



HAL
open science

Conditional VaR and expected shortfall: a new functional approach

Frédéric Ferraty, Alejandro Quintela-Del-Río

► **To cite this version:**

Frédéric Ferraty, Alejandro Quintela-Del-Río. Conditional VaR and expected shortfall: a new functional approach. *Econometric Reviews*, 2016, 35 (2), pp.263-292. 10.1080/07474938.2013.807107 . hal-01980180

HAL Id: hal-01980180

<https://hal.science/hal-01980180>

Submitted on 25 Jan 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Conditional VaR and expected shortfall: a new functional approach

FERRATY, FRÉDÉRIC

Institut de Mathématiques de Toulouse, Université de Toulouse,

31062 Toulouse cedex 9, France. e-mail: ferraty@cict.fr

QUINTELA-DEL-RÍO, ALEJANDRO

Departamento de Matemáticas, Facultad de Informática,

Universidad de A Coruña, A Coruña, Spain. e-mail: aquintela@udc.es

Abstract

We estimate two well-known risk measures, the Value-at-risk and the expected shortfall, conditionally to a functional variable (i.e., a random variable valued in some semi(pseudo)-metric space). We use nonparametric kernel estimation for constructing estimators of these quantities, under general dependence conditions. Theoretical properties are stated whereas practical aspects are illustrated on simulated data: nonlinear functional and GARCH(1,1) models. Some ideas on bandwidth selection using bootstrap are introduced. Finally, an empirical example is given through data of the S&P 500 time series.

Corresponding author: Alejandro Quintela-del-Río.

Key words: asymptotic properties; conditional expected shortfall; conditional value-at-risk; functional process; functional kernel estimator; functional nonparametric estimation; international financial index.

1 Introduction

A major concern for regulators and owners of financial institutions is the risk analysis. The Value-at-Risk (VaR) (Embrechts et al., 1997) is one of the most common risk measures used in finances. It measures down-side risk and is determined for a given probability level p . In a typical situation, measuring losses, the VaR is the lowest value which exceeds this level (that is, the quantile of the loss distributions). The expected shortfall (ES) (Acerbi, 2002) is the expected loss given the loss exceeds the VaR threshold. From the Basel Accords (1996, 2006), the VaR (and more recently the ES) forms the essential basis of the determination of market risk capital. Many banks compute VaR for managing the financial hazard of their portfolios (Gilli and K ellezi, 2006). Several methods have been developed to calculate the VaR. See, for example, the paper of Bao et al. (2006) for an exhaustive description of up to 16 methods to estimate the VaR, with applications to different financial time series.

Usually, one collects additional information \mathcal{I} (i.e., past observed returns, economical exogenous covariates, etc.) and to take into account such relevant information \mathcal{I} , the conditional VaR (CVaR) of the variable Y_t (being Y_t the risk or loss variable which can be the negative logarithm of returns at time t) is defined, for a fixed level p , as the value $\nu_p(t)$ such that $P(Y_t > \nu_p(t)|\mathcal{I}) = p$. The conditional ES (CES) is defined as $\mu_p(t) = E[Y_t|Y_t > \nu_p(t), \mathcal{I}]$. Most studies estimate CVaR through quantile estimation (see for instance Gaglianone et al., 2009 and references therein). When the conditional information is a finite-dimensional predictor, Scaillet (2004 and 2005) proposed to estimate CES and CVaR nonparametrically by using kernel estimators. Other works based on nonparametric estimators are those of Chen (2007), Chen and Tang (2005) or Cai and Wang (2008). A related approximation is that of Cosma et al. (2007), where they introduce a new approach on shape preserving estimation of cumulative distribution functions and probability density functions, using the wavelet methodology for multivariate dependent data (these estimators preserve shape constraints such as monotonicity, positivity and integration to one, and allow

for low spatial regularity of the underlying functions). The paper also discusses CVaR estimation for financial time series data. Other noteworthy works are those of Fermanian and Scaillet (2005), discussing nonparametric estimation of the VaR and the expected shortfall in a credit environment where netting agreements are present; or Linton and Xiao (2011) and Zhu and Galbraith (2011), estimating the expected shortfall in the heavy tail case. Comparable studies of nonparametric functional techniques applied to quantile estimation (but in an environmental setting) can be seen in Quintela and Francisco (2011).

In this paper, we consider the estimation of the CES and CVaR when one has at hand a covariate \mathcal{X} valued in some semi-metric¹ space $(\mathcal{F}, d(\cdot, \cdot))$. Such a random variable is called functional covariate. This amounts equivalently to considering an infinite-dimensional covariate instead of a finite one as in Cai and Wang (2008). This situation may occur when a continuous process of returns $\{Z_t\}_{t \in [0, (n+1)\tau]}$ is cut into $n + 1$ pieces $\mathcal{X}_i = \{\mathcal{X}_i(\delta) = Z_{(i-1)\tau+\delta}; 0 \leq \delta < \tau\}$ ($i = 1, \dots, n + 1$). This mechanism which consists in building the process $\{\mathcal{X}_i\}_{i=1, \dots, n}$ of functional variables (also called functional process) from the continuous process $\{Z_t\}_{t \in (0, n\tau]}$ is quite standard now (for more details, see the monograph of Bosq, 2000, and precursor references therein). Moreover, for $i = 1, \dots, n$, the building of a \mathcal{X}_{i+1} -measurable real random variable (r.r.v.) Y_i allows us to consider a new statistical sample of n pairs $(\mathcal{X}_1, Y_1), \dots, (\mathcal{X}_n, Y_n)$ identically distributed but not necessarily independent. Various choices for the Y_i 's are possible: a future return (i.e., $Y_i = \mathcal{X}_{i+1}(\delta)$ with $\delta \in [0, \tau)$), or the supremum return over a future range (i.e., $Y_i = \sup_{\delta \in [0, \tau)} \mathcal{X}_{i+1}(\delta)$), or any other interesting scalar quantity computed from \mathcal{X}_{i+1} . It makes sense now to study features of the distribution of the r.r.v. Y conditionally to the functional variable \mathcal{X} . This statistical issue has been thoroughly investigated in the functional data literature, and the reader can find numerous works dealing with conditional cumulative distribution, conditional density, conditional mean, conditional mode or conditional quantiles (for general overviews, see the monograph of Ferraty

¹A semi-metric (also called pseudo-metric) $d(\cdot, \cdot)$ is a metric allowing $d(\chi_1, \chi_2) = 0$ for some $\chi_1 \neq \chi_2$

and Vieu, 2006, and the collective book edited by Ferraty and Romain, 2011; for more recent advances, see for instance Lemdani et al., 2009, or Ferraty et al., 2010). In such a conditional functional situation, the CVaR is the quantity $\nu_p(\chi)$ such that $1 - F(\nu_p(\chi)|\chi) = p$ where $F(\cdot|\chi)$ is the cumulative distribution function (c.d.f.) conditionally to $\mathcal{X} = \chi$ and the CES is given by

$$\mu_p(\chi) = E(Y|Y > \nu_p(\chi), \mathcal{X} = \chi) = \int_{\nu_p(\chi)}^{\infty} y f(y|\chi) dy/p, \quad (1)$$

where $f(y|\chi)$ is the conditional probability distribution function of Y given $\mathcal{X} = \chi$.

