 1971, page 316). That is, we partition $\{1,2,\ldots,n\}$ into $2b_n+1$ subset with large-block of size $r=r_n$ and small block of size $s=s_n$. Let $b=b_n=INT\left(\frac{n}{r_n+s_n}\right)$, where $INT\left(x\right)$ denotes the integer part of x. Define the random variables, for $0\leq j\leq b-1$,

$$\nu_j=\sum_{i=j(r+s)}^{j(r+s)+r-1}\overline{\eta}_i,\qquad \gamma_j=\sum_{i=j(r+s)+r}^{(j+1)(r+s)-1}\overline{\eta}_i,\qquad \text{and}\quad \nu_b=\sum_{i=b(r+s)}^{n-1}\overline{\eta}_i$$

Then

$$\begin{aligned} J_1 &= \frac{1}{\sqrt{n}}\Big\{\sum_{j=0}^{b-1}\nu_j+\sum_{j=0}^{b-1}\gamma_j+\nu_b\Big\} \\ &= \frac{1}{\sqrt{n}}\Big\{T_{n,1}+T_{n,2}+T_{n,3}\Big\} \end{aligned}$$

We then need to show the following, as $n\to\infty$:

1. The sum over the residual blocks $T_{n,2}$ and $T_{n,3}$ are asymptotically negligible in probability,

2. The summands ν_j in large blocks, $T_{n,1}$ are asymptotically independent and

3. The standard Lindeberg-Feller conditions for asymptotic normality of $T_{n,1}$ under independence assumption hold.

To accomplish this, we define the large-block size r_n by $r_n=INT\left(\left(n\left|\mathbf{h}^{(i)}\right|\right)^{\frac{1}{2}}\right)$ and the small-block size $s_n=INT\left(\frac{\left(n\left|\mathbf{h}^{(i)}\right|\right)^{\frac{1}{2}}}{\log n}\right)$. Then as $n\to\infty$,

$$\frac{s_n}{r_n}\to 0\qquad\text{and}\qquad\frac{n}{r_n}\alpha\left(s_n\right)\to 0 \tag{2.1.1.20}$$

Now, to establish (1), we need to show that

$$\frac{1}{n}E\Big[T_{n,2}\Big]^2 \to 0, \qquad \frac{1}{n}E\Big[T_{n,3}\Big]^2 \to 0, \quad \text{as} \quad n \to \infty \tag{2.1.1.21}$$

The variance of $T_{n,2}$ is

$$E\Big[T_{n,2}\Big]^2 = \sum_{j=0}^{b-1} var(\gamma_j) + 2 \sum_{0 \le i < j \le b-1} cov\Big(\gamma_i, \gamma_j\Big) \equiv T_{21} + T_{22} \tag{2.1.1.22}$$

From stationarity and lemma 2.1, it follows that

$$T_{21} = b_n var\Big(\gamma_1\Big) = b_n var\Big(\sum_{j=1}^{s_n} \overline{\eta}_j\Big) = b_n s_n \Big[\mathbf{V}^2\Big(y\Big) g^2\Big(\mathbf{x}_i\Big) + o\Big(1\Big)\Big] \tag{2.1.1.23}$$

Consider the second term T_{22} on the right -hand side of (2.1.1.22). Let $r_j^* = j(r_n + s_n)$, then $r_j^* - r_i^* \ge r_n$ for all $j > i$ and therefore

$$\begin{aligned} T_{22} &= 2 \sum_{0 \le i < j \le b-1} cov\Big(\sum_{j1=1}^{s} \overline{\eta}_{r^*_{i+r+j1}}, \sum_{j2=1}^{s} \overline{\eta}_{r^*_{j+r+j2}}\Big) \\ &\le 2 \sum_{0 \le i < j \le b-1} \sum_{j1}^{s_n} \sum_{j2}^{s_n} \Big| cov\Big(\overline{\eta}_{r_i^*+r_n+j1}, \overline{\eta}_{r_j^*+r_n+j2}\Big) \Big| \\ &\le 2 \sum_{j1}^{n-r_n} \sum_{j2=j1+rn}^{n} \Big| cov\Big(\overline{\eta}_{j1}, \overline{\eta}_{j2}\Big) \Big|. \end{aligned}$$

By stationarity and lemma 2.1, we have

$$\Big|T_{22}\Big| \le 2n \sum_{j=r_n+1}^{n} \Big| cov\Big(\overline{\eta}_1, \overline{\eta}_j\Big) \Big| + o\Big(n\Big) \tag{2.1.1.24}$$

Hence, by (2.1.1.20)-(2.1.1.24), we have

$$\frac{1}{n}E\Big[T_{n,2}\Big]^2 = O\Big(b_n s_n n^{-1}\Big) + o\Big(1\Big) \tag{2.1.1.25}$$

It follows from stationarity, (2.1.1.20) and lemma 2.1 that

$$var\Big[T_{n,3}\Big] = var\left(\sum_{j=1}^{n-b_n(r_n+s_n)} \overline{\eta}_j\right) = O\Big(n - b_n(r_n + s_n)\Big) = o\Big(n\Big). \tag{2.1.1.26}$$

and therefore $\frac{1}{n}E\Big[T_{n,3}\Big] \to 0$. Combining (2.1.1.20)-(2.1.1.26), (1) is established.
To show (2) we need to show that

$$\Big| E\Big[\exp\Big(itT_{n,1}\Big)\Big] - \prod_{j=0}^{b-1} E\Big[\exp\Big(it\nu_j\Big)\Big] \Big| \to 0, \quad \text{as} \quad n \to \infty$$

We use (2.1.1.20) and lemma 1.2 to obtain

$$\left|E\left[\exp\left(itT_{n,1}\right)\right] - \prod_{j=0}^{b-1} E\left[\exp\left(it\nu_j\right)\right]\right| 16\left(\frac{n}{r_n}\right)\alpha\left(s_n\right)$$

which tend to zero by (2.1.1.20).

To establsh the third statement, we need to show that

$$\frac{1}{n}\sum_{j=0}^{b-1} E\left(\nu^2\right) \to \mathbf{V}^2\left(y\right)g^2\left(\mathbf{x}_i\right), \qquad \frac{1}{n}\sum_{j=0}^{b-1} E\left[\nu_j^2 \mathbf{I}_{\{|\nu_j| \geq \epsilon \mathbf{V}^2(y) g^2(\mathbf{x}_i)\sqrt{n}\}}\right] \to 0$$

for every $\epsilon > 0$. By stationarity, (A12) and lemma 2.1, we obtain

$$\frac{1}{n}\sum_{j=0}^{b_n-1} E\left(\nu_j^2\right) = \frac{b_n}{n} E\left(\nu_1^2\right) = \frac{b_n r_n}{n} . \frac{1}{r_n} var\left(\sum_{j=1}^{r_n} \overline{\eta}_j\right) \to \mathbf{V}^2\left(y\right)g^2\left(\mathbf{x}_i\right)$$

To establish the last part, we employ a truncation argument as follows. Let $U_{L,i} = U_i \mathbf{I}_{\{|U_i| \leq L\}}$, where L is a fixed truncation point. Correspondingly, let us denote the superscript L to indicate the quantities that involve $\left\{U_{L,i}\right\}$ instead of $\left\{U_i\right\}$. Then $J_1 = J_1^L + \widetilde{J}_1^L$ where $\widetilde{J}_1^L = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\overline{\eta}_t - \overline{\eta}_t^L\right)$. Since $\mathbf{K}$ is bounded with compact support, we have

$$\begin{aligned} \left|\overline{\eta}_t^L\right| &= \left(\left|\mathbf{h}^{(i)}\right|\right)^{-\frac{1}{2}} \left|\mathbf{K}\left(\mathbf{x}_i - \mathbf{X}_t; \mathbf{h}^{(i)}\right)\left(\mathbf{I}_{\{Y_t \leq y\}} - F_{\mathbf{X}_t}\left(y\right)\right)\right. \\ &\leq c\left(\left|\mathbf{h}^{(i)}\right|\right)^{-\frac{1}{2}}, \end{aligned}$$

for some constant c. Then using (2.1.1.20), it follows that $\max_{0 \leq t \leq b-1} \frac{1}{\sqrt{n}}\left|\overline{\eta}_t^L\right| \leq c r_n \left(n\left|\mathbf{h}^{(i)}\right|\right)^{-\frac{1}{2}} \to 0$. Therefore, for $n \to \infty$, the set $\left\{\left|\overline{\eta}_t^L\right| \geq \epsilon \mathbf{V}_L\left(y\right)g\left(\mathbf{x}_i\right)\sqrt{n}\right\}$ becomes an empty set and hence $\frac{1}{n}\sum_{j=1}^{b-1} E\left[\nu_j^2 \mathbf{I}_{\{|\nu_j| > \epsilon \mathbf{V}_L(y) g(\mathbf{x}_i)\sqrt{n}\}}\right] \to 0$.

□

2.1.2 The asymptotic properties of the QAR function estimator

This subsection establishes the asymptotic normality of the QAR estimator, $\widehat{\mu}_\theta\left(\mathbf{x}_i\right)$.

Lemma 2.3 *Assume conditions (B1)-(B5),(C1)-(C6) and (E1). Then for* $\delta_n \to 0$, *we have*

$$\widehat{F}_{\mathbf{x}_i}\left(y + \delta_n\right) - \widehat{F}_{\mathbf{x}_i}\left(y\right) = \delta_n f_{\mathbf{x}_i}\left(y\right) + o_p\left(\delta_n\right) + o_p\left(\left(n\left|\mathbf{h}^{(i)}\right|\right)^{-1}\right) \qquad (2.1.2.1)$$

Proof of lemma 2.3 :

Express the left hand side of (2.1.2.1) as

$$\widehat{F}_{\mathbf{x}_i}\Big(y+\delta_n\Big)-\widehat{F}_{\mathbf{x}_i}\Big(y\Big)=\Big(n\Big|\mathbf{h}^{(i)}\Big|\Big)^{-1}\frac{\sum_{t=1}^{n}\mathbf{K}\Big(\mathbf{x}_i-\mathbf{X}_t;\mathbf{h}^{(i)}\Big)\Big(\mathbf{I}_{\{Y_t\le y+\delta_n\}}-\mathbf{I}_{\{Y_t\le y\}}\Big)}{\widehat{g}\Big(\mathbf{x}_i\Big)} \tag{2.1.2.2}$$

We first use equation (2.1.1.14) to simplify (2.1.2.2). Then follow the same arguments as in lemma 2.2, by taking the expectation on both sides. Lastly, expand the resulting $F_{\mathbf{x}_i}\Big(y+\delta_n\Big)$ and other terms involving δ_n about their corresponding functions of y. We arrive at $E\Big[\widehat{F}_{\mathbf{x}_i}\Big(y+\delta_n\Big)-\widehat{F}_{\mathbf{x}_i}\Big(y\Big)\Big]=\delta_n f_{\mathbf{x}_i}\Big(y\Big)+o\Big(\delta_n\Big|\Big|\mathbf{h}^{(i)}\Big|\Big|\Big)$. Similarly the variance of (2.1.2.2) becomes $var\Big[\widehat{F}_{\mathbf{x}_i}\Big(y+\delta_n\Big)-\widehat{F}_{\mathbf{x}_i}\Big(y\Big)\Big]=O\Big(\delta_n\Big(n\Big|\mathbf{h}^{(i)}\Big|\Big)^{-1}\Big)$. The mean squared error goes to zero as $n\to\infty$, hence (2.1.2.2) holds. □

Note from this lemma that, we can approximate the variance of $\widehat{F}_{\mathbf{x}_i}\Big(y+\delta_n\Big)$ by the variance of the sum of $\widehat{F}_{\mathbf{x}_i}\Big(y\Big)$ and $\delta_n f_{\mathbf{x}_i}\Big(y\Big)$. This fact will become important in deriving the asymptotic properties of the scale functional estimator.

Theorem 2.2 *Assume conditions (B1)-(B5), (C1)-(C6), (D1)-(D2) and (E1) hold. Then* $\widehat{\mu}_\theta\Big(\mathbf{x}_i\Big)$ *is consistent*

$$\widehat{\mu}_\theta\Big(\mathbf{x}_i\Big)\to^p \mu_\theta\Big(\mathbf{x}_i\Big). \tag{2.1.2.3}$$

Furthermore if conditions (C7) hold, then it is asymptotically normal:

$$\Big(n\Big|\mathbf{h}^{(i)}\Big|\Big)^{\frac{1}{2}}\Bigg(\widehat{\mu}_\theta\Big(\mathbf{x}_i\Big)-\mu_\theta\Big(\mathbf{x}_i\Big)-B\Big(\mu_\theta\Big(\mathbf{x}_i\Big)\Big)\Bigg)\to^D N\left(0,\frac{\mathbf{V}^2\Big(\mu_\theta\Big(\mathbf{x}_i\Big)\Big)}{f^2_{\mathbf{x}_i}\Big(\mu_\theta\Big(\mathbf{x}_i\Big)\Big)}\right), \tag{2.1.2.4}$$

where

$$B\Big(\mu_\theta\Big(\mathbf{x}_i\Big)\Big)=-\frac{B_n\Big(\mu_\theta\Big(\mathbf{x}_i\Big)\Big)}{f_{\mathbf{x}_i}\Big(\mu_\theta\Big(\mathbf{x}_i\Big)\Big)}. \tag{2.1.2.5}$$

Proof of theorem 2.2 :

First we proof (2.1.2.3). From lemmas 2.1 and 2.2, we have for all $\mathbf{x}_i\in\mathbf{R}^d$ and y,
$\widehat{F}_{\mathbf{x}_i}\Big(y\Big)\to F_{\mathbf{x}_i}\Big(y\Big)$, in probability.
Because $F_{\mathbf{x}_i}\Big(y\Big)$ is a distribution function it follows from Glivenko-Cantelli theorem, in Krishnaiah [78], for generalized[25] empirical processes based on strong mixing sequences

[25] Where in the present case we choose $C_t=\frac{1}{n}$, for $t=1,\ldots,n$.

that

$$\sup_{y\in\mathbf{R}}\left|\widehat{F}_{\mathbf{x}_i}\left(y\right)\rightarrow F_{\mathbf{x}_i}\left(y\right)\right|\rightarrow 0,\quad \text{in probability} \tag{2.1.2.6}$$

The uniqueness assumption of $\mu_\theta\left(\mathbf{x}_i\right)$ implies that, for any fixed $\mathbf{x}_i \in \mathbf{R}^d$, there exist a $\epsilon > 0$ and $\delta\left(\epsilon\right) > 0$ such that

$$\delta=\delta\left(\epsilon\right)=\min\Big\{\theta-F_{\mathbf{x}_i}\Big(\mu_\theta\left(\mathbf{x}_i\right)-\epsilon\Big),F_{\mathbf{x}_i}\Big(\mu_\theta\left(\mathbf{x}_i\right)+\epsilon\Big)-\theta\Big\}>0$$

This implies that

$$\begin{aligned}P\Big\{\left|\widehat{\mu}_\theta\left(\mathbf{x}_i\right)-\mu_\theta\left(\mathbf{x}_i\right)\right|>\epsilon\Big\} &\leq P\Big\{\left|F_{\mathbf{x}_i}\Big(\widehat{\mu}_\theta\left(\mathbf{x}_i\right)\Big)-F_{\mathbf{x}_i}\Big(\mu_\theta\left(\mathbf{x}_i\right)\Big)\right|>\delta\Big\}\\ &\leq P\Big\{\left|F_{\mathbf{x}_i}\Big(\widehat{\mu}_\theta\left(\mathbf{x}_i\right)\Big)-\widehat{F}_{\mathbf{x}_i}\Big(\widehat{\mu}_\theta\left(\mathbf{x}_i\right)\Big)\right|>\delta-\frac{1}{n}\Big\}\\ &\leq P\Big\{\sup_y\left|\widehat{F}_{\mathbf{x}_i}\left(y\right)-F_{\mathbf{x}_i}\left(y\right)\right|>\delta'\Big\}\end{aligned} \tag{2.1.2.7}$$

for arbitrary $\delta' < \delta$ and n large enough. Here, we used $F_{\mathbf{x}_i}\Big(\mu_\theta\left(\mathbf{x}_i\right)\Big)=\theta$ and $\theta \leq \widehat{F}_{\mathbf{x}_i}\Big(\widehat{\mu}_\theta\left(\mathbf{x}_i\right)\Big) \leq \theta+\frac{1}{n}$. (2.1.2.7) tends to zero by (2.1.2.6). Hence (2.1.2.3) holds true. To prove (2.1.2.4),
let $b_n=-\frac{B_n\left(\mu_\theta\left(\mathbf{x}_i\right)\right)}{f_{\mathbf{x}_i}\left(\mu_\theta\left(\mathbf{x}_i\right)\right)}$ and $v=\frac{\mathbf{V}\left(\mu_\theta\left(\mathbf{x}_i\right)\right)}{f_{\mathbf{x}_i}\left(\mu_\theta\left(\mathbf{x}_i\right)\right)}$. For any w

$$\begin{aligned}q_n\left(w\right) &= P\Big(\frac{\widehat{\mu}_\theta\left(\mathbf{x}_i\right)-\mu_\theta\left(\mathbf{x}_i\right)-b_n}{v}\leq w\Big)\\ &= P\Big(\widehat{\mu}_\theta\left(\mathbf{x}_i\right)\leq\mu_\theta\left(\mathbf{x}_i\right)+b_n+vw\Big)\end{aligned}$$

As $\widehat{F}_{\mathbf{x}_i}\left(y\right)$ is increasing, but not necessarily strictly, we have

$$\begin{aligned}P\Big(F_{\mathbf{x}_i}\Big(\widehat{\mu}_\theta\left(\mathbf{x}_i\right)\Big)<\widehat{F}_{\mathbf{x}_i}\Big(\mu_\theta\left(\mathbf{x}_i\right)+b_nv+w\Big)\Big) &\leq q_n\left(w\right)\\ &\leq P\Big(F_{\mathbf{x}_i}\Big(\widehat{\mu}_\theta\left(\mathbf{x}_i\right)\Big)\leq\widehat{F}_{\mathbf{x}_i}\Big(\mu_\theta\left(\mathbf{x}_i\right)+b_nv+w\Big)\Big)\end{aligned}$$

By the same argument as in (2.1.2.7), we may replace $\widehat{F}_{\mathbf{x}_i}\Big(\widehat{\mu}_\theta\left(\mathbf{x}_i\right)\Big)$ by $F_{\mathbf{x}_i}\Big(\mu_\theta\left(\mathbf{x}_i\right)\Big)$ up to an error of n^{-1} at most, and we get, neglecting the n^{-1}-term which is asymptotically negligible anyhow,

$$\begin{aligned}q_n\left(w\right) &\approx P\Big(F_{\mathbf{x}_i}\Big(\mu_\theta\left(\mathbf{x}_i\right)\Big)\leq\widehat{F}_{\mathbf{x}_i}\Big(\mu_\theta\left(\mathbf{x}_i\right)+b_n+vw\Big)\\ &\approx P\Big(-f_{\mathbf{x}_i}\Big(\mu_\theta\left(\mathbf{x}_i\right)\Big).\delta_n\leq\widehat{F}_{\mathbf{x}_i}\Big(\mu_\theta\left(\mathbf{x}_i\right)\Big)-F_{\mathbf{x}_i}\Big(\mu_\theta\left(\mathbf{x}_i\right)\Big)\Big)\end{aligned} \tag{2.1.2.8}$$

with $\delta_n = b_n + vw$, where we have used lemma 2.3 and neglected the $o\left(\delta_n\right)$ and $o\left(\left(n\left|\mathbf{h}^{(i)}\right|\right)^{-1}\right)$. Using theorem 2.1 with $y_\theta = \mu_\theta\left(\mathbf{x}_i\right)$, we get

$$\begin{aligned} q_n\left(w\right) &\sim P\left(\left(n\left|\mathbf{h}^{(i)}\right|\right)^{\frac{1}{2}} \frac{\widehat{F}_{\mathbf{x}_i}\left(y_\theta\right) - F_{\mathbf{x}_i}\left(y_\theta\right) - B_n\left(y_\theta\right)}{\mathbf{V}\left(y_\theta\right)} \right. \\ &\geq \left. \frac{-f_{x_i}\left(y_\theta\right)\delta_n - B_n\left(y_\theta\right)}{\mathbf{V}\left(y_\theta\right)} \left(n\left|\mathbf{h}^{(i)}\right|\right)^{\frac{1}{2}}\right) \\ &\sim \Phi\left(\left(n\left|\mathbf{h}^{(i)}\right|\right)^{\frac{1}{2}} \frac{f_{\mathbf{x}_i}\left(y_\theta\right).\left(b_n + vw\right) + B_n\left(y_\theta\right)}{\mathbf{V}\left(y_\theta\right)}\right) \\ &= \Phi\left(w\right) \end{aligned}$$

by our choice of b_n and v. This proofs the theorem. □

This result could be used to construct confidence interval for the estimators as well as other relevant inferences. In the estimation of the scale function, both asymptotic properties for the conditional distribution and QAR estimator will be important. We will use the definition of $B\left(\mu_\theta\left(\mathbf{x}_i\right)\right)$ throughout.

2.2 Bandwidth selection

In nonparametric kernel estimation, the bandwidth play an important role in the behaviour of the estimates. This can be seen for example in theorem 2.1, where the consistency of the estimators are basically based on the sum of the bias and variance. Since the bias is proportional to $\left|\left|\mathbf{h}^{(i)}\right|\right|^2$ and variance proportional to $\left(\left|\mathbf{h}^{(i)}\right|\right)^{-1}$, the bandwidth has to be taken neither too large nor too small so as not to increase respective bias and variance of the estimates. The problem can be solved theoretically by choosing a bandwidth that balances the trade-off between the bias and variance components. For instance, if we assumed that all bandwidths are equal and set $\left|\left|\mathbf{h}^{(i)}\right|\right| = ch$, the theoretical optimal bandwidth can be taken to be proportional to $n^{-\frac{1}{d+4}}$, as seen in the proof of lemma 2.2. The importance of the appropriate bandwidth is illustrated by the following example. Consider the data generated from the autoregressive process,

$$Y_t = \mu\Big(X_t\Big) + \sigma\Big(X_t\Big)e_t, \quad t = 1, \ldots, 500 \tag{2.2.0.9}$$

where e_t are independent random variables from student-t distribution with 4 degrees of freedom. Let the true quantile autoregression function $\mu_\theta\Big(X_t\Big)$ at $\theta = 0.75$ be

$$\mu_{0.75}\Big(X_t\Big) = \begin{cases} 4 + 3X_t + \frac{1}{X_t}\phi\Big(2\log X_t - \frac{3}{2}\Big) + \Big(0.007 + 2X_t^2\Big)q^e_{0.75} & : \quad X_t > 0 \\ 0 & : \quad \text{Otherwise} \end{cases} \tag{2.2.0.10}$$

with ϕ being standard normal density. The data is shown in figure 4(a). In order to estimate (2.2.0.10) nonparametrically, we used a bisquare (biweight) kernel function,

$$K\Big(u\Big) = \frac{15}{16}\Big(1 - u^2\Big)^2_+$$

The effect of the bandwidth on the behavior of the estimate are shown in figures 4(b-d). Figure 4(b) shows the graph of true function $\mu_{0.75}\Big(X_t\Big)$ defined by (2.2.0.10) with its kernel estimate at $\theta = 0.75$ and constant bandwidth $h = 0.013$. Clearly the bandwidth used is too small and the estimate is undersmoothed; the estimate has a marked variance. Figure 4(c) shows the same curves, but with the bandwidth taken to be $h = 0.065$. In this case the bandwidth was too big, resulting in oversmoothed estimation; the corresponding estimate has high bias. Figure 4(d) shows what happens when the bandwidth is $h = 0.031$. This value of the smoothing parameter gives a fairly better estimation, than the other two, because it tends to balance the effects of variance and bias (or between undersmoothing and oversmoothing respectively).

Obviously, practical situations require automatic determination of the appropriate smoothing parameter, as a bad choice may lead to poor estimation (see again figure (4)). The bandwidth that depends only on data and is easy to compute (a data-driven bandwidth) is prefered. One of the practical methods that is used to solve the smoothness problem, is the cross-validation procedure. This aims at finding a data-driven smoothing parameter (bandwidth) that asymptotically minimizes some loss function of the error. In particular Haerdle and Marron [59] and Haerdle et al. [60], have used the square loss

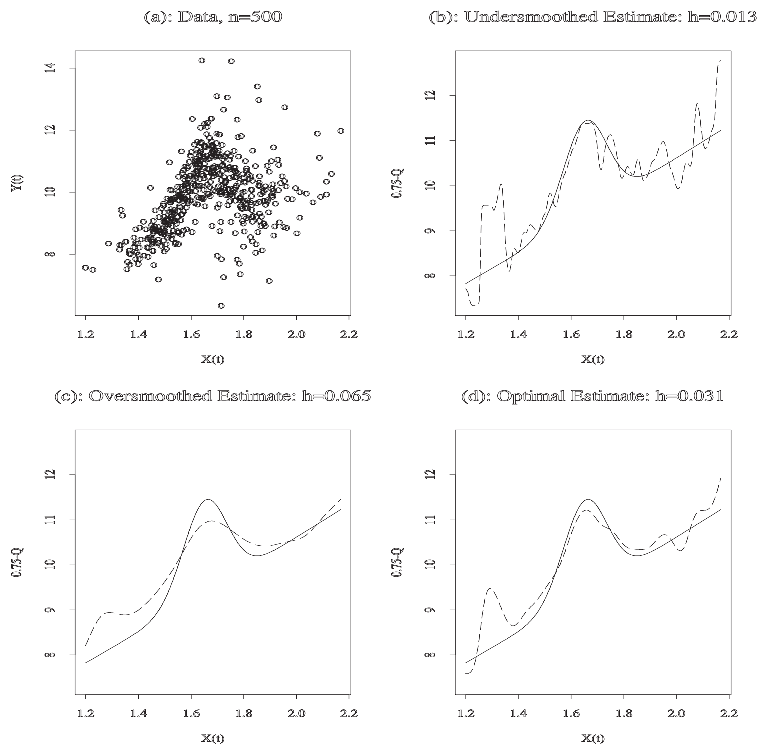

Figure 4: Bandwidth problem

function to investigate the parameters in the setting of independent variables and Haerdle et al. [61] under dependence.

2.2.1 Cross-validation

For α-mixing $\Big\{Y_t, \mathbf{X}_t\Big\}_{t=1}^{n}$, a pair of random variables in $\mathbf{R}^{1+d}$ and the conditional distribution estimator $\widehat{F}_{\mathbf{x}_i}\Big(y\Big)$, Abbegger [1] proposed a bandwidth selection procedure that minimizes an expression of the form

$$CV\Big(\mathbf{h}^{(i)}\Big) = \frac{1}{n - n(w)} \sum_{t=1}^{n} M_\theta\Big(Y_t, \widehat{\mu}_\theta^{(-t)}\Big(\mathbf{X}_i\Big)\Big) w\Big(\mathbf{X}_t\Big).$$

where $w : \mathbf{R}^d \to \mathbf{R}^d$ is some nonnegative weight function used to omit observation at the boundaries, and $n(w)$, the number of observation that take values of zero in $w(\mathbf{X}_t)$ and $\widehat{\mu}_\theta^{(-t)}\Big(\mathbf{X}_i\Big)$, is the leave one(block) out estimate obtained as

$$\widehat{\mu}_\theta^{(-t)}\Big(\mathbf{X}_i\Big) = \inf\Big\{y \in \mathbf{R} \Big| \widehat{F}_{\mathbf{X}_i}^{(-t)}\Big(y\Big) \geq \theta\Big\}, \quad 0 < \theta < 1$$

where

$$\widehat{F}^{(-t)}_{\mathbf{X}_i}\Big(y\Big) = \frac{\sum_{|t'-t|>d+bl_n} \mathbf{K}\Big\{\Big(\mathbf{X}_i - X_{t'}; \mathbf{h}^{(i)}\Big)\mathbf{I}_{\{Y_{t'}\le y\}}\Big\}\imath(t-t')}{\sum_{|t'-t|>d+bl_n} \mathbf{K}\Big\{\Big(\mathbf{X}_i - \mathbf{X}_{t'}; \mathbf{h}^{(i)}\Big)\Big\}\imath(t-t')}$$

The function $\imath$ is such that

$$\begin{aligned} &\imath\Big(0\Big) = 0 \\ &\imath\Big(t-t^{'}\Big) = 1, \quad \text{if} \quad t-t^{'} > d+bl_n \\ &0 \le \imath\Big(t-t^{'}\Big) \le 1, \quad \text{if} \quad t-t^{'} \le d+bl_n. \end{aligned} \tag{2.2.1.1}$$

It gives less weight to data closer in time to $\Big(Y_t, \mathbf{X}_t\Big)$ than those which are further away in time. The positive sequence of integers $\Big(bl_n\Big)$, indicates the number of observations (or a block of observations) left out in the t^{th} estimation point. In particular, $\imath(t-t^{'})$ is considered as a weight of the form

$$\imath(t-t^{'}) = \mathbf{I}_{[-lb_n,+lb_n]} \tag{2.2.1.2}$$

whose role is to classify the blocks of data according to their closseness in time. The estimator of the bandwith, $\mathbf{h}^{(i)}$, is then given by

$$\widehat{\mathbf{h}}^{(i)} = \min_{(h_{i1},h_{i2},\ldots,h_{id})} CV\Big(\mathbf{h}^{(i)}\Big) \tag{2.2.1.3}$$

We adopt this selection procedure in this chapter.

2.3 Uniform convergence

In this section, the uniform convergence for the estimators of the conditional distribution and QAR functions are presented. These results will be used later in the estimation of the scale functions in QA-QARCH in section (2.4).

2.3.1 Conditional distribution

In the usual mean regression based on Nadaraya-Watson estimation, the following assumptions obtained from Gyoerfi et al. [56], pages 24-25 are used to show uniform convergence of the conditional mean function. Let G_n be a compact subset of $\mathbf{R}^d$, and G be a ϵ-neighborhood of G_n ($G_n \subset G$).

Conditions 2.3.1 *(On probability distribution)*
(A1) $\exists \Gamma < \infty, \quad \forall \mathbf{B} \in \mathbf{B}\Big(\mathbf{R}^d\Big) \qquad P\Big(\mathbf{X}_t \in \mathbf{B}\Big) \leq \Gamma \lambda \Big(\mathbf{B}\Big)$
and $\exists \gamma, \epsilon > 0, \quad \forall \mathbf{B} \in \mathbf{B}\Big(G\Big) \qquad P\Big(\mathbf{X}_t \in \mathbf{B}\Big) \geq \gamma \lambda \Big(\mathbf{B}\Big)$,
where $\mathbf{B}\Big(\mathbf{R}^d\Big)$ *(respectively* $\mathbf{B}\Big(\mathbf{G}\Big)$*) is the* σ*-algebra of the Borel sets on* $\mathbf{R}^d$ *(respectively on* G*), and* λ *is the Lebesgue measure on* $\mathbf{R}^d$.
(A2) Some absolute moment for random variables Y_t *of degree 2 exists, i.e*
$\exists \upsilon > 0, \quad \exists c < \infty \qquad E\Big[\Big|Y_t\Big|^{2+\upsilon}\Big] \leq c.$
(A3) The conditional variances are bounded on G*, i.e*
$\exists V < \infty, \quad \forall \mathbf{x} \in G \qquad E\Big[\Big(Y_t - \mu\Big(\mathbf{x}\Big)\Big)^2 \Big| \mathbf{X}_t = \mathbf{x}\Big] \leq V.$

Conditions 2.3.2 *(On kernel)*
(K1) Assume (B2),i.e $\exists \overline{\mathbf{K}}, \quad \forall \mathbf{u} \in \mathbf{R}^d, \quad \Big|\mathbf{K}\Big(\mathbf{u}\Big)\Big| \leq \overline{\mathbf{K}} < \infty.$
(K2) $\Big|\Big|\mathbf{u}\Big|\Big|^d \mathbf{K}\Big(\mathbf{u}\Big) \to 0$ *as* $\Big|\Big|\mathbf{u}\Big|\Big| \to \infty.$
(K3) $\exists \widehat{K}, \quad \Big|\int \mathbf{K}^2\Big(\mathbf{u}\Big)d\mathbf{u}\Big| \leq \widehat{\mathbf{K}} < \infty.$
(K4) $\mathbf{K}$ *is Hoelder continuous of order* γ *on* $\mathbf{R}^d$*, for* $\gamma \in (0,1)$*. That is*
$\exists \gamma > 0, \quad c_k < \infty$ *such that* $\Big|\mathbf{K}\Big(\mathbf{u}\Big) - \mathbf{K}\Big(\underline{\mathbf{u}}\Big)\Big| \leq c_k \Big|\mathbf{u} - \underline{\mathbf{u}}\Big|^\gamma, \quad \forall \mathbf{u}, \overline{\mathbf{u}} \in \mathbf{R}^d.$

Conditions 2.3.3 *(On process)*
(E2) $\Big(Y_t, \mathbf{X}_t\Big)$ *is* α*-mixing with mixing coefficients* $\Big\{\alpha(n), n \in \mathbf{N}\Big\}$ *and let* $\Big\{s_n, n \in \mathbf{N}\Big\}$ *be an increasing sequence of integers such that*

$$\exists \mathbf{A} < \infty, \quad \forall n \in \mathbf{N}, \quad 1 \leq s_n \leq \frac{n}{2} \quad \textit{and} \quad \frac{n\alpha_{s_n}^{\frac{2s_n}{3n}}}{s_n} \leq A. \tag{2.3.1.1}$$

Denote $\mathbf{I}_{\{Y_t \leq y\}}$ in (2.1.0.8) by $\mathbf{I}_{t,y}$ and observe that $\widehat{F}_{\mathbf{x}_i}\Big(y\Big)$ estimates $E\Big[\mathbf{I}_{t,y}\Big|\mathbf{X}_t = \mathbf{x}_i\Big] = F_{\mathbf{x}_i}\Big(y\Big)$. We consider $\widehat{F}_{\mathbf{x}_i}\Big(y\Big)$ as the Nadaraya-Watson estimate of the conditional expectation of $\mathbf{I}_{t,y}$ given $\mathbf{X}_t$. As in Gyoerfi et al. [56], in order to deal with possible high[26] values for the random variables Y_t, let M_n be an increasing sequence of real numbers satisfying $M_n = n^\zeta$, for some $\zeta \in \Big(4\Big(\upsilon + 4\Big)^{-1}, 1\Big)$ and υ defined as in condition (A2). The following theorem gives the uniform convergence of the conditional distribution estimator for $F_{\mathbf{x}_i}\Big(y\Big)$.

[26] Although this is trivial in the case of $\mathbf{I}_{t,y}$ as it is bounded in $(0,1)$. Henceforth we will take $M_n = 1$.

Theorem 2.3 *Assume that* $\left(Y_t, \mathbf{X}_t\right)$ *is* α*-mixing with* $\alpha(n)$, s_n *satisfying (E2) and that conditions (A1)-(A3) and (K1)-(K4) hold. If the function* $F_{\mathbf{x}_i}\left(y\right)$ *is continuous in* $\mathbf{x}_i$ *on* G *and if the bandwidth* $\left|\mathbf{h}^{(i)}\right|$ *is such that* $n\left|\mathbf{h}^{(i)}\right|\left(s_n M_n \log n\right)^{-1} \to \infty$ *as* $n \to \infty$*, then* $\widehat{F}_{\mathbf{x}_i}\left(y\right)$ *converges completely*[27] *(co), i.e*

$$\sup_{\mathbf{x}_i \in G}\left|\widehat{F}_{\mathbf{x}_i}\left(y\right) - F_{\mathbf{x}_i}\left(y\right)\right| \to 0, \qquad co. \tag{2.3.1.2}$$

The proof of theorem 2.3 is as a direct consequence of theorem 3.3.5 page 37 in Gyoerfi et al. [56]. We only show that conditions (A1)-(A3) are satisfied.
Condition (A1) is equivalent to saying that the law of $\mathbf{X}_t$ is equivalent to the Lebesgue measure λ on G, i.e $\mathbf{X}_t$ have a density $g\left(\mathbf{x}\right)$ at $\mathbf{X}_t = \mathbf{x}$ and λ has a density with respect to the law of $\mathbf{X}_t$. This is satisfied if $\mathbf{X}_t$ has a bounded almost everywhere (a.e) continuous density which is bounded away from 0 on any finite interval. For (A2), $E\left[\left|\mathbf{I}_{t,y}\right|^{2+\upsilon}\right] < \infty$ for some $\upsilon > 0$ is trivial as $0 \leq \mathbf{I}_{t,y} \leq 1$ and in condition (A3), $E\left[\left(\mathbf{I}_{t,y} - F_{\mathbf{X}_t}\left(y\right)\right)^2 \middle| \mathbf{X}_t = \mathbf{x}\right]$ is bounded for $\mathbf{x} \in G$ for any set G. This is trivially satisfied as $\left|\mathbf{I}_{t,y}\right| \leq 1$ and $F_{\mathbf{x}}\left(y\right) \leq 1$. Note that it is not assumed in the theorem of Gyoerfi et al. that $\mathbf{I}_{t,y}$ should have a density. The rate of convergence of $\widehat{F}_{\mathbf{x}_i}\left(y\right)$ will be discussed in the proof of the theorem 2.4 in the next section.

2.3.2 QAR function

To show the uniform convergence of the estimator for the quantile function $\mu_\theta\left(\mathbf{x}_i\right)$, we combine the ideas from the Nadaraya-Watson type estimation of the conditional distribution and quantiles already presented in the previous sections of this chapter with the concepts of M-estimation. The function $\mu_\theta\left(\mathbf{x}\right)$, being defined for all $\mathbf{x} \in \mathbf{R}^d$, can be seen as a zero in the argument μ of the following function,

[27] A sequence $\left(W_n\right)_{\mathbf{N}}$ of random variables is said to converge completely to 0 if there exist some positive real number a such that we have $\sum_{n=1}^{\infty} P\left(W_n > a\right) < \infty$, see Gyoerfi et al. [56]. The complete convergence implies convergence in probability as well as a.s.

$$\begin{aligned}\widetilde{H}\Big(\mathbf{x},\mu\Big) &= F_{\mathbf{x}}\Big(\mu\Big) - F_{\mathbf{x}}\Big(\mu_\theta\Big(\mathbf{x}\Big)\Big) \\ &= E\Big[\mathbf{I}_{t,\mu} - F_{\mathbf{x}}\Big(\mu_\theta\Big(\mathbf{x}\Big)\Big)\Big|\mathbf{X}_t = \mathbf{x}\Big] \\ &= E\Big[\Psi_{\mathbf{x}}\Big(Y_t - \mu\Big)\Big|\mathbf{X}_t = \mathbf{x}\Big] \qquad (2.3.2.1)\end{aligned}$$

where $\Psi_{\mathbf{x}}\Big(u\Big) = \mathbf{I}_{(-\infty,0]}\Big(u\Big) - F_{\mathbf{x}}\Big(\mu_\theta\Big(\mathbf{x}\Big)\Big)$, i.e. $\Psi_{\mathbf{x}}\Big(u\Big) = \mathbf{I}_{(-\infty,0]}\Big(u\Big) - \theta$ if $F_{\mathbf{x}}\Big(\mu_\theta\Big(\mathbf{x}\Big)\Big) = \theta$. The estimate of $\widetilde{H}\Big(\mathbf{x},\mu\Big)$ at $\mathbf{x} = \mathbf{x}_i$ is

$$\begin{aligned}\widetilde{H}_n\Big(\mathbf{x}_i,\mu\Big) &= \widehat{F}_{\mathbf{x}_i}\Big(\mu\Big) - F_{\mathbf{x}_i}\Big(\mu_\theta\Big(\mathbf{x}_i\Big)\Big) \\ &= \frac{\Big(n\Big|\mathbf{h}^{(i)}\Big|\Big)^{-1}\sum_{t=1}^n \mathbf{K}\Big(\mathbf{x}_i - \mathbf{X}_t;\mathbf{h}^{(i)}\Big)\Big(\mathbf{I}_{t,\mu} - F_{\mathbf{x}_i}\Big(\mu_\theta\Big(\mathbf{x}_i\Big)\Big)}{\Big(n\Big|\mathbf{h}^{(i)}\Big|\Big)^{-1}\sum_{t=1}^n \mathbf{K}\Big(\mathbf{x}_i - \mathbf{X}_t;\mathbf{h}^{(i)}\Big)} \\ &= \sum_{t=1}^n w_t\Big(\mathbf{x}_i\Big)\Big(\mathbf{I}_{t,\mu} - F_{\mathbf{x}_i}\Big(\mu_\theta\Big(\mathbf{x}_i\Big)\Big)\Big) \\ &= \sum_{t=1}^n w_t\Big(\mathbf{x}_i\Big)\Psi_{\mathbf{x}_i}\Big(Y_t - \mu\Big) \qquad (2.3.2.2)\end{aligned}$$

It follows immediately that the solution of (2.3.2.2) satisfies

$$\widetilde{H}_n\Big(\mathbf{x}_i,\mu\Big) = 0. \qquad (2.3.2.3)$$

Mark that since $\Psi_{\mathbf{x}}\Big(Y_t - \mu\Big)$ is nondecreasing and right-continuous in μ so is $\widetilde{H}\Big(\mathbf{x},\mu\Big)$ and $\widetilde{H}_n\Big(\mathbf{x},\mu\Big)$. As $\Psi_{\mathbf{x}}\Big(Y_t - \mu\Big) \to -\theta$ for $\mu \to -\infty$ and $\Psi_{\mathbf{x}}\Big(Y_t - \mu\Big) \to 1-\theta$ for $\mu \to \infty$ if $F_{\mathbf{x}}\Big(\mu_\theta\Big(\mathbf{x}\Big)\Big) = \theta$, we have for n large that $\inf_\mu \widetilde{H}_n\Big(\mathbf{x},\mu\Big) < 0 < \sup_\mu \widetilde{H}_n\Big(\mathbf{x},\mu\Big)$, c.f (2.1.2.7) of the proof of theorem 2.2.

We will assume the following further conditions.

Conditions 2.3.4

(Q1) The density, $g\Big(\mathbf{x}\Big)$, *is bounded by* $\widetilde{\Gamma}$, *i.e.* $\widetilde{\Gamma} = \sup_{\mathbf{x}\in\mathbf{R}^d} g\Big(\mathbf{x}\Big) < \infty$.

(Q2) $\sup_{\mathbf{x}\in G_n}\Big|\widehat{g}\Big(\mathbf{x}\Big) - g\Big(\mathbf{x}\Big)\Big| \to 0$.

(Q3) The conditional density, $f_{\mathbf{x}}\Big(\mu\Big)$, *is bounded in* $\mathbf{x}$ *and* μ *by* c_Ψ *which is independent of* $\mathbf{x}$ *and* μ.

(Q4) For some compact neighborhood Θ_n *of* 0 *and a constant* c_0,
$\inf_{\mu\in\Theta_n}\inf_{\mathbf{x}\in G_n} f_{\mathbf{x}}\Big(\mu_\theta\Big(\mathbf{x}\Big) + \mu\Big) \geq c_0 > 0$.

Mark that (Q2) follows from lemma 2.4. In (Q3) $E\Big[\Psi_{\mathbf{x}}\Big(Y_t-\mu\Big)\Big|\mathbf{X}_t=\mathbf{x}\Big]$ is uniformly bounded in absolute value by 1 by the definition of $\Psi_{\mathbf{x}}$. As $\frac{d}{d\mu}E\Big[\Psi_{\mathbf{x}}\Big(Y_t-\mu\Big)\Big|\mathbf{X}_t=\mathbf{x}\Big]=f_{\mathbf{x}}\Big(\mu\Big)$, it is also strictly increasing and continuously differentiable in $\mathbf{x}$ and μ. Therefore $F_{\mathbf{x}}\Big(\mu\Big)$ is assumed to be Lipschitz in $\mathbf{x}$ and μ.