This work is organized in the following way. In Section 2 a nonparametric estimation of the CES is proposed based on a functional kernel estimator of the CVaR. In Section 3 the asymptotic optimality properties for these estimates are established. Section 4 proposes a discussion about asymptotic distribution and more complex situations. In Section 5, some advantages of our functional proposals in practice are highlighted, through two simulation studies. The first one is in line to that developed in Cai and Wang (2008), regarding the CVaR and CES estimation in a nonlinear functional regression model (Subsection 5.1). The second one investigates the functional estimates for a widely considered model in financial time series, like the GARCH(1,1) (Subsection 5.3). In Subsection 5.2, a quick look at bandwidth selection is taken by a bootstrap procedure. The performance in real applications is investigated with a data sample of the financial index S&P 500 in Subsection 5.4. Section 6 is devoted to conclusions. In a final Appendix some details of the proofs are exposed.

2 Kernel estimation of CVaR and CES

In this section, we focus on the nonparametric estimation of the conditional value-at-risk and the conditional expected shortfall. To this end, let us first consider n pairs $(\mathcal{X}_i, Y_i)_{i=1, \dots, n}$ identically distributed as (\mathcal{X}, Y) where \mathcal{X} is the functional covariate and Y the r.r.v. of interest. The adopted estimating procedure

is based on the conditional kernel estimator $\widehat{F}(\cdot|\chi)$ of the conditional c.d.f. $F(\cdot|\chi)$ (see Ferraty et al., 2005 and 2006) which is defined by the following weighted average:

$$\widehat{F}(y|\chi) = \sum_{i=1}^n W_h(\chi, \mathcal{X}_i) H(g^{-1}(y - Y_i)), \quad (2)$$

where $W_h(\chi, \mathcal{X}_i) = K_a(h^{-1}d(\chi, \mathcal{X}_i)) / \sum_{i=1}^n K_a(h^{-1}d(\chi, \mathcal{X}_i))$, $K_a(\cdot)$ is an asymmetrical kernel function, $H(\cdot)$ is a cumulative distribution function such that $H(v) = \int_{-\infty}^v K_s(u) du$ with $K_s(\cdot)$ a symmetrical kernel function, and h and g are two positive smoothing parameters (bandwidths).

According to the definition of the CVaR, it is clear that $\nu_p(\chi) = F^{-1}((1-p)|\chi)$. So, from $\widehat{F}(\cdot|\chi)$, it is easy to derive the nonparametric estimator of $\nu_p(\chi)$ as

$$\widehat{\nu}_p(\chi) = \widehat{F}^{-1}((1-p)|\chi). \quad (3)$$

Once an estimator of the CVaR is obtained, and by remarking that $\widehat{f}(y|\chi) = \partial\widehat{F}(y|\chi)/\partial y$ is a kernel estimator of the p.d.f. $f(y|\chi)$, one can define the nonparametric estimator of $\mu_p(\chi)$ as:

$$\widehat{\mu}_p(\chi) = \int_{\widehat{\nu}_p(\chi)}^{\infty} y \widehat{f}(y|\chi) dy / p. \quad (4)$$

This nonparametric functional data estimator can be written without an integral term, in a more computationally tractable form. From the expression of $\widehat{f}(y|\chi)$ one can write:

$$\widehat{\mu}_p(\chi) = (pg)^{-1} \sum_{i=1}^n W_h(\chi, \mathcal{X}_i) \int_{\widehat{\nu}_p(\chi)}^{+\infty} y K_s\left(\frac{y - Y_i}{g}\right) dy \quad (5)$$

$$= (pg)^{-1} \sum_{i=1}^n W_h(\chi, \mathcal{X}_i) \left\{ g \int_{\widehat{\nu}_p(\chi)}^{+\infty} \left(\frac{y - Y_i}{g}\right) K_s\left(\frac{y - Y_i}{g}\right) dy + Y_i \int_{\widehat{\nu}_p(\chi)}^{+\infty} K_s\left(\frac{y - Y_i}{g}\right) dy \right\}, \quad (6)$$

which allows us to get

$$\widehat{\mu}_p(\chi) = p^{-1} \sum_{i=1}^n W_h(\chi, \mathcal{X}_i) \left\{ g G\left(\frac{\widehat{\nu}_p(\chi) - Y_i}{g}\right) + Y_i \left[1 - H\left(\frac{\widehat{\nu}_p(\chi) - Y_i}{g}\right) \right] \right\}, \quad (7)$$

with $G(v) = \int_v^{+\infty} u K_s(u) du$.

3 Almost complete asymptotic properties

In this section, we focus on the almost complete ² asymptotic properties of the CVaR and the CES. Let us first introduce assumptions before giving the main results. Let $B(\chi, h)$ be the ball centered at χ and of radius h , let S be a fixed compact subset of \mathbb{R} such that $\exists \eta > 0, \nu_p(\chi) + \eta \in S$ and let N_χ be a fixed neighborhood of χ .

Assumptions on the distributions of the variables.

(H0) $\forall h > 0, P(\mathcal{X} \in B(\chi, h)) = \phi_\chi(h) > O$,

(H1) $(\mathcal{X}_i, Y_i)_{i \in \mathbb{N}}$ is a strong mixing³ sequence with arithmetic coefficient:

$$\exists \kappa > 3, \exists C > 0 : \forall n \in \mathbb{N}, \alpha(n) \leq C n^{-\kappa}, \quad (8)$$

(H2) $\sup_{i \neq j} P\{(\mathcal{X}_i, \mathcal{X}_j) \in B(\chi, h) \times B(\chi, h)\} = O\left(\phi_\chi^{2-1/(\kappa+1)}(h)\right)$,

(H3) $\forall i \neq i'$, the conditional density of $(Y_i, Y_{i'})$ given $(\mathcal{X}_i, \mathcal{X}_{i'})$ is continuous at $(\nu_p(\chi), \nu_p(\chi))$ uniformly on $(\mathcal{X}_i, \mathcal{X}_{i'}) \in N_\chi \times N_\chi$. In addition, one assumes that, for $m \in \{1, 2\}$, $E(|Y|^m | \mathcal{X}) = \sigma_m(\mathcal{X})$ where $\sigma_m(\cdot)$ is a bounded continuous function and $E(|Y_1| | Y_2 | \mathcal{X}_1, \mathcal{X}_2) = \sigma_{12}(\mathcal{X}_1, \mathcal{X}_2)$ where $\sigma_{12}(\cdot, \cdot)$ is also a bounded continuous (with respect to both arguments) function.