Theorem 2.4 *Let Θ_n ba a compact neighborhood of 0 in $\mathbf{R}$ and assume (A1)-(A3),(K1)-(K3), (E2) and (Q1)-(Q4). Suppose $h_{i,j}, j=1,2,\dots,d$ is a sequence of bandwidths depending on $n\in\mathbf{N}$ such that $\widetilde{S}_n=n\Big|\mathbf{h}^{(i)}\Big|\Big(s_n\log n\Big)^{-1}\to\infty$ as $n\to\infty$, $S_n=\Big|\Big|\mathbf{h}^{(i)}\Big|\Big|^2+\widetilde{S}_n^{-\frac{1}{2}}$ satisfies $S_n\to 0$ as $n\to\infty$ for all $\mathbf{x}_i\in G_n$ and there*

$$\exists C_\alpha>0,\quad C_\alpha<\infty\quad \textit{such that}\quad \widetilde{S}_n^{\frac{1}{2}}\Big|\Big|\mathbf{h}^{(i)}\Big|\Big|^2\le C_\alpha, \tag{2.3.2.4}$$

$\forall n\in\mathbf{N}$, then we have

$$\sup_{\mathbf{x}_i\in G_n}\Big|\widehat{\mu}_\theta\Big(\mathbf{x}_i\Big)-\mu_\theta\Big(\mathbf{x}_i\Big)\Big|=O\Big(S_n\Big)\quad a.s. \tag{2.3.2.5}$$

The prove of theorem 2.4 is close in lines with the proof in Collomb and Haerdle [29] and Gyoerfi et al. [56] chapter III. Note that it is complicated to deal with $\widetilde{H}\Big(\mathbf{x},\mu\Big)$ directly, so we decompose the difference in the following manner:

Let

$$H_n\Big(\mathbf{x},\mu\Big)=\Big(n\Big|\mathbf{h}^{(i)}\Big|\Big)^{-1}\sum_{t=1}^{n}\mathbf{K}\Big(\mathbf{x}-\mathbf{X}_t;\mathbf{h}^{(i)}\Big)\Psi_{\mathbf{x}}\Big(Y_t-\mu\Big) \tag{2.3.2.6}$$

and $\widehat{g}\Big(\mathbf{x}\Big)=\Big(n\Big|\mathbf{h}^{(i)}\Big|\Big)^{-1}\sum_{t=1}^{n}\mathbf{K}\Big(\mathbf{x}-\mathbf{X}_t;\mathbf{h}^{(i)}\Big)$, as in (2.1.0.5). Then the difference can be expressed as $\widetilde{H}_n-\widetilde{H}=\frac{H_n-H}{\widehat{g}}+\frac{H(g-\widehat{g})}{g\widehat{g}}$. Observe that

$$\begin{aligned}\sup_{\mathbf{x}\in G_n}\sup_{\mu\in\Theta_n}\Big|\widetilde{H}_n-\widetilde{H}\Big| &\le \sup_{\mathbf{x}\in G_n}\sup_{\mu\in\Theta_n}\frac{|H_n-H|}{\widehat{g}}+\sup_{\mathbf{x}\in G_n}\sup_{\mu\in\Theta_n}\frac{|H|}{g\widehat{g}}\Big|\widehat{g}-g\Big| \\ &\le \sup_{\mathbf{x}\in G_n}\sup_{\mu\Theta_n}\frac{|H_n-H|}{\widehat{g}}+\frac{1}{\delta}\sup_{\mathbf{x}\in G_n}\frac{|\widehat{g}-g|}{\widehat{g}},\end{aligned} \tag{2.3.2.7}$$

as $|\widetilde{H}|\le 1$ and $g\ge\delta>0$ on G_n. Now, if $\sup_{\mathbf{x}\in G_n}\Big|\widehat{g}-g\Big|\le\epsilon$, we have

$$\frac{1}{\widehat{g}}=\frac{1}{g+(\widehat{g}-g)}\le\frac{1}{g-|\widehat{g}-g|}\le\frac{1}{\delta-\epsilon}$$

on G_n. Therefore, to proof that $\widetilde{H}_n\to\widetilde{H}$ uniformly in $\mathbf{x}\in G_n,\mu\in\Theta_n$, it suffice to show that $H_n\to H$ and $\widehat{g}\to g$ uniformly in $\mathbf{x}\in G_n,\mu\in\Theta_n$, and also the rate of convergence will be given by the slower of the two rates of convergence of $\sup_{\mathbf{x}\in G_n}\sup_{\mu\in\Theta_n}\Big|H_n - H\Big|$ and $\sup_{\mathbf{x}\in G_n}\Big|\widehat{g} - g\Big|$. The following lemma gives the rate of convergence of $\widehat{g}\Big(\mathbf{x}_i\Big)$.

Lemma 2.4 *Under the assumptions of theorem 3.3.6 of Gyoerfi et al. [56], for G_n compact*

(i) $\sup_{\mathbf{x}_i\in G_n}\Big|\widehat{g}\Big(\mathbf{x}_i\Big) - E\Big|\widehat{g}\Big(\mathbf{x}_i\Big)\Big|\Big| = O\Big(\widetilde{S}_n^{-\frac{1}{2}}\Big)$ *a.s.*

(ii) $\sup_{\mathbf{x}_i\in G_n}\Big|E\Big|\widehat{g}\Big(\mathbf{x}_i\Big)\Big| - g\Big(\mathbf{x}\Big)\Big| = O\Big(\Big|\Big|\mathbf{h}^{(i)}\Big|\Big|^2\Big)$

The proof of this result follows directly from the proof of theorem 3.3.6 of Gyoerfi et al. [56]. The sum of (i) and (ii) gives the rate. Therefore, we consider only the convergence of $H_n\Big(\mathbf{x}_i, \mu\Big)$.

In this regard, the following exponential inequality for α-mixing variables obtained by Carbon (1983) and stated in Gyoerfi et al. [56] and the lemmas on convergence of $H_n\Big(\mathbf{x}_i, \mu\Big)$ that follow thereafter, will be used.

Theorem 2.5 *If $\Big(\Delta_t\Big)_{\mathbf{N}}$ is α-mixing with $E\Big[\Delta_t\Big] = 0$, $\Big|\Delta_t\Big| \le c$ and $E\Big[\Delta_t^2\Big] \le D$, then we have*

$$P\Big(\Big|\sum_{t=1}^{n}\Delta_t\Big| > \epsilon_n\Big) \le 2\exp\Big\{-\alpha\epsilon_n + 6\alpha^2 e(D + 8c^2\sum_{t=1}^{s}\alpha_t)n + 2\sqrt{e}ns^{-1}\alpha_s^{\frac{2s}{3n}}\Big\} \qquad (2.3.2.8)$$

where α is a real number and s an integer satisfying $1 \le s \le n$ and $0 \le \alpha \le \frac{sce}{4}$.

Lemma 2.5 *Under assumptions (K1)-(K4),(Q1)-(Q3) and (E2) we have for any compact $G_n \subseteq \mathbf{R}^d, \widetilde{\Theta}_n \in \mathbf{R}$,*

$$\sup_{\mathbf{x}_i\in G_n}\sup_{\mu\in\widetilde{\Theta}_n}\Big|H_n\Big(\mathbf{x}_i, \mu\Big) - E\Big[H_n\Big(\mathbf{x}_i, \mu\Big)\Big]\Big| = O\Big(\widetilde{S}_n^{-\frac{1}{2}}\Big) \quad a.s.$$

with $\widetilde{S}_n = n\Big|\mathbf{h}^{(i)}\Big|\Big(s_n \log n\Big)^{-1} \to \infty, \quad s_n \to \infty.$

Proof of lemma 2.5

We follow essentially the proof of theorem 5.2.6 of Gyoerfi et al. [56] which gives a rate for the Glivenko-Cantelli theorem in the case of α-mixing random variables. Denote

$$\Delta_t = \Big(n\Big|\mathbf{h}^{(i)}\Big|\Big)^{-1}\Big(\mathbf{K}\Big(\mathbf{x}_i - \mathbf{X}_t; \mathbf{h}^{(i)}\Big)\Psi_{\mathbf{x}_i}\Big(Y_t - \mu\Big) - E\Big[\mathbf{K}\Big(\mathbf{x}_i - \mathbf{X}_t; \mathbf{h}^{(i)}\Big)\Psi_{\mathbf{x}_i}\Big(Y_t - \mu\Big)\Big]\Big),$$

then $H_n\Big(\mathbf{x}_i,\mu\Big) - E\Big[H_n\Big(\mathbf{x}_i,\mu\Big)\Big] = \sum_{t=1}^n \Delta_t$ and $E\Big[\Delta_t\Big] = 0$. We have by (K1) and boundedness of $\Psi_{\mathbf{x}}$

$$\Big|\Delta_t\Big| \leq \Big(n\Big|\mathbf{h}^{(i)}\Big|\Big)^{-1} 2\overline{\mathbf{K}} = c < \infty.$$

We also have

$$E\Big[\Delta_t^2\Big] \leq 2\Big(n\Big|\mathbf{h}^{(i)}\Big|\Big)^{-2} E\Big[\mathbf{K}^2\Big(\mathbf{x}_i - \mathbf{X}_t;\mathbf{h}^{(i)}\Big)\Big|\Psi_{\mathbf{x}_i}\Big(Y_t-\mu\Big)\Big|^2\Big].$$

Using conditions (K3) and (Q1) and the fact that $\Big|\Psi_{\mathbf{x}_i}\Big| \leq 1$, we have

$$\begin{aligned} E\Big[\Delta_t^2\Big] &\leq 2\Big(n^2\Big|\mathbf{h}^{(i)}\Big|\Big)^{-1} E\Big[\Big|\mathbf{h}^{(i)}\Big|^{-1}\mathbf{K}^2\Big(\mathbf{x}_i - \mathbf{X}_t;\mathbf{h}^{(i)}\Big]\\ &\leq 2\Big(n^2\Big|\mathbf{h}^{(i)}\Big|\Big)^{-1}\widetilde{\Gamma}\widehat{\mathbf{K}} = D < \infty. \end{aligned} \tag{2.3.2.9}$$

As in the proof of lemma 3.3.3 in Gyoerfi et al. [56], choose $\alpha = c_2 n\Big|\mathbf{h}^{(i)}\Big|s_n^{-1}$ and $C_\alpha = \alpha s_n\Big(n\Big|\mathbf{h}^{(i)}\Big|\Big)^{-1}\overline{\mathbf{K}} > \frac{e}{4}$ and get, by applying theorem 2.5 for any sequence $\Big(\epsilon_n\Big)_{\mathbf{N}}$,

$$P\Big(\Big|\sum_{t=1}^n \Delta_t\Big| > \epsilon_n\Big) \leq c_1 \exp\Big\{-c_2 n|\mathbf{h}^{(i)}|\epsilon_n^2 s_n^{-1}\Big\}, \tag{2.3.2.10}$$

uniformly in $\mathbf{x}_i \in \mathbf{R}^d$ and $\mu \in \mathbf{R}$ with some constants $c_1, c_2 > 0$.

Next, using the compactness of $\widetilde{\Theta}_n$, we cover it with M intervals $\mathbf{I}_m$ of length C_M and: $\widetilde{\Theta}_n \subset \cup_{m=1}^M \mathbf{I}_m$, $\quad \mathbf{I}_m = \Big[\mu_{m-1},\mu_m\Big]$, $\quad \Big|\mu_m - \mu_{m-1}\Big| = C_M, m = 1,\ldots,M.$

Mark that for all $m = 1,\ldots,M$,

$E\Big[H_n\Big(\mathbf{x}_i,\mu_{m-1}\Big)\Big] \leq \sup_{\mu\in\mathbf{I}_m} E\Big[H_n\Big(\mathbf{x}_i,\mu\Big)\Big] = E\Big[H_n\Big(\mathbf{x}_i,\mu_m\Big)\Big]$ and

$H_n\Big(\mathbf{x}_i,\mu_{m-1}\Big) \leq \sup_{\mu\in\mathbf{I}_m} H_n\Big(\mathbf{x}_i,\mu\Big) = H_n\Big(\mathbf{x}_i,\mu_m\Big)$.

Therefore, we have for any $\mu \in \mathbf{I}_m$, using monotonicity of $\mathbf{I}_{t,\mu}$ and $F_{\mathbf{x}_i}\Big(\mu\Big)$ in μ:

$$\begin{aligned} H_n\Big(\mathbf{x}_i,\mu\Big) - E\Big[H_n\Big(\mathbf{x}_i,\mu\Big)\Big] &\leq H_n\Big(\mathbf{x}_i,\mu_m\Big) - E\Big[H_n\Big(\mathbf{x}_i,\mu_m\Big)\Big]\\ &+ E\Big[H_n\Big(\mathbf{x}_i,\mu_m\Big)\Big] - E\Big[H_n\Big(\mathbf{x}_i,\mu\Big)\Big]\\ &\leq H_n\Big(\mathbf{x}_i,\mu_m\Big) - E\Big[H_n\Big(\mathbf{x}_i,\mu_m\Big)\Big]\\ &+ E\Big[H_n\Big(\mathbf{x}_i,\mu_m\Big)\Big] - E\Big[H_n\Big(\mathbf{x}_i,\mu_{m-1}\Big)\Big]. \end{aligned}$$

and

$$\begin{aligned} E\Big[H_n\Big(\mathbf{x}_i,\mu\Big)\Big] - H_n\Big(\mathbf{x}_i,\mu\Big) &\leq E\Big[H_n\Big(\mathbf{x}_i,\mu\Big)\Big] - E\Big[H_n\Big(\mathbf{x}_i,\mu_{m-1}\Big)\Big] \\ &+ E\Big[H_n\Big(\mathbf{x}_i,\mu_{m-1}\Big)\Big] - H_n\Big(\mathbf{x}_i,\mu_{m-1}\Big) \\ &\leq E\Big[H_n\Big(\mathbf{x}_i,\mu_m\Big)\Big] - E\Big[H_n\Big(\mathbf{x}_i,\mu_{m-1}\Big)\Big] \\ &+ E\Big[H_n\Big(\mathbf{x}_i,\mu_{m-1}\Big)\Big] - H_n\Big(\mathbf{x}_i,\mu_{m-1}\Big). \end{aligned}$$

Using condition (Q3),

$$\begin{aligned} E\Big[H_n\Big(\mathbf{x}_i,\mu_m\Big)\Big] - E\Big[H_n\Big(\mathbf{x}_i,\mu_{m-1}\Big)\Big] &\leq c_\Psi\Big|\mu_m-\mu_{m-1}\Big|\Big(n\Big|\mathbf{h}^{(i)}\Big|\Big)^{-1}E\Big[\sum_{t=1}^{n}\mathbf{K}\Big(\mathbf{x}_i-\mathbf{X}_t;\mathbf{h}^{(i)}\Big)\Big] \\ &= c_\Psi C_M E\Big[\hat{g}\Big(\mathbf{x}_i\Big)\Big]. \end{aligned}$$

We get for all $\mu \in \mathbf{I}_m$,

$$\begin{aligned} \Big|H_n\Big(\mathbf{x}_i,\mu\Big) - E\Big[H_n\Big(\mathbf{x}_i,\mu\Big)\Big]\Big| &\leq \max\Big\{\Big|H_n\Big(\mathbf{x}_i,\mu_{m-1}\Big) - E\Big[H_n\Big(\mathbf{x}_i,\mu_{m-1}\Big)\Big]\Big| \\ &\quad , \Big|H_n\Big(\mathbf{x}_i,\mu_m\Big) - E\Big[H_n\Big(\mathbf{x}_i,\mu_m\Big)\Big]\Big|\Big\} + c_\Psi C_M E\Big[\hat{g}\Big(\mathbf{x}_i\Big)\Big], \end{aligned}$$

and, therefore,

$$\sup_{\mu\in\tilde{\Theta}_n}\Big|H_n\Big(\mathbf{x}_i,\mu\Big)-E\Big[H_n\Big(\mathbf{x}_i,\mu\Big)\Big]\Big| \leq \max_{m=0,1,\ldots,M}\Big|H_n\Big(\mathbf{x}_i,\mu_m\Big)-E\Big[H_n\Big(\mathbf{x}_i,\mu_m\Big)\Big]\Big|+c_\Psi C_M E\Big[\widehat{g}\Big(\mathbf{x}_i\Big)\Big] \tag{2.3.2.11}$$

We first consider the first term and get, using (2.3.2.10),

$$\begin{aligned} P\Big(\max_{m=0,\ldots,M}\Big|H_n\Big(\mathbf{x}_i,\mu_m\Big) &- E\Big[H_n\Big(\mathbf{x}_i,\mu_m\Big)\Big]\Big| > \epsilon_n\Big) \\ &\leq \sum_{m=0}^{M} P\Big(\Big|H_n\Big(\mathbf{x}_i,\mu_m\Big) - E\Big[H_n\Big(\mathbf{x}_i,\mu_m\Big)\Big]\Big| > \epsilon_n\Big) \\ &\leq (M+1)c_1\exp\Big\{-c_2 n\Big|\mathbf{h}^{(i)}\Big|\epsilon_n^2 s_n^{-1}\Big\}. \end{aligned}$$

We choose $C_M = c_3 n^{-1}$ for some $c_3 > 0$ and, therefore, $M+1 \leq c_4 n$. Using the definition of $\widetilde{S}_n$, we have

$$\begin{aligned} P\Big(\max_{m=0,\ldots,M}\Big|H_n\Big(\mathbf{x}_i,\mu_m\Big)-E\Big[H_n\Big(\mathbf{x}_i,\mu_m\Big)\Big]\Big|>\epsilon_n\Big) &\leq c_1c_4n\exp\Big\{-c_2\widetilde{S}_n\log n\epsilon_n^2\Big\} \\ &= c_1c_4n^{1-c_2\widetilde{S}_n\epsilon_n^2} \\ &= c_1c_4n^{1-c_2\epsilon^2a_n^2} \\ &\leq const.n^{-r} \end{aligned}$$

for arbitrary $r>0$ if n is choosen large enough. Here, we have chosen $\epsilon_n=\epsilon\widetilde{S}_n^{-\frac{1}{2}}a_n$ for some arbitrary sequence $a_n\to\infty(n\to\infty)$. Choosing, e.g $r=2$, we get

$$\sum_{n=1}^{\infty}P\Big(\frac{\widetilde{S}_n^{\frac{1}{2}}}{a_n}\max_{0,\ldots,M}\Big|H_n\Big(\mathbf{x}_i,\mu_m\Big)-E\Big[H_n\Big(\mathbf{x}_i,\mu_m\Big)\Big]\Big|>\epsilon\Big)<\infty$$

which implies $\frac{\widetilde{S}_n^{\frac{1}{2}}}{a_n}\max_{m=0,\ldots,M}\Big|H_n\Big(\mathbf{x}_i,\mu_m\Big)-E\Big[H_n\Big(\mathbf{x}_i,\mu_m\Big)\Big]\Big|\to 0$ a.s. by the Borel-Cantelli lemma. As $a_n\to\infty$ arbitrarily slowly, this implies that

$$\widetilde{S}_n^{\frac{1}{2}}\max_{m=0,\ldots,M}\Big|H_n\Big(\mathbf{x}_i,\mu_m\Big)-E\Big[H_n\Big(\mathbf{x}_i,\mu_m\Big)\Big]\Big| \quad \text{is bounded a.s.} \tag{2.3.2.12}$$

Now as $\Big|\mathbf{h}^{(i)}\Big|\to 0, \quad s_n\to\infty$, we have $\widetilde{S}_n^{\frac{1}{2}}C_M=\widetilde{S}_n^{\frac{1}{2}}c_3n^{-1}\to 0$ for $n\to\infty$. By lemma 2.4, $E\Big[\widehat{g}\Big(\mathbf{x}_i\Big)\Big]$ converges a.s to $g\Big(\mathbf{x}_i\Big)$ uniformly in $\mathbf{x}_i\in G_n$, and therefore, it is bounded. This implies $c_\Psi C_M E\Big[\widehat{g}\Big(\mathbf{x}_i\Big)\Big]=c_\Psi c_3n^{-1}E\Big[\widehat{g}\Big(\mathbf{x}_i\Big)\Big]\to 0$. Combining (2.3.2.11) with (2.3.2.12) and the boundness of $\widehat{g}\Big(\mathbf{x}_i\Big)$ we finally get

$$\sup_{\mu\in\widetilde{\Theta}_n}\Big|H_n\Big(\mathbf{x}_i,\mu\Big)-E\Big[H_n\Big(\mathbf{x}_i,\mu\Big)\Big]\Big|=O\Big(\widetilde{S}_n^{-\frac{1}{2}}\Big) \quad \text{a.s. uniformly in} \quad \mathbf{x}_i\in G_n.$$

□

Define $H_n\Big(\mathbf{x}_i,\mu_\theta\Big(\mathbf{x}_i\Big)+\mu\Big)=\Big(n\Big|\mathbf{h}^{(i)}\Big|\Big)^{-1}\sum_{t=1}^{n}\mathbf{K}\Big(\mathbf{x}_i-\mathbf{X}_t;\mathbf{h}^{(i)}\Big)\Psi_{\mathbf{x}_i}\Big(Y_t-\mu_\theta\Big(\mathbf{x}_i\Big)-\mu\Big)$

Lemma 2.6 *In addition to the assumptions of lemma 2.5 assume* $\mu_\theta\Big(\mathbf{x}_i\Big)$ *is continuous, then we have for any compact* $G_n\subseteq\mathbf{R}^d,\Theta_n\subseteq\mathbf{R}$

$$\sup_{\mathbf{x}_i\in G_n}\sup_{\mu\in\Theta_n}\Big|H_n\Big(\mathbf{x}_i,\mu_\theta\Big(\mathbf{x}_i\Big)+\mu\Big)-E\Big[H_n\Big(\mathbf{x}_i,\mu_\theta\Big(\mathbf{x}_i\Big)+\mu\Big)\Big]\Big|=O\Big(\widetilde{S}_n^{-\frac{1}{2}}\Big) \quad a.s.$$

Proof of lemma 2.6

As $\mu_\theta\Big(\mathbf{x}_i\Big)$ is continuous, $\widetilde{\Theta}_n = \Big\{\nu = \mu_\theta\Big(\mathbf{x}_i\Big) + \mu, \mathbf{x}_i \in G_n, \mu \in \Theta_n\Big\}$ is compact too. Therefore

$$\sup_{\mathbf{x}_i \in G_n} \sup_{\mu \in \Theta_n} \Big|H_n\Big(\mathbf{x}_i, \mu_\theta\Big(\mathbf{x}_i\Big)+\mu\Big) - E\Big[H_n\Big(\mathbf{x}_i, \mu_\theta\Big(\mathbf{x}_i\Big)+\mu\Big)\Big]\Big| \leq \sup_{\mathbf{x}_i \in G_n} \sup_{\nu \in \widetilde{\Theta}_n} \Big|H_n\Big(\mathbf{x}_i, \nu\Big) - E\Big[H_n\Big(\mathbf{x}_i, \nu\Big)\Big]\Big|.$$

Hence, the assertion follows from lemma 2.5. □

Lemma 2.7 *Under assumptions (A1)-(A3) and (K1)-(K3), we have*

$$\sup_{\mathbf{x}_i \in G_n} \sup_{\mu \in \Theta_n} \Big|E\Big[H_n\Big(\mathbf{x}_i, \mu_\theta\Big(\mathbf{x}_i\Big) + \mu\Big)\Big] - H\Big(\mathbf{x}_i, \mu_\theta\Big(\mathbf{x}_i\Big) + \mu\Big)\Big| = O\Big(\Big\|\mathbf{h}^{(i)}\Big\|^2\Big)$$

Proof of lemma 2.7

Since the bias term does not depend on probability distribution of the time series $\Big\{Y_t, \mathbf{X}_t\Big)\Big\}_{t=1}^n$, it can be treated exactly as in the independent case. Its manipulation is based on Taylor expansion of $E\Big[\Psi_{\mathbf{x}_i}\Big(Y_t - \mu_\theta\Big(\mathbf{x}_i\Big) - \mu\Big)\Big|\mathbf{X}_t = \mathbf{x}\Big]$ up to order two (say),

$$\Big|E\Big[H_n\Big(\mathbf{x}_i, \mu_\theta\Big(\mathbf{x}_i\Big) + \mu\Big)\Big] - H\Big(\mathbf{x}_i, \mu_\theta\Big(\mathbf{x}_i\Big) + \mu\Big)\Big| \leq c_5\Big\|\mathbf{h}^{(i)}\Big\|^2 \quad \text{for some} \quad 0 < c_5 < \infty, \tag{2.3.2.13}$$

uniformly for $\mathbf{x}_i \in G_n$ and $\mu \in \Theta_n$, c.f lemma 2.2 and also see Haerdle and Luckhaus [57]. For the constant on the right hand of (2.3.2.13), we have assumed that $\sup_{\mathbf{x}_i} \sup_{\mu \in \Theta} \Big|\nabla^2 E\Big[\Psi_{\mathbf{x}_i}\Big(Y_t - \mu_\theta\Big(\mathbf{x}_i\Big) - \mu\Big)\Big|\mathbf{X}_t = \mathbf{x}_i\Big]\Big| \leq c_5 < \infty$

□

Proof of theorem 2.4

Observe that for $n \to \infty$, lemmas 2.4,2.5 and 2.6 and the fact that by lemma 2.7 the bias

$$\sup_{\mathbf{x}_i \in G_n} \Big|E\Big[H_n\Big(\mathbf{x}_i, \mu_\theta\Big(\mathbf{x}_i\Big) + \mu\Big)\Big] - H\Big(\mathbf{x}_i, \mu_\theta\Big(\mathbf{x}_i\Big) + \mu\Big)\Big| \to 0 \quad \text{as} \quad n \to \infty, \tag{2.3.2.14}$$

imply that under conditions (A1)-(A3),(K1)-(K3),(E2),(Q1)-(Q2) and $\widetilde{S}_n \to \infty$ as $n \to \infty$, we have

$\sup_{\mathbf{x}_i \in G_n} \Big|H_n\Big(\mathbf{x}_i, \mu_\theta\Big(\mathbf{x}_i\Big) + \mu\Big) - H\Big(\mathbf{x}_i, \mu_\theta\Big(\mathbf{x}_i\Big) + \mu\Big)\Big| \to 0$ as $n \to \infty$ a.s. and

$$\sup_{\mathbf{x}_i \in G_n} \Big|\widehat{g}\Big(\mathbf{x}_i\Big) - g\Big(\mathbf{x}_i\Big)\Big| \to 0 \quad \text{as} \quad n \to \infty \quad \text{a.s} \tag{2.3.2.15}$$

Hence, by the remarks following (2.3.2.7),

$$\sup_{\mathbf{x}_i \in G_n} \left| \widetilde{H}_n\Big(\mathbf{x}_i, \mu_\theta\Big(\mathbf{x}_i\Big) + \mu\Big) - \widetilde{H}\Big(\mathbf{x}_i, \mu_\theta\Big(\mathbf{x}_i\Big) + \mu\Big) \right| \to 0 \quad \text{as} \quad n \to \infty \quad \text{a.s.} \tag{2.3.2.16}$$

Using a similar technique as in Collomb and Haerdle [29], fix $\epsilon > 0$. The strict monotonicity[28] of $E\Big[\Psi_{\mathbf{x}}\Big]$ and positivity of g on G_n imply $\forall \mathbf{x}_i \in G_n \quad \widetilde{H}\Big(\mathbf{x}_i, \mu_\theta\Big(\mathbf{x}_i\Big) - \epsilon\Big) < 0 < \widetilde{H}\Big(\mathbf{x}_i, \mu_\theta\Big(\mathbf{x}_i\Big) + \epsilon\Big)$. The convergence in (2.3.2.16) implies that for all sufficiently large n, $\forall \mathbf{x}_i \in G_n \quad \widetilde{H}_n\Big(\mathbf{x}_i, \mu_\theta\Big(\mathbf{x}_i\Big) - \epsilon\Big) < 0 < \widetilde{H}_n\Big(\mathbf{x}_i, \mu_\theta\Big(\mathbf{x}_i\Big) + \epsilon\Big)$ a.s. and by the definition (2.3.2.3), the positivity of the weights, and the monotonicity of $\widetilde{H}_n$ in μ we have $\forall \mathbf{x}_i \in G_n \quad \mu_\theta\Big(\mathbf{x}_i\Big) - \epsilon < \widehat{\mu}_\theta\Big(\mathbf{x}_i\Big) < \mu_\theta\Big(\mathbf{x}_i\Big) + \epsilon$, which can be written as

$$\sup_{\mathbf{x}_i \in G_n} \left| \hat{\mu}_\theta\Big(\mathbf{x}_i\Big) - \mu_\theta\Big(\mathbf{x}_i\Big) \right| \to 0 \quad \text{a.s. as} \quad n \to \infty. \tag{2.3.2.17}$$

Also observe that lemmas 2.6 and 2.7 and the Borel-Cantelli lemma show that under the additional conditions, using also $S_n^{-1} \le \widetilde{S}_n^{\frac{1}{2}}$ by definition of S_n,

$\sup_{\mathbf{x}_i \in G_n} \sup_{\mu \in \Theta_n} \left| H_n\Big(\mathbf{x}_i, \mu_\theta\Big(\mathbf{x}_i\Big) + \mu\Big) - H\Big(\mathbf{x}_i, \mu_\theta\Big(\mathbf{x}_i\Big) + \mu\Big) \right| = O\Big(S_n\Big)$ and therefore, with (2.3.2.15), we have

$$\sup_{\mathbf{x}_i \in G_n} \sup_{\mu \in \Theta_n} \left| \widetilde{H}_n\Big(\mathbf{x}_i, \mu_\theta\Big(\mathbf{x}_i\Big) + \mu\Big) - \widetilde{H}\Big(\mathbf{x}_i, \mu_\theta\Big(\mathbf{x}_i\Big) + \mu\Big) \right| = O\Big(S_n\Big). \tag{2.3.2.18}$$

The definition of $\mu_\theta\Big(\mathbf{x}_i\Big)$ and $\widehat{\mu}_\theta\Big(\mathbf{x}_i\Big)$ shows that for all $\mathbf{x}_i \in \mathbf{R}^d$, we have

$$\widetilde{H}_n\Big(\mathbf{x}_i, \widehat{\mu}_\theta\Big(\mathbf{x}_i\Big)\Big) - \widetilde{H}\Big(\mathbf{x}_i, \widehat{\mu}_\theta\Big(\mathbf{x}_i\Big)\Big) = \widetilde{H}\Big(\mathbf{x}_i, \mu_\theta\Big(\mathbf{x}_i\Big)\Big) - \widetilde{H}\Big(\mathbf{x}_i, \widehat{\mu}_\theta\Big(\mathbf{x}_i\Big)\Big).$$

A Taylor expansion of $\widetilde{H}\Big(\mathbf{x}_i, .\Big) = F_{\mathbf{x}_i}\Big(.\Big) - \theta$ gives

$$\widetilde{H}_n\Big(\mathbf{x}_i, \widehat{\mu}_\theta\Big(\mathbf{x}_i\Big)\Big) - \widetilde{H}\Big(\mathbf{x}_i, \widehat{\mu}_\theta\Big(\mathbf{x}_i\Big)\Big) = -\Big(\widehat{\mu}_\theta\Big(\mathbf{x}_i\Big) - \mu_\theta\Big(\mathbf{x}_i\Big)\Big) f_{\mathbf{x}_i}\Big(\widetilde{\mu}_\theta\Big(\mathbf{x}_i\Big)\Big), \tag{2.3.2.19}$$

where $\widetilde{\mu}_\theta\Big(\mathbf{x}_i\Big)$ is between $\mu_\theta\Big(\mathbf{x}_i\Big)$ and $\widehat{\mu}_\theta\Big(\mathbf{x}_i\Big)$. Using result (2.3.2.17) and for n_0 sufficiently large, we have $\sup_{\mathbf{x}_i \in G_n} \left| \widehat{\mu}_\theta\Big(\mathbf{x}_i\Big) - \mu_\theta\Big(\mathbf{x}_i\Big) \right| \to 0$, i.e $\widehat{\mu}_\theta\Big(\mathbf{x}_i\Big) - \mu_\theta\Big(\mathbf{x}_i\Big) \in \Theta_n$ a.s. $\forall n \ge n_0$. From condition (Q4), we have $\inf_{\mathbf{x}_i \in G_n} \frac{d}{d\mu} E\Big[\Psi_{\mathbf{x}_i}\Big(Y_t - \widetilde{\mu}_\theta\Big(\mathbf{x}_i\Big)\Big)\Big|\mathbf{X}_t = \mathbf{x}_i\Big] \inf_{\mathbf{x}_i} f_{\mathbf{x}_i}\Big(\widetilde{\mu}_\theta\Big(\mathbf{x}_i\Big)\Big) \ge c_0 > 0$ a.s. and

$$\sup_{\mathbf{x}_i \in G_n} \left| \widetilde{H}_n\Big(\mathbf{x}_i, \widehat{\mu}_\theta\Big(\mathbf{x}_i\Big)\Big) - \widetilde{H}\Big(\mathbf{x}_i, \widehat{\mu}_\theta\Big(\mathbf{x}_i\Big)\Big) \right| \le \sup_{\mathbf{x}_i \in G_n} \sup_{\mu \in \Theta_n} \left| \widetilde{H}_n\Big(\mathbf{x}_i, \mu_\theta\Big(\mathbf{x}_i\Big) + \mu\Big) - \widetilde{H}\Big(\mathbf{x}_i, \mu_\theta\Big(\mathbf{x}_i\Big) + \mu\Big) \right|. \tag{2.3.2.20}$$

[28] It is monotone and right continuous.

Using (2.3.2.19) in (2.3.2.20), we get

$$\begin{aligned}\sup_{\mathbf{x}_i\in G_n}\left|\widehat{\mu}_\theta\left(\mathbf{x}_i\right)-\mu_\theta\left(\mathbf{x}_i\right)\right| &\leq c_0^{-1}\sup_{\mathbf{x}_i\in G_n}\sup_{\mu\in\Theta_n}\left|\widetilde{H}_n\left(\mathbf{x}_i,\mu_\theta\left(\mathbf{x}\right)+\mu\right)-\widetilde{H}\left(\mathbf{x}_i,\mu_\theta\left(\mathbf{x}\right)+\mu\right)\right| \\ &= O\left(S_n\right), \quad \text{by (2.3.2.18).} \qquad (2.3.2.21)\end{aligned}$$

This completes the proof.

□

Note that the difference $\widehat{F}_{\mathbf{x}_i}\left(y\right)-F_{\mathbf{x}_i}\left(y\right)=\widetilde{H}_n\left(\mathbf{x}_i,y\right)-\widetilde{H}\left(\mathbf{x}_i,y\right)$ and hence the rate of convergence of the former expression is implied by the latter one.

2.4 Scale function in $QAR - QARCH$

This section derives results on consistency and asymptotic normality for various scale functional value estimators for the process given in (1.4.0.2). The results are based on prior knowledge of the results in section (2.1) and (2.3).

2.4.1 Consistency and asymptotic normality of the estimator in $QARCH$

We begin by considering only the heteroscedastic part of process (1.4.0.2), while assuming the QAR, $\quad \mu_{t,\theta}=0$. That is we consider

$$Y_t=\sigma_{t,\theta}Z_t, \quad t=1,\dots,n. \qquad (2.4.1.1)$$

where now $M_\theta\left(Y_t\right) = M_\theta\left(Y_t,0\right)$. Observe that (2.4.1.1) can be written in terms of additive[29] noise as follows

$$M_\theta\left(Y_t\right)=\sigma_{t,\theta}+\sigma_{t,\theta}\left(M_\theta\left(Z_t\right)-1\right) \qquad (2.4.1.2)$$

where the last term on the right hand can be considered stationary with zero conditional θ-quantile. It then follows that $Q_{t,\theta}\left(M_\theta\left(Y_t\right)\right)=\sigma_{t,\theta}$. Thus , $\sigma_{t,\theta}$ can be obtained as a solution to the equation, $P\left(M_\theta\left(Y_t\right)-\sigma\leq 0\middle|\mathbf{X}_t=\mathbf{x}\right)=\theta$, where we assume $\sigma:\mathbf{R}^d\to\mathbf{R}_+$ to be a smooth but unknown nonparametric function at point $\mathbf{x}$. At a fixed

[29]See chapter 3 for more formulations.

θ and a fixed point $\mathbf{x}_i \in \mathbf{R}^d$ the kernel estimator, $\widehat{\sigma}_\theta\left(\mathbf{x}_i\right)$, of $\sigma_\theta\left(\mathbf{x}_i\right)$ will be defined as a solution to the equation $\widehat{F}_{\mathbf{x}_i}\left(\sigma\right) = \theta$. Hence we only need to estimate the conditional distribution of $M_\theta\left(Y_t, 0\right)$ at point y given $\mathbf{x}_i$, and obtain the scale function estimator as the inverse, $\widehat{\sigma}_\theta\left(\mathbf{x}_i\right) = \widehat{F}^{-1}_{\mathbf{x}_i}\left(\theta\right)$. Since the error term in (2.4.1.2) has a zero conditional θ-quantile, the stochastic process (2.4.1.2) can be modelled with the method described in section 2, by replacing the dependent variable Y_t by $M_\theta\left(Y_t\right)$. The following theorem establishes consistency and asymptotic normality of the conditional scale function estimator when $\mu_{t,\theta} = 0$.

Theorem 2.6 *Let $F_{\mathbf{x}_i}\left(y\right)$ be the conditional distribution function of $M_\theta\left(Y_t, 0\right)$ given $\mathbf{X}_t = \mathbf{x}_i$. In model (2.4.1.1), assume conditions (B1)-(B5),(C1)-(C6), (D1)-(D2) and (E1) and that (1.4.0.3) and (1.4.0.4) hold. Then*

$$\widehat{\sigma}_\theta\left(\mathbf{x}_i\right) - \sigma_\theta\left(\mathbf{x}_i\right) \rightarrow^p 0 \quad \textit{in probability} \tag{2.4.1.3}$$

In addition, if conditions (C7) are satisfied, then

$$\left(n\left|\mathbf{h}^{(i)}\right|\right)^{\frac{1}{2}}\left(\widehat{\sigma}_\theta\left(\mathbf{x}_i\right) - \sigma_\theta\left(\mathbf{x}_i\right) - B\left(\sigma_\theta\left(\mathbf{x}_i\right)\right) + o_p\left(\left|\left|\mathbf{h}^{(i)}\right|\right|^2\right)\right) \rightarrow^D N\left(0, \frac{\mathbf{V}^2\left(\sigma_\theta(\mathbf{x}_i)\right)}{f^2_{\mathbf{x}_i}\left(\sigma_\theta\left(\mathbf{x}_i\right)\right)}\right) \tag{2.4.1.4}$$

where the bias $B\left(\sigma_\theta\left(\mathbf{x}_i\right)\right)$ and $\mathbf{V}^2\left(\sigma_\theta\left(\mathbf{x}_i\right)\right)$ are as defined in theorem 2.2 and lemma 2.1 respectively.

The prove of theorem 2.6 proceeds in the same lines as the prove of theorem 2.2 with an initial estimator of a consistent conditional distribution function as in theorem 2.1. If the errors e_t, in (1.1.1.1), are symmetrically distributed, the relationship between the conditional distribution function of $M_\theta\left(Y_t\right)$ given $\mathbf{X}_t = \mathbf{x}_i$ and Y_t given $\mathbf{X}_t = \mathbf{x}_i$ at a fixed $\theta = 0.5$, becomes $P\left(M_{0.5}\left(Y_t\right) \leq \sigma_{0.5}\left(\mathbf{x}_i\right)\middle|\mathbf{x}_i\right) = 2P\left(Y_t \leq \sigma_{0.5}\left(\mathbf{x}_i\right)\middle|\mathbf{x}_i\right) - 1 = 0.5$, by lemma 1.3. This implies the conditional median absolute deviation, $\sigma_{0.5}\left(\mathbf{x}_i\right)$, can be estimated by $\widehat{\mu}_{0.75}\left(\mathbf{x}_i\right)$, when the conditional median is zero.

2.4.2 Consistency of the scale function estimator in $QAR - QARCH$

Consider again (1.4.0.2). When the QAR, $\mu_{t,\theta}$, is unknown, modeling the heteroscedasticity part can be based on the estimated residuals after removing the effect of the $\mu_{t,\theta}$ component. In the mean-variance models or in time series $AR - ARCH$ model, Engle [40] and Koenker and Zhao [76], have carried out the study in two stages: In the former the first step involves estimating the mean component by least squares and computing the residuals and then estimating the $ARCH$ part by regressing the squared residuals on the lagged squared residuals. The latter paper also studies the asymptotic behaviour of quantile regression estimator when applied to the estimated absolute residuals in (1.4.1.2). In this subsection we will present the consistency and asymptotic normality results of the kernel estimator of $\sigma_\theta\Big(\mathbf{X}_t\Big)$, at $\mathbf{x}_i$, when applied to the residuals in model (1.4.0.2). Our first step involves estimating the $\mu_\theta\Big(\mathbf{x}_i\Big)$ as in section (2.1) and computing the residuals. In the second step, we pass the residuals through the loss function M_θ. Finally, we estimate the function by applying the methods in section 2.1 to the transformed residuals conditional on $\mathbf{X}_t$.

In section (2.1), it was proved that $\widehat{\mu}_\theta\Big(\mathbf{x}_i\Big)$ is $\Big(n\Big|\mathbf{h}^{(i)}\Big|\Big)^{\frac{1}{2}}$-consistent estimator for $\mu_\theta\Big(\mathbf{x}_i\Big)$. Let $\widehat{\mu}_\theta\Big(\mathbf{X}_t\Big)$ be the estimated conditional θ-quantile of Y_t on $\mathbf{X}_t$ at point $\Big(y, \mathbf{x}_i\Big)$ and define the residuals as $\Big(Y_t - \widehat{\mu}_\theta\Big(\mathbf{X}_t\Big)\Big)$. First note that the consistency of $\widehat{\mu}_\theta\Big(\mathbf{X}_t\Big)$ implies that $\widehat{\mu}_\theta\Big(\mathbf{X}_t\Big) = \mu_\theta\Big(\mathbf{X}_t\Big) + \delta_n$ with $\delta_n = O\Big(\Big(n\Big|\mathbf{h}^{(i)}\Big|\Big)^{-1}\Big) > 0$, which is constant for fixed n. The estimated residuals can then be written as
$Y_t - \widehat{\mu}\Big(\mathbf{X}_t\Big) = \sigma_\theta\Big(\mathbf{X}_t\Big)Z_t - \delta_n$ and the transformed ones, which we denote as $\widehat{R}_t$, as

$$\widehat{R}_t = M_\theta\Big(Y_t - \widehat{\mu}_\theta\Big(\mathbf{X}_t\Big), 0\Big) = M_\theta\Big(\sigma_\theta\Big(\mathbf{X}_t\Big)Z_t - \delta_n, 0\Big). \tag{2.4.2.1}$$

Let

$$R_t = M_\theta\Big(Y_t - \mu_\theta\Big(\mathbf{X}_t\Big), 0\Big) \tag{2.4.2.2}$$

be the true residuals, where $\mu_\theta\Big(\mathbf{X}_t\Big)$ is known . The estimator of the conditional distribution can be written as

$$\widehat{F}_{\mathbf{x}_i}\Big(r'\Big) = \Big(n\Big|\mathbf{h}^{(i)}\Big|\Big)^{-1}\sum_{t=1}^{n} w_t\Big(\mathbf{x}_i\Big)\mathbf{I}_{\{\widehat{R}_t \le r'\}}, \tag{2.4.2.3}$$

where $w_t\Big(\mathbf{x}_i\Big)$ is defined as in (2.3.2.2) and r' is fixed real-valued on $\mathbf{R}_+$ in the neighborhood of r, based on (2.4.2.2). For the asymptotic properties of $\widehat{F}_{\mathbf{x}_i}\left(r^{'}\right)$ and subsequent properties of the scale function estimator, we will deal with r instead of $r^{'}$. In this case we make use of the uniform convergence and express $\widehat{R}_t$ in terms of R_t as $\widehat{R}_t = \widehat{R}_t - R_t + R_t = R_t + \delta_n$, now with $\delta_n = O\left(S_n\right)$ being the bound of the error. Observe that the indicator function $\mathbf{I}_{\{R_t \leq r\}}$ implies $\mathbf{I}_{\{R_t+\delta_n \leq r+\delta_n\}} = \mathbf{I}_{\{\widehat{R}_t \leq r+\delta_n\}}$. For $n \to \infty$, (2.4.2.1) and (2.4.2.3) suggest that the conditional distribution estimator of $\widehat{R}_t$ on $\mathbf{X}_t$ is consistent with bias of similar form as in the zero conditional θ-quantile case. Theorem 2.7 below gives the consistency and asymptotic normalitiy of the conditional scale function estimator as the inverse of (2.4.2.3) at a fixed θ.