Nonparametric regularity assumptions.

The value $j > 1$ exists such that:

²We consider the almost complete convergence: Let $(z_n)_n$ be a sequence of real r.v.'s; we say that z_n converges almost completely (a.co.) to zero if and only if, $\forall \epsilon > 0, \sum_{n=1}^{\infty} P(|z_n| > \epsilon) < \infty$. Moreover, let $(u_n)_n$ be a sequence of positive real numbers; we say that $z_n = O(u_n)$ a.co. if and only if $\exists \epsilon > 0, \sum_{n=1}^{\infty} P(|z_n| > \epsilon u_n) < \infty$. This kind of convergence implies both almost sure convergence and convergence in probability.

³A sequence $(Z_n)_{n \in \mathbb{Z}}$ is called strong mixing sequence if and only if $\lim_{n \rightarrow +\infty} \alpha(n) = 0$ where $\alpha(n) = \sup_{k \in \mathbb{Z}} \sup_{A \in \mathcal{T}_{-\infty}^k} \sup_{B \in \mathcal{T}_{k+n}^{+\infty}} |P(A \cap B) - P(A)P(B)|$ and for any $(i, j) \in \mathbb{Z}^2$ such that $-\infty \leq i \leq j \leq +\infty$, \mathcal{T}_i^j is the σ -algebra generated by $(Z_l, i \leq l \leq j)$

(H4) $F(\cdot|\chi)$ is strictly increasing and j -times continuously differentiable in S ,

(H5) $\exists C > 0/\forall(y_1, y_2) \in S \times S, \forall(\chi_1, \chi_2) \in N_\chi \times N_\chi$,

$$|F^{(j)}(y_1|\chi) - F^{(j)}(y_2|\chi)| \leq C \left\{ d(\chi_1, \chi_2)^{\beta_1} + |y_1 - y_2|^{\beta_2} \right\},$$

(H6) $\forall l, 1 \leq l < j, F^{(l)}(\nu_p(\chi)|\chi) = 0$ and $|F^{(j)}(\nu_p(\chi)|\chi)| > 0$.

Assumptions on the kernel estimator.

(H7) The asymmetrical kernel $K_a(\cdot)$ is a bounded continuous function on its support $(0, 1)$ such that

$$0 < C_1 < K_a(t) < C_2 < +\infty,$$

(H8) The symmetrical kernel function $K_s(\cdot)$ has a compact support $[-1, 1]$, satisfies $m_0(K_s) = 1$,

$$m_1(K_s) = 0, m_2(K_s) < \infty \text{ with } m_k(K_s) = \int u^k K_s(u) du, \text{ and } \int |u|^{\beta_2} K_s(u) du < \infty. \text{ In addition,}$$

$$\forall l \geq j, K_s^{(l-1)}(\cdot) \text{ exists and is bounded,}$$

(H9) The integrated kernel function $H(\cdot)$ is strictly increasing such that:

- $\forall(y_1, y_2) \in S \times S, |H(y_1) - H(y_2)| \leq C |y_1 - y_2|$ and $\int_{\mathbb{R}} |t|^{\beta_2} H^{(1)}(t) dt < +\infty$,

- $\forall(y_1, y_2) \in S \times S, |H^{(j)}(y_1) - H^{(j)}(y_2)| \leq C |y_1 - y_2|$ and

$$\exists \omega > 0/\forall j' \leq j - 1, \lim_{y \rightarrow +\infty} |y|^{1+\omega} \left| H^{(j')}(y) \right| = 0, \text{ and } H^{(j')}(\cdot) \text{ is bounded,}$$

(H10) $\lim_{n \rightarrow \infty} h = 0, \lim_{n \rightarrow \infty} \log n (n \phi_\chi(h))^{-1} = 0$ and $g^2 = O((n \phi_\chi(h))^{-1})$.

Theorem 3.1 *If the conditions (H0)-(H10) hold, and if $\eta > 0$ exists such that $g \phi_\chi(h) \geq C n^{\frac{3-\kappa}{\kappa+1}+\eta}$, we*

get that:

(i) *the conditional estimated c.d.f. is consistent with the rate*

$$\sup_{y \in S} |\widehat{F}(y|\chi) - F(y|\chi)| = O(h^{\beta_1}) + O(g^{\beta_2}) + O_{a.co.} \left(\sqrt{\frac{\log n}{n \phi_\chi(h)}} \right), \quad (9)$$

(ii) the successive derivatives of the conditional estimated c.d.f. are consistent with the rate

$$\sup_{y \in S} |\widehat{F}^{(j)}(y|\chi) - F^{(j)}(y|\chi)| = O(h^{\beta_1}) + O(g^{\beta_2}) + O_{a.co.} \left(\sqrt{\frac{\log n}{n h^{2j-1} \phi_\chi(h)}} \right), \quad (10)$$

(iii) the estimated CVaR is consistent with the rate

$$\widehat{\nu}_p(\chi) - \nu_p(\chi) = O(h^{\beta_1/j}) + O(g^{\beta_2/j}) + O_{a.co.} \left(\left(\frac{\log n}{n \phi_\chi(h)} \right)^{1/(2j)} \right). \quad (11)$$

The proof of this theorem is a straightforward extension of the work of Ferraty et al. (2005). They obtained the almost sure pointwise convergence of the conditional c.d.f. and its successive derivatives as well as the conditional quantiles under arithmetic strong mixing dependence. Because one is interested in getting uniformity over a compact subset of \mathbb{R} , by using standard techniques (which involves a recovery of S ; see for instance Ferraty et al., 2006), it is easy to derive a uniform version of the results given in Ferraty et al. (2005). However, the reader will find in the Appendix the main guidelines to have an idea on how to obtain these results.

Theorem 3.2 *Under the Assumptions (H0)-(H10), and if the following inequalities*

$$\exists \eta > 0, C n^{\frac{3-\kappa}{\kappa+1} + \eta} \leq g \phi_\chi(h) \text{ and } \phi_\chi(h) \leq C' n^{\frac{1}{1-\kappa}} \quad (12)$$

are satisfied with $\kappa > (5 + \sqrt{17})/2$, we get that:

$$\widehat{\mu}_p(\chi) - \mu_p(\chi) = O(g^2) + O_{a.co.} \left(\sqrt{\frac{\log n}{n \phi_\chi(h)}} \right) + O \left((\widehat{\nu}_p(\chi) - \nu_p(\chi))^2 \right). \quad (13)$$

4 Towards asymptotic normality and more complex situation

Towards asymptotic normality. Asymptotic distribution is an interesting issue but not central in our work. Indeed, detailing such theoretical property needs additional technical conditions which would affect the clarity of our purpose and mask the main idea of this work: the usefulness of introducing a

functional data approach in estimating CVaR and CES. Nevertheless, one can give the scheme used to derive the asymptotic normality for the CVaR. On the one hand, one combines the uniform convergence of the conditional kernel estimator $\widehat{f}(\cdot|\chi)$ (Theorem 3.1 (ii) with $j = 1$), the almost complete convergence of CVaR (Theorem 3.1-(iii)) and the pointwise asymptotic normality of the conditional kernel estimator $\widehat{F}(\cdot|\chi)$ stated in Theorem 6 in Quintela-del-Rio (2008). On the other hand, it is possible to extend Theorem 2.1 in Berline et al. (2001) to the case of functional covariate. Then, as soon as $f(\nu_p(\chi)|\chi) \neq 0$ and under additional conditions that the reader can find in Quintela-del-Rio (2008), it suffices to combine this last result with the previous ones to get the asymptotic normality of the CVaR:

$$\sqrt{\frac{n \phi_\chi(h)}{\log n}} (\widehat{\nu}_p(\chi) - \nu_p(\chi)) \xrightarrow{d} \mathcal{N} \left(0, \frac{F(\nu_p(\chi)|\chi)(1 - F(\nu_p(\chi)|\chi))}{f^2(\nu_p(\chi)|\chi)} \right).$$

For the asymptotic normality of the conditional expected shortfall $\widehat{\mu}_p(\chi)$, things are not so easy and it is still an open problem. However, one can expect deriving such result with the following guidelines. The first step consists in stating the exact dominant terms in the bias and variance of $\widehat{\mu}_p(\chi)$. In a second step, one has to use the standard small/large blocks decomposition (see for instance Masry, 2005, or Quintela-del-Rio, 2008) to show that: 1) the asymptotic dominant term is deduced from the large blocks, 2) the variables based on the summations over the large blocks are asymptotically independent and fulfill the Lindeberg-Feller condition. Once all these results are stated, one will be able to directly conclude the asymptotic normality.

Towards more complex situation. Note that the pairs $(\mathcal{X}_i, Y_i)_{i=1, \dots, n}$ fulfill a strong mixing dependency condition (see **H1**); the pairs are identically distributed but dependent. As explained in the Introduction, such pairs may be built from some continuous process Z_t , and supposing the pairs identically distributed remain to consider Z_t as a stationary process. An interesting question would be: what happens when the root continuous process Z_t is not stationary? In other words, are we able to extend our results when the pairs $(\mathcal{X}_i, Y_i)_{i=1, \dots, n}$ are not identically distributed? Of course, if for any (i, i') the distribution of

(\mathcal{X}_i, Y_i) differs from $(\mathcal{X}_{i'}, Y_{i'})$, it is not possible to state the consistency of the kernel estimators of CVaR or CES because our method is based on the estimator of the conditional distribution. Nevertheless, our estimators and results can be extended in some cases when the data are not identically distributed. One can consider the situation when one observes a dataset where the pairs (\mathcal{X}_i, Y_i) are blockwisely identically distributed. For instance, the root continuous process Z_t (from which the pairs are extracted) is not stationary but it is possible to cut it into K pieces where the stationarity is fulfilled. Then, the sample of data can be split into K subsamples, in each subsample the data being identically distributed (each subsample corresponding to a stationary piece of the process Z_t). To estimate CES and CVaR conditionally to some χ , it suffices to identify the subsample corresponding to this χ . Then, use only the data contained in this subsample for estimating CES and CVaR.

5 Simulations and applications

5.1 Autoregressive functional regression model

In the first part of our simulation experiments, we focus on an autoregressive nonlinear regression model, used in Ferraty et al. (2012). Let us consider the process Z_t generated as

$$Z_t = 0.9Z_{t-1} + a_t, \quad a_t \sim \mathcal{N}(0, 0.1), \quad (14)$$

and next the average

$$Z_t^* = \frac{Z_t + Z_{t+1} + Z_{t+19}}{20}, \quad t = 1, \dots, N \quad (15)$$

We consider a total sample size of $N = 100 \times n$ consecutive times. Splitting the total series into n paths of size 100, we get n trajectories $\chi_i = \{\chi_i(t) = Z^*(t + 100 \times (i - 1)), t = 1, \dots, 100\}$, $i = 1, \dots, n$, that are the n observations of the functional explanatory variable in our experiment. In Figure 1 we show an example of the resulting process; in Figure 1(a) the total time series with size N , and in Figure 1(b)

some of the functional data χ_i .

[FIGURE 1 OVER HERE]

Some real valued response Y can be associated with each path of the process, and a natural statistical question could be to predict Y given some new trajectory of the process. This is really a functional problem in the sense that the explanatory variable (the whole past path) can be seen as a realization of a continuous random function. Such a modelling has many practical impacts including various fields of applied statistics. The reader may find for instance in Linton and Sancetta (2009) some specific real data problem, linked with risk analysis, for which there is real evidence for using infinite dimensional regressors in time series analysis.

In our case, the responses are generated by

$$Y_i = r(\mathcal{X}_i) + \varepsilon_i, \quad i = 1, \dots, n, \quad \text{with} \quad r(\mathcal{X}) = \int \mathcal{X}(t)^2 \cos(t) dt \quad \text{and} \quad \varepsilon_i \sim N(0, \sigma_0). \quad (16)$$

Each time a sample data $\mathcal{E} = \{(\chi_i, Y_i)\}_{i=1}^n$ is generated, we also generate one new functional fixed datum χ_0 , in which we calculate the CVaR and the CES. From model (16) we can check that, for fixed χ_0 , the conditional density $f(y/\chi_0)$ is Gaussian with mean $r(\chi_0)$ and standard deviation σ_0 . Several econometric models for volatility dynamics assume the conditional normality, as for instance J.P. Morgan's Riskmetrics (Riskmetrics, 1995) (see Subsection 5.3 below).

Thus, the CVaR and the CES can be exactly computed as

$$\nu_p(\chi_0) = F^{-1}((1-p)/\chi_0), \quad F \sim N(0, 1) \quad (17)$$

and

$$\mu_p(\chi_0) = p^{-1} \int_{\nu_p(\chi_0)}^{+\infty} y \frac{1}{\sigma_0 \sqrt{2\pi}} e^{-(y-r(\chi_0))^2/2\sigma_0^2} dy. \quad (18)$$

Therefore, the accuracy of our estimates can be calculated by means of the absolute errors between the true quantities and their estimations, that is,

$$AE(\mathcal{E}, \nu_p) = |\nu_p(\chi_0) - \hat{\nu}_p(\chi_0)| \quad AE(\mathcal{E}, \mu_p) = |\mu_p(\chi_0) - \hat{\mu}_p(\chi_0)|. \quad (19)$$

The various parameters of the methods were chosen in the following ways:

- The kernel function K_a was taken to be the uniform density on $(0, 1)$ and K_s the usual Epanechnikov kernel.
- The semi-metric $d(\cdot, \cdot)$ was taken to be the usual L_2 distance between curves.