Theorem 2.7 *Suppose conditions (B1)-(B5), (C1)-(C7),(D1)-(D2), (E1) and (1.4.1) hold for $y = r \in \mathbf{R}_+$, then under conditions of theorem 2.4*

$$\left(n\left|\mathbf{h}^{(i)}\right|\right)^{\frac{1}{2}}\left(\widehat{\sigma}_\theta\left(\mathbf{x}_i\right) - \sigma_\theta\left(\mathbf{x}_i\right) \quad - \quad B\left(\sigma_\theta\left(\mathbf{x}_i\right)\right) - O\left(S_n\left(n\left|\mathbf{h}^{(i)}\right|\right)^{-\frac{1}{2}}\right)f_{\mathbf{x}_i}\left(\sigma_\theta\left(\mathbf{x}_i\right)\right)\right)$$
$$\to^D N\left(0, \frac{\mathbf{V}^2\left(\sigma_\theta\left(\mathbf{x}_i\right)\right)}{f_{\mathbf{x}_i}\left(\sigma_\theta\left(\mathbf{x}_i\right)\right)}\right) \tag{2.4.2.4}$$

where the conditional density functions, $f_{\mathbf{x}_i}\left(.\right)$, are based on appropriate(response) random variables.

Proof of theorem 2.7
We first proof that $\widehat{F}_{\mathbf{x}_i}\left(r^{'}\right)$ is a consistent estimator for $F_{\mathbf{x}_i}\left(r\right)$. Let $r^{'} = r + O\left(S_n\right)$ with $O\left(S_n\right) \to 0$ as $n \to \infty$. Then by lemma 2.3,

$$\widehat{F}_{\mathbf{x}_i}\left(r^{'}\right) - F_{\mathbf{x}_i}\left(r\right) = \widehat{F}_{\mathbf{x}_i}\left(r\right) - F_{\mathbf{x}_i}\left(r\right) + O\left(S_n\right)f_{\mathbf{x}_i}\left(r\right) \tag{2.4.2.5}$$

and taking expectation on both sides and making use of lemma 2.2, it results in the bias

$$E\left[\widehat{F}_{\mathbf{x}_i}\left(r^{'}\right) - F_{\mathbf{x}_i}\left(r\right)\right] = B_n\left(r\right) + O\left(S_n\right)f_{\mathbf{x}_i}\left(r\right) \tag{2.4.2.6}$$

In similar lines as in lemma 2.2, we obtain the variance as

$$var\left[\widehat{F}_{\mathbf{x}_i}\left(r^{'}\right) - F_{\mathbf{x}_i}\left(r\right)\right] \sim \left(n\left|\mathbf{h}^{(i)}\right|\right)^{-1}\mathbf{V}^2\left(r\right) \tag{2.4.2.7}$$

where $var\left[\widehat{F}_{\mathbf{x}_i}\left(r\right) - F_{\mathbf{x}_i}\left(r\right)\right]$ is obtained from lemma 2.2. In both the bias and variance, terms of smaller order in probability have been left out. Because $B_n\left(r\right)$ is of order $O\left(\left|\left|\mathbf{h}^{(i)}\right|\right|^2\right)$, the mean squared error is seen to go to zero as n goes to infinity and hence $\widehat{F}_{\mathbf{x}_i}\left(r^{'}\right) \rightarrow F_{\mathbf{x}_i}\left(r\right)$ in probability with a rate which implies $\widehat{F}_{\mathbf{x}_i}\left(r^{'}\right)$ is consistent. To show that

$$\left(n\left|\mathbf{h}^{(i)}\right|\right)^{\frac{1}{2}}\left(\widehat{F}_{\mathbf{x}_i}\left(r\right) - F_{\mathbf{x}_i}\left(r\right) - B_n\left(r\right) - O\left(S_n\left(n\left|\mathbf{h}^{(i)}\right|\right)^{-\frac{1}{2}}\right)f_{\mathbf{x}_i}\left(r\right)\right) \tag{2.4.2.8}$$

is asymptotically normal, we proceed as in theorem 2.1, by replacing $var\left[\widehat{F}_{\mathbf{x}_i}\left(y\right)\right]$ by the estimated variance, $var\left[\widehat{F}_{\mathbf{x}_i}\left(r^{'}\right)\right]$. Finally, to proof that the left hand side of (2.4.2.4) is asymptotically normally distributed with mean zero, we note that by Krishnaiah [78], $\sup_{r\in\mathbf{R}}\left|\widehat{F}_{\mathbf{x}_i}\left(r\right) - F_{\mathbf{x}_i}\left(r\right)\right| \rightarrow 0$, in probability. The uniqueness assumption of $\sigma_\theta\left(\mathbf{x}_i\right)$, for any fixed $\mathbf{x}_i \in \mathbf{R}^d$, implies that there exist a $\epsilon > 0$ and $\delta\left(\epsilon\right)$, such that

$$\begin{aligned} P\left\{\left|\widehat{\sigma}_\theta\left(\mathbf{x}_i\right)\right.\right. &- \left.\left.\sigma_\theta\left(\mathbf{x}_i\right)\right| > \epsilon\right\} \leq P\left\{F_{\mathbf{x}_i}\left(\widehat{\sigma}_\theta\left(\mathbf{x}_i\right)\right) - F_{\mathbf{x}_i}\left(\sigma_\theta\left(\mathbf{x}_i\right)\right)\right| > \delta\right\} \\ &\leq P\left\{\sup_r\left|\widehat{F}_{\mathbf{x}_i}\left(r\right) - F_{\mathbf{x}_i}\left(r\right)\right| > \delta\right\} \end{aligned} \tag{2.4.2.9}$$

which goes to zero by above argument. At the same time $O\left(S_n\right)$ goes to zero, from theorem 2.4. Lastly, observe that since $\widehat{F}_{\mathbf{x}_i}\left(r^{'}\right)$ is asymptotically normally distributed, so is $\widehat{F}_{\mathbf{x}_i}\left(\widehat{\sigma}_\theta\left(\mathbf{x}_i\right)\right)$ and hence using the same arguments again as in theorem 2.2, we arrive at (2.4.2.4) with the specified quantities. □

Note that equation (2.4.2.5) shows that under conditions of theorem 2.4, with Y_t replaced by $M_\theta\left(Y_t - \widehat{\mu}_\theta\left(\mathbf{X}_t\right), 0\right)$, $\widehat{F}_{\mathbf{x}_i}\left(r^{'}\right)$ converges uniformly over $\mathbf{x}_i \in G_n$, i.e

$$\begin{aligned} \sup_{\mathbf{x}_i\in G_n}\left|\widehat{F}_{\mathbf{x}_i}\left(r^{'}\right) - F_{\mathbf{x}_i}\left(r\right)\right| &= \sup_{\mathbf{x}_i\in G_n}\left|\widehat{F}_{\mathbf{x}_i}\left(r\right) - F_{\mathbf{x}_i}\left(r\right)\right| + O\left(S_n\right) \\ &= O\left(S_n^{'}\right) + O\left(S_n\right) \qquad \text{a.s.} \end{aligned}$$

where $S_n^{'}$, apart from the bandwidths, is of the same form as S_n. Further, equations (2.3.2.20) and (2.3.2.21) shows that $\sup_{\mathbf{x}_i\in G_n}\left|\widehat{\sigma}_\theta\left(\mathbf{x}_i\right) - \sigma_\theta\left(\mathbf{x}_i\right)\right| = O\left(S_n^{'} + S_n\right)$.

Suppose $F^{(2)}$ is the conditional distribution of $M_\theta\Big(Y_t, \mu_\theta\big(\mathbf{X}_t\big)\Big)$ at point $\mathbf{x}_i$ and $F^{(1)}$ the conditional distribution of Y_t at point $\mathbf{x}_i$. Then as in remark of theorem 2.6, observe that for a symmetric $F^{(1)}$ and $\theta = 0.5$

$$F^{(2)}_{\mathbf{x}_i}\Big(\sigma_{0.5}\big(\mathbf{x}_i\big)\Big) = 2F^{(1)}_{\mathbf{x}_i}\Big(\mu_{0.5}\big(\mathbf{x}_i\big) + \sigma_{0.5}\big(\mathbf{x}_i\big)\Big) - 1 = 0.5$$

and $F_{\mathbf{x}_i}\Big(\mu_{0.5}\big(\mathbf{x}_i\big) + \sigma_{0.5}\big(\mathbf{x}_i\big)\Big) = 0.75$ implies $\mu_{0.5}\big(\mathbf{x}_i\big) + \sigma_{0.5}\big(\mathbf{x}_i\big) = \mu_{0.75}\big(\mathbf{x}_i\big)$. The estimator of $\sigma_{0.5}\big(\mathbf{x}_i\big)$ can be obtained as $\widehat{\sigma}_{0.5}\big(\mathbf{x}_i\big) = \widehat{\mu}_{0.75}\big(\mathbf{x}_i\big) - \widehat{\mu}_{0.5}\big(\mathbf{x}_i\big)$.
To assess the performance of $\widehat{\sigma}_\theta\big(\mathbf{x}_i\big)$, just as $\widehat{\mu}_\theta\big(\mathbf{x}_i\big)$, use its asymptotic mean squared error (AMSE). Using theorem 2.7 and equal bandwidths, only for brevity,

$$AMSE\Big(\widehat{\sigma}_\theta\big(\mathbf{x}_i\big)\Big) = B_1^2\Big(\sigma_\theta\big(\mathbf{x}_i\big)\Big) + \frac{\mathbf{V}^2\Big(\sigma_\theta\big(\mathbf{x}_i\big)\Big)}{nh^d f^2_{\mathbf{x}_i}\Big(\sigma_\theta\big(\mathbf{x}_i\big)\Big)} \tag{2.4.2.10}$$

where $B_1\Big(\sigma_\theta\big(\mathbf{x}_i\big)\Big) = B\Big(\sigma_\theta\big(\mathbf{x}_i\big)\Big) + O\big(S_n\big) f_{\mathbf{x}_i}\Big(\sigma_\theta\big(\mathbf{x}_i\big)\Big)$. The optimal bandwidth can be choosen such that the $AMSE\Big(\widehat{\sigma}_\theta\big(\mathbf{x}_i\big)\Big)$ is minimum.

The estimator of the QAR obtained from the conditional distribution can be considered as an estimator of conditional θ-quantile using the kernel based method of the implicit equation in Huber [67]. If we consider a minimization problem involving the kernel estimator for the loss function (1.1.2.1), we immediately note that the explicit equation is

$$\begin{aligned} H_n\big(\mathbf{x}_i, \mu\big) &= \widehat{F}_{\mathbf{x}_i}\big(\mu\big) - F_{\mathbf{x}_i}\Big(\mu_\theta\big(\mathbf{x}_i\big)\Big) \\ &= -\widehat{g}^{-1}\big(\mathbf{x}_i\big)\Big(n\big|\mathbf{h}^{(i)}\big|\Big)^{-1}\sum_{t=1}^{n}\mathbf{K}\Big(\mathbf{X}_t - \mathbf{x}_i; \mathbf{h}^{(i)}\Big)\Big(\mathbf{I}_{\{Y_t-\mu\le 0\}} - F_{\mathbf{x}_i}\Big(\mu_\theta\big(\mathbf{x}_i\big)\Big)\Big) \end{aligned}$$

which is measurable in $\big(Y_t, \mathbf{X}_t\big)$ and monotonically decreasing in μ. The equation also satisfies further assumptions on page 49 of Huber [67]. If we are interested in the consistency and asymptotic normality for $\Big(n\big|\mathbf{h}^{(i)}\big|\Big)^{\frac{1}{2}}\Big(\widehat{\mu}_\theta\big(\mathbf{x}_i\big) - \mu_\theta\big(\mathbf{x}_i\big)\Big)$ or $\Big(n\big|\mathbf{h}^{(i)}\big|\Big)^{\frac{1}{2}}\Big(\widehat{\sigma}_\theta\big(\mathbf{x}_i\big) - \sigma_\theta\big(\mathbf{x}\big)\Big)$ in the case of zero conditional θ-quantile, we only have to impose a few assumptions for dependent data and then directly show the estimator to be consistent and asymptotically normal using some standard rules. In the case of unknown conditional θ-quantile, and using explicit equations, the sandwich theorem, corollary 3.2 in Huber [67], continues to apply so long as the additional set of assumptions on page 131 are met. The advantage here is that the asymptotic properties of the simultaneous conditional θ-quantile and scale function could be carried out and investigated using again some prespecified rules.

2.5 Alternative methods

It is known that the Nadaraya-Watson estimator of the conditional distribution, can be considered as a local constant estimator obtained by the least square method. It has poor behaviour at the boundaries for unbalanced design matrix among other disadvantages, see Fan et al. [46]. The estimation could be improved by using either of the alternative methods described below.

(1) The first alternative is to use the Weighted Nadaraya-Watson estimator of the conditional distribution proposed in Hall et al. [62]. The conditional distribution estimator at the design point $\mathbf{x}_i$, can be written as

$$\widehat{F}_{\mathbf{x}_i}\left(y\right) = \frac{\sum_{t=1}^{n} P_t\left(\mathbf{x}_i\right)\mathbf{K}\left(\mathbf{X}_t - \mathbf{x}_i; \mathbf{h}^{(i)}\right)\mathbf{I}_{\{Y_t \leq y\}}}{\sum_{t=1}^{n} P_t\left(\mathbf{x}_i\right)\mathbf{K}\left(\mathbf{X}_t - \mathbf{x}_i; \mathbf{h}^{(i)}\right)}$$

where the weight function $P_t\left(\mathbf{x}_i\right)$ at point $\mathbf{x}_i$ is such that $P_t\left(\mathbf{x}_i\right) > 0$, $\sum_{t=1}^{n} P_t\left(\mathbf{x}_i\right) = 1$ and

$$\left(n\left|\mathbf{h}^{(i)}\right|\right)^{-1}\sum_{t=1}^{n}\left(\mathbf{X}_t - \mathbf{x}_i\right)P_t\left(\mathbf{x}_i\right)\mathbf{K}\left(\mathbf{X}_t - \mathbf{x}_i; \mathbf{h}^{(i)}\right) = 0 \qquad (2.5.0.11)$$

The $P_t\left(\mathbf{x}_i\right)$ is choosen such that $\sum_{t=1}^{n}\log\left(P_t\left(\mathbf{x}_i\right)\right)$ is minimized subject to the constraints $\sum_{t=1}^{n} P_t\left(\mathbf{x}_i\right) = 1$ and (2.5.0.11) through the Langrange multiplier rule. Then $\left\{P_t\left(\mathbf{x}_i\right)\right\}$ is given as

$$P_t\left(\mathbf{x}_i\right) = n^{-1}\left\{1 + \lambda\left(\left|\mathbf{h}^{(i)}\right|\right)^{-1}\mathbf{K}\left(\mathbf{x}_i - \mathbf{X}_t; \mathbf{h}^{(i)}\right\}^{-1}$$

where λ, which is a function of the data and design point $\mathbf{x}_i$, is uniquely defined by (2.5.0.11). For detailed set up, we make reference to the above mentioned paper.

(2) The second method is based on local polynomial approximation and extends a local constant fit to a local linear estimator of the conditional distribution, see Yu and Jones [114]. Its motivation lies on the fact that any smooth function can locally be expanded into a Taylor series around a point, say $\mathbf{x}_i$ in a compact subset $G_n \subset \mathbf{R}^d$. Consider the conditional mean regression curve $\mu\left(\mathbf{X}_t\right) = E\left[Y_t \middle| \mathbf{X}_t\right]$, as an example. If we assume that the p^{th} derivative exist at point $\mathbf{x}_i$, then $\mu\left(\mathbf{X}_t\right)$ can be expressed as a local polynomial of degree $p-1$ centered around $\mathbf{x}_i$ as

$$\mu\left(\mathbf{X}_t\right) = \sum_{j=0}^{p-1} \beta_j\left(\mathbf{x}_i\right)\left(\mathbf{X}_t - \mathbf{x}_i\right)^j + Rem\left(\mathbf{X}_t\right)$$

for $\mathbf{X}_t$ sufficiently close to $\mathbf{x}_i$, where $\beta_j\left(\mathbf{x}_i\right) = \frac{1}{j!}\frac{d^j\mu\left(\mathbf{X}_t\right)}{d\mathbf{X}_t^j}\Big|_{\mathbf{X}_t=\mathbf{x}_i} = \frac{\mu^{(j)}\left(\mathbf{x}_i\right)}{j!}, j = 1, 2, \ldots, p-1$ and $Rem\left(\mathbf{X}_t\right)$ is the remainder term consisting of terms with derivative greater than $p-1$. The local polynomial approximation is then given by disregarding the remainder term,

$$\mu\left(\mathbf{X}_t\right) \approx \sum_{j=0}^{p-1} \beta_j\left(\mathbf{x}_i\right)\left(\mathbf{X}_t - \mathbf{x}_i\right)^j$$

For practical purpose, the polynomial is usually restricted to degree 1, i.e. local linear. The estimation of conditional density and its derivatives with local quadratic for strictly stationary process is investigated in Fan et al. [47], where a double-kernel idea is proposed. Let $\widetilde{K}$ be a kernel function (a nonnegative density function) , then as $\widetilde{h} \to 0$, $E\left[\widetilde{K}_{\widetilde{h}}\left(Y_t - y\right)\middle|\mathbf{X}_t\right] \approx \widetilde{f}_{\mathbf{X}_t}\left(y\right)$ where the left hand side can be regarded as the regression of $\widetilde{K}_{\widetilde{h}}\left(Y_t - y\right)$ on $\mathbf{X}_t$. It can be seen that the Taylor series expansion about $\mathbf{x}_i = \left(\mathbf{x}_{t,1}, \ldots, \mathbf{x}_{t,d}\right)^T$ yields

$$\begin{aligned} E\left[\widetilde{K}_{\widetilde{h}}\left(Y_t - y\right)\middle|\mathbf{X}_t\right] &\approx \widetilde{f}_{\mathbf{x}_i}\left(y\right) + \nabla\widetilde{f}_{\mathbf{x}_i}\left(y\right)\left(\mathbf{X}_t - \mathbf{x}_i\right)^T + \frac{1}{2}\left(\mathbf{X}_t - \mathbf{x}_i\right)^T\nabla^2\widetilde{f}_{\mathbf{x}_i}\left(y\right)\left(\mathbf{X}_t - \mathbf{x}_i\right) \\ &\equiv \beta_0\left(\mathbf{x}_i\right) + \beta_1\left(\mathbf{x}_i\right)^T\left(\mathbf{X}_t - \mathbf{x}_i\right) + \beta_2\left(\mathbf{x}_i\right)\left\{\left(\mathbf{X}_t - \mathbf{x}_i\right)\left(\mathbf{X}_t - \mathbf{x}_i\right)^T\right\} \end{aligned}$$

The estimators are then given by $\widehat{\widetilde{f}}_{\mathbf{x}_i}\left(y\right) = \widehat{\beta}_0\left(\mathbf{x}_i\right)$ and $\nabla\widehat{\widetilde{f}}_{\mathbf{x}_i}\left(y\right) = \widehat{\beta}_1\left(\mathbf{x}_i\right)$ which are solutions of a convex loss function. In similar approach local linear double-kernel smoothing can be used in the context of conditional quantile estimation, see Yu and Jones [114], for the asymptotic properties of the resulting estimators in the case of scalar design variable and iid data including the bandwidth choice. If $\widetilde{K}$ is the second kernel function, the corresponding cdf is given as $\widetilde{F}\left(v\right) = \int_{-\infty}^{v} \widetilde{K}\left(u\right)du$. For observation Y_t, we have $\int_{-\infty}^{y} \widetilde{K}_{\widetilde{h}}\left(Y_t - v\right)dv = \widetilde{F}\left(\frac{y-Y_t}{\widetilde{h}}\right)$ and as $\widetilde{h} \to 0$ we have the conditional expectation giving $E\left[\widetilde{F}\left(\frac{y-Y_t}{\widetilde{h}}\right)\middle|\mathbf{X}_t\right] \approx \widetilde{F}_{\mathbf{X}_t}\left(\frac{y-Y_t}{\widetilde{h}}\right)$ and therefore the local linear approximation at $\mathbf{x}_i$ becomes

$$\widetilde{F}_{\mathbf{X}_t}\left(y\right) \approx \widetilde{F}_{\mathbf{x}_i}\left(y\right) + \nabla\widetilde{F}_{\mathbf{x}_i}\left(y\right)\left(\mathbf{X}_t - \mathbf{x}_i\right)^T + \frac{1}{2}\left(\mathbf{X}_t - \mathbf{x}_i\right)^T \nabla^2\widetilde{F}_{\mathbf{x}_i}\left(y\right)\left(\mathbf{X}_t - \mathbf{x}_i\right)$$

This motivates a local linear regression estimator of the conditional distribution at point $\mathbf{x}_i$ as

$$\widehat{F}_{\mathbf{h}^{(i)},\widetilde{h}}\left(y\middle|\mathbf{x}_i\right) = \widehat{\beta}_0\left(\mathbf{x}_i\right)$$

where

$$\left(\widehat{\beta}_0\left(\mathbf{x}_i\right), \widehat{\beta}_1\left(\mathbf{x}_i\right)\right) = \arg\min_{\beta_0,\beta_1}\left\{\left(n\left|\mathbf{h}^{(i)}\right|\right)^{-1}\sum_{t=1}^{n}\left[\widetilde{F}\left(\frac{y-Y_t}{\widetilde{h}}\right)-\beta_0-\beta_1\left(\mathbf{X}_t-\mathbf{x}_i\right)\right]^2\mathbf{K}\left(\mathbf{X}_t-\mathbf{X}_t;\mathbf{h}^{(i)}\right)\right\}$$

The conditional θ-quantile estimator is then obtained as the inverse of the estimated distribution, $\widehat{\mu}_\theta\left(\mathbf{x}_i\right) = \widehat{F}^{-1}_{\mathbf{h}^{(i)},\widetilde{h}}\left(\theta\middle|\mathbf{x}_i\right)$, if $\widehat{\mu}_\theta\left(\mathbf{x}_i\right)$ is uniquely defined. The asymptotics in Yu and Jones [114] holds also for the scale functions based on zero conditional θ-quantile for iid data. It would also be interesting to apply the double kernel method in the case of non-zero conditional θ-quantile, as was done in subsection 2.4.2, in the case of Nadaraya-Watson estimator. We must also point out that there has not been attempts to use the distributional based local polynomial to estimate the conditional θ-quantile function for the time series data. However, we do not discuss this further but leave it for future research prospects.

2.6 Conclusion

In this chapter, we derived the QAR estimator by inverting conditional distribution function estimator. It has been shown that the estimator is consistent and asymptotically normally distributed and that, under suitable conditions, it converges uniformly with an appropriate rate. We have derived the scale function estimator by inverting conditional distribution estimator. The consistency and asymptotic normality for the estimators based on both known and unknown QAR have also been established.

3 Direct estimation method

In this chapter, we present results based on direct estimation of the scale (or QAR) function, using direct minimization of a loss function. In section (3.1) various forms of the estimators which could be derived from model (1.4.0.2) are outlined. Section (3.2) treats the estimation as a local constant problem and presents consistency and asymptotic normality for the estimator. In section (3.3) we base the estimation of the scale function on local polynomials, which also provides estimators for the derivative of the scale function. The consistency of the scale estimate is provided through the work in Honda [65]. The rest of the sections in the chapter provide numerical comparison results and proposes a method for standardizing the scale function. Lastly, we extend the approach to nonparametric GQARCH type models similar to GARCH models.
Throughout we will assume equal bandwidths, for simplicity, and concentrate on the modeling scale function.

3.1 The estimators

For intuitive understanding make reference to quantile-scale model (1.4.0.2). We can estimate $\mu_{t,\theta}$ and $\sigma_{t,\theta}$ such that the following equation is simultaneously satisfied,

$$P\Big(Y_t \leq \mu_{t,\theta}\Big|\mathbf{X}_t = \mathbf{x}\Big) = P\Big(M_\theta\Big(Y_t, \mu_{t,\theta}\Big) \leq \sigma_{t,\theta}\Big|\mathbf{X}_t = \mathbf{x}\Big) = \theta \tag{3.1.0.12}$$

The second equation is just another way of expressing that $\sigma_{t,\theta}$ is the conditional scale function of the variable Y_t. The structure of model (1.4.0.2), provides various estimating directions: A two stage procedure can be used to estimate $\sigma_{t,\theta}$ by first obtaining the consistent estimator of $\mu_{t,\theta}$, as in chapter 2, which is then substituted in the second part of (3.1.0.12) to derive the estimate of $\sigma_{t,\theta}$ nonparametrically. Note that $M_\theta\Big(Y_t, \mu_{t,\theta}\Big) = \sigma_{t,\theta} M_\theta\Big(Z_t, 0\Big)$ and observe the following alternatives for estimating $\sigma_{t,\theta}$.
Write

$$\frac{1}{\sigma_{t,\theta}} M_\theta\Big(Y_t, \mu_{t,\theta}\Big) = 1 + Z_{t,1} \tag{3.1.0.13}$$

where $Z_{t,1} = M_\theta\Big(Z_t, 0\Big) - 1$ has zero θ-quantile. This leads to an asymmetric least absolute deviations estimator of $\sigma_{t,\theta}$,

$$\overline{\sigma}_1 = \arg\min_{\sigma} \sum_{t=1}^{n} \mathbf{K_h}\Big(\mathbf{x}_i - \mathbf{X}_t\Big) M_\theta\Big(\frac{1}{\sigma} M_\theta\Big(Y_t, \mu_{t,\theta}\Big), 1\Big) \tag{3.1.0.14}$$

where $\mathbf{K_h}\Big(\mathbf{u}\Big) = \frac{1}{h^p}\mathbf{K}\Big(\mathbf{u};\mathbf{h}\Big)$ and $\sigma \in \mathbf{R}_+$.

The second estimator is motivated by the regression equation

$$M_\theta\Big(Y_t, \mu_{t,\theta}\Big) = \sigma_{t,\theta} + Z_{t,2} \tag{3.1.0.15}$$

where $Z_{t,2} = \sigma_{t,\theta}\Big(M_\theta\Big(Z_t, 0\Big) - 1\Big)$, which also has zero θ-quantile. The asymmetric absolute least deviations estimator is then obtained as

$$\overline{\sigma}_2 = \arg\min_{\sigma} \sum_{t=1}^{n} \mathbf{K_h}\Big(\mathbf{x}_i - \mathbf{X}_t\Big) M_\theta\Big(M_\theta\Big(Y_t, \mu_{t,\theta}\Big), \sigma\Big) \tag{3.1.0.16}$$

Our third absolute deviations estimator is motivated by the regression relationship

$$\log\Big(M_\theta\Big(Y_t, \mu_{t,\theta}\Big)\Big) = \log\Big(\sigma_{t,\theta}\Big) + Z_{t,3} \tag{3.1.0.17}$$

where $Z_{t,3} = \log\Big(M_\theta\Big(Z_t, 0\Big)\Big)$ again has zero θ-quantile. The asymmetric least absolute deviation estimator for $\log\Big(\sigma_{t,\theta}\Big)$ is defined by

$$\widehat{\varsigma} = \arg\min_{\varsigma} \sum_{t=1}^{n} \mathbf{K_h}\Big(\mathbf{x}_i - \mathbf{X}_t\Big) M_\theta\Big(\log\Big(M_\theta\Big(Y_t, \mu_{t,\theta}\Big)\Big), \varsigma\Big) \tag{3.1.0.18}$$

where the estimator for $\sigma_{t,\theta}$ is $\exp\Big(\widehat{\varsigma}\Big)$. We consider all the errors $Z_{t,1}, Z_{t,2}$ and $Z_{t,3}$ as stationary time series and (3.1.0.13),(3.1.0.15) and (3.1.0.17), as general nonparametric quantile regression problems. However, we will later base our estimation on (3.1.0.17). The reason is two fold: First, because of its intuitive conformation with the common assumptions underlying regression models, i.e the error $Z_{t,3}$ are iid. Secondly, the fact that the unknown function is on $\mathbf{R}$ implies the properties of the loss function would be identical to the one we would have used for estimating the the QAR of Y_t and thus avoiding repetitions. Whereas considering (3.1.0.14) with also iid $Z_{t,1}$ leads to a more complicated function to minimize, and we would have to study it separately. On the contrary, the errors in (3.1.0.15) are clearly not independent and therefore some additional weights reflecting the dependence would need to be incorporated on to the weight function. At the same time the unknown function σ is only defined on $\mathbf{R}_+$. However, we would like to remark that the structure of the model (1.4.0.2) also allows direct estimation of $\sigma^2_{t,\theta}$ in all the alternatives. For instance the square would motivate the regression of the form

$$\frac{1}{\sigma^2_{t,\theta}} M^2_\theta\Big(Y_t, \mu_{t,\theta}\Big) = 1 + Z^2_{t,1}$$

where $Z^2_{t,1} = \Big(M^2_\theta\Big(Z_t, 0\Big) - 1\Big)$ which still has zero θ-quantile. The estimator is then obtained as

$$\overline{\sigma}^2_1 = \arg\min_\sigma \sum_{t=1}^n \mathbf{K_h}\Big(\mathbf{x}_i - \mathbf{X}_t\Big) M_\theta\Big(\frac{1}{\sigma} M^2_\theta\Big(Y_t, \mu_{t,\theta}\Big), 1\Big)$$

which is just the square of the estimator obtained in (3.1.0.14). We will assume throughout the following section that Y_t has zero conditional θ-quantile, $\mu_{t,\theta}$ and we write $\widetilde{Y}_t = \log\Big(M_\theta\Big(Y_t, 0\Big)\Big)$.

3.2 Local constant estimator of the scale function

Now, we consider the transformed data $\widetilde{Y}_t$ which satisfy the QAR-model

$$\widetilde{Y}_t = \sigma_{t,\theta} + Z_t, \quad t = 1, 2, \dots, \tag{3.2.0.19}$$

where $Z_t = Z_{t,3}$ are as defined in the previous section and $\sigma_{t,\theta}$ corresponds to $\log\Big(\sigma_{t,\theta}\Big)$ in model (1.4.0.2), i.e $\sigma_{t,\theta} = \exp(\varsigma_{t,\theta})$ from the previous section. We assume the quantile regression function $\sigma_\theta\Big(\mathbf{X}_t\Big)$ is at least once differentiable and therefore can be expressed as $\sigma_\theta\Big(\mathbf{X}_t\Big) = \sigma_\theta\Big(\mathbf{x}_i\Big) + r\Big(\mathbf{X}_t, \mathbf{x}_i\Big)$, for $\mathbf{X}_t \in \mathbf{R}^d$ in the neighborhood of $\mathbf{x}_i \in \mathbf{R}^d$ using Taylor expansion. Here $r\Big(\mathbf{X}_t, \mathbf{x}_i\Big)$ is the remainder term comprising of terms of derivatives of $\sigma_\theta\Big(\mathbf{x}_i\Big)$ of order one and higher. In this section we will discard the remainder term and approximate the quantile regression $\sigma_\theta\Big(\mathbf{X}_t\Big)$ locally at $\mathbf{x}_i$ as $\sigma_\theta\Big(\mathbf{X}_t\Big) \approx \sigma\Big(\mathbf{x}_i\Big)$. This reduces the problem to a local constant approximation. We fit the local constant by using the weighted asymmetric least absolute regression. For $\mathbf{x}_i = \Big(x_{i,1}, \dots, x_{i,d}\Big)$, a point in $\mathbf{R}^d$ and $\mathbf{X}_t = \Big(X_{i,1}, \dots, X_{i,d}\Big)$, we define a local constant estimator for $\sigma_\theta\Big(\mathbf{x}_i\Big)$ as the minimizer of

$$\sum_{t=1}^n \mathbf{K_h}\Big(\mathbf{x}_i - \mathbf{X}_t\Big) M_\theta\Big(\widetilde{Y}_t, \sigma\Big) \tag{3.2.0.20}$$

with respective to σ. Denote the estimator of $\sigma_\theta\left(\mathbf{x}_i\right)$ as $\widehat{\sigma}_\theta\left(\mathbf{x}_i\right) = \widehat{\sigma}$ and define the true objective function at point $\mathbf{x}_i \in \mathbf{R}^d$ as

$$Q\left(\mathbf{x}_i, \sigma\right) = E\left[\left(M_\theta\left(\widetilde{Y}_t, \sigma\right) - M_\theta\left(\widetilde{Y}_t, 0\right)\right)\Big|\mathbf{X}_t = \mathbf{x}_i\right], \tag{3.2.0.21}$$

where the additional last term ensures that $Q\left(\mathbf{x}_i, \sigma\right)$ is finite. To estimate $\sigma_\theta\left(\mathbf{x}_i\right)$, we define the kernel estimator of $Q\left(\mathbf{x}_i, \sigma\right)$ as

$$\widehat{Q}_n\left(\mathbf{x}_i, \sigma\right) = \frac{1}{n}\sum_{t=1}^{n} \mathbf{K_h}\left(\mathbf{x}_i - \mathbf{X}_t\right)\left(M_\theta\left(\widetilde{Y}_t, \sigma\right) - M_\theta\left(\widetilde{Y}_t, 0\right)\right), \tag{3.2.0.22}$$

where now (3.2.0.22) replaces (3.2.0.20) . The kernel estimator of $\sigma_\theta\left(\mathbf{x}_i\right)$ is then given by

$$\widehat{\sigma}_\theta\left(\mathbf{x}_i\right) = \arg\min_{\sigma \in \Theta_n} \widehat{Q}_n\left(\mathbf{x}_i, \sigma\right) \tag{3.2.0.23}$$

where $\Theta_n \subset \mathbf{R}$, is a compact subset of $\mathbf{R}$. The following section derives the consistency for the estimator $\widehat{\sigma}_\theta\left(\mathbf{x}_i\right)$.

3.2.1 Consistency and asymptotic distribution

Let $Q_n\left(\mathbf{x}_i, \sigma\right) = \sum_{t=1}^{n} \mathbf{K_h}\left(\mathbf{x}_i - \mathbf{X}_t\right) Q\left(\mathbf{X}_t, \sigma\right)$ where $Q\left(\mathbf{X}_t, \sigma\right) = E\left[\left(M_\theta\left(\widetilde{Y}_t, \sigma\right) - M_\theta\left(\widetilde{Y}_t, 0\right)\right)\Big|\mathbf{X}_t\right]$, is a deterministic function. To show that $\widehat{\sigma}_\theta\left(\mathbf{x}_i\right)$ is weakly consistent to $\sigma_\theta\left(\mathbf{x}_i\right)$, the idea is first to show that $\widehat{Q}_n\left(\mathbf{x}_i, \sigma\right)$ is weakly consistent to $Q\left(\mathbf{x}_i, \sigma\right)$ for all $\mathbf{x}_i$. We shall need a few technical properties of $M_\theta\left(.\right)$, expressed in the following lemmas.

Lemma 3.1 *Let $\left(y, \mu\right)$ be real-valued random variables and define $q\left(y, \mu\right)$ as*

$$q\left(y, \mu\right) = M_\theta\left(y, \mu\right) - M_\theta\left(y, 0\right).$$

Then for all y, $q\left(y, \mu\right)$ is Lipschitz continuous in μ with Lipschitz constant 1, i.e $\left|q\left(y, \mu\right) - q\left(y, \mu'\right)\right| \leq \left|\mu - \mu'\right|$ for all y, μ, μ'.

Proof of lemma 3.1 :

Note that

$$q\Big(y,\mu\Big) - q\Big(y,\mu^{'}\Big) = \theta\Big(\mu^{'} - \mu\Big) - \Big(\Big(y-\mu\Big)\mathbf{I}_{\{y-\mu\leq 0\}} - \Big(y-\mu^{'}\Big)\mathbf{I}_{\{y-\mu^{'}\leq 0\}}\Big). \quad (3.2.1.1)$$

For $\mu < y < \mu^{'}$, we have $\mathbf{I}_{\{y-\mu\leq 0\}} = 0, \quad \mathbf{I}_{\{y-\mu^{'}\leq 0\}} = 1$, and (3.2.1.1) becomes

$$\begin{aligned} q\Big(y,\mu\Big) - q\Big(y,\mu^{'}\Big) &= \theta\Big(\mu^{'}-\mu\Big) - \Big(y-\mu^{'}\Big) \\ &= \Big(y-\mu\Big) - \Big(1-\theta\Big)\Big(\mu^{'}-\mu\Big) \end{aligned} \quad (3.2.1.2)$$

For $(y-\mu^{'}) > 0$ and $(y-\mu) > 0$, the last two expressions on the right of (3.2.1.2) both imply

$$-\Big(1-\theta\Big)\Big(\mu^{'}-\mu\Big) \leq q\Big(y,\mu\Big) - q\Big(y,\mu^{'}\Big) \leq \theta\Big(\mu^{'}-\mu\Big).$$

and therefore $\Big|q\Big(y,\mu\Big) - q\Big(y,\mu^{'}\Big)\Big|$ is bounded from above by at least one of $\theta\Big(\mu^{'}-\mu\Big)$ and $\Big(1-\theta\Big)\Big(\mu^{'}-\mu\Big)$. Similarly, for $\mu \leq \mu^{'} < y$ and $y < \mu \leq \mu^{'}$, we have respectively $\mathbf{I}_{\{y-\mu\leq 0\}} = 0, \quad \mathbf{I}_{\{y-\mu^{'}\leq 0\}} = 0$ implying $q\Big(y,\mu\Big) - q\Big(y,\mu^{'}\Big) = \theta\Big(\mu^{'}-\mu\Big)$ and $\mathbf{I}_{\{y-\mu\leq 0\}} = 1, \quad \mathbf{I}_{\{y-\mu^{'}\leq 0\}} = 1$ implying $q\Big(y,\mu\Big) - q\Big(y,\mu^{'}\Big) = \Big(1-\theta\Big)\Big(\mu-\mu^{'}\Big)$. Hence

$$\begin{aligned} \Big|q\Big(y,\mu\Big) - q\Big(y,\mu^{'}\Big)\Big| &\leq \max\Big(\theta, 1-\theta\Big)\Big|\mu-\mu^{'}\Big| \\ &\leq \Big|\mu-\mu^{'}\Big| \end{aligned} \quad (3.2.1.3)$$

immediately implies the assertion. □

Lemma 3.1 implies that the function $q\Big(y,\mu\Big)$ is not only convex, but also continuous in $\mu \in \mathbf{R}$. In the case of $y \in \mathbf{R}$ and $u = \Big(\mu_1, \mu_2, \ldots, \mu_d\Big) \in \mathbf{R}^d$, the lemma continues to apply. In this case equation (3.2.1.3) becomes $q\Big(y,u\Big) - q\Big(y,u^{'}\Big) \leq \sum_{i=1}^{d}\Big|\mu_i - \mu_i^{'}\Big|$ which is an L_1 norm. We will denote the norm as $\rho\Big(u,u^{'}\Big)$ and use it whenever it is necessary in the rest of this chapter. The following lemma will be used to establish the existance and uniqueness of the minimum, when we take $E\Big[\Big(q\Big(y,\mu\Big) - q\Big(y,0\Big)\Big)\Big|\mathbf{X}_t\Big]$ as an objective function for minimization.

Lemma 3.2 *Let $q\Big(y,\mu\Big)$ be defined as in lemma 3.1. Let $\Big(Y_t, \mathbf{X}_t\Big) \in \mathbf{R} \times \mathbf{R}^d$ with conditional density and quantile functions, of Y_t on $\mathbf{X}_t$ given as $f_{\mathbf{X}_t}\Big(y\Big) : \mathbf{R}^{1+d} \to \mathbf{R}$ and $y_\theta = y_\theta\Big(\mathbf{X}_t\Big)$ respectively. Then*

1.

$$E\Big[q\Big(Y_t,\mu\Big)\Big|\mathbf{X}_t\Big]-E\Big[q\Big(Y_t,y_\theta\Big)\Big|\mathbf{X}_t\Big]=\begin{cases}E\Big[\Big(Y_t-\mu\Big)\mathbf{I}_{[\mu,y_\theta]}\Big|\mathbf{X}_t\Big], & :\quad\forall\mu\le y_\theta\\ E\Big[\Big(\mu-Y_t\Big)\mathbf{I}_{[y_\theta,\mu]}\Big|\mathbf{X}_t\Big], & :\quad\forall\mu\ge y_\theta\end{cases}\tag{3.2.1.4}$$

2. Let $\Big|\mu-y_\theta\Big|\ge\delta>0$. *Then for a suitable lower bound* $c\Big(\mathbf{X}_t\Big)$ *of* $f_{\mathbf{X}_t}\Big(y\Big)$ *on* $\Big[y_\theta-\delta,y_\theta+\delta\Big]$,

$$E\Big[q\Big(Y_t,\mu\Big)\Big|\mathbf{X}_t\Big]-E\Big[q\Big(Y_t,y_\theta\Big)\Big|\mathbf{X}_t\Big]\ge c\Big(\mathbf{X}_t\Big)\frac{\delta^2}{2}$$

3. Assume $f\Big(\mathbf{X}_t,y\Big)$ *is continuous and positive in the neighborhood of* $\Big(\mathbf{x}_i,y_\theta\Big(\mathbf{x}_i\Big)\Big)$ *and let* $\Big|\mu-y_\theta\Big(\mathbf{x}_i\Big)\Big|\ge\delta>0$ *for some* $\mathbf{x}_i$. *Then*

$$E\Big[q\Big(Y_t,\mu\Big)\Big|\mathbf{X}_t=\mathbf{x}_i\Big]-E\Big[q\Big(Y_t,y_\theta\Big(\mathbf{x}_i\Big)\Big)\Big|\mathbf{X}_t=\mathbf{x}_i\Big]\ge c\frac{\delta^2}{2}\tag{3.2.1.5}$$

for some constant $c>0$ *which is uniform lower bound of* $f_{\mathbf{X}_t}\Big(y\Big)$ *on* $\Big[y_\theta\Big(\mathbf{x}_i\Big)-\delta^*\Big]$ *for all* $\mathbf{X}_t$ *in a neighborhood around* $\mathbf{x}_i$ *and some* $\delta^*>0$.

Proof of lemma 3.2 :

For part (1), consider first the case when $\mu<y_\theta$. Then

$$\begin{aligned}E\Big[q\Big(Y_t,\mu\Big)\Big|\mathbf{X}_t\Big] &- E\Big[q\Big(Y_t,y_\theta\Big)\Big|\mathbf{X}_t\Big]=\Big(E\Big[q\Big(Y_t,\mu\Big)\Big|\mathbf{X}_t\Big]-E\Big[q\Big(Y_t,y_\theta\Big)\Big|\mathbf{X}_t\Big]\Big)\mathbf{I}_{\{\mu\le y_\theta\le Y_t\}}\\ &+ \Big(E\Big[q\Big(Y_t,\mu\Big)\Big|\mathbf{X}_t\Big]-E\Big[q\Big(Y_t,y_\theta\Big)\Big|\mathbf{X}_t\Big]\Big)\mathbf{I}_{\{Y_t<\mu\le y_\theta\}}\\ &+ \Big(E\Big[q\Big(Y_t,\mu\Big)\Big|\mathbf{X}_t\Big]-E\Big[q\Big(Y_t,y_\theta\Big)\Big|\mathbf{X}_t\Big]\Big)\mathbf{I}_{\{\mu\le Y_t\le y_\theta\}}\\ &= \theta y_\theta\mathbf{I}_{(y_\theta,\infty)}+\Big(1-\theta\Big)\Big(\mu-y_\theta\Big)\mathbf{I}_{(-\infty,\mu)}\\ &+ \Big(Y_t-\theta\mu-\Big(1-\theta\Big)y_\theta\Big)\mathbf{I}_{[\mu,y_\theta]}\\ &= \theta\Big(y_\theta-\mu\Big)\Big)\mathbf{I}_{(y_\theta,\infty)}-\Big(1-\theta\Big)\Big(y_\theta-\mu\Big)\mathbf{I}_{(-\infty,y_\theta)}\\ &+ \Big(Y_t-\theta\mu-\Big(1-\theta\Big)y_\theta\Big)+\Big(1-\theta\Big)\Big(y_\theta-\mu\Big)\Big)\mathbf{I}_{[\mu,y_\theta]}\end{aligned}$$

where $\mathbf{I}$ is an indicator function with respect to Y_t, i.e $\mathbf{I}_{(-\infty,y_\theta)}=\mathbf{I}_{(-\infty,y_\theta)}\Big(Y_t\Big)$. The expectation of the first sum goes to zero as the law of iterated expections gives

$$\begin{aligned} E\Big[\Big(y_\theta &- \mu\Big)\Big(\theta \mathbf{I}_{(y_\theta,\infty)} - \Big(1-\theta\Big)\mathbf{I}_{(-\infty,y_\theta)}\Big)\Big|\mathbf{X}_t\Big] \\ &= E\Big[\Big(y_\theta - \mu\Big)E\Big[\Big(\theta \mathbf{I}_{(y_\theta,\infty)} - \Big(1-\theta\Big)\mathbf{I}_{(-\infty,\mu_\theta)}\Big)\Big|\mathbf{X}_t\Big]\Big] \\ &= 0 \end{aligned} \tag{3.2.1.6}$$

The second part is just $\Big(Y_t - \mu\Big)\mathbf{I}_{[\mu,y_\theta]}$, and the assertion follows for $\mu \le y_\theta$. The statement for $y_\theta \le \mu$ follows completely analogously.