A general discussion about the choice of the kernel functions and the semi-metrics is given in the book of Ferraty and Vieu (2006) and its companion R package from which all the routines used here have been extracted.

- The bandwidth parameters were selected in the intervals $[d_1, d_2]$, being $d_1 = \min \{d(\chi_i, \chi_j)\}_{i,j=1}^n$ and $d_2 = \max \{d(\chi_i, \chi_j)\}_{i,j=1}^n$ for h , and g in $[g_1, g_2]$ with $g_1 = (\max(Y_i) - \min(Y_i))/20$ and $g_2 = (\max(Y_i) - \min(Y_i))/2$ (see Quintela et al., 2011). We constructed an equidistant partition (20 values) of these intervals and we chose the bandwidths as those minimizing the error (19) defined before.
- σ_0 was chosen so it corresponds to a signal-to-noise ratio of 0.1.
- The selected value for p was 0.1 here and throughout the simulation study.

Results. For different sample sizes, from $n = 100$ to $n = 500$, we replicate this experiment $r = 500$ times (i.e., 500 samples $\mathcal{E}_1, \dots, \mathcal{E}_{500}$ are built). In Figure 2 we show the boxplot of the distributions of the absolute errors for three sample sizes (100, 250 and 500).

[FIGURE 2 OVER HERE]

From Figure 2, even if the averaged errors are quite stable, we can observe the improvement of the estimations in terms of variance when the sample size increases, confirming the asymptotic theory about the consistency of the proposed estimators.

In a second step, instead of a single datum, we generated a grid of m new functional points $\chi_0^1, \dots, \chi_0^m$, and we calculated the performance of the estimators in terms of the mean absolute deviation error, defined as $MAE(\mathcal{E}, \nu_p) = m^{-1} \sum_{k=1}^m |\nu_p(\chi_0^k) - \hat{\nu}_p(\chi_0^k)|$ for the CVaR (the same definition for the CES). We also ran the estimating procedure with several values for n , for m (from 10 to 50), and we replicated the experiment $r = 500$ times. The obtained results largely agree with the above conclusions for only one datum χ_0 , and are available upon request (we omit them to save space).

5.2 Some advances in bandwidth selection

We follow the ideas of Beirlant et al. (2004) in a related setting, and we use the bootstrap to choose the bandwidth for a data set at hand. From a functional data sample $\{(\chi_i, Y_i)\}_{i=1}^n$, generated as in (16), we calculate the estimates $\hat{\nu}_p(\chi_0)$ and $\hat{\mu}_p(\chi_0)$. Next, we throw with replacement a bootstrap sample $\{(\chi_i^b, Y_i^b)\}_{i=1}^n$ and we calculate $\hat{\nu}_p(\chi_0)^b$ and $\hat{\mu}_p(\chi_0)^b$. This procedure is replicated from 1 to B times, calculating thus B estimates.

Because each estimate $\hat{\nu}_p(\chi_0)$ (resp. $\hat{\mu}_p(\chi_0)$) depends on two bandwidths h and g , we choose the bootstrap bandwidths as

$$(h_{OPT}, g_{OPT}) = \arg \min_{h,g} MSE(\hat{\nu}_p(\chi_0)(h, g)) \quad (resp. \quad \arg \min_{h,g} MSE(\hat{\mu}_p(\chi_0)(h, g))), \quad (20)$$

being

$$MSE(\hat{\nu}_p(\chi_0)(h, g)) = \frac{1}{B} \sum_{b=1}^B (\hat{\nu}_p(\chi_0)^b(h, g) - \hat{\nu}_p(\chi_0))^2 \quad (21)$$

and

$$MSE(\hat{\mu}_p(\chi_0)(h, g)) = \frac{1}{B} \sum_{b=1}^B (\hat{\mu}_p(\chi_0)^b(h, g) - \hat{\mu}_p(\chi_0))^2. \quad (22)$$

To test this method, we repeat the simulation experiment of Section 5.1. The sample size was $n = 50$, the replication number 500 and the size of the bootstrap resamples was $B = 200$. The boxplot of the results for the bootstrap errors and the minimal ones is presented in Figure 3.

[FIGURE 3 OVER HERE]

Both the form of the distributions and the mean values are closer, indicating the potential interest of the bootstrap selection method. However, the high effort in computation time (we need to minimize a two-dimensional function by some numerical algorithm) means that the bootstrap methodology should be explored more exhaustively in subsequent works. See Ferraty et al. (2010b) for the first theoretical advances in bootstrap for nonparametric functional regression.

5.3 Garch models

Our second simulation experiment is based on the generation of GARCH models. Let $\{X_t, t \in \mathbb{Z}\}$ be a strictly stationary real time series, representing daily observations of a financial asset price. Suppose that the dynamics of X_t is given by

$$X_t = \mu + \sigma_t Z_t, \quad (23)$$

where the innovations Z_t are i.i.d. variables with zero mean and unit variance, and σ_t are non-negative random variables such that σ_t is a function of $X_t, X_{t-1}, X_{t-2} \dots$. Then, $EX_t = \mu$ and $Var(X_t/X_{t-1}, X_{t-2} \dots) = \sigma_t^2$. The *ARCH* model was introduced by Engle (1982), where it is supposed that the conditional variance σ_t^2 follows an autoregressive process of the ε_t^2 . The *GARCH* model (Bollerslev, 1986) arises specifying σ_t^2 not only as a function of past values of ε_t^2 , but also from σ_t^2 . In a *GARCH*(1,1) process, the conditional

variance of the mean-adjusted series $\varepsilon_t = X_t - \mu$ is given by

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2, \quad (24)$$

with the three parameters α_0, α_1 and $\beta > 0$. One main interest of these kinds of models is that the volatilities can be relatively well predicted (Danielsson and de Vries, 2000). These can be used both to forecast the returns and to predict quantiles (Engle and Patton, 2001). Berkowitz and O'Brien (2002) established that the *GARCH*(1, 1) is precise in fitting several financial series, against more complicated models, and the CVaR and the CES in these models are significantly correlated with the volatility of the returns. The intensive use of this class of models in the econometric literature makes appealing the exploration of nonparametric functional techniques with the same.

Here, with the aim of comparing the functional techniques with other classical ones, we select two widely known methods to estimate the CVaR: the empirical quantile (named historical simulation) and the parametric fitting. We are interested in the quantiles and expected shortfall for the 1-step prediction distribution, that is, $F_{X_t/\mathcal{I}}(\cdot)$.

1. *Historical Simulation (HS)*. The idea consists in assuming that the distribution of the time series X_t will remain the same in the past and in the future, and hence, the empirical distribution of historical values will be used in forecasting the CVaR (see, e.g., Jorion, 2000). The interest of this technique lies in the fact that 75% of financial institutions in Europe, Canada and US that disclose their value-at-risk use historical simulation methods (Perignon and Smith, 2008). The historical simulation uses the 1-year data to compute the CVaR with the sample quantile of value p , $\hat{\nu}_t(p)$, of the 250 previous data.
2. *Garch fitting*. If we suppose that the data are generated by a *GARCH*(1, 1) model, we can proceed

to estimating the CVaR and the CES by the following formulas (McNeil and Frey, 2000):

$$\hat{\nu}_p(t) = \hat{\mu}_t + \hat{\sigma}_t z_p, \quad (25)$$

$$\hat{\mu}_p(t) = \hat{\mu}_t + \hat{\sigma}_t E[Z/Z > z_p], \quad (26)$$

where z_p is the $(1 - p)$ -quantile for Z_t . The parameters can be estimated by a pseudo-maximum-likelihood approach. Also, if we assume some specific distribution for the innovations, such as the Normal or T, the values z_p and $E[Z/Z > z_p]$ can be estimated by calculating the residuals (after parameters estimation). Some type of extreme value analysis could be also done with the estimated residuals to avoid suppositions in this sense. Here, we will work with Gaussian innovations. An initial parameter estimation uses the first 250 data, and a model re-estimation is done every 50 days. This is known as a rolling window scheme (Giot and Laurent, 2003; Raggi and Bordignon, 2006). In these last papers, shorter lags have been experimented and show very little differences.

3. *NFDE techniques.* A general prediction problem in time series can be tackled by functional methods (Ferraty et al., 2005). If a real time series X_t is observed at N discretized times, X_1, \dots, X_N , we can construct a $(d + 1)$ -dimensional sample $\left\{(\vec{X}_i, Y_i)\right\}_{i=1}^R$ (of size $R = N - d + 1$) as follows:

$$\vec{X}_i = (X_{i-d}, \dots, X_{i-1}), Y_i = X_i. \quad (27)$$

and, for predicting a future value, the problem consists of a standard prediction problem of response Y given a d -dimensional explanatory variable. The goodness of classical parametric methods (e.g., *ARIMA* models) or nonparametric ones is seriously damaged by the curse of dimensionality. This drawback can be avoided using a functional model, considering a functional data sample $\{(\chi_i, Y_i)\}_{i=1}^R$, where $\chi_i = \vec{X}_i$ as in (27) and $Y_i = X_i$. Next, we can consider $\chi_0 = (X_{t-d+1}, \dots, X_t)$, and calculate $\hat{\nu}_p(t) = \hat{\nu}_p(\chi_0)$ and $\hat{\mu}_p(t) = \hat{\mu}_p(\chi_0)$ as in (3) and (4) in Section 2.

We tested several values of $d = 10, \dots, 50$. Here, we will only show $d = 25$ because similar results were always obtained. It is important to remark that, in practice, we would not be able to carry out this procedure with classical nonparametric regression (as in Cai and Wong, 2008), because of the curse of the dimensionality problem. Here, the bandwidths of the nonparametric estimators were selected using the bootstrap method of Section 5.2 above.

Results. Firstly, one particular simulation example is shown. We generate 1000 data corresponding to model (23) with the variance adjusted by (24). The parameters are $\mu = 0, \alpha_0 = \alpha_1 = 0.05$ and $\beta = 0.9$ and the innovations Z_t standard Gaussian (Berkowitz and O'Brien, 2002). We use the 250 first data as the in-sample, and the 750 remaining data as the out-of-sample to backtest for the CVaR estimate. Figure 4 displays the simulated data and the estimated values for the CVaR: the top one corresponding to the HS method; the middle one to the GARCH fitting, and the bottom for the NFDE. As expected, the best estimation is that provided by the GARCH parametric fitting. The NFDE estimation provides a stable one that is always close to the GARCH values. The high variability of the HS estimates provides, in general, extreme results.

[FIGURE 4 OVER HERE]

For the CES estimates, we show, in Figure 5, the GARCH estimates (top) and the NFDE estimates (bottom). The first estimates faithfully reflect the volatility of the time series, and the NFDE shows a similar behavior to that in the CVaR estimation above.

[FIGURE 5 OVER HERE]

A detailed analysis of the sequence of the CES nonparametric estimates shows that they are clearly concentrated around zero, revealing also good properties as they must accomplish (see McNeill and Frey,

2000). Concretely, tests of randomness and zero mean were carried out for the residuals, in those index $t \in T$ such that $X_t > \hat{\nu}_p(t)$. The p-value was high in both cases, in favor of the null hypothesis.

Now, we proceed to replicate this experiment $r = 1000$ times, calculating the CVaR estimates, and computing the errors (19) for each sample. Next, we show in Table 1 the average over the 1000 replications of the total differences $(N - R)^{-1} \sum_{t=R+1}^N |\hat{\nu}_p(t) - \nu_p(t)|$ between the real and estimated values (the standard deviation appears in parentheses) for each method.

[TABLE 1 OVER HERE]

In Table 2, we compare the CES estimates for Methods 2 and 3. As in the above occasions, we show the performance of the estimates by computing $(N - R)^{-1} \sum_{t=R+1}^N |\hat{\mu}_p(t) - \mu_p(t)|$ and averaging over the 1000 replications.

[TABLE 2 OVER HERE]

The results agree with those obtained in the case of the CVaR estimates. The overestimated results for the NFDE method (Figure 5) is reflected in a higher value for the average of the errors. We want to emphasize the upcoming possibilities, because the results could probably be improved by better selecting the semi-metric considered and its corresponding parameters, and by having a good automatic bandwidth selection method. Some ideas in this direction were outlined in Section 5.2 below, but a complete study would require a more exhaustive analysis, which remains open for further research.

5.4 A real data application

Our empirical analysis is based on the daily returns of the international financial index S&P 500. This index has been considered several times in the literature (Gaglianone et al., 2010, and references therein). This time series is an example of heavy tailed distribution, and has illustrative features of financial

data (volatility, skewness, excess kurtosis, non-normality, etc.). We will consider the dates between 07/April/1986 and 31/August/2010 (6156 data). With this period we have a long enough time series that contains some financial crisis periods, including the market crash of "Black Monday" (19/October/1987) (see the long spike at the left of Figure 6 below). The bull market of the second half of the 1990s, the bear market between 2000 and 2003, and the recent critical episodes since 2007 also appear.

For the sake of comparison with the other well-studied methodologies in calculating the CVaR and CES of the Section 5.3 before, we consider as the time series $X_t = -100 \log(y_t/y_{t-1})$ (y_t are the prices of the index series), and our results could be compared with others using the same index data (McNeill and Frey, 2000; Dettling and Bühlmann, 2004). Figure 6 displays the time series data (where the trend of the original series is removed).