In part (2), observe that, if $\mu \le y_\theta$, we have even $\mu \le y_\theta - \delta$. Then using part(1) of this lemma

$$\begin{aligned} E\Big[q\Big(Y_t,\mu\Big)\Big|\mathbf{X}_t\Big] &- E\Big[q\Big(Y_t,y_\theta\Big)\Big|\mathbf{X}_t\Big] \\ &= E\Big[\Big(Y_t - \mu\Big)\mathbf{I}_{[\mu,y_\theta]}\Big|\mathbf{X}_t\Big] \\ &= \int_{\mu}^{y_\theta}\Big(u-\mu\Big)f_{\mathbf{X}_t}\Big(u\Big)du \\ &\ge \int_{y_\theta-\delta}^{y_\theta}\Big(u-\mu\Big)f_{\mathbf{X}_t}\Big(u\Big)du \\ &\ge c\Big(\mathbf{X}_t\Big)\int_{y_\theta-\delta}^{y_\theta}\Big(u-\mu\Big)du \\ &\ge c\Big(\mathbf{X}_t\Big)\delta\Big(y_\theta - \mu - \frac{\delta}{2}\Big) \\ &\ge c\Big(\mathbf{X}_t\Big)\frac{\delta^2}{2} \end{aligned}$$

where for the first inequality, the integrand is assumed to be nonnegative. In the second inequality, we assumed that $f_{\mathbf{X}_t}\Big(u\Big) \ge c\Big(\mathbf{X}_t\Big) > 0$ for $\Big|\mu - y_\theta\Big| \le \delta$ and for the third one, that $y_\theta - \mu \ge \delta$. The case when $\mu \ge y_\theta$ can be delt with analogously.

In the third part, observe that under the given assumption together with condition (C2) on continuity of $g\Big(\mathbf{x}_i\Big)$, we have that $f_{\mathbf{x}}\Big(y\Big) \ge c > 0$ for all $\mathbf{x}, y$ in the neighborhood of $\Big(\mathbf{x}_i, y_\theta\Big(\mathbf{x}_i\Big)\Big)$. Therefore, the arguments in the proof of part (2) can be made uniform.

□

From lemmas 3.1 and 3.2 and the convexity of $q\Big(y,.\Big)$, we conclude that the minimum, $\arg\min_{\mu\in\Theta_n} E\Big[\Big(q\Big(Y_t,\mu\Big) - \Big(q\Big(Y_t,0\Big)\Big|\mathbf{X}_t\Big]$, in a compact convex subset $\Theta_n \subset \mathbf{R}$ exists and is unique. The results of lemmas 3.1 and 3.2 also applies to the loss function defined in Chaudhuri [24], i.e for any $\left(y,\mu\right)$ as defined above, $\widetilde{M}_\theta\left(y,\mu\right) = \left|y-\mu\right| + \left(2\theta-1\right)\left(y-\mu\right)$. This is because for any fixed θ the relationship, $M_\theta\left(y,\mu\right) = \frac{1}{2}\widetilde{M}_\theta\left(y,\mu\right)$, holds. In the following two results, it is shown that the kernel estimator, (3.2.0.22), converges uniformly in probability to $Q\left(\mathbf{x}_i,\sigma\right)$ with respect to σ.

Lemma 3.3 *Suppose conditions (B1),(B4),(C2) for $g\left(\mathbf{x}_i\right) > 0$ on a compact subset $G_n \in \mathbf{R}^d$ hold and that the uniform equicontinuity condition $\forall\epsilon > 0, \exists\delta > 0$ such that*

$$\sup_{\mathbf{x}_i \in G_n} \sup_{\mathbf{x}:|\mathbf{x}-\mathbf{x}_i|\leq\delta} \left|Q\left(\mathbf{x},\sigma\right) - Q\left(\mathbf{x}_i,\sigma\right)\right| \leq \epsilon, \tag{3.2.1.7}$$

is satisfied for all fixed σ. If condition (D1) for $h_{i,j} = h$ hold, then for all $\mathbf{x}_i \in G_n$ and sufficiently large n, we have

$$\sup_{\mathbf{x}_i \in G_n} \left|E\left[\widehat{Q}_n\left(\mathbf{x}_i,\sigma\right)\right] - Q\left(\mathbf{x}_i,\sigma\right)\right| \rightarrow 0 \tag{3.2.1.8}$$

Proof of lemma 3.3

The equicontinuity of the couples $\left(Y_t,\mathbf{X}_t\right)$ implies that

$$\begin{aligned} E\left[\widehat{Q}_n\left(\mathbf{x}_i,\sigma\right)\right] - Q\left(\mathbf{x}_i,\sigma\right) &= E\Big\{\mathbf{K_h}\left(\mathbf{x}_i-\mathbf{X}_t\right)\left(Q\left(\mathbf{X}_t,\sigma\right) - Q\left(\mathbf{x}_i,\sigma\right)\right)\Big\} \\ &= \int \mathbf{K_h}\left(\mathbf{x}_i-\mathbf{u}\right)\left(Q\left(\mathbf{u},\sigma\right)g\left(u\right) - Q\left(\mathbf{x}_i,\sigma\right)\right)d\mathbf{u} \end{aligned}$$

using Bochner's theorem in Parzen [94] as applied in Collomb and Haerdle [29] completes the proof. □

Let Θ_n be a compact subset of an open subset Θ of $\mathbf{R}$, then since both $E\left[\widehat{Q}_n\left(\mathbf{x}_i,\sigma\right)\right]$ and $Q\left(\mathbf{x}_i,\sigma\right)$ are Lipschitz continuous functions with respect to σ, it follows from lemma 3.3, see also lemma 2.7, that for each $\mathbf{x}_i \in G_n$ and fixed θ

$$\sup_{\mathbf{x}_i \in G_n} \sup_{\sigma\in\mathbf{\Theta}_n} \left|E\left[\widehat{Q}_n\left(\mathbf{x}_i,\sigma\right)\right] - Q\left(\mathbf{x}_i,\sigma\right)\right| = O\left(h^2\right) \tag{3.2.1.9}$$

This is clearly seen by taking Taylor expansion of (3.2.1.9) up to the second order and using the symmetric condition (B5), we get $\left|E\left[\widehat{Q}_n\left(\mathbf{x}_i,\sigma\right)\right] - Q\left(\mathbf{x}_i,\sigma\right)\right| \leq \frac{h^2}{2}a\int \mathbf{u}^2\mathbf{K}\left(u\right)d\mathbf{u}$ uniformly for $\mathbf{x}_i$ in G_n and σ in Θ_n, where

$$\sup_{\sigma\in\Theta_n} \sup_{\mathbf{x}_i\in G_n} \sup_{\mathbf{x}\in\mathbf{R}} \left|\nabla^2\Big\{\mathbf{K_h}\left(\mathbf{x}_i-\mathbf{X}_t\right)Q\left(\mathbf{X}_t,\sigma\right)\Big\}\right| \leq a < \infty$$

and ∇^2 is taken with respect to $\mathbf{x}_i$.

The following lemma establishes uniform convergence in probability of $\widehat{Q}_n\Big(\mathbf{x}_i,\sigma\Big)$ to $Q\Big(\mathbf{x}_i,\sigma\Big)$.

Lemma 3.4 *Suppose conditions in lemma 3.3 hold, then for* $\mathbf{x}_i \in G_n$, $\widehat{Q}_n\Big(\mathbf{x}_i,\sigma\Big)$ *converges to* $Q\Big(\mathbf{x}_i,\sigma\Big)$ *in probability uniformly on any compact set* Θ_n *of* $\mathbf{R}$ *containing* $\sigma_\theta\Big(\mathbf{x}_i\Big)$*, i.e for all* $\epsilon > 0$*,*

$$\lim_{n\to\infty} P\Big(\sup_{\mathbf{x}_i\in G_n}\sup_{\sigma\in\Theta_n}\Big|\widehat{Q}_n\Big(\mathbf{x}_i,\sigma\Big)-Q\Big(\mathbf{x}_i,\sigma\Big)\Big|\geq\epsilon\Big)=0 \tag{3.2.1.10}$$

Proof of lemma 3.4

We use the triangle inequality and lemma 3.3:

$$\begin{aligned}
\widehat{Q}_n\Big(\mathbf{x}_i,\sigma\Big) \;&-\; Q_n\Big(\mathbf{x}_i,\sigma\Big)+Q_n\Big(\mathbf{x}_i,\sigma\Big)-Q\Big(\mathbf{x}_i,\sigma\Big)\\
&\leq \Big|\widehat{Q}_n\Big(\mathbf{x}_i,\sigma\Big)-Q_n\Big(\mathbf{x}_i,\sigma\Big)\Big|+\Big|Q_n\Big(\mathbf{x}_i,\sigma\Big)-Q\Big(\mathbf{x}_i,\sigma\Big)\Big|\\
&\leq \sup_{\mathbf{x}_i\in G_n}\sup_{\sigma\in\Theta_n}\Big|\widehat{Q}_n\Big(\mathbf{x}_i,\sigma\Big)-Q_n\Big(\mathbf{x}_i,\sigma\Big)\Big|+\sup_{\mathbf{x}_i\in G_n}\sup_{\sigma\in\Theta_n}\Big|Q_n\Big(\mathbf{x}_i,\sigma\Big)-Q\Big(\mathbf{x}_i,\sigma\Big)\Big|\\
&= \sup_{\mathbf{x}_i\in G_n}\sup_{\sigma\in\Theta_n}\widehat{\overline{Q}}_n\Big(\mathbf{x}_i,\sigma\Big)+\sup_{\mathbf{x}_i\in G_n}\sup_{\sigma\in\Theta_n}\overline{Q}_n\Big(\mathbf{x}_i,\sigma\Big)
\end{aligned}$$

where then by (3.2.1.9) and for all $\frac{\epsilon}{2}>0$, $\lim_{n\to\infty} P\Big(\sup_{\mathbf{x}_i\in G_n}\sup_{\sigma\in\Theta_n}\Big|\overline{Q}_n\Big(\mathbf{x}_i,\sigma\Big)\Big|\geq\frac{\epsilon}{2}\Big)=0$

Next, we show that

$$P\Big(\sup_{\mathbf{x}_i\in G_n}\sup_{\sigma\in\Theta_n}\Big|\widehat{\overline{Q}}_n\Big(\mathbf{x}_i,\sigma\Big)\Big|\geq\frac{\epsilon}{2}\Big)\to 0,\quad n\to\infty \tag{3.2.1.11}$$

The left hand side of (3.2.1.11) is equivalent to

$$\begin{aligned}
P\Big(\Big|\widehat{\overline{Q}}_n\Big(\mathbf{x}_i,\sigma\Big)\Big| \;&>\; \frac{\epsilon}{2},\quad \text{for some}\quad \sigma\Big)\\
&= P\Big(\bigcup_{\sup_{\mathbf{x}_i\in G_n}\big|\sigma\in\Theta_n}\Big\{\Big|\widehat{\overline{Q}}_n\Big(\mathbf{x}_i,\sigma\Big)\Big|>\frac{\epsilon}{2}\Big\}\Big)\\
&\leq \sum_{\sigma\in\Theta_n} P\Big(\widehat{\overline{Q}}_n\Big(\mathbf{x}_i,\sigma\Big)>\frac{\epsilon}{2}\Big)
\end{aligned}$$

Since Θ is not countable, we use the following trivial form:

$$\sup_{\mathbf{x}_i \in G_n} \sup_{\sigma \in \Theta_n} \left| \widehat{\overline{Q}}_n \left(\mathbf{x}_i, \sigma \right) \right|$$
$$\leq \max_{k=1,2,\ldots,m(\delta)} \sup_{\mathbf{x}_i} \left| \widehat{\overline{Q}}_n \left(x_i, \sigma_k \right) \right| + \sup_{k=1,2,\ldots,m(\delta)} \sup_{|\sigma - \sigma_k| \leq \delta} \sup_{\mathbf{x}_i} \left| \widehat{\overline{Q}}_n \left(\mathbf{x}_i, \sigma_k \right) - \widehat{\overline{Q}}_n \left(\mathbf{x}_i, \sigma \right) \right|$$

where for each $\sigma \in \Theta_n$, σ_k denotes the nearest neighbor for σ such that $\left| \sigma - \sigma_k \right| \leq \delta$ and $m(\delta) = card\left\{ \sigma : \left| \sigma - \sigma_k \right| \leq \delta, \quad k = 1, \ldots, n \right\}$. We have then

$$P\Big(\sup_{\mathbf{x}_i \in G_n} \sup_{\sigma \in \Theta_n} \left| \widehat{\overline{Q}}_n \left(\mathbf{x}_i, \sigma \right) \right| > \frac{\epsilon}{2} \Big) \leq P\Big(\sup_{k=1,2,\ldots,m(\delta)} \sup_{\mathbf{x}_i \in G_n} \left| \widehat{\overline{Q}}_n \left(\mathbf{x}_i, \sigma_k \right) \right| > \frac{\epsilon}{2} \Big)$$
$$+ \; P\Big(\sup_{\mathbf{x}_i \in G_n} \sup_{|\sigma - \sigma_k| \leq \delta, k=1,2,\ldots,m(\delta)} \left| \widehat{\overline{Q}}_n \left(\mathbf{x}_i, \sigma_k \right) - \widehat{\overline{Q}}_n \left(\mathbf{x}_i, \sigma \right) \right| > \frac{\epsilon}{2} \Big) \tag{3.2.1.12}$$

By lemma 3.1 and condition (B2),

$\sup_{\mathbf{x}_i \in G_n} \sup_{\sigma \in \Theta_n} \left| \widehat{\overline{Q}}_n \left(\mathbf{x}_i, \sigma' \right) - \widehat{\overline{Q}}_n \left(\mathbf{x}_i, \sigma \right) \right| \leq 2\overline{\mathbf{K}} \left| \sigma' - \sigma \right| \leq 2\overline{\mathbf{K}}\delta,$

with probability say, $\pi_n \to 1$, as $n \to \infty$ uniformly in $(\sigma', \sigma) \in \Theta_n$. Therefore the second part on the right of (3.2.1.12) becomes

$$P\Big(\sup_{\mathbf{x}_i \in G_n} \sup_{|\sigma - \sigma_k| \leq \delta, k=1,2,\ldots,m(\delta)} \left| \widehat{\overline{Q}}_n \left(\mathbf{x}_i, \sigma_k \right) - \widehat{\overline{Q}}_n \left(\mathbf{x}_i, \sigma \right) \right| > \frac{\epsilon}{2} \Big) \leq 1 - \pi_n$$

which goes to zero as $n \to \infty$. The first part is the same as

$$\begin{aligned} P\Big(\sup_{\mathbf{x}_i \in G_n} \left| \widehat{\overline{Q}}_n \left(\mathbf{x}_i, \sigma_k \right) \right| \; &> \; \frac{\epsilon}{2}, \quad \text{for some} \quad k \Big) \\ &\leq \; \sum_{k=1}^{m(\delta)} P\Big(\sup_{\mathbf{x}_i \in G_n} \left| \widehat{\overline{Q}}_n \left(\mathbf{x}_i, \sigma_k \right) > \frac{\epsilon}{2} \Big) \\ &\leq \; m(\delta) \max_k P\Big(\sup_{\mathbf{x}_i \in G_n} \left| \widehat{\overline{Q}}_n \left(\mathbf{x}_i, \sigma_k \right) \right| > \frac{\epsilon}{2} \Big) \end{aligned}$$

Following the argument of lemma 3 in Collomb and Haerdle [29], their exist a constant a_2 such that $P\Big(\left| \widehat{\overline{Q}}_n \left(\mathbf{x}_i, \sigma_k \right) \right| > \frac{\epsilon}{2} \Big) \leq a_2 m(\delta) n^{-3}, \forall n \in \mathbf{N}$. By choosing $m(\delta) = n$, we obtain $P\Big(\left| \widehat{\overline{Q}}_n \left(\mathbf{x}_i, \sigma_k \right) \right| > \frac{\epsilon}{2} \Big) \to 0$. This completes the proof.

□

The uniform convergence of $\widehat{Q}_n\left(\mathbf{x}_i, \sigma\right)$ with an appropriate rate can be established using similar arguments as in lemmas 2.5,2.6 and 2.7.

Theorem 3.1 *Suppose conditions (B1)-(B6), (C1)-(C2), (D1) and (E1) hold for* $\widetilde{Y}_t$ *and* $\sigma_{t,\theta}$ *of model (3.2.0.19). Then* $\widehat{\sigma}_\theta\left(\mathbf{x}_i\right)$ *is weakly consistent, i.e* $\widehat{\sigma}_\theta\left(\mathbf{x}_i\right) \to^p \sigma_\theta\left(\mathbf{x}_i\right)$ *for each* $\mathbf{x}_i \in G_n$.

Proof of theorem 3.1:

The theorem follows if we can show the following three properties, see corollary 2.6 of White and Wooldridge [113] in the case of dependent observations and lemma (A) of Newey and Powell [93] in the case of independent obsrvations:

(1) $\widehat{Q}_n\left(\mathbf{x}_i, \sigma\right)$ converges to $Q\left(\mathbf{x}_i, \sigma\right)$ in probability uniformly on any compact set $\Theta_n \in \Theta$ containing $\sigma_\theta\left(\mathbf{x}_i\right)$,

(2) $Q\left(\mathbf{x}_i, \sigma\right)$ has a unique minimum on Θ_n at $\sigma_\theta\left(\mathbf{x}_i\right)$,

(3) $Q\left(\mathbf{x}_i, \sigma\right)$ is continuous and convex in σ.

Now (1) has been shown in lemma 3.4, whereas (3) follows immediately from lemma 3.1, the definition of $Q\left(\mathbf{x}, \sigma\right)$ and the convexity of $M_\theta\left(y, \sigma\right)$ in σ. It remains to show (2). Obseve that

$$
\begin{aligned}
& E\left[\widehat{Q}_n\left(\mathbf{x}_i, \sigma\right)\right] - E\left[\widehat{Q}_n\left(\mathbf{x}_i, \sigma_\theta\left(\mathbf{x}_i\right)\right)\right] \\
= & \frac{1}{n}\sum_{t=1}^{n} E\left[\mathbf{K_h}\left(\mathbf{x}_i - \mathbf{X}_t\right) E\left[\left(M_\theta\left(Y_t, \sigma\right) - M\left(Y_t, \sigma_\theta\left(\mathbf{x}_i\right)\right)\right) \middle| \mathbf{X}_t\right]\right] \\
= & \frac{1}{n}\sum_{t=1}^{n} E\left[\mathbf{K_h}\left(\mathbf{x}_i - \mathbf{X}_t\right)\left(Q\left(\mathbf{X}_t, \sigma\right) - Q\left(\mathbf{X}_t, \sigma_\theta\left(\mathbf{x}_i\right)\right)\right)\right]
\end{aligned}
$$

From lemma 3.2 part (2), $Q\left(\mathbf{X}_t, \sigma\right) - Q\left(\mathbf{X}_t, \sigma_\theta\left(\mathbf{x}_i\right)\right) \geq c\left(\mathbf{X}_t\right)\frac{\delta^2}{2}$, where $c\left(\mathbf{X}_t\right) > 0$ is the lower bound of $f_{\mathbf{X}_t}\left(y\right)$ on $\left[\sigma_\theta\left(\mathbf{X}_t\right) - \delta, \sigma_\theta\left(\mathbf{X}_t\right) + \delta\right]$. Therefore, we have, using stationarity for the second line,

$$
\begin{aligned}
E\Big[\widehat{Q}_n\Big(\mathbf{x}_i,\sigma\Big)\Big]-E\Big[\widehat{Q}_n\Big(\mathbf{x}_i,\sigma_\theta\Big(\mathbf{x}_i\Big)\Big)\Big] &\geq \frac{1}{n}\sum_{t=1}^{n}E\Big[\mathbf{K_h}\Big(\mathbf{x}_i-\mathbf{X}_t\Big)c\Big(\mathbf{X}_t\Big)\Big]\frac{\delta^2}{2} \\
&= E\Big[\mathbf{K_h}\Big(\mathbf{x}_i-\mathbf{X}_t\Big)c\Big(\mathbf{X}_t\Big)\Big]\frac{\delta^2}{2} \\
&\geq c\frac{\delta^2}{2} \qquad (3.2.1.13)
\end{aligned}
$$

for some $c>0$ and all sufficiently large n as K_h has support $[-h,h]$ and, therefore, the expectation runs only over those $\mathbf{X}_1$ with $\Big|\mathbf{x}_i-\mathbf{X}_1\Big|\leq h$, i.e for $\mathbf{X}_1$ arbitrary close to $\mathbf{x}_i$ for $n\to\infty$. Now, using part (3) of lemma 3.2, the assertion follows. □

The fact that $\widehat{Q}_n\Big(\mathbf{x}_i,\sigma\Big)$ is stricly convex, continuous as a function of σ and bounded for any two $\sigma^{'}$s, implies that the minimizer, $\widehat{\sigma}_\theta\Big(\mathbf{x}_i\Big)$, of $\widehat{Q}_n\Big(\mathbf{x}_i,\sigma\Big)$ exists uniquely and is a solution of

$$\frac{d}{d\sigma}\widehat{Q}_n\Big(\mathbf{x}_i,\sigma\Big)=0 \qquad (3.2.1.14)$$

Re-arranging equation (3.2.1.14), we immediately note that the bias and asymptotic variance for $\widehat{\sigma}_\theta\Big(\mathbf{x}_i\Big)$, are precisely of the same form as those presented in chapter 2, with the appropriate bandwidth. The asymptotic normality of the estimator, $\widehat{\sigma}_\theta\Big(\mathbf{x}_i\Big)$, can therefore be shown by proceeding in the same lines as in chapter 2. We therefore state the following theorem without proof.

Theorem 3.2 *Suppose conditions (B1)-(B5),(C1)-(C7), (D1)-(D3), (E1) and (1.4.1) hold for $\widetilde{Y}_t$. Then we have*

$$\Big(nh^d\Big)^{\frac{1}{2}}\Big(\widehat{\sigma}_\theta\Big(\mathbf{x}_i\Big)-\sigma_\theta\Big(\mathbf{x}_i\Big)-B\Big(\sigma_\theta\Big(\mathbf{x}_i\Big)\Big)\Big)\sim N\left(0,\frac{\mathbf{V}^2\Big(\sigma_\theta(\mathbf{x}_i)\Big)}{f^2_{\mathbf{x}_i}\Big(\sigma_\theta\Big(\mathbf{x}_i\Big)\Big)}\right) \qquad (3.2.1.15)$$

where the bias $B\Big(\sigma_\theta\Big(\mathbf{x}_i\Big)\Big)$ and $\mathbf{V}^2\Big(\sigma_\theta\Big(\mathbf{x}_i\Big)\Big)$ are defined in similar forms, with appropriate bandwidth, as in theorem 2.3 and lemma 2.1 respectively.

3.3 Local polynomial estimator of the scale function

Let $\lambda=\Big(\lambda_1,\ldots,\lambda_d\Big)$ be a d-dimensional vector of nonnegative integers and $[\lambda]=\sum_{j=1}^{d}\lambda_j$. Let G be some fixed open neighborhood of $\mathbf{x}_i\in\mathbf{R}^d$. For a fixed nonnegative integer p and

real number c_1 and c_2 such that $c_1 > 0$ and $0 < c_2 \leq 1$, let $H\Big(c_1, p, c_2\Big)$ be a collection of all real valued functions $\sigma_\theta\Big(\mathbf{X}_t\Big)$ on G such that
(i) $\nabla^\lambda \sigma_\theta\Big(\mathbf{X}_t\Big)$ exists and is continuous in $\mathbf{X}_t$ for all $\mathbf{X}_t \in G$ and $[\lambda] \leq p$.
(ii) For any $\mathbf{X}_t, \mathbf{x}_i \in G$ and $[\lambda] = p$, we have $\nabla^\lambda \sigma_\theta\Big(\mathbf{X}_t\Big) - \nabla^\lambda \sigma_\theta\Big(\mathbf{x}_i\Big) \leq c_1 \Big|\Big|\mathbf{X}_t - \mathbf{x}_i\Big|\Big|^{c_2}$ for $c_2 \geq \frac{1}{2}$, where $\Big|\Big|.\Big|\Big|$ is an Euclidean norm.
Then the functions $H\Big(c_1, p, c_2\Big)$, with the order of smoothness of $\Big(p + c_2\Big)$ at $\mathbf{x}_i = \Big(\mathbf{x}_{i,1}, \ldots, \mathbf{x}_{i,d}\Big)$, are continuously differentiable up to order p on G and their p-th derivative are uniformly Hoelder continuous at $\mathbf{x}_i$ with exponent at least $\frac{1}{2}$. We will assume the conditional θ-quantile of $\widetilde{Y}_t$ on $\mathbf{X}_t$ is an element of $H\Big(c_1, p, c_2\Big)$ for some fixed c_1, p and c_2. Let us now consider a sequence of positive real numbers $h = cn^{-\frac{1}{2p+d}}$ and let G_n denote a cube $[-h, h]^d$ in $\mathbf{R}^d$. Here h is the bandwidth which depends on n such that as $n \to \infty$, the cube shrinks and becomes completely contained in open subset G. Henceforth, it is assumed that n is such that $G_n \subseteq G$. Let Λ be the set of all d-dimensional vectors λ with nonnegative integer components such that $[\lambda] < p$ and denote $[\Lambda]$ to be the size of of the set Λ. Given $\mathbf{X}_t \in \mathbf{R}^d$ in the neighborhood of $\mathbf{x}_i$, the Taylor expansion of $\sigma_\theta\Big(\mathbf{X}_t\Big)$ up to $(p-1)$-th order can be written as

$$\begin{aligned}
\sigma_\theta\Big(\mathbf{X}_t\Big) &= \sum_{[\lambda]=0}^{p-1} \sum_{\lambda_1+\ldots..+\lambda_d=[\lambda]} \sigma_{\theta,\lambda_1,\ldots,\lambda_d} \prod_{j=1}^{d} h^{-\lambda_j} \Big(X_{t,j} - x_{i,j}\Big)^{\lambda_j} - r\Big(\mathbf{X}_t, \mathbf{x}_i\Big) \\
&= \sum_{[\lambda]=0}^{p-1} \sum_{\lambda_1+\ldots..+\lambda_d=[\lambda]} \sigma_{\theta,\lambda_1,\ldots,\lambda_d} h^{-[\lambda]} \prod_{j=1}^{d} \Big(X_{t,j} - x_{i,j}\Big)^{\lambda_j} - r\Big(\mathbf{X}_t, \mathbf{x}_i\Big) \\
&= \sum_{\lambda \in \Lambda} \mathbf{D}_{\theta,\lambda}\Big(\mathbf{x}_i\Big) h^{-[\lambda]} \Big(\mathbf{X}_t - \mathbf{x}_i\Big)^{\lambda} - r\Big(\mathbf{X}_t, \mathbf{x}_i\Big) \\
&= P_h\Big(\mathbf{D}_{\theta,\lambda}\Big(\mathbf{x}_i\Big), \mathbf{X}_t - \mathbf{x}_i\Big) - r\Big(\mathbf{X}_t, \mathbf{x}_i\Big) \\
&\approx P_h\Big(\mathbf{D}_{\theta,\lambda}\Big(\mathbf{x}_i\Big), \mathbf{X}_t - \mathbf{x}_i\Big) \qquad (3.3.0.16)
\end{aligned}$$

where $\mathbf{D}_{\theta,\lambda}\Big(\mathbf{x}_i\Big) = \Big\{\mathbf{D}_{\theta,\lambda_1,\ldots,\lambda_d}\Big(\mathbf{x}_i\Big) : \lambda_1 + \ldots.. + \lambda_d = [\lambda] \quad \text{and} \quad [\lambda] = 0, 1, \ldots, p-1\Big\}$ is a $[\Lambda]$-dimensional vector of coefficients, $\Big(\mathbf{X}_t - \mathbf{x}_i\Big)^\lambda = \prod_{j=1}^{d}\Big(\mathbf{X}_{t,j} - x_{i,j}\Big)^{\lambda_j}$ with the usual convention $0^0 = 1$. Observe that when $\mathbf{X}_t = \mathbf{x}_i$, then $P_h\Big(\mathbf{D}_{\theta,\lambda}\Big(\mathbf{x}_i\Big), 0\Big) = \mathbf{D}_{\theta,0,\ldots,0}\Big(\mathbf{x}_i\Big)$ and therefore we estimate $\sigma_\theta\Big(\mathbf{x}_i\Big)$ by the local estimator of $\mathbf{D}_{\theta,\lambda}\Big(\mathbf{x}_i\Big)$.
Let $\widehat{\mathbf{D}}_\theta\Big(\mathbf{x}_i\Big)$ be the minimizer of

$$\widehat{Q}_n\Big(\mathbf{x}_i, \mathbf{D}\Big) = \sum_{t=1}^{n} \mathbf{K_h}\Big(\mathbf{X}_t - \mathbf{x}_i\Big)\Big[M_\theta\Big(\widetilde{Y}_t, P_h\Big(\mathbf{D}, \mathbf{X}_t - \mathbf{x}_i\Big)\Big) - M_\theta\Big(\widetilde{Y}_t, 0\Big)\Big]$$

where $M_\theta\Big(y, \mu\Big) = \frac{1}{2}\widetilde{M}_\theta\Big(y, \mu\Big)$. Because we are using only the values of $\mathbf{X}_t$ that fall in G_n, then for a fixed value say $n^{'}$ of the number of $\mathbf{X}_t, t = 1, \ldots, n$ that fall in G_n, the minimization is a problem of minimizing continuous function over bounded and closed subset of linear subspace $\mathbf{R}^{n^{'}}$. Futhermore from lemma 3.2, the minimization problem has a unique solution. The estimator for $\sigma_\theta\Big(\mathbf{x}_i\Big)$ is obtained as the first element of the estimate of $\mathbf{D}_{\theta,\lambda}\Big(\mathbf{x}_i\Big)$. The derivatives for $\widehat{\sigma}_\theta\Big(\mathbf{x}_i\Big)$ are obtained by multiplying the corresponding elements of $\widehat{\mathbf{D}}_{\theta,\lambda}\Big(\mathbf{x}_i\Big)$ by $C_\lambda h^{-[\lambda]}$, where C_λ depends on λ.

The results in Honda [65] provides Bahadur-type expansions of the estimator of the form $\widehat{\mathbf{D}}_{\theta,\lambda}\Big(\mathbf{x}_i\Big)$ as well as the derivatives, which we will readily adopt for our purpose. Arrange $h^{-[\lambda]}\Big(\mathbf{X}_t - \mathbf{x}_i\Big)^{\lambda}$ and $\mathbf{D}_{\theta,\lambda}\Big(\mathbf{x}_i\Big)$, where both quantities depend on h, in ascending order with respect to λ and denote them by $\overline{\mathbf{X}_t - \mathbf{x}_i} \in \mathbf{R}^{[\Lambda]}$ and $\mathbf{D}_\theta\Big(\mathbf{x}_i\Big) \in \mathbf{R}^{[\Lambda]}$ respectively.

3.3.1 Consistency and asymptotic distribution

Assume the following conditions

Conditions 3.3.1

(L1) For any $j_1 < \ldots.. < j_{[\Lambda]}$, $\overline{\mathbf{X}_{j_1} - \mathbf{x}_i}, \ldots, \overline{\mathbf{X}_{j_{[\Lambda]}} - \mathbf{x}_i}$ *is linearly independent for any* $\mathbf{x}_i$ *with probability 1.*

(L2) $\mathbf{K}\Big(u\Big)$ *is bounded nonnegative kernel function with the compact support* $\Big\{\Big|u\Big| \leq 1\Big\} \subset \mathbf{R}^d$ *and Lipschitz continuous. Assume the bandwidth to be* $h = cn^{-\frac{1}{2p+d}}, \quad c > 0.$

(L3) $\mathbf{X}_t$ *has a density,* $g\Big(\mathbf{x}_i\Big)$ *which is bounded for* $\mathbf{x}_i \in G_n$

(L4) $c_3 h^d \mathbf{I}_{[\Lambda]} < E\Big\{\mathbf{K}_h\Big(\mathbf{X}_1 - \mathbf{x}_i\Big)\Big(\overline{\mathbf{X}_1 - \mathbf{x}_i}\Big)\Big(\overline{\mathbf{X}_1 - \mathbf{x}_i}\Big)^T g\Big(\mathbf{X}_1, \mathbf{x}_i\Big)\Big\} < c_4 h^d \mathbf{I}_{[\Lambda]}$
$E\Big\{\mathbf{K}_h\Big(\mathbf{X}_t - \mathbf{x}_i\Big)\mathbf{K}_h\Big(\mathbf{X}_k - \mathbf{x}_i\Big)\Big\} < c_5 h^{d+1}$ *for* $t \neq k$, *where* $\mathbf{I}_{[\Lambda]}$ *is an identity matrix of size* $[\Lambda]$.

(L5) The conditional distribution of $\widetilde{Y}_t$ *given* $\mathbf{F}_{t-1}$ *has no atom with probability 1.*

The following theorems give the uniform Bahadur representation and uniform convergence of the estimator for $\mathbf{D}_\theta\Big(\mathbf{x}_i\Big)$.

Theorem 3.3 *Let* $\left\{\left(Y_t, \mathbf{X}_t\right)\right\}$ *satisfy the assumptions of model (1.4.0.2). Suppose assumptions (L1)-(L5) hold and* $\alpha\left(s\right) \leq cs^{-r}$ *and for some* $\frac{p}{2} < \eta < p, r > \frac{d+p}{2\eta-p} \vee (\frac{(3dp+4p+3d-d\eta)}{\eta} - 1) \vee (\frac{3dp+2p+3d}{p})$ *. Then*

$$\begin{aligned}\widehat{\mathbf{D}}_\theta\left(\mathbf{x}_i\right) \quad - \quad & \mathbf{D}_\theta\left(\mathbf{x}_i\right) = \left(2\sum_t^n \mathbf{K}_h\left(\mathbf{X}_t - \mathbf{x}_i\right)\left(\overline{\mathbf{X}_t - \mathbf{x}_i}\right)\left(\overline{\mathbf{X}_t - \mathbf{x}_i}\right)^T g\left(\mathbf{X}_t, \mathbf{x}_i\right)\right\}\Big)^{-1} \\ \times \quad & \sum_{t=1}^n \mathbf{K}_h\left(\mathbf{X}_t - \mathbf{x}_i\right)\overline{\mathbf{X}_t - \mathbf{x}_i}\left\{\left(sgn(Z_{t,3}) + 2\theta - 1\right) + 2\left(\theta - G\left(\mathbf{X}_t, r\left(\mathbf{X}_t, \mathbf{x}_i\right)\right)\right)\right\} \\ + \quad & O\left(h^{2p-\eta}\right), \qquad \textit{uniformly on} \quad G_n \quad \textit{almost sure (a.s).} \qquad (3.3.1.1)\end{aligned}$$

where $sgn(u) = 1(u > 0), \quad 0(u = 0),$ *and* $-1(u < 0)$ *and* $G\left(\mathbf{x}, z\right) = P\left(Z_{t,3} \leq z \middle| \mathbf{X}_t = \mathbf{x}\right) = F_{\mathbf{x}}\left(\sigma_\theta\left(\mathbf{x}_i\right) + z\right)$

Theorem 3.4 *Suppose that assumptions (L1)-(L4) hold and that* $\alpha\left(s\right) \leq c_1 s^{-r}$*, for* $r > \frac{3dp+2p+3d}{p}$*. Then*

$$\left|\widehat{\mathbf{D}}_\theta\left(\mathbf{x}_i\right) - \mathbf{D}_\theta\left(\mathbf{x}_i\right)\right| = O\left(h^p\left(\log n\right)^{\frac{1}{2}}\right), \qquad \textit{uniformly on} \quad G_n \qquad \textit{a.s}$$

The proofs for theorems 3.3 and 3.4 are found in Honda [65], where also the asymptotic distribution for $\widehat{\mathbf{D}}_\theta\left(\mathbf{x}_i\right) - \mathbf{D}_\theta\left(\mathbf{x}_i\right)$ is given. The consistency of $\widehat{\mathbf{D}}_\theta\left(\mathbf{x}_i\right)$ at a fixed point shares the same representation with the uniform convergence. That is, at a fixed point say $\mathbf{x}_i \in G_n$, we have $\left|\widehat{\mathbf{D}}_\theta\left(\mathbf{x}_i\right) - \mathbf{D}_\theta\left(\mathbf{x}_i\right)\right| = O\left(h^d\left(\log n\right)^{\frac{1}{2}}\right)$ a.s.
The estimation of the QARCH in the presence of unknown QAR, $\mu_\theta\left(\mathbf{X}_t\right)$, using local polynomial proceeds in a similar manner as in chapter 2. So long as the estimator for the QAR is uniformly consistent, then under similar arguments as in chapter (2), the estimator for the scale function would be consistent. Simultaneous estimation of the QAR and scale function in QAR-QARCH processes could be investigated using methods similar to simultaneous M-estimation in Huber [67]. For instance, taking $\left(\mu, \sigma\right) \in \mathbf{R} \times \mathbf{R}_+$ one could base the theoretical objective function on $E\left[\left(M_\theta\left(M_\theta\left(Y_t, \mu_{\mathbf{x}}\right), \sigma\right) - M_\theta\left(Y_t, 0\right)\right)\middle|\mathbf{X}_t = \mathbf{x}\right]$,where $\mu_{\mathbf{x}}$ depends on $\mathbf{x}$ through some weight functions, and maximize its kernel estimate with respect to μ and σ. Koenker and Zhao [76] discusses analogous parametric version applied to AR-ARCH with absolute errors.
In the remaining part of this chapter, we will adopt local polynomials of degree one for various estimations.

3.4 Numerical comparison

In this section we compare numerically estimators (3.1.0.16) and (3.1.0.18). We use the bandwidth of the form $h_\theta = h_{mean}\Big\{\theta\Big(1-\theta\Big)/\phi\Big(\Phi^{-1}(\theta)\Big)^2\Big\}^{\frac{1}{5}}$ and proposed in Yu and Jones [114], where ϕ and Φ are standard normal density and distribution functions respectively. We select h_{mean} using the leave-block out cross-validation based on local constant fit of the conditional mean. It should be noted that this bandwidth only serves as a thumb of rule and better procedures may need to be studied in the context of local polynomial in time series set up. We simulated the data from a heteroscedastic ARCH process

$$Y_t = \Big(0.075 + 0.3Y_{t-1}^2 + 0.62Y_{t-2}^2\Big)^{\frac{1}{2}} e_t, \quad t = 1, 2, \ldots, \tag{3.4.0.2}$$

with iid Student's t-distributed error e_t and took $t = 3, \ldots 1002$. The true volatility is shown in figure (5). Because our interest was to obtain the conditional scale function using model (1.4.0.2), we adjusted Y_t in (3.4.0.2) of its conditional 0.75-quantile and estimated the scale at $\theta = 0.75$ from the assumed nonparametric model, $Y_t = \sigma_{0.75}\Big(Y_{t-1}, Y_{t-2}\Big)Z_t$, where Z_t is a zero 0.75-quantile. Figures (6) and (7) presents the surfaces of the estimated regression using (3.1.0.16) and (3.1.0.18) respectively, with the latter transformed accordingly. In both cases 400 data points were used for the estimation. Visually, they do not seem to be quite different from each other.

In order to provide a quantitative assessment of the accuracy of these estimators, we generated 500 replicates of size 1000 from the process (3.4.0.2). Then we calculated the mean average squared error (MASE) for a few values of θ using the following formula,

$$MASE_\theta\Big(\widehat{\sigma}_\theta\Big(\mathbf{x}_i\Big)\Big) = \frac{1}{500}\sum_{j=1}^{500}\Big[\frac{1}{1000}\sum_{i=1}^{1000}\Big(\widehat{\sigma}_\theta^{(j)}\Big(\mathbf{x}_i\Big) - \sigma_\theta^{(j)}\Big(\mathbf{x}_i\Big)\Big)^2\Big],$$

The $MASE_\theta\Big(\widehat{\sigma}_\theta\Big(\mathbf{x}_i\Big)\Big)$ are depicted in table (1), where in each distributional error, the second rows represent results obtained by using the estimator (3.1.0.16). As expected the MASE for both estimators increase as one moves away from $\theta = 0.5$, with the difference between them growing faster outside the range $(0.4, 0.6)$. In particular, for heavy tailed and asymmetric distributions, the $MASE$ at $\theta = 0.8$ is more than 12% smaller for the estimator based on (3.1.0.18). This could be attributed to the dependence effect of the

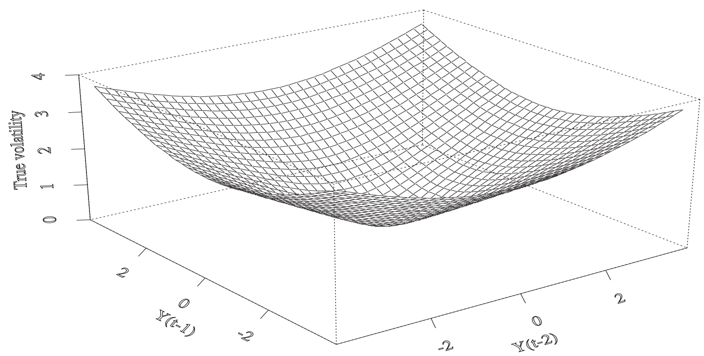

Figure 5: True volatility

scale function $\sigma_{t,\theta}$ in the errors.

3.5 Estimation of the true volatility

In this section the original variable, Y_t instead of the log-transformed, will be used. Recall that using model (1.4.0.2) with reference to (1.1.1.1), we have

$$\sigma_{t,\theta} = \sigma_t M_\theta^e \tag{3.5.0.3}$$

as the volatility (conditional standard deviation) up to a multiplicative constant. We consider the constant as a nuisance parameter and use the standardization method similar to the one in Huber [67] to remove it. Let $b > 0$ be a rescaling constant for the conditional scale function $\sigma_{t,\theta}$. Then b will be such that when multiplied across (3.5.0.3) leaves us with the volatility, i.e $b = \left(M_\theta^e\right)^{-1}$. Lets us consider the conditional median absolute deviation (CMAD) from a symmetric distribution F. This can be expressed as $\sigma_{t,0.5} = \mu_{t,0.75} - \mu_{t,0.5} = \sigma_t\left(q_{0.75}^e - q_{0.5}^e\right)$, with $q_{0.5}^e = 0$. To standardize it, we will need to multiply by a rescaling constant $b = \left(q_{0.75}^e\right)^{-1}$. Observe that if F is a normal distribution with

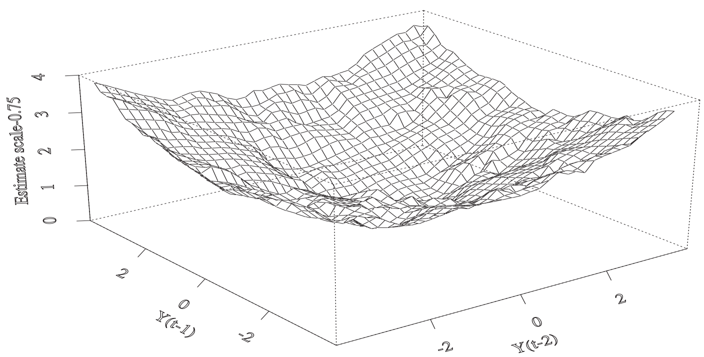

Figure 6: Surface of scale estimate based on (3.1.0.16)

conditional mean zero, then b reduces to $b = \left[\Phi_e^{-1}\Big(0.75\Big)\right]^{-1} \approx 1.482$. For a student t-distributed random variable e_t with v degrees of freedom, we have

$$b = \left[t_v^{-1}\Big(0.75\Big) - t_v^{-1}\Big(0.5\Big)\right]^{-1} \tag{3.5.0.4}$$

For general symmetrically distributed random variables, $Y_t, t = 1, 2, \ldots$ and assuming the existence of the first two moments, it is not difficult to see that the constant b can be obtained as

$$b = \left|\left|\frac{Y_t - \mu_t}{\sigma_{t,0.5}}\right|\right|_2 \tag{3.5.0.5}$$

where μ_t is the conditional expectation and $||y||_2$ is the L_2-norm; $||.||_2 = \left(E\Big[y^2\Big]\right)^{\frac{1}{2}}$. Based on a realization $\Big(Y_t, \mathbf{X}_t\Big), t = 1, 2 \ldots, n$, we can estimate the volatility as

$$\widehat{\sigma}\Big(\mathbf{X}_t\Big) = \widehat{b}\widehat{\sigma}_{0.5}\Big(\mathbf{X}_t\Big), \quad \text{for} \quad t = 1, \ldots, n \tag{3.5.0.6}$$

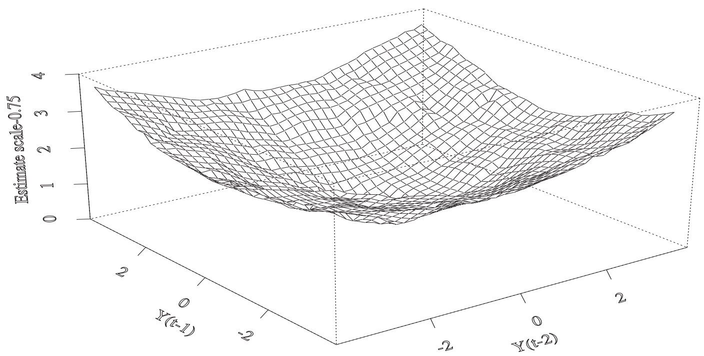

Figure 7: Surface of scale estimate based on (3.1.0.18)

where

$$\widehat{b} = \left(\frac{1}{n-1} \sum_{t=1}^{n} \left(\frac{Y_t - \widehat{\mu}\left(\mathbf{X}_t\right)}{\widehat{\sigma}_{0.5}\left(\mathbf{X}_t\right)} \right)^2 \right)^{\frac{1}{2}}.$$

and $\widehat{\mu}\left(\mathbf{X}_t\right)$ is a least square estimate based on local linear, see (4.2.0.14) in chapter (4). For general distribution, one can use conditional quantile range (CQR) to estimate the volatility. For $\theta > 0.5$, it can be defined as

$$\begin{aligned} CQR_{t,\theta} &= \mu_{t,\theta} - \mu_{t,1-\theta} \\ &= \sigma_t \left(q_\theta^e - q_{1-\theta}^e \right) \end{aligned} \tag{3.5.0.7}$$

Like in symmetric case, σ_t can be obtained by multiplying through (3.5.0.7) by a constant $b = \left(q_\theta^e - q_{1-\theta}^e \right)^{-1}$, which is the same as $\left\| \frac{Y_t - \mu_t}{CQR_{t,\theta}} \right\|_2$. The estimator of the volatility then becomes

$$\widehat{\sigma}\left(\mathbf{X}_t\right) = \widehat{b}\widehat{CQR}_\theta\left(\mathbf{X}_t\right) \tag{3.5.0.8}$$

where

Table 1: $MASE_\theta$ for two methods. Second row is $MASE_\theta$ for (3.1.0.16)

θ	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**
Normal	0.88200	0.54700	0.46800	0.24800	0.47200	0.50100	1.01700
	0.93400	0.76700	0.46700	0.24900	0.45300	0.88200	1.02300
Student-t(4)	1.24700	0.70400	0.56100	0.45200	0.60500	0.94600	1.56800
	1.69600	1.03600	0.57300	0.49100	0.61400	0.92100	1.77800
Cauchy	4.25500	1.64500	0.94500	0.62500	0.91400	1.42200	3.84700
	5.93400	1.93700	0.96600	0.69000	0.91900	1.83000	5.02600
Gamma(2,2)	3.35700	1.53400	0.90300	0.59100	0.88100	1.34700	3.09400
	4.33400	1.79300	0.95100	0.68700	0.93100	1.96400	4.85600

$$\widehat{b} = \left(\frac{1}{n-1} \sum_{t=1}^{n} \left(\frac{Y_t - \widehat{\mu}\Big(\mathbf{X}_t\Big)}{\widehat{CQR}_\theta\Big(\mathbf{X}_t\Big)} \right)^2 \right)^{\frac{1}{2}}$$

Now going back to our general scale function in (3.5.0.3), the following proposition can easily be checked by the method of moments.

Proposition 3.1 *Let* $\Big(Y_t, \mathbf{X}_t\Big), t = 1, 2, \ldots,$ *be* α*-mixing with* $E\Big|Y_t\Big|^{4+\delta} < \infty$ *and* $\delta > 0$, *then for any real random variable* Y_t, *the constant* b *is given by*

$$b = \left\| \frac{Y_t - \mu_t}{\sigma_{t,\theta}} \right\|_2. \tag{3.5.0.9}$$

The estimator of (3.5.0.9) follows as

$$\widehat{b} = \left(\frac{1}{n-1} \sum_{t=1}^{n} \left(\frac{Y_t - \widehat{\mu}\Big(\mathbf{X}_t\Big)}{\widehat{\sigma}_\theta\Big(\mathbf{X}_t\Big)} \right)^2 \right)^{\frac{1}{2}}$$

This gives an estimator of the true volatility function as

$$\widehat{\sigma}\Big(\mathbf{X}_t\Big) = \widehat{b}\widehat{\sigma}_\theta\Big(\mathbf{X}_t\Big) \tag{3.5.0.10}$$

The estimator, $\widehat{b}$, is a sample unconditional standard deviation of the mean adjusted-scaled random variable Y_t. It is not a robust or reliable estimator in small sample samples, but in a large sample it can be expected to provide a reliable estimate as will be seen in the simulation that follows. It can, however, be improved by omiting the 2.5% of the largest and smallest values of the scaled variable. That is, if we let

$$s_t = \left(\frac{Y_t - \widehat{\mu}\Big(\mathbf{X}_t\Big)}{\widehat{\sigma}_\theta\Big(\mathbf{X}_t\Big)} \right), \quad t = 1, 2, \ldots, n$$

Then the estimator for b can be defined as $\widehat{b} = \frac{1}{nx-1} \sum_{t=1}^{n} s_t^2 \mathbf{I}_{\{F_{s_t}^{-1}(0.025) \le s_t \le F_{s_t}^{-1}(0.975)\}}$ where $nx = card\Big\{t : F_{s_t}^{-1}(0.025) \le s_t \le F_{s_t}^{-1}(0.975), \quad t = 1, 2, \ldots, n\Big\}$, as are robust estimators like trimmed mean in Jaeckel [72].

To conclude this section, we generated 500, 800 and 1000 from an AR(1)-TARCH(1) process

$$Y_t = 0.5 + 0.3Y_{t-1} + \sqrt{0.01 + 0.1Y_{t-1}^2 + 0.35\Big(\frac{|Y_{t-1}| - Y_{t-1}}{2}\Big)^2}\, e_t, t = 2, \ldots, \qquad (3.5.0.11)$$

under four different distribution[30] of the error e_t, as shown in tables (2)-(4). The samples were then replicated 500 times. We estimated the conditional median absolute deviation (CMAD), the conditional quantile range (CQR)[31] and QARCH[32]. The true volatilities were estimated using the formula (3.5.0.6),(3.5.0.8) and (3.5.0.10). The performance was then assessed by their average mean absolute proportionate error (AMAPE)

$$AMAPE\Big(\widehat{\sigma}\Big(\mathbf{X}_t\Big)\Big) = \frac{1}{500} \sum_{j=1}^{500} \left[\frac{1}{n_j} \sum_{i=1}^{n_j} \left| \frac{\widehat{\sigma}^{(j)}\Big(x_i\Big) - \sigma^{(j)}\Big(x_i\Big)}{\sigma^{(j)}\Big(x_i\Big)} \right| \right], \quad n_j = 500, 800, 1000$$

The results are shown in tables (2),(3) and (4).

The AMAPE tends to be the same for all the estimators at a given distributional error and sample size. This indicates some sort of standardization of the scale functional estimators into the same quantity (volatility). As the sample size, n, increases, the AMAPE decreases confirming the theoretical result on convergence. Thus it is expected that as $n \to \infty, \quad \widehat{b} \to b$ and hence $\widehat{\sigma}\Big(X_t\Big) \to \sigma\Big(X_t\Big)$. This result indicate further investigations

[30] With the errors from stdudent's-t and Gamma distribution adjusted and scaled appropriately.

[31] At $\theta = 0.75$-level

[32] At $\theta = 0.75$-level

Table 2: $n_j = 500$: AMAPE

Error	**CMAD**	**QARCH**	**CQR**
Normal	0.2472	0.2480	0.2424
Student-t(4)	0.2974	0.3068	0.3069
Cauchy	0.7005	0.6940	0.6733
Gamma(2,2)	0.5946	0.5890	0.5704

Table 3: $n_j = 800$: AMAPE

Error	**CMAD**	**QARCH**	**CQR**
Normal	0.2005	0.2005	0.2004
Student-t(4)	0.2469	0.2467	0.2470
Cauchy	0.6846	0.6745	0.6654
Gamma(2,2)	0.5592	0.5434	0.5346

Table 4: $n_j = 1000$: AMAPE

Error	**CMAD**	**QARCH**	**CQR**
Normal	0.1840	0.1844	0.1842
Student-t(4)	0.2068	0.2495	0.1985
Cauchy	0.5476	0.5406	0.5434
Gamma(2,2)	0.5067	0.4956	0.4900

could be carried out to determine among others things of interest, the rate of convergence. It is important to mention that in some case, it is of interest to determine b that varies with time t. This brings about the phenomenon of scale changes observed in Beran and Ocker [10]. In this case we could use the least square estimator based on local linear approximation to obtain,

$$\widehat{b}_t^2 = \arg \min_{(b_0,b_1)\in\mathbf{R}_+\times\mathbf{R}} \sum_{t=1}^{n} \Big(s_t^2 - b_0 - b_1\Big(\frac{t-0.5}{n} - \widetilde{t}\Big)\Big)^2 k_h\Big(\frac{t-0.5}{n} - \widetilde{t}\Big)$$

where k is a univariate kernel function, h the bandwidth and $\widetilde{t}$ is a fixed point in $(0,1)$. The asymptotic properties of $\widehat{b}_t^2$ can be obtained precisely in the same lines and in exact forms as the ones outlined in Feng [48].

3.6 Extensions to GQARCH

Let $\Big\{Y_t, t \in \mathbf{Z}\Big\}$ be stationary stochastic process adopted to filtration $\Big\{\mathbf{F}_t; t \in \mathbf{Z}\Big\}$ and having the form

$$\begin{aligned} Y_t &= \sigma_{t,\theta} Z_t \\ \sigma_{t,\theta}^2 &= \sigma_\theta^2\Big(\mathbf{X}_t, S_t\Big) \end{aligned} \tag{3.6.0.12}$$

where we take $S_t = \Big(\sigma_{t-1}^2, \ldots, \sigma_{t-\tau}^2\Big)$ with $\sigma_{t-i}^2 = b_i^2\sigma_{t-i,\theta}^2, \quad i = 1, 2, \ldots, \tau$. The b_i's are the rescaling constants which may be constant or time dependent within some periods and $\sigma_{t-i,\theta}$ are the lagged values of the conditional scale function based on QARCH. Let $\Big\{Z_t, t \in \mathbf{Z}\Big\}$ be iid innovations with zero θ-quantile and finite fourth moment[33]. Assume all other assumptions specified in model (1.4.0.2). Observe that (3.6.0.12) can be written in terms of an additive noise $M_\theta^2\Big(Y_t, 0\Big) = \sigma_\theta^2\Big(\mathbf{X}_t, S_t\Big) + Z_{t,2}$ where as in (3.1.0.15), $Z_{t,2} = \sigma_\theta^2\Big(\mathbf{X}_t, S_t\Big)\Big(M_\theta^2\Big(Z_t, 0\Big) - 1\Big)$ and has zero θ-quantile. The stochastic function $\sigma_\theta^2\Big(\mathbf{X}_t, S_t\Big)$ can be estimated by regressing $M_\theta^2\Big(Y_t, 0\Big)$ on $\mathbf{X}_t$ and S_t using the asymmetric least absolute based on local linear.

Using the method in section (3.5), denote $\widehat{S}_t$ to be a variable consisting of the estimates given by $\widehat{\sigma}_{t-i}^2 = \widehat{b}_i^2\widehat{\sigma}_{t-i,\theta}^2, \quad i = 1, \ldots, \tau$. The consistency of this extension can be given through the contraction property with respect to the hidden variable, c.f Buehlmann and McNeil [18]. That is for $\mathbf{x}_i \in \mathbf{R}^{d-\tau}$ and $S_t, \widehat{S}_t \in \mathbf{R}_+^\tau$ we assume

[33] This is for the purpose of estimationg $b^{'}$s

$$\sup_{\mathbf{x}_i \in \mathbf{R}^{d-\tau}} \left| \sigma_\theta\Big(\mathbf{x}_i, \sigma_1^2, \ldots, \sigma_\tau^2\Big) - \sigma_\theta\Big(\mathbf{x}_i, \widehat{\sigma}_1^2, \ldots, \widehat{\sigma}_\tau^2\Big)\right| \leq \sum_{j=1}^{\tau} c_j \left|\sigma_j^2 - \widehat{\sigma}_j^2\right|$$

for some $0 < c_1, \ldots, c_\tau < 1$ with $\sum_{j=1}^{\tau} c_j < 1$. By assuming that for any $\delta > 0$, and $\sup_{1<j<\tau} \left|\widehat{\sigma}_j^2 - \sigma_j^2\right| \leq \delta \to 0$ with n, similar arguments used in section (3.2) or (3.3) could be applied to show consistency.

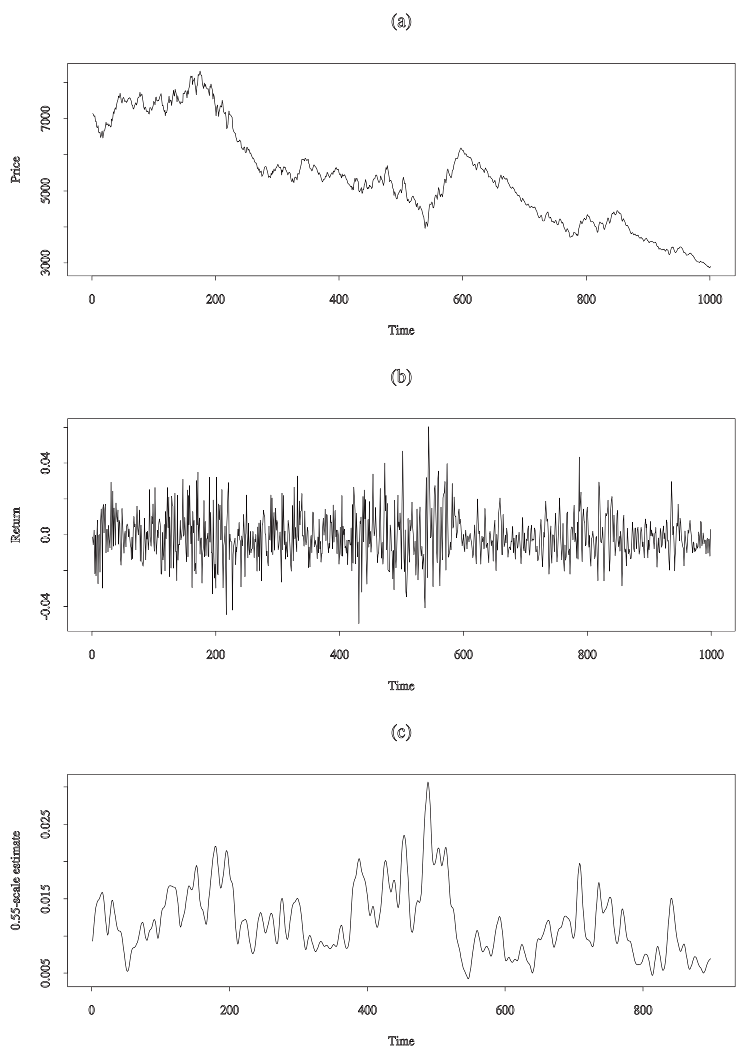

Figure 8: Real data: Scale function estimate in (c)

In the following example, we applied the method to real data consisting of DAX prices for the period ranging from (1/1/1997) to (6/11/2000). The data is shown in figure 8(a) with a decreasing trend. The calculated returns shown in figure 8(b) contain periods of high volatilites around the times 200 and 500. The nonparametric scale function estimate, shown in figure 8(c) at $\theta = 0.55$, is high for both low and high values of the returns.

3.7 Conclusion

This chapter provided results on consistency and asymptotic distribution for the estimator of QARCH based on local constant and polynomials, under the assumption that the QAR of Y_t is zero. The comparison between (3.1.0.16) and (3.1.0.18) reveals the estimator based on the former could be suffering from correlation problems. Numerical results based on the proposed standardization method indicates the CMAD, QARCH and CQR could be used to estimate the volatility. However, extracting the volatility from a too asymmetric distribution, the CQR is more appropriate.

4 Extreme Quantile Autoregression (Extreme QAR)

Usually the applications of quantiles to financial risk analysis are not only restricted to moderate or relatively high probability levels, but also high and sometimes beyond the maximum observation, or in other words, out-of-sample. We call the quantiles, which are located among the largest observations or even beyond the data maximum, extreme quantiles. In this chapter we combine the QAR results presented in chapters 2 and 3, for the interior parts of the data, with results from extreme value theory for the extreme parts to provide approximate extreme QAR and its estimate.

4.1 Result from extreme value theory

Extreme value theory is a classical topic in probability theory. For a survey on the subject, see for example Leadbetter et al. [80], or Embrechts et al. [39]. In this section we give some intuition and basic results of extreme value theory (EVT) which can be considered as a complement of central limit (for cumulative sums) that deals with fluctuations of sample maxima.

4.1.1 Generalized Extreme Value distribution

The limiting behaviour of sample extrema (maxima) are studied under the family of extreme value distributions. One of the main results is due to Fisher and Tippett (1928) who specified the form of the limit distributions[34] for an appropriately normalized maxima, as summarized in theorem 4.1.

Theorem 4.1 *Suppose* $e_1, e_2, \ldots,$ *is a sequence of iid random variables from unknown distribution* F *and* $M_n = \max\Big(e_1, e_2, \ldots e_n\Big)$ *denotes the maximum of the first* n *observations. If a sequence of real numbers* $a_n > 0$ *and* $b_n \in \mathbf{R}$ *can be found such that the sequence of normalized maxima,* $\frac{M_n - b_n}{a_n}$*, converges in distribution(law or weakly),i.e*

$$\lim_{n\to\infty} Pr\left\{\frac{M_n - b_n}{a_n} \leq e\right\} = \lim_{n\to\infty} F^n\Big(a_n e + b_n\Big) = H(e), \quad e \in \mathbf{R},$$

[34] And was generalized by von Mises [90]

for some non-degenerate d.f H, then H belongs to one of the three distribution types:

$$H_\xi(e) = \begin{cases} \exp\{-(1+\xi e)^{-\frac{1}{\xi}}\} & : \quad \xi \neq 0 \\ \exp\{-\exp(-e)\} & : \quad \xi = 0 \end{cases} \tag{4.1.1.1}$$

where[35] *e is such that $1+\xi e > 0$, ξ is the shape parameter and the special case $H_0(e)$ is interpreted as $\lim_{\xi \to 0} H_\xi(e)$.*

H_ξ is called the Generalized Extreme Value distribution (GEV). An important concept for the application of extreme value theory to VaR (or extreme quantile) estimation is the Maximum Domain of Attraction (MDA). In simple terms, a random variable e_t is said to belong to the maximum domain of attraction of the extreme value distribution H ($\{e_t\} \in MDA(H)$) if and only if the Fisher-Tippet theorem holds for $\{e_t\}$. The result is very significant, since the asymptotic distribution of the maxima always belongs to one of these distributions, whatever the underlying distribution function and therefore the asymptotic distribution of the maxima can be estimated without making strict assumptions about the nature of the underlying distribution function of the observation.

The shape parameter ξ is crucial in determining the class (type) of the GEV distribution;

(i) The distribution H_ξ for $\xi = \frac{1}{\alpha} > 0$ is known as the Fréchet. The distributions in $MDA\Big(H_\xi, \xi > 0\Big)$ are fat-tailed and their tails decay like a power function (see green and red curves in figure (9), for $\xi > 0$). The tail index, α, can be related to the number of finite moments. For t-distribution, α is the degrees of freedom and for stable distribution α is the characteristic exponent. The red and green curve depicts the tail of distributions with first and fourth moments being finite respectively. The relatively slow decline in the tails generates moments that are not necessarily finite. The class includes the Pareto, Burr, Loggamma, Cauchy and t-distribution and has been found to be the most appropriate for modelling fat-tailed financial data.

(ii) For $\xi < 0$ the distributions are said to be in the maximum domain of attraction of the Weibull $\Big(H_\xi, \quad \xi < 0\Big)$. Such processes are short tailed or bounded; this is depicted by the

[35] Here and throughout this thesis, e denotes a real value of random variables, say $e_t, t = 1, \ldots$ and not the usual convention for the exponential.

light blue curve in figure 9, for $\xi < 0$. They include the uniform and beta distributions. (iii) The distributions with $\xi \to 0$ or $\alpha \to \infty$, belong to the maximum domain of attraction of the Gumbel distribution $MDA(H_0)$. This class is characterized by medium tails, shown by dark blue curve in figure (9). They include gamma, normal, lognormal.

In connection to theorem 4.1 is that the upper tail (also applies trivially to the lower

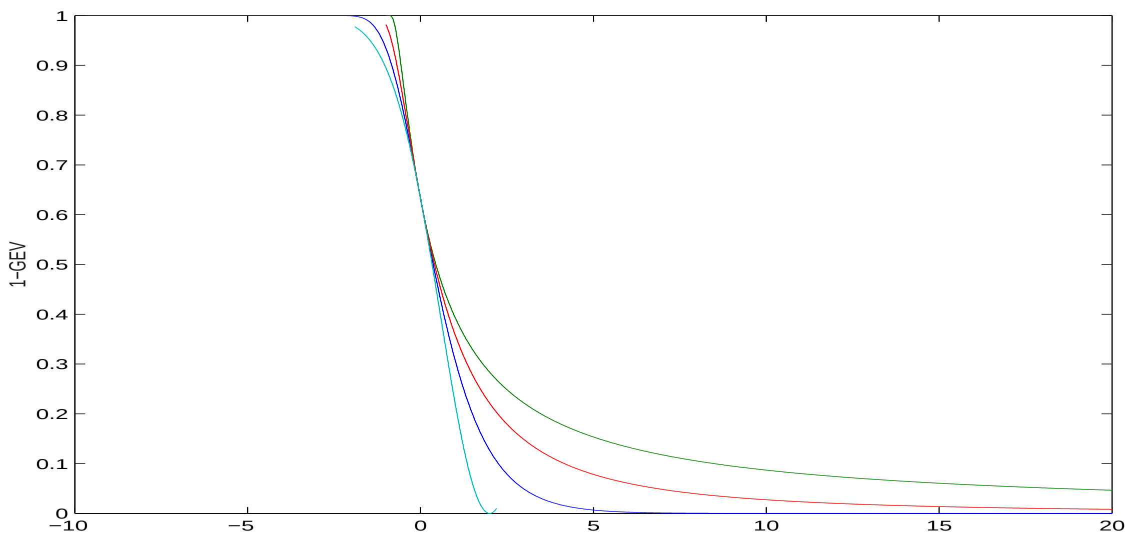

Figure 9: A plot of excess distribution of the GEV against a sequence of real numbers. Green($\xi = 1$), Red($\xi = 0.25$), Dark blue ($\xi \to 0$), Light blue ($\xi = -0.5$)

tail as well) of any fat tailed random variable e_t has the following property:

$$\lim_{e \to \infty} \frac{1 - F\Big(ce\Big)}{1 - F\Big(e\Big)} = c^{-\frac{1}{\xi}}, \quad \xi, c, e > 0.$$

where F can be interpreted as any distribution function which varies regularly at infinity with tail index $\alpha = \frac{1}{\xi}$. From this, it is important to note that regardless of the underlying distribution of e_t, the tails have the same general shape, where only the shape parameter is important. If the data are generated by a heavy tailed distribution, then to a first order approximation, its distribution has a Pareto type tail,

$$P\Big\{e_t > e\Big\} = \overline{F}\Big(e\Big) \sim ae^{-\frac{1}{\xi}}, \quad a \in \mathbf{R}_+, \quad \xi > 0, \quad \text{for} \quad e \to \infty. \tag{4.1.1.2}$$

Theorem 4.1 was proved by Gnedenko (1946) who showed that for $\xi > 0$, $F \in MDA(H_\xi, \xi > 0)$ if and only if

$$\overline{F}\Big(e\Big) = e^{-\frac{1}{\xi}} L, e > 0 \tag{4.1.1.3}$$

for some slowly varying functions L^{36}. This is a necessary and sufficient condition for the tail of any distribution function F to belong to the maximum domain of attraction of a Fréchet distribution.

4.2 Extreme QAR functon: Part I

In this section, we explore a semiparametric estimation procedure for extreme QAR. We take the QAR in the interior parts of a data based on relatively high probability level, say θ, to be an initial as well as the beginning of the right-hand tail of a heavy tailed distribution. This is then combined with quantiles obtained by using Gnedenko's result and a Hill's estimator of the tail index to arrive at an approximate extreme QAR function at high probability levels, say $\varphi > \theta$.

For intuitive understanding, we first consider the iid random variables $e_t, \dots$ based on the process given in (1.1.1.1). To derive an estimate of an extreme quantile q^e_φ for $\varphi \approx 1$, we also consider a high quantile q^e_θ where $\theta < \varphi$ is large but not so close to 1 as φ. Later on, we choose θ large, but still small enough that the procedure of the previous chapters provide reliable estimates of q^e_θ. φ is so large that we have non or only few data in our sample around q^e_φ, and the purely nonparametric approaches do no longer provide good estimates of q^e_φ. The quantiles, q^e_θ and q^e_φ, correspond, respectively, to the excess probabilities $\bar{F}\Big(q^e_\theta\Big) = 1-\theta$ and $\bar{F}\Big(q^e_\varphi\Big) = 1-\varphi$. Then, using Gnedenko's result (4.1.1.3), the excess probabilities also satisfy

$$\bar{F}\Big(q^e_\theta\Big) = \Big(q^e_\theta\Big)^{-\frac{1}{\xi}} L\Big(q^e_\theta\Big) \tag{4.2.0.4}$$

$$\bar{F}\Big(e\Big) = e^{-\frac{1}{\xi}} L\Big(e\Big) \quad , \quad e > q^e_\theta. \tag{4.2.0.5}$$

Dividing (4.2.0.5) by (4.2.0.4) and noting that for large θ, $\frac{L(e)}{L(q^e_\theta)} \approx 1$, we obtain,

[36] A positive, Lebesgue measurable function L on $(0, \infty)$ is slowly varying if $\lim_{e\to\infty} \frac{L(te)}{L(e)} = 1, \quad t > 0$. See theorem 3.3.7 page 131 Embrechts et al. [39] for more details and other variations.

$$e \approx \left(\frac{1 - F\left(e\right)}{1-\theta}\right)^{-\xi} q_\theta^e, \quad \text{for large} \quad \theta \tag{4.2.0.6}$$

If $F\left(e\right) = \varphi > \theta$, then the φ-quantile can be obtained as the inverse,

$$q_\varphi^e \approx \left(\frac{1-\varphi}{1-\theta}\right)^{-\xi} q_\theta^e, \quad \text{for large} \quad \theta \quad \text{and} \quad \varphi > \theta \tag{4.2.0.7}$$

whose estimate is

$$\widehat{q}_\varphi^e = \left(\frac{1-\varphi}{1-\theta}\right)^{-\widehat{\xi}} \widehat{q}_\theta^e.$$

Under the assumption that the threshold q_θ^e is known and that $\overline{F}\left(e\right) = ce^{-\frac{1}{\xi}}$ for $e > q_\theta^e$ and an appropriate constant $c > 0$, the maximum likelihood estimator of the reciprocal of the tail index $\xi = \frac{1}{\alpha}$, is easily obtained by

$$\widehat{\xi} = \frac{1}{N_\theta} \sum_{t=1}^{n} \log\left(\frac{e_t}{q_\theta^e}\right) \mathbf{I}_{\{e_t > q_\theta^e\}} \tag{4.2.0.8}$$

with N_θ, the number of exceedances. This is known as the Hill estimator, introduced and shown, in Hill [64] that it is consistent. In practice the threshold level needs to be determined. We have to recall that we are not necessarily dealing with a Pareto distribution, but rather with a distribution whose tail belongs to $MDA\left(H_\xi, \xi > 0\right)$ and therefore looks like a Pareto tail. Consequently, we are looking for some level, say q_θ^e, above which the Pareto law applies to a good approximation. So long as we know that $\xi > 0$, θ can be set high enough, and we obtain q_θ^e parametrically by minimizing $M_\theta\left(e_t, \mu\right)$, defined in (1.1.2.1), with respect to μ. This results in nothing but a sample quantile estimator at θ, which is consistent and asymptotically normal as shown in Koenker and Basset [75]. The empirical study of the daily log-returns have shown that the frequently encountered values of $\alpha = \frac{1}{\xi}$ based on the excesses are between 3 and 4, see Longin [82] and Embrechts et al. [39] for example.

The properties of the quantile and tail probability estimators follow directly from the properties of the Hill estimator of the tail index $\widehat{\alpha} = \frac{1}{\widehat{\xi}}$. The consistency of the shape parameter, $\widehat{\xi}$ and $\widehat{q}_\theta^e$ implies that $\widehat{q}_\varphi^e$ is consistent. An inverse estimator can be obtained in a similar way from (4.2.0.6) as

$$\widehat{F}\Big(e\Big) = 1 - (1-\theta)\Big(\frac{e}{\widehat{q}^e_\theta}\Big)^{-\frac{1}{\xi}} \tag{4.2.0.9}$$

The above derivations enable us to present our first result in this chapter. The proposition below extends the estimation of extreme quantiles in the iid case to the dependent case, by augmenting the QAR with Gnedenko's result and Hill's estimator of the shape parameter in (4.2.0.8).

Proposition 4.1 *Assume the random variable $Y_t, t = 1, \ldots,,$ in model (1.1.1.1) and the iid errors e_t with d.f $F \in MDA\Big(H_\xi, \xi > 0\Big)$. The conditional extreme time varying quantile is given by*

$$\mu_{t,\varphi} = \mu_{t,\theta} + \sigma_t q^e_\theta\left(\left(\frac{1-\varphi}{1-\theta}\right)^{-\xi} - 1\right) \tag{4.2.0.10}$$

Proof of Proposition (4.1

From (1.4.0.2) and lemma 1.3

$$\frac{Y_t - \mu_{t,\theta}}{\sigma_{t,\theta}} = Z_t = \frac{e_t}{M^e_\theta} - \frac{q^e_\theta}{M^e_\theta} \tag{4.2.0.11}$$

Consider $\left\{\frac{e_t}{M^e_\theta}\right\}$ to be iid random variables and $\frac{q^e_\theta}{M^e_\theta}$ to be the threshold, then equation (4.2.0.6) and (4.2.0.7) give $\frac{\mu_{t,\varphi}-\mu_{t,\theta}}{\sigma_{t,\theta}} + \frac{q^e_\theta}{M^e_\theta} = \left(\frac{1-\varphi}{1-\theta}\right)^{-\xi} . \frac{q^e_\theta}{M^e_\theta}$ resulting in

$$\frac{\mu_{t,\varphi} - \mu_{t,\theta}}{\sigma_{t,\theta}} = q^z_\varphi = \frac{q^e_\theta}{M^e_\theta}.\left(\left(\frac{1-\varphi}{1-\theta}\right)^{-\xi} - 1\right) \tag{4.2.0.12}$$

Rearrangement completes the proof.

□

We now assume again that $\mu_{t,\theta} = \mu_\theta\Big(\mathbf{X}_t\Big)$ is the conditional θ-quantile of Y_t given $\mathbf{X}_t$. We also assume that $\mu_t = \mu\Big(\mathbf{X}_t\Big)$, the conditional expectation of Y_t given $\mathbf{X}_t$, exists. Because the estimator of $\sigma_t q^e_\theta$ would involve the second moment, we replace it by $(\mu_{t,\theta} - \mu_t)$, which requires only the first one. The extreme conditional φ-quantile estimator is then

$$\widehat{\mu}_\varphi\Big(\mathbf{x}_i\Big) = \widehat{\mu}_\theta\Big(\mathbf{x}_i\Big) + \Big(\widehat{\mu}_\theta\Big(\mathbf{x}_i\Big) - \widehat{\mu}\Big(\mathbf{x}_i\Big)\Big)\Big(\Big(\frac{1-\varphi}{1-\theta}\Big)^{-\widehat{\xi}} - 1\Big) \tag{4.2.0.13}$$

where $\widehat{\xi} = \frac{1}{N_\theta}\sum_{t=1}^{n}\log\left(\frac{Y_t-\widehat{\mu}_\theta(X_t)}{\widehat{\mu}_\theta(X_t)-\widehat{\mu}(X_t)}\right)\mathbf{I}_{\{Y_t>\widehat{\mu}_\theta(X_t)\}}$ and N_θ is the number of exceedances of Y_t over $\widehat{\mu}_\theta\left(X_t\right)$. We obtain the estimator for $\mu\left(\mathbf{X}_t\right)$ by using local linear method in Fan and Gijbels [45]. With $\left\{\left(Y_t, \mathbf{X}_t\right)\right\}_{t=1}^{n}$, we assume that the function $\mu\left(\mathbf{x}_i\right)$ has the second partial derivative so that it can be approximated by a linear function in the neighborhood of a point $\mathbf{x}_i$ as $\mu\left(\mathbf{X}_t\right) \approx b_0\left(\mathbf{x}_i\right) + \sum_{j=1}^{d} b_j\left(\mathbf{x}_i\right)\left(X_{t,j} - x_{i,j}\right)$ with $\mu\left(\mathbf{x}_i\right) = b_0\left(\mathbf{x}_i\right)$ and $\frac{\partial\mu\left(\mathbf{x}_i\right)}{\partial x_{i,j}} = b_j\left(\mathbf{x}_i\right), \quad j = 1,\ldots,d$. The estimate, $\widehat{\mu}\left(\mathbf{x}_i\right)$, is obtained as the first element of the minimizer of

$$\sum_{t=1}^{n}\left(Y_t - b_0 - \sum_{j=1}^{d} b_j\left(X_{t,j} - x_{i,j}\right)\right)^2 \mathbf{K}_{\mathbf{h}}\left(\mathbf{X}_t - \mathbf{x}_i\right) \tag{4.2.0.14}$$

with respect to a vector $\mathbf{b} = \left(b_0,\ldots,b_d\right)^{'}$. Masry [86] has shown under the assumptions $h \to 0, \quad nh^d \to \infty$ and boundedness of nh^{d+4}, among others that

$$\widehat{\mu}\left(\mathbf{x}_i\right) - \mu\left(\mathbf{x}_i\right) = O\left(\left(\frac{\log^4(n)}{nh^d}\right)^{\frac{1}{2}}\right) \tag{4.2.0.15}$$

almost surely for each point $\mathbf{x}_i$. Note that if the conditional expectation, μ_t is assumed to be equal to zero, then $\widehat{\mu}_\theta\left(\mathbf{X}_t\right)$ can be taken as the scale function.

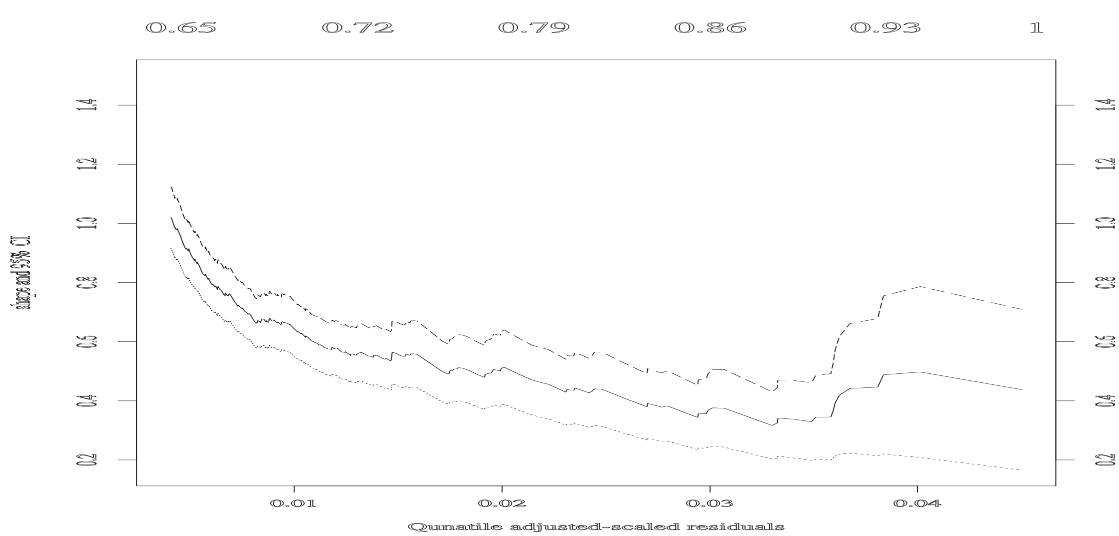

Figure 10: Shape estimate against threshold

We applied formula (4.2.0.13) in the estimation of the extreme conditional quantile on real data (negative returns), from BASF in the period ranging from 1/1990 to 12/1992. First we estimated the QAR of Y_t at $\theta = 0.9$ by the procedure described in chapter (3)

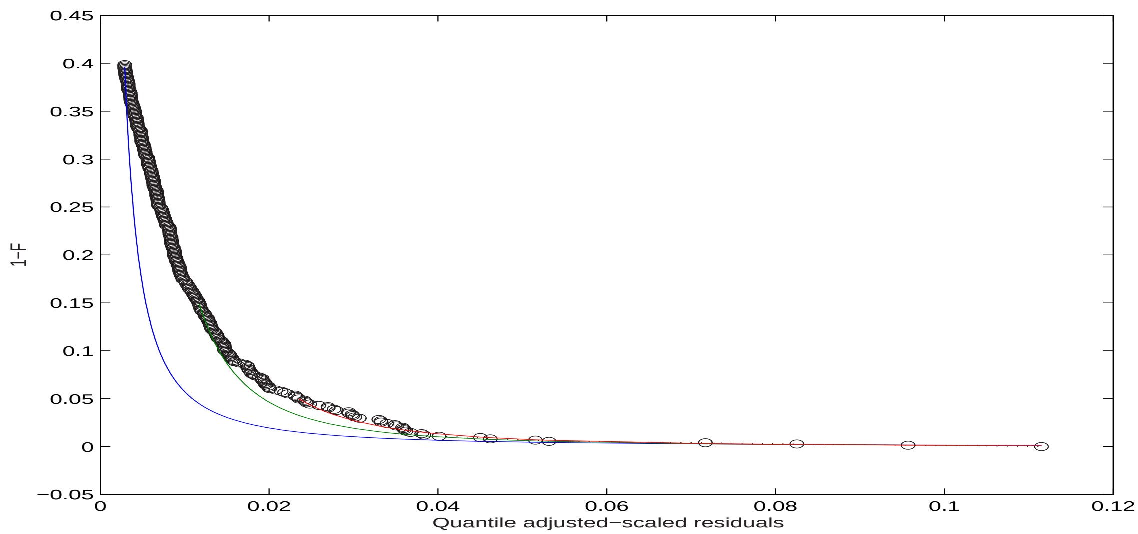

Figure 11: Hill estimates of tail distribution

and the conditional mean. Then we adjusted and scaled the resulting residuals. We then estimated the shape parameter from the scaled excesses over the QAR using Hill's estimator. Figure (10) depicts a plot of the Hill's estimates of the shape against the excesses. We choose $\xi = 0.31$ (corresponding to the stable areas) and estimated the tail distribution, which is depicted in figure (11) as a red curve. The circle represent the empirical distribution. The blue and green curves represent the Hill estimates of the tail distribution when the threshold is fixed at $\theta = 0.6$ and 0.85. Estimation for the latter two were done with their respective estimates of the shape parameter. For low threshold, the Hill's estimator underestimate the degree of heavy tailedness of the distribution. At $\theta = 0.90$, it produces almost the same distribution as the empirical. For higher threshold, we expect it to produce even heavier tail than the empirical. The estimate of the QAR at $\varphi = 0.95$ and 0.99 are shown in figure (12) with blue and red colored curves respectively. We deffer all other comments to section (4.6), where a detailed simulation study is carried out.

4.3 Parametric estimation of extreme quantile(iid case)

This section presents some results on parametric fitting of distribution to a series of iid excesses beyond a high threshold. This results are useful for the work in the following section.

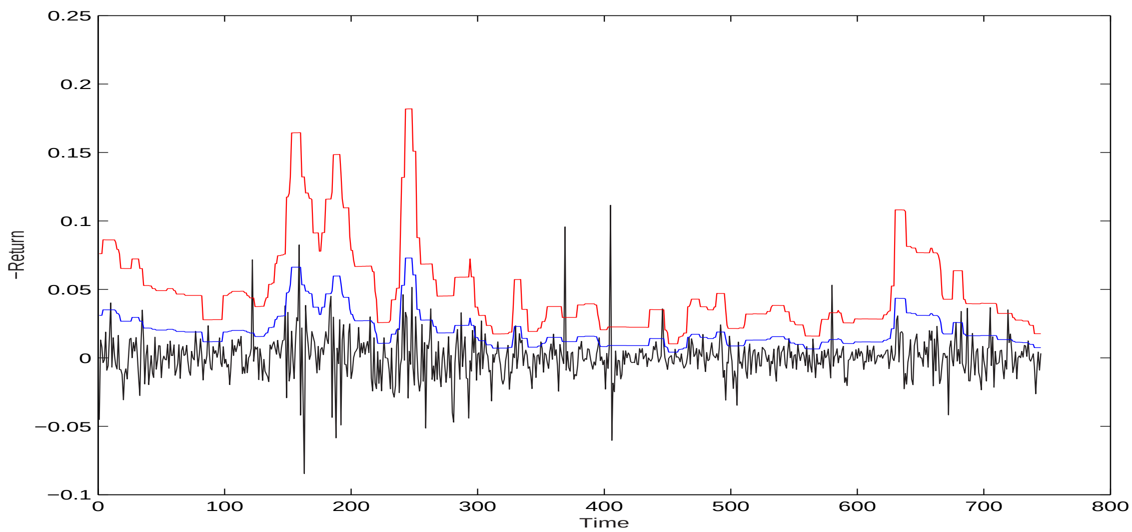

Figure 12: Hill estimates(blue at 0.95 and red at 0.99) superimposed on daily negative returns on BASF.

Definition 4.3.1 *Let* e_t *be iid random variables and* $z_1, \dots, z_{N_\theta}$ *be the series of exceedances over the threshold* $u = q^e_\theta$*. The excess distribution function of the random variable* e_t *with the distribution function* F *over the threshold* u *is defined as*

$$F_u\Big(z\Big) = Pr\Big(e_t - u \leq z \Big| e_t > u\Big), \quad z > 0 \tag{4.3.0.16}$$

It is assumed that the excesses are iid with distribution function F_u and u is less than e_F[37] . In terms of distribution function, (4.3.0.16) can be written as

$$F_u\Big(z\Big) = \frac{F\Big(u+z\Big) - F\Big(u\Big)}{1 - F\Big(u\Big)} \tag{4.3.0.17}$$

which, when rearranged, one arrives at the tail distribution of the random variable e_t above the threshold u

$$\overline{F}(u+z) = \overline{F}(u).\overline{F}_u(z) \tag{4.3.0.18}$$

This result makes it possible to estimate the tail of the original distribution, by separately estimating F and F_u in (4.3.0.18). The Peak Over Threshold (POT), due to

[37] $e_F = \sup_{e \in \mathbf{R}}\{F\Big(e\Big) < 1\}$, the right- hand endpoint of the distribution, usually but not necessarily, assumed to be $+\infty$

Todorovic and Zelenhasic (1970), can be used to model all large observations exceeding a high threshold u. In this context a fully parametric model based on the generalized Pareto distribution (GPD), defined below, can be fitted to the excesses.