[FIGURE 6 OVER HERE]

Backtesting. The precision of one particular model to estimate the CVaR of a financial time series (where we do not know the true theoretical value of this quantity) will be measured by a *backtesting* procedure. Basel Accords recommends financial institutions to use an *in-sample* of R past data (e.g., $R = 250$ past days or one year) to "backtest" the accuracy of their methods of loss evaluation of the CVaR. More specifically, they must estimate a maximal possible loss over a period of 1 or 10 days at $p\%$ confidence level. The backtesting consists in the quantification, in some way, of how many times the level equal to $p\%$ -CVaR is exceeded. Because of the difficulty in calculating, in practice, the loss level over a 10-day period, the financial institutions only compute a daily estimation of the CVaR, or at most, some estimation based on the "square-root-of-time" method, i.e., obtaining a multi-period prediction by multiplying the one day prediction by the square root of the length of the time horizon.

Backtesting is based on comparing the estimated values with the theoretical ones, for the *out-of-sample* data from $t = R + 1, \dots, N$. Formally, this is usually achieved by means of the "check" function of

Koenker and Bassett (1978), consisting in calculating the expected loss of the quantile for a given level p by

$$Q(p) = E[1_{(X_t > \nu_p(t))} - p][X_t - \nu_p(t)]. \quad (28)$$

Following Bertail et al. (2004), this function can be regarded as a "predictive" quasi-likelihood, and measures the lack-of-fit of a quantile model. From a real time series, it can be evaluated from the out-of-sample CVaR forecasts by

$$\hat{Q}(p) = \frac{1}{N-R} \sum_{t=R+1}^N |1_{(X_t > \hat{\nu}_p(t))} - p| |X_t - \hat{\nu}_p(t)|. \quad (29)$$

For a concrete p , the model that provides the forecasts $\{\hat{\nu}_p(t)\}_{t=R+1}^N$ with the minimum total value for $\hat{Q}(p)$ will be considered the best model.

To check the accuracy of the CES estimators, we have to analyze the sequence of the CES estimators for the *out-of-sample* data, in those indexes $t \in T$ such that $X_t > \hat{\nu}_p(t)$. If the model is correctly specified, the average of the observed values greater than $\nu_p(t)$ should be approximately equal to the predicted values. The sample quantile will be used as a benchmark for the comparison (Zhu and Galbraith, 2011) between the parametric (GARCH) and nonparametric (NFDE) estimators. Thus, we will compute the mean absolute error, as a measure of predictive out-of-sample performance:

$$MAE(p) = \frac{1}{J} \sum_{t=R+1}^N |\hat{\nu}_p(t) - OES(p)| 1_{(X_t > \nu_p(t))}, \quad (30)$$

where $OES(p)$ corresponds to the observed expected shortfall

$$OES(p) = \frac{1}{J} \sum_{t=R+1}^N X_t 1_{(X_t > \nu_p(t))}, \text{ where } J = \sum_{t=R+1}^N 1_{(X_t > \nu_p(t))},$$

that is, the sample value for the true CES (considering as the unknown $\nu_p(t)$ the sample quantile). Therefore, the mean absolute error (30) allows us to directly compare the results between parametric and nonparametric predictions.

Results. We compute the SH, GARCH(1,1) and NFDE estimates and their respective loss (29) for typical values of $p = 0.1$ and 0.05 . The results appear in Table 3. For both values of p the values of the expected loss are similar, with a slight advantage for the nonparametric procedure. Despite the big differences with the better performance of the GARCH fitting to estimate the real quantiles (see Table 1 above), NFDE can perform even better in terms of the expected loss $\hat{Q}(p)$, showing the good expectations for this methodology in this financial field.

[TABLE 3 OVER HERE]

A plot of some estimates appears in Figures 7 and 8 (dotted lines). Figure 7 shows the SH and GARCH estimates of the CVaR. Figure 8 shows the NFDE estimates of CVaR and CES.

[FIGURE 7 OVER HERE]

[FIGURE 8 OVER HERE]

With respect to the CVaR estimation, the GARCH model shows more variability, better adjusting to the real values, versus the conservative behavior of the nonparametric estimates. This is pointed out by the smoothness of the nonparametric estimator, while the parametric one is more affected by high short-run variations. The good performance of the GARCH fitting for this particular time series was indicated by several authors (e.g., Berkowitz and O'Brien, 2002). In general, GARCH fitting behaves better in crisis periods, while NFDE predicts the quantiles well in calm periods. The behavior in the risk overestimation or underestimation can also be studied checking that, except in high crisis periods, both models capture, in general, the quantile violations at the time when they occur.

The form of the CES estimates is similar to that of the CVaR above. The results for the mean absolute errors (30) are shown in Table 4, with better results for the nonparametric estimates in both

values of p . The nonparametric estimates of the CES rarely overestimate the quantile threshold, and this is the principal reason why the nonparametric estimates get better results in the averaged absolute error.

[TABLE 4 OVER HERE]

From the results of Tables 3 and 4, we verify the competitiveness of our nonparametric procedure with other classical ones. It is clear that more detailed studies could be performed using other more complex parametric models (Bao et al., 2006; Zhu and Galbraith, 2011), and that the effectiveness of functional methods could be extended through new bandwidth selection methods. Both issues should be investigated further.

6 Conclusions

CVaR and CES are, nowadays, two valuable factors in decision making of market analysis. Many mathematical methods have been studied to approximate these two parameters. When no knowledge of the shape of their theoretical form is assumed, nonparametric estimation can be performed directly through a time series of one or several portfolio returns. Here, we have worked with nonparametric methods combined with functional data. Based on a functional scheme, nonparametric techniques can consider more information of past series without being affected by dimensionality problems. The functional nonparametric estimators of the CVaR and CES have been proven to be asymptotically optimal, under general dependence conditions (mixing) and in an almost complete sense. The performance in practice of the proposed estimators has been examined under a nonlinear functional regression model and for a parametric GARCH(1,1) time series model. In general, the approaches proposed in this paper yielded good estimates of the parameters of interest, checked by the absolute errors and the backtesting results. Moreover, as seen in related works (Ferraty et al., 2010b), bootstrap appears as a promising approach to

obtain bandwidth parameters for a data set at hand, perhaps penalized by a high computation time. Of course, this work aims to be just a teaser of what appears to be a new way of estimating parameters of interest in financial series. Further work should be focused on comparing this method with different and more advanced parametric models, as well as on the development of computationally efficient techniques for bandwidth selection.

Acknowledgements

All the participants of the STAPH group on Functional Statistics in the Mathematical Institute of Toulouse are greatly acknowledged for their helpful comments on this work (activities of this group are available at the webpage <http://www.math-univ.toulouse.fr/staph>).

The authors are also grateful to two anonymous Reviewers and the Associate Editor, whose constructive remarks greatly helped improve the manuscript.