Definition 4.3.2 (Standard generalized Pareto distribution (GPD)) .
The generalized Pareto d.f G_ξ, *is defined by*

$$G_\xi(z) = \begin{cases} 1-(1+\xi z)^{-\frac{1}{\xi}} & : \quad \text{if} \quad \xi \neq 0 \\ 1-\exp(-z) & : \quad \text{if} \quad \xi = 0 \end{cases} \tag{4.3.0.19}$$

where $z \geq 0$ *if* $\xi \geq 0$ *and* $0 \leq z < -\frac{1}{\xi}$, *if* $\xi < 0$.

The location-scale family, denoted as $G_{\xi,\nu,\beta}(e)$, of (4.3.0.19) is obtained by replacing z by $\frac{e-\nu}{\beta}$ for $\nu \in \mathbf{R}$, $\beta > 0$, i.e

$$G_{\xi,\nu,\beta}(e) = \begin{cases} 1-\left(1+\frac{\xi}{\beta}(e-\nu)\right)^{-\frac{1}{\xi}} & : \\ 1-\exp(-\frac{e-\nu}{\beta}) & : \end{cases} \tag{4.3.0.20}$$

where

$$e \in D(\xi,\beta) = \begin{cases} [0,\infty) & : \quad \text{if} \quad \xi \geq 0 \\ \left[0, -\frac{\beta}{\xi}\right] & : \quad \text{if} \quad \xi < 0 \end{cases}$$

and $G_{0,\nu,\beta} = \lim_{\xi \to 0} G_{\xi,\nu,\beta}\Big(e\Big)$. For $0 < \xi < \frac{1}{2}$, the random variable e_t which follows a $G_{\xi,\nu,\beta}$ has the mean and variance equal to

$$E\Big[e_t\Big] = \nu + \frac{\beta}{1-\xi}$$

and

$$var\Big[e_t\Big] = 2\frac{\beta^2}{\xi^3}\frac{\Gamma(\xi^{-1}-2)}{\Gamma(\xi^{-1}+1)}$$

respectively. As an abbreviation, we write $G_{\xi,\beta} \equiv G_{\xi,0,\beta}$ for the GPD with location parameter $\nu = 0$.

Two approaches can be considered when fitting a GPD. The first one is based on the assumption that the unknown distribution function F has an exact GPD tail: Let $e_1, \ldots, e_n$ be iid with distribution function F whose tail above a threshold u follows exactly a GPD tail, i.e.

$$F_u\Big(z\Big) = G_{\xi,\beta(u)}\Big(z\Big) = 1 - \left(1 + \frac{\xi}{\beta(u)} z\right)^{-\frac{1}{\xi}}. \tag{4.3.0.21}$$

The estimates for ξ and $\beta(u)$ can be otained by maximum likehood estimation. The estimates exist so long as $\xi > -1$ and are asymptotically normal and efficient when $\xi > -\frac{1}{2}$, see Smith [103]. For high u, the density of the excesses can be approximated at an arbitrary Z_i by

$$f_{\xi,\beta(u)}\Big(z\Big) = \frac{\xi}{\beta(u)}\Big(1 + \frac{\xi z}{\beta(u)}\Big)^{-\frac{1}{\xi}-1},$$

whose log-likelihood function is

$$L\Big(\xi, \beta(u)\Big) = N_\theta \log\Big(\beta(u)\Big) - \Big(\frac{1}{\xi} + 1\Big) \sum_{i=1}^{N_\theta} \log\Big(1 + \frac{\xi Z_i}{\beta(u)}\Big)$$

as given in Embrechts et al. [39]. Hosking and Wallis (1987) have shown that the MLE, although asymptotically most efficient, it is not as efficient as the method of moment even in samples as large as 500. The GPD estimators based on the method of moments are of the form

$$\begin{aligned}
\widehat{\xi} &= \frac{1}{2}\Big(1 - \frac{\overline{z} - u}{s^2}\Big) \\
\widehat{\beta}(u) &= \frac{\overline{z} - u}{2}\Big(\frac{\overline{z} - u}{s^2} + 1\Big)
\end{aligned}$$

where $\overline{z}$ and s^2 are empirical mean and variance respectively.

To relax the strictness of the exact type of the distribution a more realistic approach is to use, that for any heavy tailed distribution F, the following result due to Balkema-de Haan and Pickands(1975) holds.

Theorem 4.2 (Limiting distribution of $\overline{F}_u(z)$)

For $F \in MDA(H_\xi, \xi > 0)$, the generalised Pareto distribution (GPD) is the limiting distribution for the distribution of excesses, as the threshold tends to the far right endpoint e_F .i.e

$$\lim_{u \to e_F} \sup_{0<z<e_F-u} \left| \overline{F}_u(z) - \overline{G}_{\xi,\beta(u)}(z) \right| = 0 \qquad (4.3.0.22)$$

So

$$\overline{F}(u+z) \approx \overline{F}(u).\overline{G}_{\xi,\beta(u)}(z) \qquad (4.3.0.23)$$

The result states that if F is in the maximum domain of attraction of a Fréchet distribution, then as the threshold u approaches the endpoint of F, the GPD asymptotically approximates the excess distribution function $F_u\Big(z\Big)$.

In order to exploit theorem 4.2 in our problem, we summarize the asymptotic properties of the ML estimates of the GPD parameter estimates in the following lemma. For the proof, see Smith [103]

Lemma 4.1 (Asymptotic properties)
Let $F \in MDA\Big(H_\xi, \xi > 0\Big)$, i.e $L\Big(e\Big) = e^{\frac{1}{\xi}}\overline{F}\Big(e\Big)$ is slowly varying at ∞ and suppose $N_u \to \infty, \quad u \to e_F$ simultaneously. Then the ML estimates, $\begin{bmatrix} \widehat{\xi} \\ \widehat{\beta}(u) \end{bmatrix}$, are consistent and asymptotically normal with

$$\sqrt{N_u}\begin{bmatrix} \widehat{\xi} - \xi \\ \frac{\widehat{\beta}(u)}{\beta(u)} - 1 \end{bmatrix} \quad \to^D N\left[\begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad , M^{-1}\right], \qquad (4.3.0.24)$$

where $M = \frac{1}{(1+\xi)(2\xi+1)}\begin{bmatrix} 2 & 1 \\ 1 & 1+\xi \end{bmatrix}$ *is the Fisher information matrix for* $(\xi, \beta(u))'$ *and* $M^{-1} = \Big[1+\xi\Big]\begin{bmatrix} 1+\xi & -1 \\ -1 & 2 \end{bmatrix}$.

This result can be used to make inferences on the estimates. The next section develops a procedure, similar to section (4.2), for dependent data.

4.4 Extreme QAR: Part II

In Chapters (2) and (3), we gave consistent estimators for $\mu_\theta\Big(\mathbf{x}_i\Big)$ and scale function, $\sigma_\theta\Big(\mathbf{x}_i\Big)$ via conditional distribution function estimators and direct minimization. These two approaches were shown to work with dependent data under strong mixing conditions. As already seen, in order to exploit results from extreme value theory, independence in the series is required. We propose to filter the trend due to stochastic location or in general, QAR, and volatility by adjusting the QAR and then scaling the difference. Let us denote the filtered excess residuals by

$$Z_t^+ = \left(\frac{Y_t - \mu_\theta(\mathbf{X}_t)}{\sigma_\theta(\mathbf{X}_t)} \right) > 0, \quad t = 1, \ldots, n \tag{4.4.0.25}$$

Note that the assumption of independence are relaxed up to some high levels of θ. That is, whereas the mean-variance method[38] assumes standardized excesses over the conditional mean are iid , our proposed approach only assumes independence for only Z_t^+ corresponding to large θ. In practice, if we replace $\mu_\theta\Big(\mathbf{X}_t\Big)$ and $\sigma_\theta\Big(\mathbf{X}_t\Big)$ by their estimates, this assumption is only approximately satisfied. However, the following discussion shows that the resulting estimates of the Z^+ are at least uncorrelated to a good degree of approximation.

We note that scaling a conditional variable, helps to reduce the dependence structure in the data. This is clearly evident from figures ((13),(14) and (15)), where the maximum autocorrelation for the first five lags[39] are plotted against the increasing levels of θ corresponding to the threshold estimate $\widehat{\mu}_\theta\Big(\mathbf{X}_t\Big)$. In all the plots, the continuous line represents the 95% confidence level of the autocorrelation, computed as $\frac{\Phi^{-1}(0.95)}{\sqrt{N_\theta}}$, of the events in excess of the threshold. The dotted line represents the maximum autocorrelation from the unscaled residuals and the thick dotted, the autocorrelation from the scaled residuals in (4.4.0.25). In all cases, we observe that as the threshold increases, the autocorrelations for the scaled excesses become statistically insignificant. It should be noted that that because of sparseness of the extremes, the level θ for the scale function should neither be too high nor too low.

Let $f_{\mathbf{X}_t}\Big(z\Big)$ be the conditional density of $\widehat{Z}_t$ on $\mathbf{X}_t$. The following assumption is imposed on the QAR adjusted-scaled residuals, $\widehat{Z}_t$.

[38] Which involves historical simulation (HS) in finding the threshold.

[39] For various excesses obtained on returns from Commerzbank, Deutsche Bank and DAX30 over three year periods.

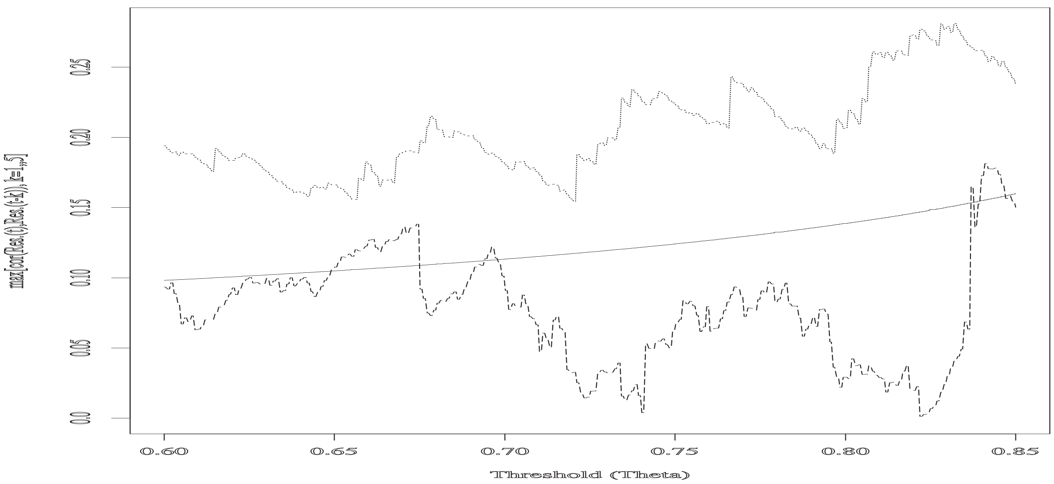

Figure 13: Commerzbank: Maximum autocorrelation for the first 5 lags against the increasing threshold(theta). The upper curve was obtained from the unscaled QAR adjusted negative returns. The lower was obtained from the QAR adjusted-scaled returns.

Conditions 4.4.1 *Let* $\widehat{Z}_t = \frac{Y_t - \widehat{\mu}_\theta(\mathbf{X}_t)}{\widehat{\sigma}_\theta(\mathbf{X}_t)}$ *be the sample residuals approximating the* Z_t. *We assume that, at least, to a good approximation the conditional density function* $g_{\mathbf{x}_i}\Big(z\Big)$ *of the excesses of* $\widehat{Z}_t$ *over the threshold* q^z_θ *given* $\mathbf{X}_t = \mathbf{x}_i$ *is such that*

$$g_{\mathbf{x}_i}\Big(z\Big) = g\Big(z\Big), \quad \forall \mathbf{x}_i \quad and \quad G\Big(z\Big) \in MDA\Big(H_\xi, \xi > 0\Big). \tag{4.4.0.26}$$

The condition states that the excess conditional distribution of the QAR adjusted-scaled residuals is heavy tailed and independent of the covariate beyond the threshold at high probability level, (i.e from the definition of Z_t, they are iid). Our main interest is now to find the distribution function of the data well above the threshold $u = q^z_\theta = 0$, whose inverse gives the QAR-scaled extreme quantile. From (4.3.0.18), the implicit form of this distribution can be written as

$$F\Big(q^z_\theta + z\Big) = F\Big(q^z_\theta\Big) + \Big(1 - F\Big(q^z_\theta\Big)\Big) F_{q^z_\theta}\Big(z\Big) \tag{4.4.0.27}$$

whose estimate we denote as

$$\widehat{F}\Big(\widehat{q}^z_\theta + z\Big) = \widehat{F}\Big(\widehat{q}^z_\theta\Big) + \Big(1 - \widehat{F}\Big(\widehat{q}^z_\theta\Big)\Big) \widehat{F}_{\widehat{q}^z_\theta}\Big(z\Big).$$

From chapter (2), see also Cai, Z and Roussas [20], we have

$$\widehat{F}\Big(\widehat{q}^z_\theta\Big) = F\Big(q^z_\theta\Big) = \theta \tag{4.4.0.28}$$

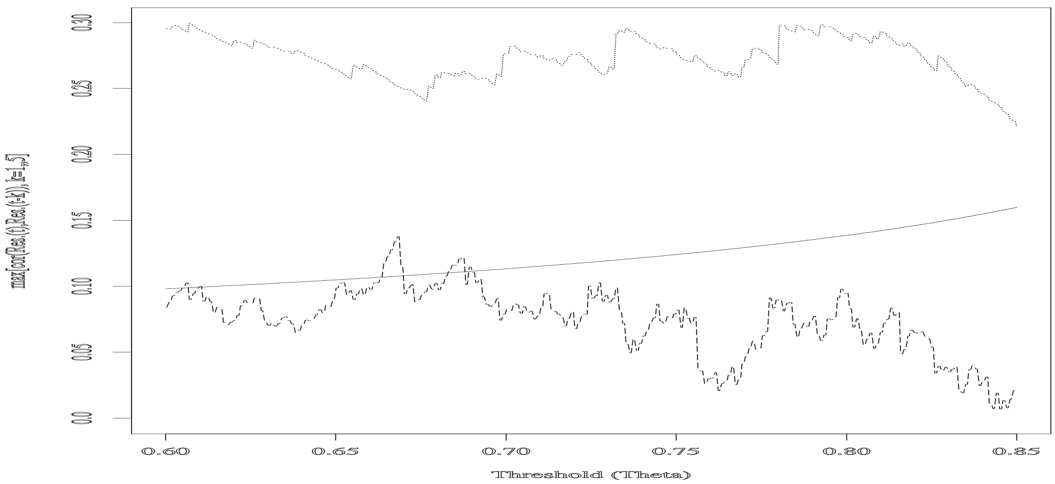

Figure 14: Deutsche Bank: Maximum autocorrelation for the first 5 lags against the increasing threshold(theta). The upper curve was obtained from the unscaled QAR adjusted negative returns. The lower was obtained from the QAR adjusted-scaled returns.

Therefore $\widehat{F}\Big(\widehat{q}^z_\theta + z\Big) = \theta + \Big(1-\theta\Big)\widehat{F}_{\widehat{q}^z_\theta}\Big(z\Big)$. Since $\mu_{t,\theta}$ and $\sigma_{t,\theta}$ have already been consistently estimated in chapters (2) and (3), we assume they are known. This simplifes our tail estimator to

$$\widehat{F}\Big(q^z_\theta + z\Big) = \theta + \Big(1-\theta\Big)\widehat{F}_{q^z_\theta}\Big(z\Big) \Leftrightarrow \widehat{F}\Big(z\Big) = \theta + \Big(1-\theta\Big)\widehat{F}_0\Big(z\Big) \qquad (4.4.0.29)$$

where $z > 0$ and $q^z_\theta = 0$ by the definition of our model. The following lemma shows that $\widehat{F}\Big(z\Big)$ is asymptotically a generalized Pareto distribution function estimator.

Lemma 4.2 (Tail distribution)

Let $Z_t, t = 1, \ldots, n$ be independent random variables with zero θ-quantile, i.e $q^z_\theta = 0$, and let their excess distribution $F_0\Big(z\Big)$ above q^z_θ be a GPD with parameters ξ and β. Then,

$$F\Big(z\Big) = \theta + \Big(1-\theta\Big)F_0\Big(z\Big) = G_{\xi,\nu,\widetilde{\beta}}\Big(z\Big)$$

is again a GPD with the same shape parameter ξ, scale parameter $\widetilde{\beta} = \beta\Big(1-\theta\Big)^{\xi}$ and location parameter $\nu = \frac{\beta}{\xi}\Big(\Big(1-\theta\Big)^{\xi} - 1\Big) = \frac{\widetilde{\beta}}{\xi}\Big(1 - \Big(1-\theta\Big)^{-\xi}\Big)$.

Proof of lemma 4.2

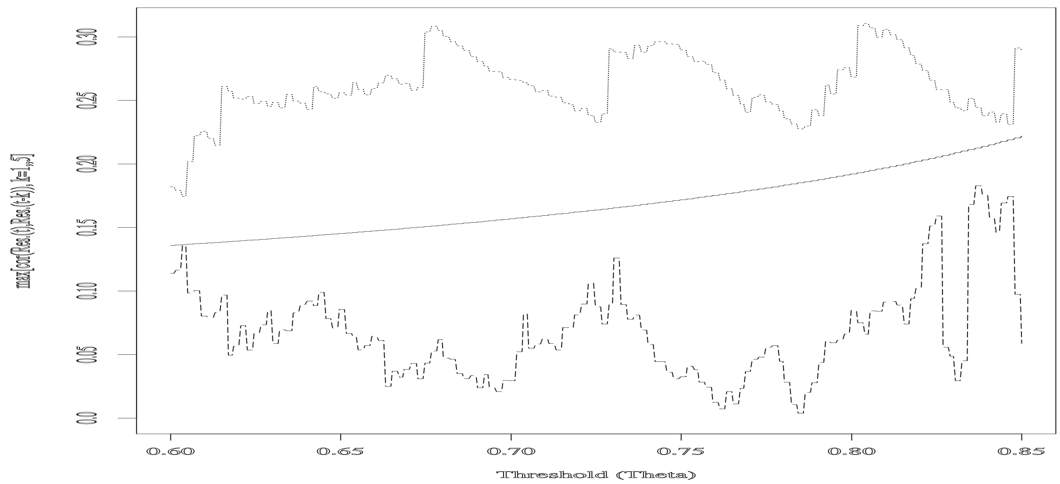

Figure 15: DAX30: Maximum autocorrelation for the first 5 lags against the increasing threshold(theta). The upper curve was obtained from the unscaled QAR adjusted negative returns. The lower was obtained from the QAR adjusted-scaled returns.

As $F_0\Big(z\Big) = G_{\xi,0,\beta}\Big(z\Big)$,

$$\begin{aligned}
F\Big(z\Big) &= \theta + (1-\theta)G_{\xi,0,\beta}\Big(z\Big) \\
&= 1 - \Big(1-\theta\Big)\Big[1 + \frac{\xi}{\beta}z\Big]^{-\frac{1}{\xi}} \\
&= 1 - \Big[\Big(\frac{1}{1-\theta}\Big)^{\xi} + \frac{\xi}{\beta(1-\theta)^{\xi}}z\Big]^{-\frac{1}{\xi}} \\
&= 1 - \Big[\frac{\xi}{\beta(1-\theta)^{\xi}}\Big(z + \frac{\beta}{\xi}\Big)\Big]^{-\frac{1}{\xi}} \\
&= 1 - \Big[1 + \frac{\xi}{\beta(1-\theta)^{\xi}}\Big(z + \frac{\beta}{\xi}\Big) - 1\Big]^{-\frac{1}{\xi}} \\
&= 1 - \Big[1 + \frac{\xi}{\beta(1-\theta)^{\xi}}\Big(z + \frac{\beta}{\xi} - \frac{\beta(1-\theta)^{\xi}}{\xi}\Big)\Big]^{-\frac{1}{\xi}} \\
&= G_{\xi,\nu,\widetilde{\beta}}\Big(z\Big) \qquad (4.4.0.30)
\end{aligned}$$

where $\widetilde{\beta} = \beta(1-\theta)^{\xi}$ is the scale parameter and $\nu = \frac{\widetilde{\beta}}{\xi}\Big((1-\theta)^{\xi}-1\Big)$, the location parameter.

□

If our original data Y_t follow a QAR-model $Y_t = \mu_{t,\theta} + \sigma_{t,\theta}Z_t$ with innovation Z_t having zero θ-quantile, then $\mu_{t,\theta}$ is the QAR of Y_t at θ. If we choose the threshold $u = \mu_{t,\theta}$, the excess distribution function of Y_t is

$$\begin{aligned} F_u^Y\Big(y\Big) &= P\Big(Y_t \le u \Big| Y_t > u\Big) \\ &= P\Big(Z_t \le \frac{y}{\sigma_{t,\theta}} \Big| Z_t > 0\Big) = F_0\Big(\frac{y}{\sigma_{t,\theta}}\Big). \end{aligned} \tag{4.4.0.31}$$

Heuristically, for $\theta \to 1$, we have $\mu_{t,\theta} = u \to \infty$, and by theorem 4.2, we can expect F_u^Y and, then, F_0 to be well approximated by a GPD. By lemma 4.2, this will also hold for F.

Let $q_{\theta,\varphi}$ be the quantile above a threshold $q_\theta^z = 0$ based on Z_t and derived by inverting the distribution $F\Big(z\Big)$ at a particular level of $\varphi > \theta$. That is for a fixed $\varphi \in (0, 1)$,

$$\begin{aligned} q_\varphi^z &= \inf_{z>0}\Big\{F\Big(z\Big) \ge \varphi\Big\} \\ &= \inf_{z>0}\Big\{1 - \Big(1-\theta\Big)\Big(1 - F_0\Big(z\Big)\Big) \ge \varphi\Big\}, \quad \text{from (4.4.0.27)} \\ &\approx \sup_{z>0}\Big\{\overline{G}_{\xi,\beta(\theta)}\Big(z\Big) \le \frac{1-\varphi}{1-\theta}\Big\}, \quad \text{for} \quad \theta \to 1 \quad \text{and} \quad \overline{G} = 1 - G \\ &= \overline{G}^{-1}_{\xi,\beta(\theta)}\Big(\frac{1-\varphi}{1-\theta}\Big) \\ &= \frac{\beta(\theta)}{\xi}\Big(\Big(\frac{1-\varphi}{1-\theta}\Big)^{-\xi} - 1\Big) \end{aligned} \tag{4.4.0.32}$$

Compare (4.2.0.12). We denote its estimate by $\widehat{q}_\varphi^z = \inf_{z>0}\Big\{\widehat{F}\Big(q_\theta^e + z\Big) \ge \varphi\Big\} = \frac{\widehat{\beta}\big(\theta\big)}{\widehat{\xi}}\Big(\Big(\frac{1-\varphi}{1-\theta}\Big)^{-\widehat{\xi}} - 1\Big)$. Intuitively, $\widehat{q}_\varphi^z$ will be a consistent estimate of q_φ^z. Since $\sqrt{N_\theta}\Big(\widehat{\xi} - \xi, \frac{\widehat{\beta}(\theta)}{\beta(\theta)} - 1\Big)$ is consistent and asymptotically normal with zero mean and covariance given in lemma 4.1, it follows that $\Big(\widehat{\xi} - \xi, \widehat{\beta}(\theta) - \beta(\theta)\Big) \to^p \Big(0, 0\Big)$ as $\theta \to 1$ and $N_\theta \to \infty$. Then by by corollary 6.3.14 (iv) in Dudewicz and Mishra [35], page 323, $\frac{\widehat{\beta}(\theta)}{\widehat{\xi}}\Big(\Big(\frac{1-\varphi}{1-\theta}\Big)^{-\widehat{\xi}} - 1\Big)$ estimates $\overline{G}^{-1}_{\xi,\beta(\theta)}\Big(\frac{1-\varphi}{1-\theta}\Big)$ consistently. The latter coincides approximately with q_φ^z by the heuristic arguments we have given above for $F\Big(z\Big)$ being approximately a GPD. For an exact proof, however, we would need a version of theorem 4.2 for dependent data, i.e. in particular for our QAR-process Y_t. As the extreme value theory for financial time series models is still in its infancy, such a result is beyond the scope of this thesis.

For random variables $Y_t, t = 1, 2, \ldots$, generated by the process (1.4.0.2) with $Z_t, t = 1, \ldots$, being iid with $F \in MDA\Big(H_\xi, \xi > 0\Big)$, the conditional QAR of Y_t at φ is given by $\mu_{\theta,\varphi} = \mu_\theta\Big(\mathbf{X}_t\Big) + \sigma_\theta\Big(\mathbf{X}_t\Big)q_\varphi^z$. In the following section, it is argued heuristically that the estimator for $\mu_{\theta,\varphi}\Big(\mathbf{X}_t\Big)$ is consistent.

4.4.1 Consistency of the extreme QAR function estimator

In the QAR model the conditional quantile estimate of Y_t given $\mathbf{X}_t$, $\widehat{\mu}_\theta\Big(\mathbf{X}_t\Big)$, becomes the initial estimate as well as the conditional threshold. The GPD is parametrically fitted to the excesses over $\widehat{\mu}_\theta\Big(\mathbf{X}_t\Big)$. The overall estimator at point, $\mathbf{X}_t = \mathbf{x}_i$, then becomes

$$\widehat{\mu}_{\theta,\varphi}\Big(\mathbf{x}_i\Big) = \widehat{\mu}_\theta\Big(\mathbf{x}_i\Big) + \widehat{\sigma}_\theta\Big(\mathbf{x}_i\Big)\widehat{q}^z_\varphi, \quad i = 1, \ldots, n, \tag{4.4.1.1}$$

where $\widehat{q}^z_\varphi$ is estimated from the residuals $\widehat{Z}_t = \frac{Y_t - \widehat{\mu}_\theta(\mathbf{X}_t)}{\widehat{\sigma}_\theta(\mathbf{X}_t)}$ by fitting a GPD as described in the previous section. The intuition for the estimate $\widehat{\mu}_{\theta,\varphi}\Big(\mathbf{x}_i\Big)$ which combines a nonparametric quantile estimate with a parametric fit in the extreme tail using theorem 4.2 is quite similar to the VaR-estimates of McNeil and Frey (2000) based on the POT-model.

The following heuristic argument shows that we can expect

$$\widehat{\mu}_{\theta,\varphi}\Big(\mathbf{x}_i\Big) - \mu_{\theta,\varphi}\Big(\mathbf{x}_i\Big) \to^p 0 \tag{4.4.1.2}$$

for $n \to \infty, \quad N_\theta \to \infty, \theta \to 1$ and $\theta < \varphi < 1$.

The left hand of (4.4.1.2) can be expressed as

$$\widehat{\mu}_{\theta,\varphi}\Big(\mathbf{x}_i\Big) - \mu_{\theta,\varphi}\Big(\mathbf{x}_i\Big) = \widehat{\mu}_\theta\Big(\mathbf{x}_i\Big) - \mu_\theta\Big(\mathbf{x}_i\Big) + \widehat{\sigma}_\theta\Big(\mathbf{x}_i\Big)\widehat{q}^z_\varphi - \sigma_\theta\Big(\mathbf{x}_i\Big)q^z_\varphi \tag{4.4.1.3}$$

Under conditions (B1)-(B6), (C1)-(C6), (D1), (E1) and for $n \to \infty$, we have $\widehat{\mu}_\theta\Big(\mathbf{x}_i\Big) - \mu_\theta\Big(\mathbf{x}_i\Big) \to^p 0$, and $\widehat{\sigma}_\theta\Big(\mathbf{x}_i\Big) - \sigma_\theta\Big(\mathbf{x}_i\Big) \to^p 0$, see chapters (2) and (3). For $F \in MDA\Big(H_\xi, \xi > 0\Big)$ and $N_\theta \to \infty, \quad \theta \to 1$, we have heuristically $\widehat{q}^z_\varphi \to^p q^z_\varphi$ by the argument in previous section. By using corollary 6.3.14(iii), page 323 in Dudewicz and Mishra [35]), we get $\widehat{\sigma}_\theta\Big(\mathbf{x}_i\Big)\widehat{q}^z_\varphi - \sigma_\theta\Big(\mathbf{x}_i\Big)q^z_\varphi \to^p 0, \quad \text{as} \quad \theta \to 1, \quad n \to \infty$. Finally, the application of corollary 6.3.14(i), Dudewicz and Mishra [35], on $\widehat{\mu}_{\theta,\varphi}\Big(\mathbf{x}_i\Big)$ completes the argument. For exact proof, again we would need theorem 4.2 for QAR-processes.

4.4.2 Estimation strategy

In all we adopt the following strategy in the estimation of the extreme QAR function:

Assume $Y_t, \quad t = 1, 2, \ldots, n + d$ are generated by the QAR-QARCH process.

(1) For m equally spaced $\theta \in \Big(0.55, 0.9\Big)$ and $j = 1, 2, \ldots, m$, use the QAR-QARCH model to estimate the conditional threshold, $\mu_{\theta_j}\Big(\mathbf{X}_t\Big)$, and scale function, $\sigma_{\theta_j}\Big(\mathbf{X}_t\Big)$ at

$\mathbf{x}_i, i = 1, \ldots, n$. Denote the estimates at $\mathbf{X}_t = \mathbf{x}_i$ as $\widehat{\mu}_{\theta_j}\left(\mathbf{x}_i\right)$ and $\widehat{\sigma}_{\theta_j}\left(\mathbf{x}_i\right)$. Adjust Y_t of $\widehat{\mu}_{\theta_j}\left(\mathbf{x}_i\right)$ and scale the resulting residuals to obtain a series of scaled excess quantile residuals over the scaled threshold; $\frac{Y_t - \widehat{\mu}_{\theta_j}(\mathbf{x}_i)}{\widehat{\sigma}_{\theta_j}(\mathbf{x}_i)} > 0, \quad i = 1, \ldots, n$ for every $j = 1, \ldots, m$.

(2) For equally spaced $\varphi \in [0.95, 1]$ fit the GPD on the excesses in (1) and graph the set of fitted quantiles for the respective thresholds.

(3) Make visual judgement to see whether the required level of θ (in this case φ) falls in an appropriate region (the peak area). Proceed to step (4) if the finding is positive, otherwise, use other appropriate method. As an illustration of this step, we generated a series of iid random variables, e_t of size 1000 from a t-distribution with 3 degrees of freedom. We then fitted a set of quantiles on the excesses over the quantile adjusted random variable e_t. The surface of the estimated quantiles is shown in figure (16). Clearly for low quantiles (corresponding to (vartheta)), the GPD underestimates as the threshold increases. For higher quantiles the GPD gives the peak quantiles as the threshold increases. Let us note such visual observation helps in deciding whether the estimation of the required quantile corresponding to a probability level needs a combination of EVT or not. If one is interested in the quantile at $\varphi = 0.98$, say, the GPD and high threshold would underestimate the quantile. On the other hand, taking a high threshold would deliver the desired quantile at high levels above 0.990, because the peak over threshold (POT) quantiles are generated. That is high threshold delivers the peak of the required quantile in heavy tailed distribution.

(4) Choose a final high level threshold $\widehat{\mu}_\theta\left(\mathbf{x}_i\right)$, by setting $\theta \in \left(0.85, 0.92\right)$, depending on the size of the series or by using the mean excess graph to select the θ where the curve appears to strech out linearly.

(5) Fit the GPD over the excesses and extract the required quantile.

Step (5) completes the main computation required. To get back the extreme QAR, rearrange $\frac{\widehat{\mu}_{\theta,\varphi}\left(\mathbf{x}_i\right) - \widehat{\mu}_\theta\left(\mathbf{x}_i\right)}{\widehat{\sigma}_\theta\left(\mathbf{x}_i\right)} = \widehat{q}^z_\varphi$.

If the underlying process is assumed to be an AR-ARCH, with the conditional mean known (or equal to zero), step (1) of the strategy can be omitted. In this case, the simplest scale function is $\widehat{\mu}_\theta\left(\mathbf{x}_i\right)$ for $\theta > 0.5$ in step (2). The scaled excesses become

$$\frac{Y_t}{\widehat{\mu}_\theta\left(\mathbf{x}_i\right)} - 1 > 0, \quad \theta > 0.5. \tag{4.4.2.1}$$

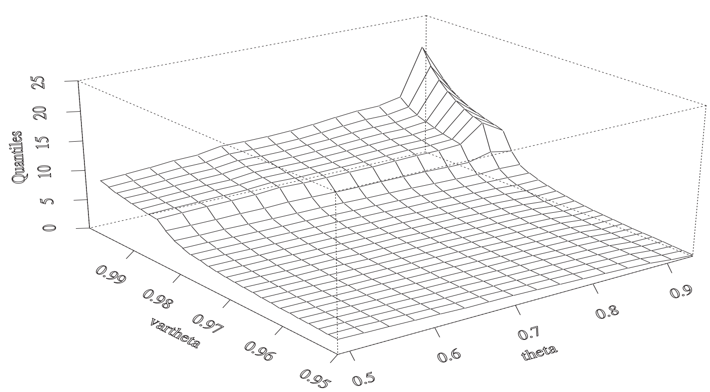

Figure 16: Surface plot of the fitted φ-quantile on the excesses over threshold θ(corresponding to q^e_θ)

We applied the above strategy in the estimation of the extreme QAR at $\varphi = 0.95$ and 0.99 from negative returns on BASF, see figure (12). The normal quantile-quantile plot in figure (17) shows the returns have heavier tail than normal distribution. In order to determine the shape parameter, ξ, we fitted the GPD on the excesses over the quantile residuals corresponding to 100 equally spaced $\theta \in (0.45, 0.92)$ and plotted the estimated shape against the level $(1-\theta)$ shown in figure (18). From the graph we chose $\widehat{\xi} = 0.25$ which correspond to the fairly stable areas (between $1-\theta = 0.10$ and 0.3). This area is also supported by the plot of mean excess function (MEF) in figure (19) which indicate a linear stretch in areas beginning $\theta = 0.65$ (given on the third axis) corresponding to the threshold value of about 0.015, on the first axis. The estimated tail distribution is given in figure (20), where the dots represent the empirical estimates. Both of the estimated tails (blue and green) are heavier than the empirical distribution in $1-F \in (0.019, 0.15)$. Below 0.019, the estimate obtained by using the threshold at $\theta = 0.85$, provides a heavier tail then the other one at $\theta = 0.6$. As compared to the Hill's estimate of the tail distribution (see figure (11)), the GPD appears to be capable of capturing large values at relatively high levels of φ as opposed to the Hill's estimator which produces tails thinner then the empirical distribution at $\theta = 0.85$ for relatively high quantiles. The estimates of the conditional quantile functions at $\varphi = 0.95$ using direct conditional quantile regression (QAR) method, QAR augmented with GPD based on unscaled QAR adjusted residuals (QAR+GPD) and QAR augmented with GPD based on QAR adjusted-scaled residuals (QAR+sc.GPD) are respectively shown in figure (21) as blue, green and red curves. The threshold was taken to be at $\theta = 0.85$. The QAR estimates for all the three approaches do not seem to be quite different from each other, with the exception of the QAR+GPD which appears to underestimate in cases of high volatlities. We again applied the three methods in the estimation of the extreme QAR, on the same data, at $\varphi = 0.99$. The result is shown in figure (22), where the blue, green and red curves represent estimates obtained by QAR, QAR+GPD and QAR+sc.GPD respectively. Clearly the estimates are different. The QAR (so is QAR+GPD), do not appear to exhibit the characteristics of the underlying volatility. On the other hand, the QAR+sc.GPD appears to adjust quite well according to the underlying volatility. Lastly, we superimposed the estimates of extreme QAR at $\varphi = 0.99$ in figure (23), obtained by using the QAR augmented with the Hill estimator (QAR+sc.Hill) and QAR+sc.GPD, represented in green and blue curves respectively. Both estimates appear to adjust according to the underlying volatility. However, there is marked diffference in the estimates where the volatility is high. The above discussion repesents only the beginning of a detailed numerical study of the performance of the introduced model and its variants in section (4.6)

4.5 Threshold problem

An important issue in the estimation of extreme quantile by EVT, is the choice of an appropriate threshold, $\mu_\theta\Big(\mathbf{X}_t\Big)$, that determines the number of order statistics to be used in the estimation procedure. If the threshold is too high, there are too few exceedances resulting in a high variable estimator. On the other hand, a low threshold produces a

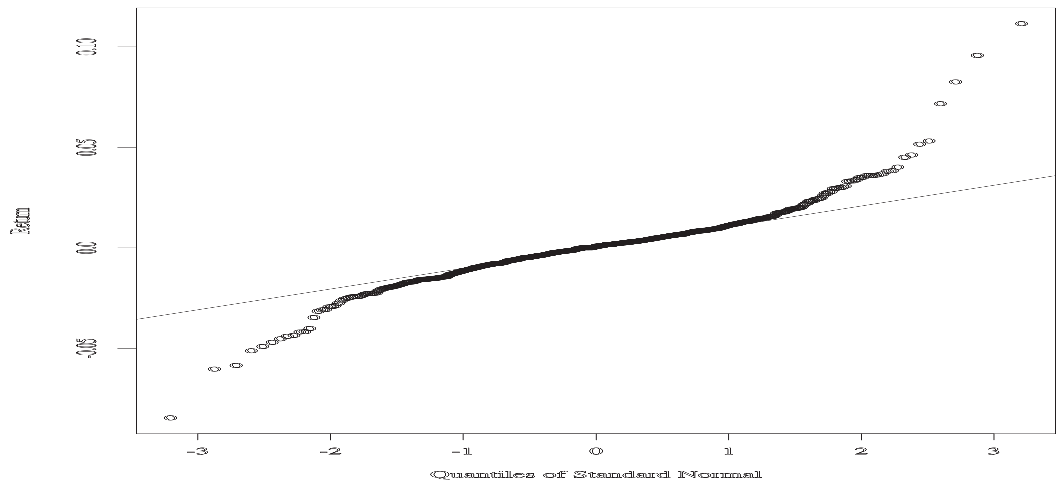

Figure 17: Returns on BASF: Quantile-quantile plot of residuals against normal distribution shows the quantile residuals to be leptokurtic

<u>*biased*</u> estimator, because the asymptotic approximation becomes very poor. This was cleary seen in figure (11), in the estimation of the tail distribution through the Hill's estimate of shape parameter; as the threshold increased the tail estimator become thicker and better for heavy tailed. A small monte-Carlo study was performed to illustrate the sensitivity of the estimator to the threshold. We generated 500 samples of size 1000 using student t-distribution with 4 degrees of freedoms. Then we estimated $95\%, 99\%$, and 99.5% quantiles using the GPD formula. The threshold values θ corresponding to q^e_θ were chosen by decreasing the proportion[40] $(1-\theta)$ of the sample exceeded from $\theta = 0.6$ to 0.95 using 200 equal intervals. The bias and variance were computed, respectively, as

$$Bias\Big(\widehat{q}_{\theta_k,\varphi}\Big) = \sum_{j=1}^{500} \frac{\widehat{q}^{(j)}_{\theta_k,\varphi}}{500} - q_\varphi\Big), \quad k = 1, \ldots, 200.$$

and

$$var\Big(\widehat{q}_{\theta_k,\varphi}\Big) = \frac{1}{500}\sum_{j=1}^{500}\Big(\widehat{q}^{(j)}_{\theta_k,\varphi} - q_\varphi\Big)^2 - Bias^2\Big(\widehat{q}_{\theta_k,\varphi}, \qquad k = 1, \ldots, 200.$$

where q_φ is the true quantile. The results were the plotted in figures (24), (25) and (26). Clearly, there seem to be a compromise between the variance and the bias, which

[40]That is by reducing the number of exceedances $n(1-\theta)$

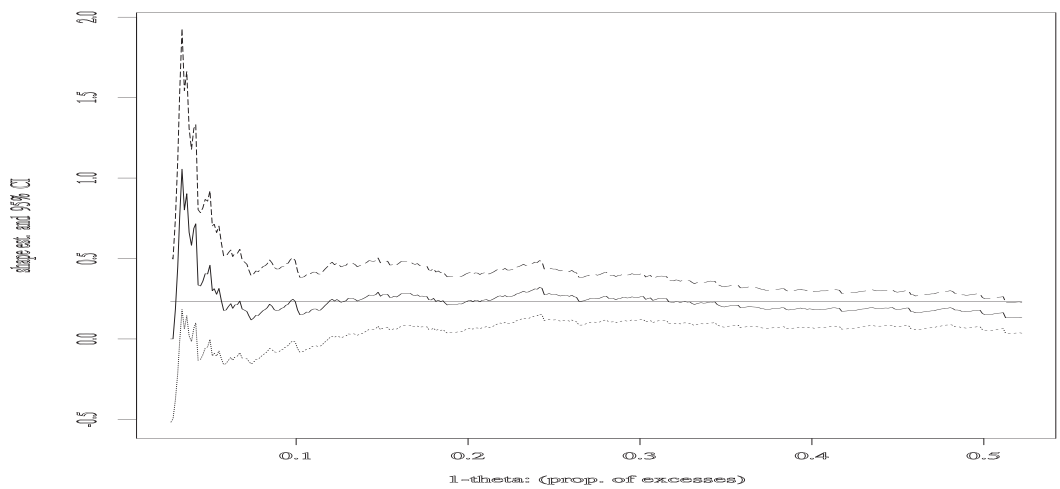

Figure 18: The shape estimate against increasing threhold $(1-\theta)$, where θ varies from high levels to low. The shape was taken as 0.25

theoretically could be obtained by minimizing the mean squared error. But because quantiles are not observable, in practice, basing the selection on minimum MSE would result in a biased estimator. Some recent statistical developments in threshold selection have been oberved in Danielsson and de Vries [31], where a two step bootstrap method to select the sample fraction on which Hill estimator is based, is proposed. To current, threshold selection when using the GPD has remained a graphical solution, see for example figure (19) and Embrechts et al. [39] for more details. Moreover there is a consensus in literature[41] that taking a very high threshold suffices. This is in line with the fact that the GPD quantile estimation are usually repeated several times to have a graph of quantiles depicting the quantile of stable areas and the POT quantiles, see again figure (16). Seemingly, the latter one is more important as in the estimation of risk, one better *overestimate* than *underestimate*.

[41]See for example McNeil [87].

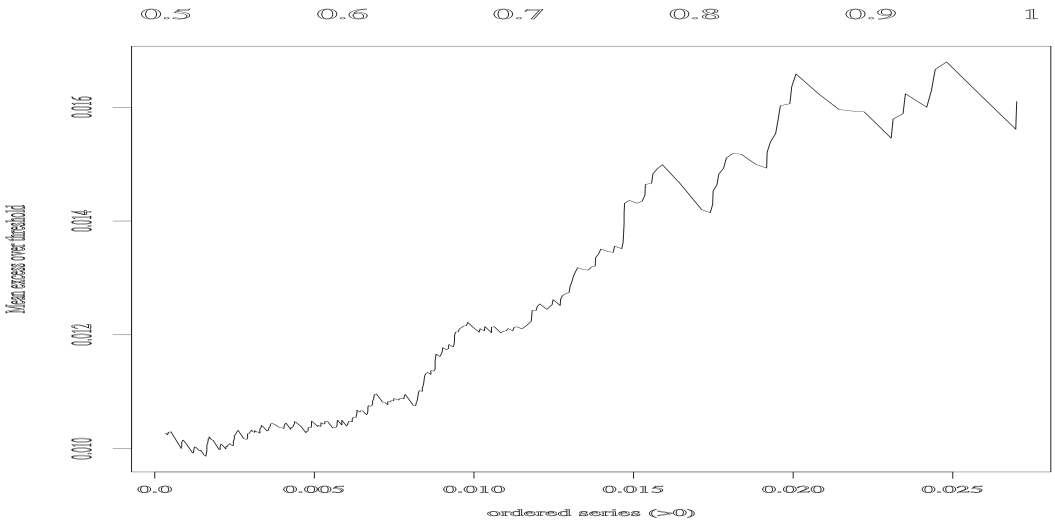

Figure 19: Returns on BASF: Plot of mean excess function against the ordered quantile ajusted-scaled excesses. The third axis indicate the increasing proportion of the ordered excesses.

4.6 The performance of extreme QAR models

In this section we evaluate the performance of various quantile models in the estimation of extreme QAR using artificial[42] and real data. In particular, we will estimate the extreme QAR using Historical simulation (HS), direct estimation by generalized Pareto distribution (GPD), direct estimation through the Hill's estimator (Hill), direct QAR-ARCH model (QAR), the QAR augmented with GPD on unscaled residuals (QAR+GPD), the QAR augmented with Hill estimator on unscaled residuals (QAR+Hill), the QAR augmented with GPD on scaled residuals (QAR+sc.GPD) and the QAR augmented with (Hill) on scaled residuals (QAR+sc.Hill). The QAR+GPD and QAR+Hill can be thought as a combination of GPD, respectively Hill, with the usual quantile regression model where the unscaled quantile residuals residuals are assumed to be iid. The QAR+sc.GPD and QAR+sc.Hill are a combination of GPD (respectively Hill) with quantile regression model where the unscaled quantile residuals are not assumed independent.

[42] The advantage of working with artificial data is that the true (extreme) QAR is known. This makes it possible to quantify the errors associated with a particular method for measuring extreme QAR.

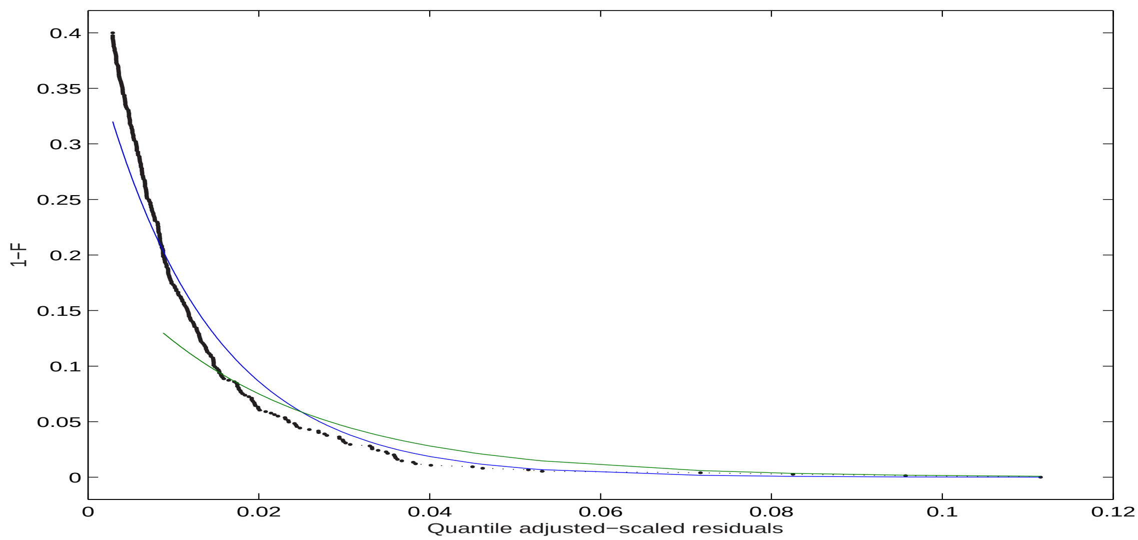

Figure 20: Reurn on BASF: Estimates of the tail distribution. Dot represent the empirical. Green and blue are estimates obtained by setting the threshold to $\theta = 0.6$ and 0.85, respectively.

4.6.1 A monte Carlo study

We perform a Monte Carlo study to compare our estimates with various direct estimators and their combination. We generated 500 samples of size 1280 observations for four different threshold processes, AR(1)-TARCH(1),

$$Y_t = 0.5 + 0.3Y_{t-1} + \sqrt{0.01 + 0.1Y_{t-1}^2 + 0.35\Big(\frac{|Y_{t-1}| - Y_{t-1}}{2}\Big)^2} e_t, \quad t = 2,,, \qquad (4.6.1.1)$$

where e_t are zero mean-unit variance iid errors. The errors were generated using random number generator with the following distributions: standard normal, student[43]-t with 3 & 4 degrees of freedom and Gamma[44] with (2,2) degrees of freedom. Considering only the student-t distributed error with 4 d.f, an example of the scale function estimate is given in figure (27), where the dotted represents an estimate of the true scale function, solid, at $\theta = 0.90$.

The performance of the models were then evaluated using the mean average squared

[43] To make the errors iid with zero mean-unit variance, we scaled the t-ditributed errors by $\sqrt{\frac{v}{v-2}}$, where v is the degrees of freedom.

[44] The mean, ab, was first subtracted from the errors and the result divided by the standard deviation $\sqrt{ab^2}$.

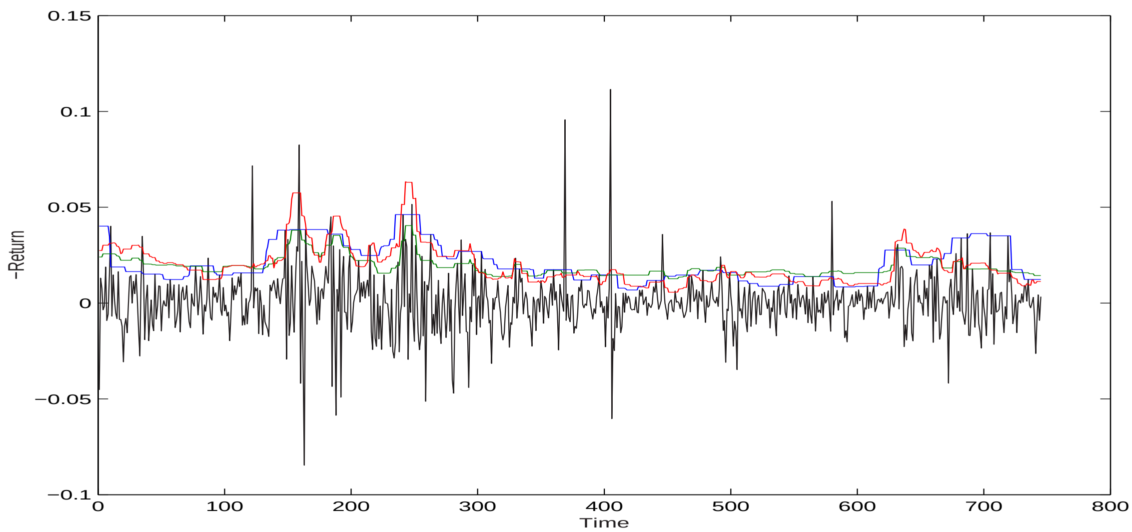

Figure 21: Returns on BASF: Plot of negative returns against time. Superimposed are the estmated conditional quantile at $\varphi = 0.95$

error (MASE):

$$MASE\Big(\widehat{\mu}_{\theta,\varphi}\Big(X_t\Big)\Big) = \frac{1}{500}\sum_{j=1}^{500}\Big[\frac{1}{1000}\Big|\Big|\widehat{\mu}^{(j)}_{\theta,\varphi}\Big(X_t\Big) - \mu^{(j)}_{\theta,\varphi}\Big(X_t\Big)\Big|\Big|^2\Big] \tag{4.6.1.2}$$

where $||.||$ denote the Euclidean norm, $\widehat{\mu}^{(j)}_{\theta,\varphi}\Big(X_t\Big)$ and $\mu^{(j)}_{\theta,\varphi}\Big(X_t\Big)$ are (1000×1) vectors of estimated and true functions of extreme QAR at φ respectively, for the j^{th} sample. We used 280 less observations in all the models, except in HS where 281 were used in the rolling window. The results are shown in table (6). The GPD, Hill and HS did worse than the QAR and the combination at the 95% level. This is partly due to poor asymptotic approximation at $\theta = 0.95$ and their underlying assumptions do not conform with the simulated process. Under normal errors, the GPD appears to perform better than HS, although both assumes iid. The stars indicate no evaluation was performed as the estimator is only suitable for heavy tailed distributions. As the level of φ increases to 0.995, the MASE for both GPD and Hill tend to decrease , under nonnormal errors. Calculating the ratios of the MASE for the direct methods;

$$\frac{MASE\Big(Hill\Big)}{MASE\Big(QAR\Big)} \quad \text{and} \quad \frac{MASE\Big(GPD\Big)}{MASE\Big(QAR\Big)}$$

at $\varphi = 0.95, 0.99, 0.995$ and for all errors, we observe the ratio goes to below 1 from above

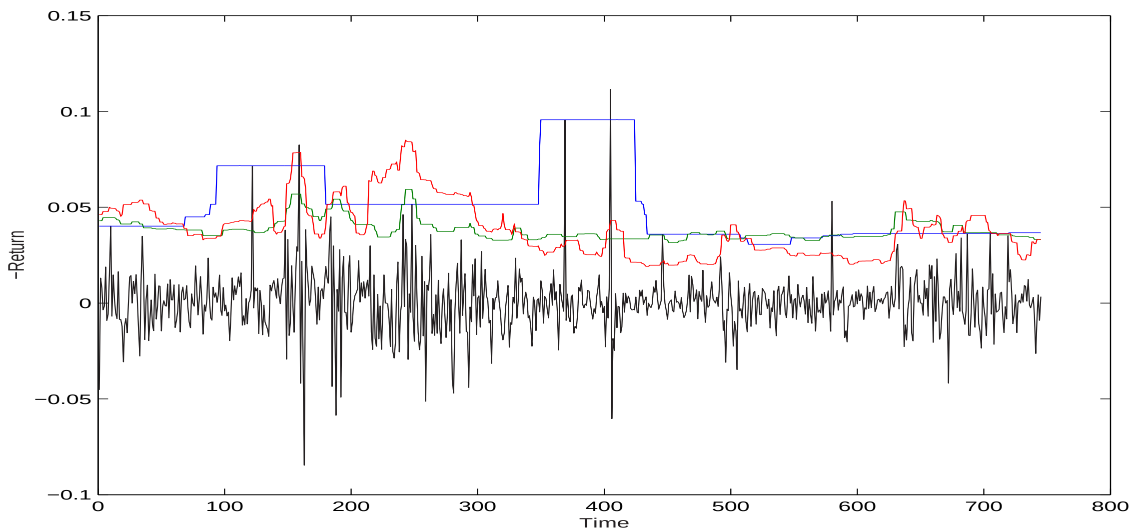

Figure 22: Returns on BASF: Plot of the negative returns against time. Superimposed are conditional 0.99-quantile estimates

1. This indicates clearly that the rate at which the MASE increase with the increasing φ is slower[45] for the estimates obtained by direct application of Hill and GPD than for the direct QAR. Thus the estimates from Hill and GPD tend to be better than QAR in the extremes than in the interior where the QAR is good. This support our idea of combining the direct quantile regression with extreme value theory for high levels. From the table, it is noted that all the QAR models and their combinations produces similar estimates at $\varphi = 0.95$, but increasingly differ as φ increases. This confirm the observations made in figures (21),(22) and (23). By introducing the scale, we observe that the $QAR + sc.GPD$ and $QAR + sc.Hill$ are clearly superior in terms of the efficiency gained, to the direct QAR and $QAR + GPD$. However, the estimates from QA+sc.GPD outperforms the corresponding ones from QAR+sc.Hill.

4.6.2 Backtesting

Many banks that use VaR[46] models routinely test the performance of the models by comparing the daily profit and losses with model-generated risk measures to guage the

[45] Given that at $\varphi = 0.95$, the MASE for the Hill and GPD were at least 4 times as large as the MASE for the QAR.

[46] See definition (1.2.1).

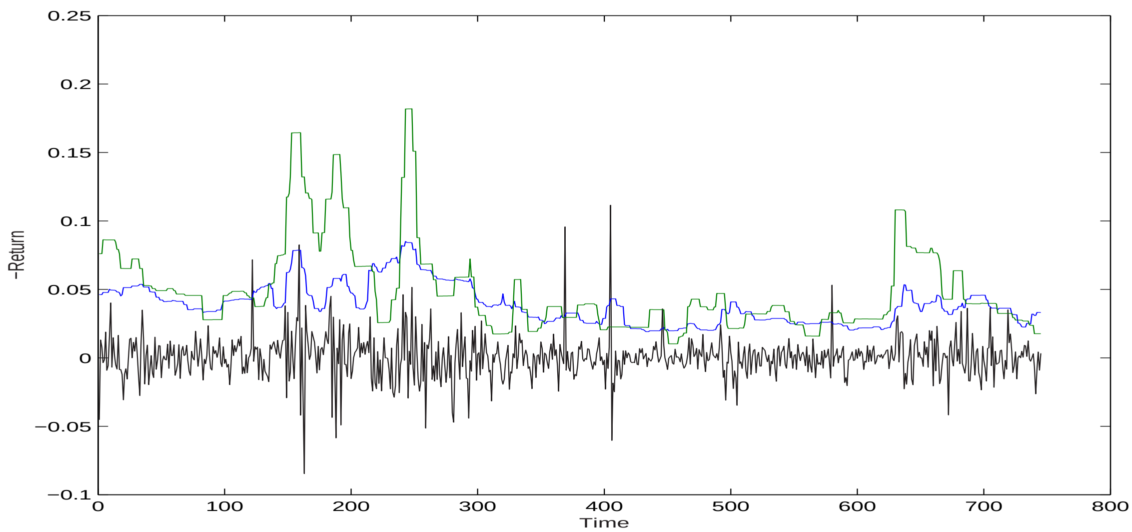

Figure 23: Returns on BASF: Plot of negative returns against time. Superimposed are the conditiona 0.99-quantile estimates obtained by QAR+sc.Hill(green) and QAR+sc.GPD (blue).

accuracy of their risk measurement systems. Such testing is refered to as *backtesting*. There are quite a number of techniques that test the performance of VaR models, see for example Kupiec [79] and Cassidy and Gizycki [23]. Kupiec [79] presents an approach to analyse exceptions[47] based on the observations that a comparison between daily profit and loss outcomes and the corresponding VaR gives rise to a binomial experiment. Under the assumption that the daily VaR measures are independent, the binomial outcomes represent a sequence of independent Bernoulli trials each with probability of failure equal to 1 minus the models specified level of confidence. For instance a 95% level gives a 5% as the probability of failure on each triall. Hence testing the accuracy of the model is equivalent to testing the null hypothesis that the probability of failure on each trials equals the model's specified probability. The test we consider is known as *Kupiec's POF-Test* which is based on the proportion of failures observed over the entire sample period. The null hypothesis test that the VaR model's stated level is equal to the realized probability level covered by the model ($H_o : \varphi = \widetilde{\varphi}$) is achieved by the Likelihood-Ratio-Tests (LR) statistics given by

[47]If the actual trading loss exceed the VaR estimate the result is recorded as a failure or *exception*.

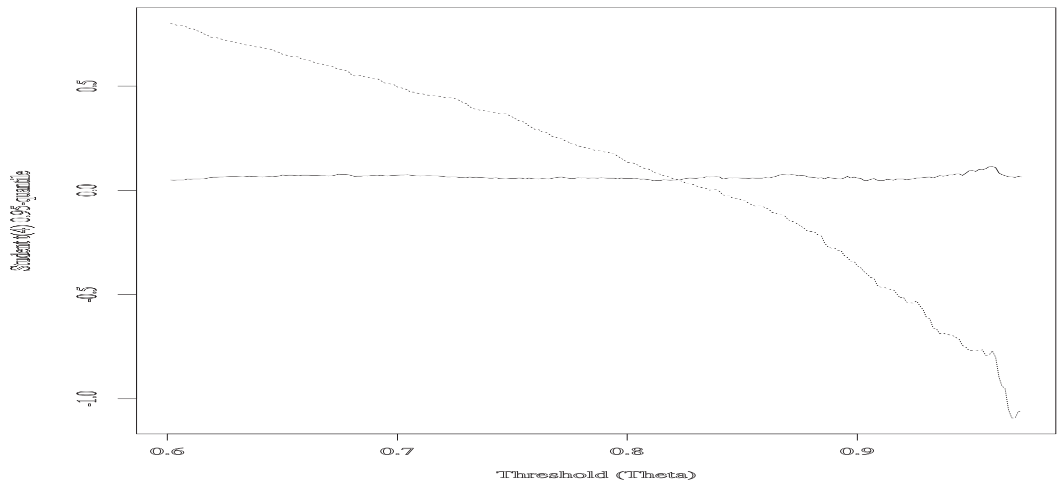

Figure 24: The variance (solid) and bias (dotted) for the estimate at $\varphi = 0.95$ against the threshold (theta)

Table 5: Model Verification: Nonrejection regions. Number of failures at 5% level

Probability level, φ	**T=250 days**	**T=500 days**	**T=750 days**	**T=1000 days**
0.0500	$7 \leq N \leq 19$	$17 \leq N \leq 35$	$27 \leq N \leq 49$	$38 \leq N \leq 64$
0.010	$1 \leq N \leq 6$	$2 \leq N \leq 9$	$3 \leq N \leq 13$	$5 \leq N \leq 16$
0.0050	$0 \leq N \leq 4$	$1 \leq N \leq 6$	$1 \leq N \leq 8$	$2 \leq N \leq 9$
0.0010	$0 \leq N \leq 1$	$0 \leq N \leq 2$	$0 \leq N \leq 3$	$0 \leq N \leq 3$
0.0001	$0 \leq N \leq 0$	$0 \leq N \leq 0$	$0 \leq N \leq 1$	$0 \leq N \leq 1$

$$LR = -2ln\left(\frac{\varphi^N\left(1-\varphi\right)^{T-N}}{\widetilde{\varphi}^N\left(1-\widetilde{\varphi}\right)^{T-N}}\right)$$

where T represents the number of backtesting points, N denotes a Bernoulli random variable representing the total number of observed failures and $\widetilde{\varphi}$ is the maximum likelihood estimator, given by $\frac{N}{T}$, for $N \geq 1$. The statistic is asymptotically distributed as a chi-square disribution with 1 degrees of freedom. If the LR statistics exceed the critical value, 99% quantile of the χ^2_1, the hypothesis $H_o : \varphi = \widetilde{\varphi}$ against a two sided is rejected. In accordance with the convention, we will set the size of the test to 5%. For a number of left tail probabilities and evaluation sample sizes, table (5) gives the nonrejection regions.

To perform the backtest we considered a one period ahead returns data and used the

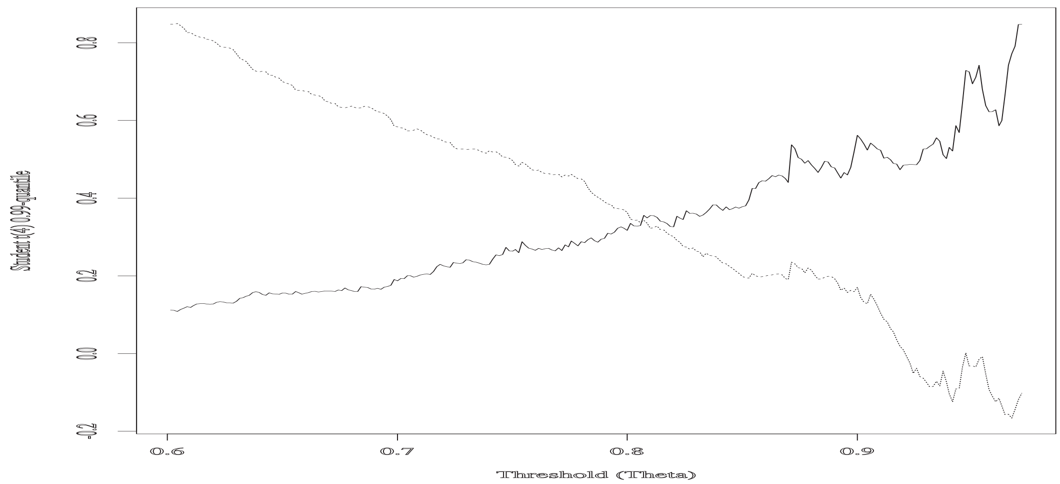

Figure 25: The variance (solid) and bias (dotted) for the estimate at $\varphi = 0.99$ against the threshold (theta)

first 501 observations to estimate $\mu_{t+1,\theta}$ and $\sigma_{t+1,\theta}$. This means we effectively had 500 pairs of observations $\left(Y_{t+1}, X_t\right)$ for the estimation of $\mu_{t+1,\theta,\varphi}$ under the assumption that the underlying process is QAR-QARCH. In the case of historical simulation, we considered the first 680 returns from which we used the first 181 as a rolling window. The result is depicted in table (7), for the levels $\varphi = 0.95, 0.99$ and 0.995.

At $\varphi = 0.95$, the Kupiec test rejects the direct GPD and Hill methods under BASF, because of overestimation. At 99%, the QAR+GPD and QAR+Hill models significantly overestimate the risk in BMW and DBK companies, whereas the HS method significantly underestimate the risks in BMW and DAX. At 99.5%, except QAR, QAR+sc.GPD and QAR+sc.Hill, all other methods either under-or overestimate the risk in most of the portfolios. According to the Basel accord directives, it is only the histotical simulation the would be penalized, for example at 99% level, because it underestimate the VaR in DAX. However, for banks, it is not only important to know whether their model underpredicts the VaR but also if the model is too conservative, because the latter would unnecessarily jeopardize their profit opportunities. Hence, for the banks, all the models which significantly overestimate would also be unfavourable. We however remark that the Kupiec's test does not seem to be reliable for high values of φ. This is seen, for example, in the case of QAR, which is not significant at $\varphi = 0.995$ but appears to be poor in terms of the MASE provided in section (4.6)

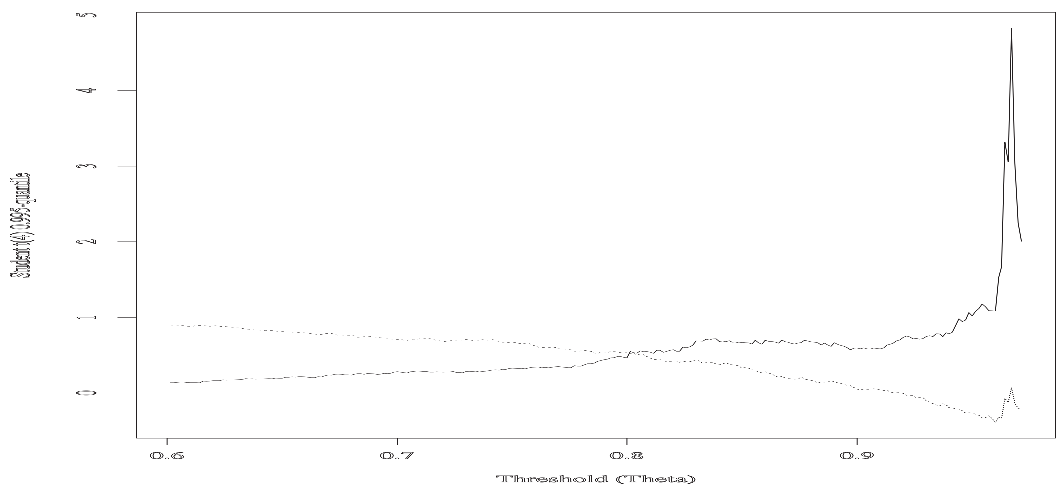

Figure 26: The variance (solid) and bias (dotted) for the estimate at $\varphi = 0.995$ against the threshold (theta)

4.6.3 T-periods extreme Value-at-Risk

For prediction purpose, usually a portfolio returns is split up into estimation and evaluation sample for each VaR technique. The estimation sample is used to estimate the model in question and predict the VaR of the portfolio and then the adequacy of the model is assessed by the means of evaluation sample. This procedure works well with parametric models, where the estimation sample is used to estimate the parameters of the model (e.g normal GARCH), which is then used in the second sample. In nonparametric set up, such parameters are not there and to our knowledge, one has to estimate directly the T-periods ahead VaR by conditionally regressing the T-periods returns on the current. The estimated function is then used for backtesting. This turns out to work well when T is small, for instance, the one period backtesting that we carried out in section (4.6.2). However as T increases such estimates are known to be unreliable. In this section, we propose an approach for T-periods estimation of VaR, in terms of negative returns, which could be thought as a partial solution to the problem.

Before we go into details, let us note that most VaR models based on variance techniques assume normality, despite the well known fact that high frequency financial data have fatter tails than can be explained by the normal distribution. For good reasons, the parameters of normal distribution are usually easier to estimate as there is often an

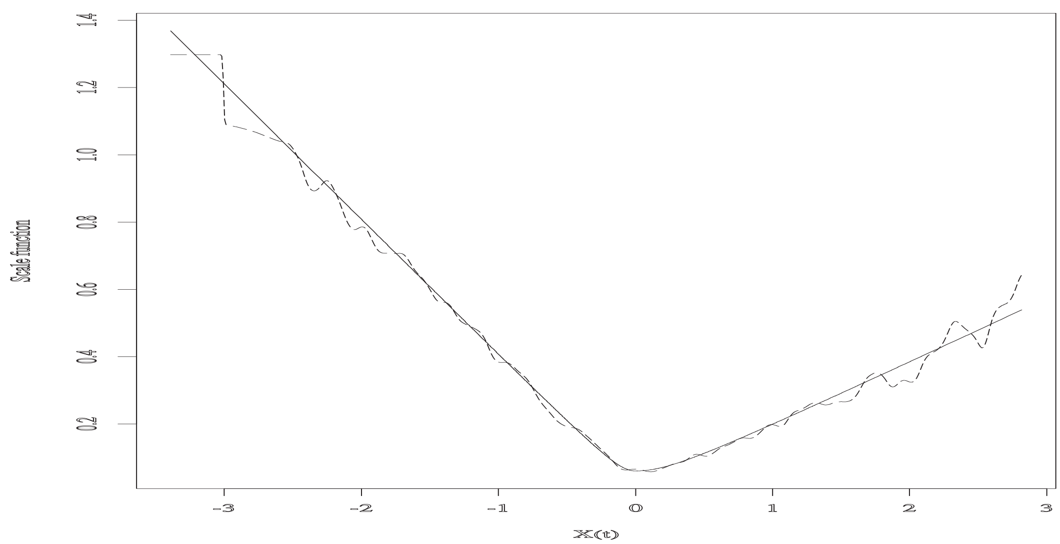

Figure 27: Scale function at $\theta = 0.9$ based on AR(1)-TARCH(1) data, $n = 1000$:-Dotted is estimate and solid, the true function.

analytical solution for them. Second and perhaps the most important, is the additivity of the normal[48] distribution: The sum of two normally distributed random variables is also normally distributed. This characteristic is very important for the calculation of multi-days VaRs based on one-day VaR- a feature of the Basel guidlines. Let us define a T-period ahead returns as

$$e_{t+T} = \log\Big(\frac{P_{t+T}}{P_t}\Big) = \sum_{j=1}^{T} \log\Big(\frac{P_{t+j}}{P_{t+j-1}}\Big) \qquad (4.6.3.1)$$

where P_t and P_{t+T} are the current and T-periods ahead stock prices respectively. If we assume that the returns are iid with zero mean-constant(or unit) variance, it is well known, see Dacorogna et al. [30], that self additivity implies the $\sqrt{T}$ scaling factor for the T-periods ahead risk of a portfolio on the basis of one-period ahead risk:

$$VaR_{t+T,\varphi} \approx \sqrt{T} VaR_{t+1,\varphi}. \qquad (4.6.3.2)$$

Approximation (4.6.3.2), which is already implimented in Riskmetrics, is known as

[48]Normal distribution belongs to a class of sum-stable distributions which is characterised by the fact that sums of random variables from a sum-stable distribution again follow that sum-stable distribution.

the *square root of time rule*.

Suppose we now assume that a one period returns are iid with fat tailed distribution. Then we know the tail distribution belongs to the $MDA\Big(H_{\frac{1}{\alpha}}, \frac{1}{\alpha} > 0\Big)$ and looks like a Pareto tail distribution. To the first order approximation, see (4.2.0.6), we have

$$1 - P\Big(e_{t+1} \leq e\Big) = \overline{F}\Big(e\Big) \approx (1-\theta)\Big(\frac{e}{q^e_\theta}\Big)^{-\alpha}, \qquad e > q^e_\theta > 0, \quad \theta \to 1$$

where q^e_θ is the threshold above which a Pareto like tail holds, e can be regarded as a portfolio's loss and α is the tail index. The method of obtaining a T-step period prediction can be based on the work in Feller [49](VIII. 8) where it is shown that the tail risk for fat tailed distributions is, to a first approximation, linearly additive. Hence for T-period ahead returns, we have

$$1 - P\Big(e_{t+T} \leq e\Big) \approx T(1-\theta)\Big(\frac{e}{q^e_\theta}\Big)^{-\alpha}. \tag{4.6.3.3}$$

Such result has been discussed in the case of nonnormal stable distribution (with $\alpha < 2$) in Fama and Miller [44], page 270. Because of additivity of the tails of heavy tailed distributions it is easy to see from (4.6.3.3), see also in Danielsson and de Vries [31], that for a T-periods ahead VaR based on one period VaR a factor of $T^{\frac{1}{\alpha}}$ is needed, i.e

$$VaR_{t+T,\varphi} \approx T^{\frac{1}{\alpha}} VaR_{t+1,\varphi}, \quad \text{based on iid data}$$

This is called *α-root of time rule*.

If we now turn back to our problem, clearly the α- root of time rule cannot be applied directly to the random variable Y_{t+1}, because it is not iid. However, if we assume the functions $\mu_{t+1,\theta}$ and $\sigma_{t+1,\theta}$ are fairly constant within a specified period, T, we can use the rule to predict a T-period ahead risk. Consider a one period QAR adjusted-scaled returns

$$\frac{Y_{t+1} - \mu_{t+1,\theta}}{\sigma_{t+1,\theta}} = Z_{t+1}$$

which we assume to be iid with Pareto like beyond $\frac{q^e_\theta}{M^e_\theta}$, see equation (4.2.0.11). The extreme unconditional quantile of Z_{t+1} is given as

$$\frac{\mu_{t+1,\theta,\varphi} - \mu_{t+1,\theta}}{\sigma_{t+1,\theta}} = q_{\varphi}^{z}$$

The T-periods prediction based on the unconditional random variable Z_{t+1} is cleary seen as

$$\frac{\mu_{t+T,\theta,\varphi} - \mu_{t+1,\theta}}{\sigma_{t+1,\theta}} \approx T^{\frac{1}{\alpha}} q_{\varphi}^{z}$$

Hence, by rearranging we obtain the a T-periods VaR as

$$VaR_{t+T,\theta,\varphi} \approx \mu_{t+1,\theta} + T^{\frac{1}{\alpha}} \sigma_{t+1,\theta} q_{\varphi}^{z}$$

where the quantile q_{φ}^{z} is based on Pareto distribution with $\frac{1}{\alpha} > 0$, as in (4.2.0.12). The T-periods estimate of VaR, $VaR_{t+T,\varphi}$ is then given by

$$\widehat{VaR}_{\theta,\varphi}^{(T)}\Big(\mathbf{x}_i\Big) = \widehat{\mu}_{\theta}\Big(\mathbf{x}_i\Big) + T^{\frac{1}{\alpha}} \widehat{\sigma}_{\theta}\Big(\mathbf{x}_i\Big) \widehat{q}_{\varphi}^{z}, \tag{4.6.3.4}$$

where the components $\widehat{\mu}_{\theta}\Big(\mathbf{x}_i\Big)$, $\widehat{\sigma}_{\theta}\Big(\mathbf{x}_i\Big)$ and $\widehat{q}_{\varphi}^{z}$ are consistent estimates.

4.7 Conclusion

We have combined the QAR-QARCH model, based on nonparametric quantile regression methodology, with extreme value theory, in the estimation of extreme QAR. We have argued that the overall estimator is heuristically consistent for the true one. The performance of different models were evaluated by using artificial and real data. The result shows that QAR-QARCH augmented with GPD performs best: It overcomes the problem in QAR alone (caused by the sparseness of data in high levels) and the direct application of EVT which do not cope with the volatility clustering and low thresholds, clearly noted in the Monte Carlo results. The problem of multiperiod estimation has been discussed and procedure proposed.

However, the method only models the behaviour of profit and loss (P& L) of a portfolio and therefore has the following disadvantages: In the case a portfolio consist of multiple risk factors (market rates), and VaR estimates based on the model indicates an increase in risk, the source of the increase will not be apparent because the approach does not model the behaviour of individual risk factors.

Table 6: Monte Carlo simulation. The thresholds were fixed at $\theta = 0.90$

$\varphi = 0.95$	**Norm.**	**Stud.-t(3)**	**Stud.-t(4)**	**Gamm.(2,2)**
HS	1.9900	12.140	7.9200	8.0400
GPD	1.8600	9.4600	4.8900	5.2100
Hill	*.****	10.787	6.7623	7.3456
QAR	0.3900	1.1400	0.8600	1.2600
QAR+ GPD	0.5800	1.6100	1.2000	1.4200
QAR+ Hill	*.****	1.8100	1.4700	1.4920
QAR+sc.GPD	0.3500	1.0200	0.8740	1.1000
QAR+sc.Hill	*.****	1.3140	1.1734	1.4672
$\varphi = 0.99$				
HS	6.2300	64.410	44.100	27.810
GPD	3.6000	16.280	15.460	13.530
Hill	*.****	18.490	14.740	15.140
QAR	1.3700	15.640	9.8400	13.010
QAR+ GPD	0.8800	6.3200	3.8000	6.8300
QAR+ Hill	*.****	6.2200	4.0000	6.7600
QAR+sc.GPD	0.7200	3.8000	2.9000	4.0000
QAR+sc.Hill	*.****	4.3670	3.4010	4.4607
$\varphi = 0.995$				
HS	26.010	94.760	80.520	82.460
GPD	10.570	70.640	54.410	57.320
Hill	**.***	71.440	69.260	59.124
QAR	19.780	80.43	70.600	67.840
QAR+ GPD	4.9400	51.46	43.660	40.470
QAR+ Hill	*.****	51.46	44.780	39.280
QAR+sc.GPD	3.9400	9.960	7.4200	7.2800
QAR+sc.Hill	*.****	11.60	7.7410	8.9978

Table 7: Backtesting on 510 points. Threshold taken at $\theta = 0.80$.

$\varphi = 0.95$ **(25)**	**BASF(745)**	**BMW(744)**	**DAX(745)**	**DBK(745)**	**BAI(745)**
HS	00028	00027	00024	00026	00024
GPD	00016	00020	00017	00022	00022
Hill	00016	00018	00018	00020	00021
QAR	00025	00024	00025	00025	00023
QAR+GPD	00020	00021	00019	00019	00020
QAR+Hill	00019	00018	00019	00021	00020
QAR+sc.GPD	00022	00023	00024	00024	00025
QAR+sc.Hill	00017	00020	00019	00020	00019
$\varphi = 0.99$ (5)					
HS	00008	00010	00010	00006	00007
GPD	00004	00003	00005	00004	00006
Hill	00002	00004	00006	00004	00003
QAR	00005	00005	00004	00005	00006
QAR+GPD	00003	00001	00004	00001	00004
QAR+Hill	00002	00001	00003	00001	00003
QAR+sc.GPD	00005	00006	00005	00005	00005
QAR+sc.Hill	00005	00006	00004	00004	00006
$\varphi = 0.995$ (3)					
HS	00007	00005	00009	00006	00004
GPD	00000	00000	00002	00002	00000
Hill	00000	00000	00000	00002	00001
QAR	00003	00002	00002	00005	00001
QAR+GPD	00000	00000	00001	00001	00000
QAR+Hill	00000	00000	00000	00001	00000
QAR+sc.GPD	00002	00003	00002	00002	00003
QAR+sc.Hill	00001	00002	00001	00002	00002

5 Conditional expected shortfall

In the last decade, there has been regulatory concerns in the financial sector on the question of how to evaluate portfolio risk. Artzner, Delbaen, Eber and Heath (1999), provides an axiomatic foundation for "coherent" risk measures.

5.1 Coherent risk measure

Definition 5.1.1 *(Artzner et al. [6])*
Consider a set V of real-valued random variables on some probability space $\Big(\Omega, \mathbf{F}, \mathbf{P}\Big)$ with finite first moment (tail index $\alpha > 1$) for all $\mathbf{e} \in V$. The function $\rho : V \to \mathbf{R}$ is a coherent risk measure if the following axioms(or properties) hold

$$
\begin{array}{lll}
(i) & \textit{Monotonicity:} & \mathbf{e}_1, \mathbf{e}_2 \in V, \textit{with} \quad \mathbf{e}_1 \le \mathbf{e}_2 \Rightarrow \rho\Big(\mathbf{e}_1\Big) \ge \rho\Big(\mathbf{e}_2\Big) \\
(ii) & \textit{Sub-additivity:} & \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_1 + \mathbf{e}_2 \in V \quad \Rightarrow \rho\Big(\mathbf{e}_1 + \mathbf{e}_2\Big) \le \rho\Big(\mathbf{e}_1\Big) + \rho\Big(\mathbf{e}_2\Big) \\
(iii) & \textit{Positive homogeneity:} & \textit{For all} \quad \lambda > 0, \quad \mathbf{e} \in V, \quad \Rightarrow \rho\Big(\lambda \mathbf{e}\Big) = \lambda \rho\Big(\mathbf{e}\Big) \quad \textit{and} \\
(iv) & \textit{Translation invariance:} & \textit{For all} \quad \lambda \in \mathbf{R} \quad \mathbf{e} \in V, \quad \Rightarrow \rho\Big(\mathbf{e} + \lambda \mathbf{1}\Big) = \rho\Big(\mathbf{e}\Big) - \lambda \\
 & \textit{for a numeraire} \quad \mathbf{1}. &
\end{array}
$$

The sub-additivity axiom is of particular interest here. It expresses the fact that a portfolio made up of sub-portfolios will risk an amount which is at most the sum of separate amounts risked by its sub-portfolios. The global risk of a portfolio will be the sum of the risks of its parts only in the case when the latter can be triggered by concurrent events, namely if the sources of these risks conspire to act together. In all other cases, the global risk of the portfolio will be strictly less than the sum of its partial risks. Thus if a measure is sub-additive, the portfolio diversification will always lead to risk reduction. If a risk measure is not sub-additive, diversification could produce an increase in their value even when partial risks arise from independent (or mutually exclusive) events, Acerbi et al. [2].

An obvious area in the financial industry, where sub-additivity plays a great role, is in capital adequacy requirements in banking supervision: If a bank is made up of several branches such that the capital requirement of each branch is dimensioned on its own,

then under sub-additivity, the regulator is confident that also the overall bank capital is an adequate one. Outside sub-additivity, the risk of the whole bank may turn out to be much bigger than the sum of the branches risks.
Unfortunately, risk measures based on second moments including the standard deviation, as well as quantile based measures like VaR are not necessarily subadditive and hence not coherent risk measure, Artzner et al. [6]. This happens in situations where a portfolio is split into subportfolios such that the sum of the VaR for individual subportfolio is required to be at least the global portfolio one. The second disadvantage is that it only gives a bound on the losses that occur with a given frequency; it tells nothing about the potential size of a loss given that it has exceeded the bound. To correct these problems, we take the (conditional) average of those events exceeding the VaR. In the following section, we consider the case for iid data.

5.2 Expected shortfall under iid case

Let $F\Big(z\Big)$ be a probability distribution function of iid random variables $Z_t, t = 1, \ldots$ based on the process (1.4.0.2) and for some probability $\varphi \in (0, 1)$ such that $0 << \theta < \varphi < 1$, consider the φ-quantile as

$$q_\varphi^z = \inf\Big\{ z \in \mathbf{R} \Big| F\Big(z\Big) \geq \varphi \Big\} \tag{5.2.0.5}$$

If F is continuous, then $P\Big(Z_t = q_\varphi^z\Big) = 0$ and $F\Big(q_\varphi^z\Big) = \varphi$, while if F is discontinuous in q_φ^z, then $P\Big(Z_t = q_\varphi^z\Big) > 0$ and $F\Big(q_\varphi^z\Big) = P\Big(Z_t \leq q_\varphi^z\Big) > \varphi$.
Consider the order statistics $z_{n,n} \leq \ldots \leq z_{k+1,n} \leq z_{k,n} \leq \ldots z_{1,n}$ as the sorted values of the n-tuple $\Big(Z_1, \ldots, Z_n\Big)$ and let $k = [n(1-\varphi)]\Big(= \max_{m\in\mathbf{N}}\Big\{m \leq n(1-\varphi)\Big\}\Big)$ be the integer part of $n(1-\varphi)$. The set of observations which constitute the $100(1-\varphi)\%$ largest of the total values in the sample is represented by the largest k observations (outcomes) $\Big\{z_{k,n} \ldots, z_{1,n}\Big\}$. As usual, $z_{k+1,n}$, denotes the empirical quantile which we may write as, $q_{\varphi,n}^z$ where φ stands for the proportion of observations below $z_{k,n}$. The average of the observations or losses in the $100(1-\varphi)\%$ worst cases, denoted as $\Im_{Z_t}\Big(q_{\varphi,n}^z\Big)$, can be obtained as

$$\Im_{Z_t}\left(q^z_{\varphi,n}\right) = \frac{1}{k}\sum_{t=k}^{1} z_{t,n}. \tag{5.2.0.6}$$

It is simply the mean of the k largest observations. By including all the observation in a sample, one derives a theoretical limit value for (5.2.0.6):

$$\begin{aligned}
\Im_{Z_t}\left(q^z_{\varphi,n}\right) &= \frac{1}{k}\sum_{t=n}^{1} z_{t,n}\mathbf{I}_{\{t\leq k\}} \\
&= \frac{1}{k}\sum_{t=n}^{1} z_{t,n}\mathbf{I}_{\{z_{t,n}>z_{k+1,n}\}} \\
&= \frac{1}{k}\sum_{t=n}^{1} Z_t\mathbf{I}_{\{Z_t>z_{k+1,n}\}} \\
&= \frac{n}{k}\frac{1}{n}\sum_{t=n}^{1} Z_t\mathbf{I}_{\{Z_t>z_{k+1,n}\}}
\end{aligned} \tag{5.2.0.7}$$

where we assume that F has a density and, therefore, $P\left(z_{k,n} > z_{k+1,n}\right) = 1$. With $n-k = n\varphi + o\left(n^{\frac{1}{2}}\right)$, i.e. $\frac{k}{n} = 1-\varphi+o\left(n^{-\frac{1}{2}}\right)$, the central limit theorem gives,

$$q^z_{\varphi,n} \sim N\left(q^z_{\varphi}, \frac{\varphi\left(1-\varphi\right)}{nf^2\left(q^z_{\varphi}\right)}\right). \tag{5.2.0.8}$$

by theorem 7.4.2.1, page 374 in Dudewicz and Mishra [35]. Moreover, if $\lim_{n\to\infty} q^z_{\varphi,n} = q^z_{\varphi}$ with probability 1, then

$$\lim_{n\to\infty}\Im_{Z_t}\left(q^z_{\varphi,n}\right) = \frac{1}{1-\varphi}E\left[Z_t\mathbf{I}_{\{Z_t>q^z_{\varphi}\}}\right] \tag{5.2.0.9}$$

with probability 1. Hence we give a general definition of the expected shortfall in the upper tail of a distribution:

Definition 5.2.1 *(Acerbi and Tasche [4])*
Let $\left\{Z_t, t=1,\ldots,n\right\}$ *be iid random variables representing negative returns of a portfolio on some specified length of time, and* $\varphi \in (0,1)$ *be some specified probability level. The expected* $100(1-\varphi)\%$ *shortfall of the random variable* Z_t *is defined as*

$$\Im_{Z_t}\Big(q^z_\varphi\Big) = \frac{1}{1-\varphi}E\Big[Z_t \mathbf{I}_{\{Z_t > q^z_\varphi\}}\Big] + q^z_\varphi\left(1 - \frac{P\Big(Z_t > q^z_\varphi\Big)}{1-\varphi}\right) \tag{5.2.0.10}$$

This definition was first introduced in Acerbi and Tasche [4] where it is shown that the simple sub-additive nature of the sample expected shortfall estimator generalizes easily to (5.2.0.10). The second part on the right hand side is a correction term to allow also for discrete distributions with a point mass at q^z_φ. If F has a density, this term vanishes as the $P\Big(Z_t > q^z_\varphi\Big) = 1-\varphi$, and the expected shortfall becomes

$$\Im_{Z_t}\Big(q^z_\varphi\Big) = \frac{1}{1-\varphi}E\left[Z_t I_{\{Z_t > q^z_\varphi\}}\right]. \tag{5.2.0.11}$$

The idea of fitting a GPD to the extreme tail, explained in chapter 4,is easily adapted to estimating the conditional expected shortfall. Since from the definition of Z_t in (1.4.0.2), $q^z_\varphi = q^z_{\theta,\varphi} > q^z_\theta = 0$, the excesses over q^z_φ given that Z_t exceeds q^z_φ, can be written as

$$Z_t - q^z_\varphi \Big| Z_t > q^z_\varphi = (Z_t - q^z_\theta) - (q^z_\varphi - q^z_\theta) \Big| (Z_t - q^z_\theta) > (q^z_\varphi - q^z_\theta) \tag{5.2.0.12}$$

The following lemma shows that the conditional distribution of $Z_t - q^z_\varphi$ given $Z_t > q^z_\varphi$, is also a GPD if $Z_t - q^z_\theta$ given $Z_t > q^z_\theta$ is.

Lemma 5.1 *Let* $F_{q^z_\theta}\Big(z\Big) = P\Big(Z_t - q^z_\theta \le z \Big| Z_t > q^z_\theta\Big) = G_{\xi,\beta(\theta)}\Big(z\Big)$ *with* $0 < \xi < 1$. *Then, for* $\varphi > \theta$, *the conditional distribution of* $Z_t - q^z_\varphi$ *given* $Z_t > q^z_\varphi$ *is also a GPD*

$$P\Big(Z_t - q^z_\varphi \Big| Z_t > q^z_\varphi\Big) = G_{\xi,\beta(\varphi)}\Big(z\Big), \tag{5.2.0.13}$$

with the same shape parameter ξ *and scale* $\beta(\varphi) = \beta(\theta) + \xi q^z_\varphi$, *and the expected shortfall is*

$$\Im_{Z_t}\Big(q^z_\varphi\Big) = \frac{q^z_\varphi(1+\xi) + \beta(\theta)}{1-\xi} = \frac{q^z_\varphi + \beta(\varphi)}{1-\xi} \tag{5.2.0.14}$$

Proof of lemma 5.1

Let F denote the conditional distribution of $Z_t - q^z_\theta$, $\overline{F} = 1 - F$. The conditional probability, $F_{q^z_\varphi}\Big(z\Big)$, of the excesses is given by

$$
\begin{aligned}
P\Big(Z_t - q^z_\varphi \le z \Big| Z_t > q^z_\varphi\Big) &= P\Big(\Big(Z_t - q^z_\theta\Big) - \Big(q^z_\varphi - q^z_\theta\Big) \le z \Big| Z_t - q^z_\theta > q^z_\varphi - q^z_\theta\Big) \\
&= \frac{F\Big(z + q^z_\varphi - q^z_\theta\Big) - F\Big(q^z_\varphi - q^z_\theta\Big)}{1 - F\Big(q^z_\varphi - q^z_\theta\Big)} \\
\text{This implies} \quad 1 - F_{q^z_\varphi}\Big(z\Big) &= \frac{\bar{F}\Big(z + (q^z_\varphi - q^z_\theta)\Big)}{\bar{F}\Big(q^z_\varphi - q^z_\theta\Big)} = \frac{\overline{F}_{q^z_\theta}\Big(z + q^z_\varphi\Big)}{\overline{F}_{q^z_\theta}\Big(q^z_\varphi\Big)}.
\end{aligned}
$$

Hence, by our assumptions,

$$
\begin{aligned}
\overline{F}_{q^z_\varphi}\Big(z\Big) &= \frac{\overline{G}_{\xi,\beta(\theta)}\Big(z + q^z_\varphi\Big)}{\overline{G}_{\xi,\beta(\theta)}\Big(q^z_\varphi\Big)} \\
&= \overline{G}_{\xi,\beta(\theta)+\xi q^z_\varphi}(z),
\end{aligned}
\tag{5.2.0.15}
$$

using $\overline{G}_{\xi,\beta}\Big(y\Big) = \Big(1 + \frac{\xi}{\beta}y\Big)^{-\frac{1}{\xi}}$.

The second part,

$$
\begin{aligned}
\Im_{Z_t}\Big(q^z_\varphi\Big) &= \frac{1}{1-\varphi}E\Big[Z_t I_{\{Z_t > q^z_\varphi\}}\Big] \\
&= q^z_\varphi + \frac{1}{1-\varphi}E\Big[\Big(Z_t - q^z_\varphi\Big)\mathbf{I}_{\{Z_t - q^z_\varphi > 0\}}\Big],
\end{aligned}
\tag{5.2.0.16}
$$

where the second term on the right is the mean excess function (MEF) over the threshold q^z_φ. We know that the mean excess function for the GPD with $\xi < 1$ and threshold u has the following expression

$$
MEF\Big(u\Big) = E\Big[Z_t - u \Big| Z_t > u\Big] = \frac{\beta + \xi u}{1-\xi}, \quad \beta + \xi u > 0 \tag{5.2.0.17}
$$

and therefore combining (5.2.0.15) and (5.2.0.17), we get

$$
\begin{aligned}
\Im_{Z_t}\Big(q^z_\varphi\Big) &= q^z_\varphi + MEF\Big(q_\varphi) \\
&= q^z_\varphi + \frac{\beta(\varphi) + \xi q^z_\varphi}{1-\xi}, \\
&= \frac{q^z_\varphi + \beta(\varphi)}{1-\xi} = \frac{q^z_\varphi\Big(1+\xi\Big) + \beta(\theta)}{1-\xi}.
\end{aligned}
\tag{5.2.0.18}
$$

□

A general result concerning the existence of moments is that if Z_t is a GPD, then for all integers r, such that $r < \frac{1}{\xi}$, the r^{th} first moments exist, see Embrechts et al. [39], page 165. Denote the estimate of the expected shortfall in (5.2.0.18) by

$$\widehat{\Im}_{Z_t}\left(\widehat{q}^z_\varphi\right) = \widehat{q}^z_\varphi \frac{1+\widehat{\xi}}{1-\widehat{\xi}} + \frac{\widehat{\beta}(\theta)}{1-\widehat{\xi}}. \tag{5.2.0.19}$$

Since $\widehat{\xi}$ and $\widehat{\beta}(\theta)$ are obtained by ordinary ML or moment methods, and $\widehat{q}^z_\varphi$ is intuitively consistent by the heuristics (from chapter (4)), then $\Im_{Z_t}\left(\widehat{q}^z_\varphi\right) \to^p \Im_{Z_t}\left(q^z_\varphi\right)$ heuristically. By the same argument as in section (4.4), $\widehat{\Im}_{Z_t}\left(\widehat{q}^z_\varphi\right)$ should also be a consistent estimate of the expected shortfall if $F_{q^z_\theta}\left(z\right)$ is not exactlty a GPD but only in the limit for $\theta \to 1$ using theorem 4.2.

5.2.1 Alternative Expected shortfall

An equivalent alternative representation of expected shortfall which reveals in a transparent way its direct dependence on φ is obtained via the inverse of a distribution function $F\left(z\right) = P\left(Z_t \leq z\right)$. Define the quantile function, q^z_ϑ, as usual as

$$q^z_\vartheta = \inf\left\{Z_t \in \mathbf{R} \middle| F\left(z\right) \geq \vartheta\right\}, \quad \vartheta \in (0,1)$$

Then $\Im_{Z_t}\left(q^z_\varphi\right)$ can be expressed as the mean of q^z_ϑ on the interval $[1-\varphi, 1)$, i.e

$$\Im_{Z_t}\left(q^z_\varphi\right) = \frac{1}{1-\varphi}\int_{1-\varphi}^{1} q^z_\vartheta d\vartheta, \tag{5.2.1.1}$$

see Acerbi and Tasche [3]. The second alternative which can be derived from (5.2.0.10) has been formulated in Rockafellar and Uryasev [99] where the expected shortfall is given by

$$ES\left(.,\varphi\right) = TCE\left(.,\varphi\right) + (\lambda-1)\left(TCE\left(.,\varphi\right) - VaR\left(.,\varphi\right)\right)$$

with $\lambda \equiv \frac{P\left(Z_t > q^z_\varphi\right)}{1-\varphi} \geq 1$ and TCE stands for tail conditional expectation. For the case that F has a density, $\lambda = 1$, and the second term vanishes.

5.3 Conditional expected shortfall for the dependent case

This section uses the extreme QAR or $VaR_{t,\varphi}$, for $\varphi \geq 0.95$, and proposes the conditional expected shortfall for dependent data. Recall the model (1.4.0.2) has VaR given as

$$VaR_{t,\varphi} = \mu_{t,\theta} + \sigma_{t,\theta} q_{\varphi}^{z} \tag{5.3.0.2}$$

where q_{φ}^{z} is the marginal φ-quantile of Z_t. The conditional extreme VaR given the past information satisfies the following probability

$$\varphi = P\Big(Y_t \leq VaR_{t,\varphi}\Big|\mathbf{F}_{t-1}\Big) \tag{5.3.0.3}$$

where $(1-\varphi)$ is the loss probability and $VaR_{t,\varphi} = \mu_{t,\varphi}$. Note that for nonnegative returns, the VaR is usually a function of the loss probability ranging from 1% to 5% while stock returns are usually measured over one day or ten day period. However, to be consistent with previous notations, we will continue working with negative log returns which means the VaR will remain a function of $\varphi \geq 0.95$. The conditional expected loss knowing that the loss is above the VaR is then defined by

$$\begin{aligned} \Im_{Y_t}\Big(VaR_{t,\varphi}\Big) &= E\Big[Y_t\Big|Y_t > \mu_{t,\varphi}; \mathbf{F}_{t-1}\Big] \\ &= E\Big[Y_t\Big|Y_t > VaR_{t,\varphi}, \mathbf{F}_{t-1}\Big] \end{aligned} \tag{5.3.0.4}$$

Proposition 5.1 *Let* $\mu : \mathbf{R}^d \to \mathbf{R}$ *be unknown function and* $(Y_t, \mathbf{X}_t) \in \mathbf{R}^{d+1}$ *be real random variables on the probability space* $(\Omega, \mathbf{F}, \mathbf{P})$ *from model (1.4.0.2). For a fixed* $\varphi \geq 0.95$*, define* $H : \mathbf{R}^{d+1} \to \mathbf{R}_+$ *as*

$$H\Big(\mu\Big) = E\Big[M_{\varphi}\Big(Y_t, \mu\Big)\Big|\mathbf{F}_{t-1}\Big]. \tag{5.3.0.5}$$

The conditional expected shortfall given that $Y_t > VaR_{t,\varphi}$ *is given as*

$$\Im_{Y_t}\Big(VaR_{t,\varphi}\Big) = \frac{1}{1-\varphi} H\Big(VaR_{t,\varphi}\Big) + \mu_t \tag{5.3.0.6}$$

with $\mu_t = E\Big[Y_t\Big|\mathbf{F}_{t-1}\Big]$ *being the conditional expectation of* Y_t *given the information in* $\mathbf{F}_{t-1}$.

Proof of Proposition (5.1)

$H\Big(\mu\Big)$ is convex (and continuous) with $\lim_{|\mu|\to\infty} H\Big(\mu\Big) = \infty$. We can take $VaR_{t,\varphi}$ as a unique minimizer of the objective function, (5.3.0.5), at a fixed φ. The expansion of $H\Big(VaR_{t,\varphi}\Big)$ yields

$$\begin{aligned} H\Big(VaR_{t,\varphi}\Big) &= E\Bigg[\Big(Y_t - VaR_{t,\varphi}\Big)\Big(I_{\{Y_t - VaR_{t,\varphi}>0\}} - (1-\varphi)\Big)\Big|\mathbf{F}_{t-1}\Bigg] \\ &= -(1-\varphi)E\Big[Y_t\Big|\mathbf{F}_{t-1}\Big] + (1-\varphi)\Bigg\{\frac{E\Big[Y_t\mathbf{I}_{\{Y_t - VaR_{t,\varphi}>0\}}\Big|\mathbf{F}_{t-1}\Big]}{1-\varphi} \\ &+ VaR_{t,\varphi}\frac{\Big(1-\varphi-P\Big(Y_t - VaR_{t,\varphi} > 0\Big|\mathbf{F}_{t-1}\Big)\Big)}{1-\varphi}\Bigg\} \\ &= -(1-\varphi)\mu_t + (1-\varphi)\Im_{Y_t}\Big(VaR_{t,\varphi}\Big) \end{aligned} \tag{5.3.0.7}$$

and hence,

$$\Im_{Y_t}\Big(VaR_{t,\varphi}\Big) = \frac{1}{1-\varphi}H\Big(VaR_{t,\varphi}\Big) + \mu_t \tag{5.3.0.8}$$

□

This result is similar to "α-risk" in Bassett et al. [8] for iid case. For Y_t having a continuous distribution, then $1-\varphi = P\Big(Y_t - VaR_{t,\varphi} > 0\Big|\mathbf{F}_{t-1}\Big)$ and putting $H\Big(VaR_{t,\varphi}\Big)$ in equation (5.3.0.8), we get

$$\Im_{Y_t}\Big(VaR_{t,\varphi}\Big) = \frac{1}{1-\varphi}E\Big[Y_t\mathbf{I}_{\{Y_t > VaR_{t,\varphi}\}}\Big|\mathbf{F}_{t-1}\Big] \tag{5.3.0.9}$$

5.3.1 Estimation under EVT framework

We assume that the excess residuals over the initial QAR adjusted-scaled threshold, q^z_θ, are iid with distribution that belong to the $MDA\Big(H_\xi, 0 < \xi < 1\Big)$. From (5.3.0.2), the VaR is given

$$VaR_{\theta,\varphi}\Big(\mathbf{X}_t\Big) = \mu_\theta\Big(\mathbf{X}_t\Big) + \sigma_\theta\Big(\mathbf{X}_t\Big)q^z_\varphi, \quad \varphi > \theta \tag{5.3.1.1}$$

with q^z_φ being the quantile of Z_t obtained from the GPD fitted to the excesses over q^z_θ. Note that we write $VaR_{\theta,\varphi}\Big(\mathbf{X}_t\Big)$ to emphasise that the estimation of $VaR_\varphi\Big(\mathbf{X}_t\Big)$ is done via an initial estimator of $\mu_\theta\Big(\mathbf{X}_t\Big)$. Because the continuity assumption holds, the conditional expected shortfall at point $\mathbf{x}_i$ can be written using (5.3.0.9) as

$$\begin{aligned}
\Im_{Y_t}\Big(VaR_{\theta,\varphi}\big(\mathbf{x}_i\big)\Big) &= \frac{1}{1-\varphi}E\Big[\Big(VaR_{\theta,\varphi}\Big(\mathbf{X}_t\Big)+\sigma_\theta\Big(\mathbf{X}_t\Big)\Big(Z_t-q^z_\varphi\Big)\Big) \\
&\cdot \quad \mathbf{I}_{\{VaR_{\theta,\varphi}(\mathbf{X}_t)+\sigma_\theta(\mathbf{X}_t)(Z_t-q^z_\varphi)>VaR_{\theta,\varphi}(\mathbf{X}_t)\}}\Big|\mathbf{X}_t=\mathbf{x}_i\Big] \\
&= VaR_{\theta,\varphi}\Big(\mathbf{x}_i\Big)+\frac{\sigma_\theta(\mathbf{x}_i)}{1-\varphi}E\Big[\Big(Z_t-q^z_\varphi\Big)\mathbf{I}_{\{Z_t-q^z_\varphi>0\}}\Big] \\
&= \mu_\theta\Big(\mathbf{x}_i\Big)+\frac{\sigma_\theta(\mathbf{x}_i)}{1-\varphi}E\Big[Z_t\mathbf{I}_{\{Z_t-q^z_\varphi>0\}}\Big] \\
&= \mu_\theta\Big(\mathbf{x}_i\Big)+\sigma_\theta(\mathbf{x}_i)\Im_{Z_t}\Big(q^z_\varphi\Big) \qquad (5.3.1.2)
\end{aligned}$$

with $\Im_Z\Big(q^z_\varphi\Big)$ given by (5.2.0.18) and $Z_t=\frac{Y_t-\mu_\theta(\mathbf{X}_t)}{\sigma_\theta(\mathbf{X}_t)}$. Denote the estimator for (5.3.1.2) as

$$\widehat{\Im}_{Y_t}\Big(\widehat{VaR}_{\theta,\varphi}\Big(\mathbf{x}_i\Big)\Big)=\widehat{\mu}_\theta\Big(\mathbf{x}_i\Big)+\widehat{\sigma}_\theta\Big(\mathbf{x}_i\Big)\widehat{\Im}_{Z_t}\Big(\widehat{q}^z_\varphi\Big). \qquad (5.3.1.3)$$

We note that by the previous heuristic arguments all the quantities in (5.3.1.3) should be consistent estimators for their respective true functions at point $\mathbf{x}_i$. Therefore, the expected shortfall estimator given by (5.3.1.3) should intuitively also be a consistent estimator of (5.3.1.2) as $n\to\infty, \quad \theta\to 1, \quad N_\theta\to\infty$.

The procedure presented holds in general set up, which include AR-(T)ARCH processes with their extensions like AR-GARCH. Simplified versions for the expected shortfall could be derived, but they may not hold in general. For example, consider the case where the conditional mean of an AR-(T)ARCH process is zero. The expected shorfall estimator, at point $\mathbf{x}_i$, can be written in a manner that does not involve $\widehat{\sigma}_\theta\Big(\mathbf{x}_i\Big)$, i.e

$$\widehat{\Im}_{Y_t}\Big(\widehat{VaR}_{\theta,\varphi}\Big(\mathbf{x}_i\Big)\Big)=\widehat{\mu}_\theta\Big(\mathbf{x}_i\Big)+\widehat{\mu}_\theta\Big(\mathbf{x}_i\Big)\Im_{\widetilde{e}_t}\Big(\widehat{q}^{\widetilde{e}}_\varphi\Big), \quad \text{for} \quad 0.5<\theta<\varphi<1,$$

where $\widetilde{e}_t=\frac{Y_t-\widehat{\mu}_\theta(\mathbf{X}_t)}{\widehat{\mu}_\theta(\mathbf{X}_t)}$ at $\mathbf{X}_t=\mathbf{x}_i$ are assumed approximately iid with zero θ-quantile. The QAR, $\widehat{\mu}_\theta\Big(\mathbf{X}_t\Big)$, becomes the threshold as well as the scale function. This setting do not hold when the conditional mean of the process is non zero.

5.3.2 Estimation under general framework

The straightforwardness of equation (5.3.0.6) motivates direct estimation of the expected shortfall without transforming the excesses over the VaR into iid. At point $\mathbf{X}_t = \mathbf{x}_i$, the expected shortfall estimator is

$$\widehat{\Im}_{Y_t}\Big(\widehat{VaR}_{\theta,\varphi}\Big(\mathbf{x}_i\Big)\Big) = \frac{1}{1-\varphi}H\Big(\widehat{VaR}_{\theta,\varphi}\Big(\mathbf{x}_i\Big)\Big) + \widehat{\mu}\Big(\mathbf{x}_i\Big) \tag{5.3.2.1}$$

where $\widehat{\mu}\Big(\mathbf{x}_i\Big)$ is a conditional mean function estimator for $\mu\Big(\mathbf{X}_t\Big)$ at $\mathbf{x}_i$ obatined as in (4.2.0.14). The expected shortfall estimator, (5.3.2.1), is consistent if $\widehat{VaR}_{\theta,\varphi}\Big(\mathbf{x}_i\Big)$ and $\widehat{\mu}\Big(\mathbf{x}_i\Big)$ are consistent.

A simple estimation algorithm would be to start by estimating the functions in QAR-QARCH model. Then fit the GPD on the scaled excesses over the QAR estimator and derive the unconditional quantile which is then used to estimate $VaR_{\varphi}\Big(\mathbf{x}_i\Big)$. Finally, obtain a local linear estimate of $\mu\Big(\mathbf{x}_i\Big)$ and put the estimated quantities in (5.3.2.1). Note that for more robust estimator, the quantity $\widehat{\mu}\Big(\mathbf{x}_i\Big)$ could be replaced by a consistent estimator for $\mu_{0.5}\Big(\mathbf{x}_i\Big)$, see chapters 2 and 3.

Under the assumption that the conditional mean function, $\mu\Big(\mathbf{x}_i\Big)$, is zero, the expected shortfall estimator reduces to the simplest form;

$$\widehat{\Im}_{Y_t}\Big(\widehat{VaR}_{\theta,\varphi}\Big(\mathbf{x}_i\Big)\Big) = \frac{1}{1-\varphi}H\Big(\widehat{VaR}_{\theta,\varphi}\Big(\mathbf{x}_i\Big)\Big) \tag{5.3.2.2}$$

which is faster to compute.

Note that both results on expected shortfall estimators in (5.3.1.3) and (5.3.2.1) assume that the first moment is finite. However, from practical point of view the latter result can be used not only with the integrable variables or variables with continuous distributions, but also in situations where discontinuities arises.

5.4 T-periods conditional expected shortfall and backtesting

This section formulates T-periods ahead prediction of the expected shortfall with either of the approaches presented in sections (5.3.1) or (5.3.2). Let us define a T-period ahead returns by

$$Y_{t+T} = -\log\Big(\frac{R_{t+T}}{R_t}\Big) = -\sum_{j=1}^{T}\log\Big(\frac{R_{t+j}}{R_{t+j-1}}\Big), \quad t = 1, \ldots, n$$

where R_t and R_{t+T} are the current and T-period ahead, for instance of stock prices, respectively.

(1) *Based on α-root of time rule*

In this case, we have a T-period ahead estimator of the expected shortfall, c.f (4.6.3.4), as

$$\Im_{Y_t}\left(\widehat{VaR}^{T}_{\theta,\varphi}\left(\mathbf{x}_i\right)\right) = \widehat{\mu}^{(1)}_{\theta}\left(\mathbf{x}_i\right) + T^{\frac{1}{\alpha}}\widehat{\sigma}^{(1)}_{\theta}\left(\mathbf{x}_i\right)\Im^{(1)}_{Z_t}\left(\widehat{q}^{z}_{\varphi}\right) \tag{5.4.0.3}$$

where $\widehat{\mu}^{(1)}_{\theta}\left(\mathbf{x}_i\right)$ and $\widehat{\sigma}^{(1)}_{\theta}\left(\mathbf{x}_i\right)$ are 1-period ahead estimates of QAR and scale function, at point $\mathbf{X}_{t+1} = \mathbf{x}_i$, respectively. The expected shortfall estimator, $\Im^{(1)}_{Z_t}\left(\widehat{q}^{z}_{\varphi}\right)$, is based on iid random variable Z_{t+1} whose quantile $\widehat{q}^{z}_{\varphi}$ is obtained using (4.2.0.12). The random variable $\mathbf{X}_{t+1}$ is $\mathbf{F}_t$-measurable.

(2) *Under GPD framework*

Here a T-periods expected shortfall, at point $\mathbf{x}_i$, is defined as

$$\begin{aligned}\Im_{Y_t}\left(\widehat{VaR}^{(T)}_{\theta,\varphi}\left(\mathbf{x}_i\right)\right) &= \frac{1}{1-\varphi}E\left[Y_{t+T}\mathbf{I}_{\{Y_{t+T}>\widehat{VaR}_{\theta,\varphi}(\mathbf{X}_{t+1})\}}\middle|\mathbf{X}_{t+1} = \mathbf{x}_i\right] \\ &= \widehat{\mu}^{(T)}_{\theta}\left(\mathbf{x}_i\right) + \widehat{\sigma}^{(T)}_{\theta}\left(\mathbf{x}_i\right)\Im^{(T)}_{Z_t}\left(\widehat{q}^{z}_{\varphi}\right) \end{aligned} \tag{5.4.0.4}$$

with $\widehat{\mu}^{(T)}_{\theta}\left(\mathbf{x}_i\right)$ and $\widehat{\sigma}^{(T)}_{\theta}\left(\mathbf{x}_i\right)$ being T-periods estimates of QAR and scale function respectively, and $\Im^{(T)}_{Z_t}\left(\widehat{q}^{z}_{\varphi}\right)$is the expected shortfall estimate derived from GPD and based on iid random variable Z_{t+T}.

(3) *Under general framework*

We define a T-periods expected shortfall as

$$\Im_{Y_t}\left(\widehat{VaR}^{(T)}_{\theta,\varphi}\left(\mathbf{x}_i\right)\right) = \frac{1}{1-\varphi}H\left(\widehat{VaR}^{(T)}_{\theta,\varphi}\left(\mathbf{x}_i\right)\right) + \widehat{\mu}^{(T)}\left(\mathbf{x}_i\right) \tag{5.4.0.5}$$

where $\widehat{\mu}^{(T)}\left(\mathbf{x}_i\right)$ is the estimate for the conditional mean function, $\mu^{(T)}\left(\mathbf{X}_{t+1}\right)$, of Y_{t+T} at $\mathbf{X}_{t+1} = \mathbf{x}_i$.

An extensive Monte Carlo study would need to be carried out to determine which of these formulae is better for periods $T > 1$. One way of performing such study would be to determine the mean average squared error (MASE) for $\Im_{Y_t}\Big(\widehat{VaR}^{(T)}_{\theta,\varphi}\Big(\mathbf{x}_i\Big)\Big)$, at levels $\varphi = 0.95, 0.99, 0.995$, against T, for $T = 2, \ldots,$. On real data situation, the performance of the expected shortfall formulae may be investigated by evaluating the magnitude of the descrepancy between Y_{t+T} and estimate, $\Im_{Y_t}\Big(\widehat{VaR}^{(T)}_{\theta,\varphi}\Big(\mathbf{X}_{t+1}\Big)\Big)$, in the event of failures in the VaR models for $T = 2\ldots, 10$ and at levels $\varphi = 0.95, 0.99, 0.995$. Define the theoretical residuals as

$$
\begin{aligned}
r_{t+T} &= \frac{Y_{t+T} - \Im_{Y_T}\Big(VaR^{(T)}_{\theta,\varphi}\Big(\mathbf{X}_{t+1}\Big)\Big)}{\sigma^{(T)}_{\theta}\Big(\mathbf{X}_{t+1}\Big)} \\
&= Z_{t+T} - \Im_{Z_t}\Big(q^z_{\varphi}\Big)
\end{aligned}
$$

Since Z_{t+T} in our model are iid with zero θ-quantile and unit scale, then $r_{t+T}\mathbf{I}_{\{Z_{t+T}-q^z_{\varphi}>0\}}$ are also iid with zero expectation and some constant variance. The empirical version of the failures can be formulated as $\widehat{r}_{t+T}\mathbf{I}_{\{Y_{t+T}-\widehat{VaR}^{(T)}_{\theta,\varphi}(\mathbf{X}_{t+1})>0\}}$, where $\widehat{r}_{t+T}$ is the estimate of r_{t+T} using the estimated quantities. Under the null hypothesis that the functions $VaR^{(T)}_{\theta,\varphi}\Big(\mathbf{X}_{t+1}\Big)$, $\sigma^{(T)}_{\theta}\Big(\mathbf{X}_{t+1}\Big)$ and the first moment of the truncated errors, $E\Big[Z_{t+T}\mathbf{I}_{\{Z_{t+T}>q^z_{\varphi}\}}\Big]$, are correctly estimated, the residuals should behave like an iid sample with zero mean and some constant variance and therefore bootstrap methods in Efron and Tibshirani [38] could be used to test one sided hypothesis against the alternative that the conditional expected shortfall is systematically underestimated. See McNeil and Frey [88], where the procedure is used in the case of AR-GARCH approach.

5.5 Conclusion

This chapter dealt with the problem of capital adequacy requirement posed by VaR as a risk measure. We have proposed two semiparametric estimator of the conditional expected shortfall. The first estimator is based on fitting the GPD to the excesses over VaR and the second one, which we consider more general, is based on Koenker-Bassett loss function. The consistency for the two estimator have also been discussed.

References

[1] Abberger, K., (1996) , Nictparametrische Schaetzung bedingter Quantile in Zeitreihen - Mit Anwendung auf Financemarketdaten ,*Hartung-Gore Verlag, Konstanz*

[2] Acerbi,C., C. Nordio and C. Sirtori (2001), Expected shortfall as a tool for risk management ,*Working paper. http://www.gloriamundi.org/var/wps.html.*

[3] Acerbi, C. and Tasche, D. (2002), On the coherence of Expected Shortfall,*Journal of Banking and Finance*, **26**(7), 1487-1503.

[4] Acerbi,C. and Tasche, D. (2002), Expected shortfall: A natural coherent alternative to VaR ,*Economic Notes* .

[5] Alexander, C. and Leigh, C (1997), On the covariance matrices used in value at risk models, *Journal of Derivatives*, **4**,50-62

[6] Artzner,P., F. Delbaen, J.-M. Eber, and D. Heath, (1999), "Coherent measures of risk", *Math.Finance* **9**, 203-228.

[7] Barone-Adesi,G., Bourgoin, F., and Giannopoulos (1998), "Don't look back", *RSIK* , **11**(8) .

[8] Bassett, G., Koenker, R. and Kordas, G. (2002), Pessimistic Portfolio allocation and choquent expected utility, Statistical Data Analysis Based on the L_1-norm and Related Methods, Yadolah Dodge ed. Birkhauser, p. 97-108 .

[9] Basel committee on Banking Supervion (1996), Amendment to the capital Accord to incorporate Market Risks.

[10] Beran, J. and Ocker, D. (2001), Volatility of stock market indices-An analysis based on SEMIFAR models,*Journal of business and Economic statistics* **19**, 103-116.

[11] Bickel, P. J and Lehmann, E. (1976), Descriptive statistics for nonparametric models.III: Dispersion, *Annals of statistics* **4**, 1139-1159.

[12] Bickel, P. J. (1978), Using residuals robustly, *Annals of statistics* **6**, 266-291.

[13] Bollerslev, T., (1986) Generalized Autoregressive Conditional Heteroskedastic *Journal of Econometrica* **31**, 307-327.

[14] Bollerslev, T. (1987), A conditionaly heteroscedastic time series model for speculative prices and rates of returns,*Review of econometric and statistics* **69**, 542-547.

[15] Bollerslev, T., Chou, R., Kroner, K.,(1992), ARCH modeling in finance: A review of the theory and empirical evidence,*Journal of Econometrics* **50**, 5-59.

[16] Bollerslev,T., Engle, R.F and Nelson, D. (1994), ARCH models. In R.F. Engle and D. L. McFaden (eds.), Handbook of Economic, vol. 4. Amsterdam: North-Holland.

[17] Bollerslev, T. and Wooldridge, J. M (1992), Quasi-Maximum Likelihood Estimation and inference in Dynamic models with Time varying covariances, *Econometric Reviews* **11**, 143-172.

[18] Buehlmann, P. and McNeil, A. , Nonparametric GARCH models,*Research Report* **No. 90**, Seminar fuer Statistik, ETH, Zuerich, Switzerland .

[19] Boudoukh, J., Richardson, M., and Whitelaw, R. (1998), The best of both worlds ,*RISK* **11**(May), 64-67.

[20] Cai, Z and Roussas, G. G, (1997), Smooth estimate of quantiles under association ,*Statistics and probability letters* **36**, 275-287.

[21] Carbon, M. (1983), Bernstein's Inequality for Strong Mixing, Non-Stationary Processes: Applications, *C. R. Acad., Scienc., Paris* **I**, 297, pp 303-306.

[22] Carrol, R. J. and Ruppert, D.(1988), Transformation and weighting in regression, London: Chapman and Hall.

[23] Cassidy, C. and Gizycki, M. (1997), Measuring traded market risk: Value-at-risk and backtesting techniques, Reserve bank of Australia.

[24] Chaudhuri, P. (1991), Nonparametric estimation of regression quantiles and their local Bahadur representation, ,*The Annals of Statistics* **19**(2), 760-777.

[25] Chen, R. and Tsay (1993), Functional coefficient autoregressive models, *Journal of the American Statistical Association* , **88**, 298-308

[26] Chesney, M. and Scott, L., (1989), Pricing European Currency: A comparison of the modified Black-scholes model and a random variance model, *Journal of finance and Quantitative analysis* **24**, 267-284.

[27] Collomb, G. (1980), Estimating nonparametriqe de probabilites conditionnelles, *C. R. Acad. Sci. Paris Sr I Math.* **291**, 427-430.

[28] Collomb, G. (1984), Propietés de convergence presque complete du predicteur añoyau, *Z. Wahr. Verw. Geb.*, **66**, 441-460.

[29] Collomb,G and Haerdle,W, (1986), Strong uniform convergence rates in robust non-parametric time series analysis and prediction: Kernel regression estimation from dependent observations, *Stochastic processes and their applications* **23**, 77-89.

[30] Dacorogna, M. M., Muller, U. A., Pictet, O. V. and de Vries, C. G., (1995) , The distribution of extremal foreign rate returns in extremely large data set,*Tinbergen Institute working paper* **TI**, 95-70.

[31] Danielsson, J. and Vries de, C.G., (1998) , Beyond the sample: extreme quantile and probability estimation, london School of Econometric, Discussion paper 298

[32] Doukhan, P. (1994), Mixing: Properties and Examples, Lecture notes in statistics, *Springer-Verlag, New-York.*

[33] Dowd, K (1998), Beyond Value at Risk: The new science of Risk managent, *Wiley, Chechester*

[34] Drost, F. C.,Klaassen, C. A. J and Werker,B.J.M (1994), Adaptiveness in time series models. In P. Mandl and M. Huskova(eds), Asymtotic statistics: Proceedings of the 5th Prague symposium. New York. Springer-verlag.

[35] Dudewicz, J. and Mishra, N., (1988), Modern mathematical statistics ,*John Wiley and Sons, Inc.*

[36] Dudewicz, E.J. (1976), Introduction to statistics and probability ,*Holt, Rinehart and Winston, New York*

[37] Duffie, D. and Pan, J. (1997), An overview of Value at Risk, *Journal of Derivatives*, **4**, 7-49

[38] Efron, B. and Tibshirani,R. (1993), *An introduction to the bootstrap*, Chapman and Hall, New York.

[39] Embrechts,P. Klueppelberg, C. and Mikosch, T (1997), Modeling Extremal Events for Insurance and Finance, *Springer, Berlin.*

[40] Engle, R.,(1982), "Autoregressive conditional heteroscedasticity with estimates of the variance of U.K inflation," *Econometrica* **50**, 987-1008.

[41] Engelking, R. (1968), Outline of general topology , North-Holland Publishing Co., Amsterdam. Translated from the Polish by K. Sieklucki.

[42] Engle, R. and Manganelli, S. (1999), CAViaR: Conditional Autoregressive Value at Risk by Regression Quantile, *DP UCSD.*

[43] Fama, E (1965), The behaviour of stock markets price, *Journal of Business* , **38**, 34-105

[44] Fama, E. T. and Miller, M. H. (1972) , The theory of finance,*Dryden Press.*

[45] Fan,J. and Gijbels, I. (1996), Local polynomial modelling and its applications, *Monographs on statistics and applied probability***66**,Chapman and Hall.

[46] Fan,J., Hu, T., C., and Truong, Y., K., (1994), " Robust nonparametric function estimation, *Scandinavian journal of statistics* **21** 433-446.

[47] Fan, J., Yao, Q. and Tong, H. (1996), Estimation of conditional densities and sensitivity measures in nonlinear dynamics system ,*Biometrika* **83**, 189-216.

[48] Feng, Y, (2002), Simultaneous modelling of conditional heteroskedasticity and scale change, *http://www.ub.uni-kostanz.de/kops/volltexte/2002/828.*

[49] Feller, W. (1971), An introduction to probability theory and its applications ,*Vol. II. 2nd ed. Willey, New York.*

[50] Fishburn, P. C., (1980), Stochastic Dominance and moments of distributions, , *Mathematics of operations Research* , **5**, 94-100

[51] Franke, J.,(1999),Nonlinear and nonparametric methods for analyzing financial time series. Kall, Peter (ed.) et al., Operations research proceedings 1998. Selected papers of the 4th annual international conference, OR98 held in Zuerich, Switzerland, August 31-September 3, 1998. *Berlin: Springer*. 271-282

[52] Franke, J., Kreiss, J-P., Mammen, E. (2002), Bootstrap of kernel smoothing in nonlinear time series, *Bernoulli* **8**(1), 1-37.

[53] Franke,J., Kreiss, J.-P., Mammen, E. and Neumann, M. H., (2000),Properties of nonparametric autoregressive bootstrap, *Manuscript*

[54] Galombos, J. (1978), The asymptotic theory of extreme order statistics , *Wiley, New York* .

[55] Gorodetskii, V.,V.(1977), On the strong mixing property for linear sequence, *Theory of probability and its Application*, **22**,411-413

[56] Gyoerfi, L., Haerdle, W., Sarda, P. and Vieu, P. (1989), Nonparametric curve estimation from time series, Lecture note in statistics, Springer-Verlag, Heidelberg.

[57] Haerdle, W and Luckhaus, S (1984), Uniform consistency of a class of regression estimators, *Ann. Statistics*, **12**, 612-623.

[58] Haerdle, W. (1990), Applied nonparametric regression, Cambrdge University Press, Cambridge.

[59] Haerdle, W. and Marron, J. S., (1985), Optimal bandwidth selection in nonparametric regression estimation, *Annal of Statistics* **13, 4**, 1465-1481.

[60] Haerdle, W., Hall, P. and Marron, J.S (1988) How far are asymptotically choosen regression smoothers from their optimum? (with discussion), *Journal of American Statistical Society* **83**, 86-101.

[61] Haerdle, W., Vieu, P. and Hart, J. (1989), Asymptitic optimal data-driven bandwidths for regression under dependence. Preprint.

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

[63] Hall, P. and Yao, Q. (2002), Inference in ARCH and GARCH models with heavy-tailed errors, *Econometrica* to appear.

[64] Hill,B.M. (1975), A simple general approach to inference about the tail of a distribution ,*Ann. Stat.* **3**, 1163-1174.

[65] Honda, T. (2000), Nonparametric estimation of a conditional quantile for α-mixing processes, ,*Ann. Inst. Statist. Math.* **52**(3), 459-470.

[66] Hosking, J.R.M. and Wallis, R. R. (1987), Parameter and quantile estimation for the generalized pareto distribution,*Technometrics* **29**, 339-349.

[67] Huber, P. J., (1981), Robust statistics, *John Wiley and Sons* .

[68] Huesler, J., and Reiss, R. D. (1987), Extreme value theory, *Lecture note in Statistics*, **51**, Springer-Verlag

[69] Hull, J. and White, A.,(1998), Value-at-Risk when daily changes in market variables are not normally distributed,*Journal of Derivatives* **5**(3), 9-19.

[70] Hull, J. and White, A.,(1987) The pricing of options on assets with stochastic volatilities *Journal of finance* **42**, 281-300.

[71] Jaeckel, L. A (1971), Some flexible estimates of location,*Annals of mathematical statistics* **42** 1540-1552.

[72] Jaeckel, L. A (1971),Robust estimates of location,*Annals of mathematical statistics* **42** 1020-1034.

[73] JP Morgan (1996), Risk Metrics-Technical Document, 4th Edition, *New-York, J. P. Morgan*

[74] Jorion, P (1997), Value-at-Risk: The new Benchmark for controlling market risk, *Irwin, Chicago*

[75] Koenker, R. and Bassett, G. (1978), Regression quantiles, *Econometrica* **46**, 33-50.

[76] Koenker, R. and Zhao, Q. (1996), Conditional Quantile Estimation and inference for ARCH models, *Econometric Theory* **12**, 793-813.

[77] Kozek, A.S (2001), On M-estimators and normal quantile, accepted for publication in the Annals of Statistics, (is also obtainable at http://wwwstat.mq.edu.au/staff/akozek/akozek).

[78] Krishnaiah, Y. S. Rama (1990), On the Glivenko-Cantelli theorem for generalized empirical processes based on strong mixing sequences,*Stat. Probab. Letter* **47**(7), 2863-2875.

[79] Kupiec, P. (1995), Techniques for varifying the accuracy of risk measurement models ,*Journal of derivatives* **2**, 73-84.

[80] Leadbetter, M., R., Lindgren, G. and Rootzen, H. (1983), Extreme and related properties of random sequence and processes *Springer-Verlag, New York-Berlin*.

[81] Linton, O.,(1994), Adaptive estimation of ARCH models, *Econometric Theory* **9**, 539-569.

[82] Longin, F. (1996), The asymptotic distribution of extreme stock market returns ,*J. Busin.* , **63**, 383-408.

[83] Mandelbrot, B. (1963), The variation of certain speculative prices *Journal of Business* , **36**, 394-419

[84] Masry, E. and Tjostheim, O., (1995), Nonparametric estimation and identification of nonlinear $ARCH$ time series: Strong convergence and asymptotic normality, *Econometric Theory* , **11**, 258-289

[85] Masry, E. and Tjostheim, O., (1997), Additive nonlinear ARX time series and projection estimates, *Econometric Theory* , **13**, 214-252

[86] Masry Elias (2001), Local linear regression estimation under long-range dependence: Strong consistency and rates,*IEEE Transactions on information Theory* **47**(7), 2863-2875.

[87] McNeil A. J. (1999), Extreme value theory for risk managers *Internal Modelling and CAD II published by RISK Books* .

[88] McNeil, A., J. and Frey, R.,(2000) Estimation of Tail-Related Risk measures for Heteroscedastic Financial Time series: An Extreme Approach, *Journal of Empirical Finance* , **7**, 271-300.

[89] Michels, P. (1992), Nichtparametrische Analyse und Prognose von Zeitreihen, *Physics-Verlag* , Heidelberg.

[90] Mises, R. von (1936), La distribution la plus grande de n valeurs. Reprint in selected papers II. *Amer. Math. Soc., Providence, R. I.* ,1954 271-294.

[91] Nelson, D.(1991), Conditional heteroscedasticity in asset returns, *Econometrica* **59**, 349-370.

[92] Nelson, D. and Foster,D. (1994), Asymptotic filtering theory for univariate ARCH models,*Econometrica* **62**, 1-42.

[93] Newey , K. W. and Powell, J. (1987), Asymmetric least squares estimation and testing,*Econometrica* **55**(4), 819-647.

[94] Parzen,E, (1962), On estimation of a probability density function and mode, *Annal of mathematical statistics* **33**, 1065-1076.

[95] Pflug, G. G., (2000), Some remarks on the Value-at-Risk and the Conditional Value-at-Risk. In: Probabilistic Constrained Optimization - Methodology and Applications (S. Uryasev editor) *Kluwer Academic Publishers*, **ISBN 0-7923-6644-1**, pp. 272 - 281.

[96] Robinson, P.M. (1983), Nonparametric estimators for time series, *Journal of time series anal.* **4**, 185-207.

[97] Rosenblatt,M (1965), Remarks on some nonparametric estimates of density function, *Annals of mathematical statistics* **27**, 832-837.

[98] Rockafellar,R., and Uryasev, S. (2000), Optimization of conditional VaR, *Journal of Risk* **2**, 21-41.

[99] Rockafellar, R. T., and Uryasev, S. (2001), Conditional VaR for general loss distributions ,*Research report 2001-5, ISE Dept., University of Florida* .

[100] Scaillet, O. (2000), Nonparametric estimation of conditional expected shortfall ,*http://www.econ.ucl.ac.be/IRES/CSSSP* .

[101] Schwert, G. W. (1989), Why does stock market volatility change over time?*Journal of finance* **44**, 1115-1154.

[102] Shephard, N., (1996)," Statistical aspects of ARCH and stochastic volatility" in time series models in Econometrics, Finance and other Fields, ed. D. Cox, D. Hinkley and O. Barndorff-Nielen, pp. 1-55, London. Chapman and Hall.

[103] Smith, R. L. (1987),Estimating tails of probability distributions ,*The annals of statistics* **15**(3), 1174-1207.

[104] Stein, E. and Stein, J (1991), Stock price distributions with stochastic volatility: An analytic approach. *The review of financial studies* **4**, 727-752.

[105] Stone, C.(1982), "Optimal global rates of convergence for nonparametric regression, *Annal of Statistics* **10**, 1040-1053.

[106] Stulz, R. (1998), Derivatives, Risk Management and Financial Engineering, *Southern Publishing*.

[107] Taylor, S.,(1986), Modeling financial time series, New York, Wiley.

[108] Todorovic, P. and Zelenhasic, E. (1970), A stochastic model for flood analysis, *Water Resources Research* **6**, 1641-1648.

[109] Turner, A. L. and Weigel,E. J. (1992), Daily stock market volatility, *Man. Sci.* **38**, 1586-1609.

[110] Volkonskii, V. A. and Rozanov, Yu. A. (1959), Some limit theorems for random functions *Theory Prob. Appl.* , **4**, 178-197.

[111] Welsh, A. H., Carrol, R. J. and Ruppert, D. (1994) Fitting heteroscedastic regression models, *Journal of the American statistical Association* **89**(425), 100-116.

[112] Withers, C.,S., Condition for linear processes to be strongly mixing, *Zeitschrift fur Wahrscheinlichkeitstheorie verwandte Gebiete*, **57**,477-480

[113] White, H. and Wooldridge, J. (1991), Some results on Sieve estimation with dependent observations, " in W. Barnett, J. Powell and G. Tauchen, eds., *Nonparametric and Semi-parametric Methods in Econometrics and Statistics.* New York: Cambridge University Press.(To appear).

[114] Yu, K. and Jones, M. C. (1998), Local linear quantile regression, *Journal of the American statistical association* **93**, 228-237.