A Details of proofs

Before going on, note the Fuk-Nagaev inequality (see for instance Rio, 2000), which is used several times throughout the proofs. Let $(W_n)_{n \in \mathcal{N} - \{0\}}$ be an α -mixing sequence of zero mean identically distributed r.r.v. with arithmetic coefficient κ . If $M < +\infty$ exists such that $|W_1| \leq M$, then, for any $r \geq 1$ and for some $C < +\infty$, one has:

$$P \left(\left| \sum_{i=1}^n W_i \right| > \epsilon \right) \leq C \left\{ \left(1 + \frac{\epsilon^2}{r s_n^2} \right)^{-r/2} + \frac{n}{r} \left(\frac{r}{\epsilon} \right)^{\kappa+1} \right\}, \quad (31)$$

where $s_n^2 = \sum_{i=1}^n \sum_{j=1}^n |Cov(W_i, W_j)|$.

An additional property used throughout the proofs is the following one:

$$\forall p > 0, \exists C_1 > 0 \text{ and } C_2 > 0, C_1 \phi_\chi(h) < \mathbb{E} K_{a,1}^p < C_2 \phi_\chi(h), \quad (32)$$

which is a direct consequence of **(H7)**.

Finally, to simplify the notation, one sets for $i = 1, \dots, n$: $G_i = G\left(\frac{\nu_p(\chi) - Y_i}{g}\right)$, $G'_i = G'\left(\frac{\nu_p(\chi) - Y_i}{g}\right)$, $H_i = H\left(\frac{\nu_p(\chi) - Y_i}{g}\right)$, $H'_i = H'\left(\frac{\nu_p(\chi) - Y_i}{g}\right)$ and $K_{a,i} = K_a(h^{-1}d(\chi, \mathcal{X}_i))$.

GUIDELINES ON THE PROOF OF THEOREM 3.1.

- Theorem 3.1-(i). We use the standard decomposition

$$\widehat{F}(y|\chi) - F(y|\chi) = \frac{1}{\widehat{F}(y|\chi)} \left\{ \left(\widehat{F}_N - \mathbb{E}\widehat{F}_N \right) - \left(F(y|\chi) - \mathbb{E}\widehat{F}_N \right) \right\} + \frac{F(y|\chi)}{\widehat{F}_D} \left(\mathbb{E}\widehat{F}_D - \widehat{F}_D \right), \quad (33)$$

with $\widehat{F}_N = (n \mathbb{E}K_{a,1})^{-1} \sum_{i=1}^n K_{a,i} H_i$ and $\widehat{F}_D = (n \mathbb{E}K_{a,1})^{-1} \sum_{i=1}^n K_{a,i}$. By remarking that

$$P\left(\left|\widehat{F}_D - \mathbb{E}\widehat{F}_D\right| > \epsilon\right) \leq P\left(\left|\sum_{i=1}^n (K_{a,i} - \mathbb{E}K_{a,i})\right| > n \phi_\chi(h) \epsilon\right), \quad (34)$$

the Fuk-Nagaev inequality (31) gives:

$$P\left(\left|\widehat{F}_D - \mathbb{E}\widehat{F}_D\right| > \epsilon\right) \leq C \left\{ \underbrace{\left(1 + \frac{\epsilon_0^2 g n \phi_\chi(h) \log n}{r s_n^2}\right)^{-r/2}}_{T_1} \right. \quad (35)$$

$$\left. + \frac{n}{r} \underbrace{\left(\frac{r}{\epsilon_0 \sqrt{\epsilon_0^2 g n \phi_\chi(h) \log n}}\right)^{\kappa+1}}_{T_2} \right\}, \quad (36)$$

with $W_i = K_{a,i} - \mathbb{E}K_{a,i}$, $\epsilon = \epsilon_0 \frac{\sqrt{n \phi_\chi(h) \log n}}{n \phi_\chi(h)}$. It is easy to control the quantity

$$s_n^2 = \sum_{i=1}^n \sum_{j=1}^n |Cov(K_{a,i}, K_{a,j})|.$$

Indeed,

$$s_n^2 \equiv s_n^2(K_{a,1}, \dots, K_{a,n}; u_n) = \sum_{0 < |i-j| < u_n} |Cov(K_{a,i}, K_{a,j})| \quad (37)$$

$$+ \sum_{|i-j| > u_n} |Cov(K_{a,i}, K_{a,j})| + \sum_{i=1}^n Var(K_{a,i}), \quad (38)$$

and taking $u_n = \phi_\chi(h)^{-1+1/(\kappa+1)}$ we conclude that $s_n^2 = O(n\phi_\chi(h))$. Now, set $r = C(\log n)^2$ in (36). We have that $T_1 = O(n^{-1-\zeta})$ ($\zeta > 0$) for large enough ϵ_0 , and the same holds for T_2 as soon as $g\phi_\chi(h) \geq n^{\frac{3-\kappa}{\kappa+1}+\eta}$. It comes:

$$\widehat{F}_D - \mathbb{E}\widehat{F}_D = O_{a.co.} \left(\sqrt{\frac{\log n}{n\phi_\chi(h)}} \right). \quad (39)$$

Similar arguments can be used to get

$$\widehat{F}_N - \mathbb{E}\widehat{F}_N = O_{a.co.} \left(\sqrt{\frac{\log n}{n\phi_\chi(h)}} \right). \quad (40)$$

The uniformity over S is obtained by involving a standard coverage of S combined with the Lipschitz feature of $H(\cdot)$ which leads to:

$$\sup_{y \in S} \left| \widehat{F}_N - \mathbb{E}\widehat{F}_N \right| = O_{a.co.} \left(\sqrt{\frac{\log n}{n\phi_\chi(h)}} \right). \quad (41)$$

Lastly, regularity assumptions on $H(\cdot)$ and $F(\cdot|\chi)$ allow us to yield:

$$\sup_{y \in S} \left| F(y|\chi) - \mathbb{E}\widehat{F}_N \right| = O(h^{\beta_1}) + O(g^{\beta_2}). \quad (42)$$

Finally, (39)-(42) achieves the proof of Theorem 3.1-(i).

- Theorem 3.1-(ii). Just remark that $\widehat{F}^{(j)}(y|\chi) = \widehat{F}_N^{(j)}/\widehat{F}_D$. It suffices to follow similar arguments as those used for stating Theorem 3.1-(i) to get Theorem 3.1-(ii).
- Theorem 3.1-(iii). As in Ferraty *et al.* (2005), it is straightforward to deduce that

$$(\nu_p(\chi) - \widehat{\nu}_p(\chi))^j = O_{a.co.} \left(\widehat{F}(\nu_p(\chi)|\chi) - F(\nu_p(\chi)|\chi) \right), \quad (43)$$

which combined with Theorem 3.1-(ii) achieves the proof of Theorem 3.1-(iii).

PROOF OF THEOREM 3.2. The proof of Theorem 3.2 is based on the following lemmas. The first one proposes a decomposition of $\widehat{\mu}_p(\chi)$ whereas the third one gives the asymptotic behavior of each term

occurring in this decomposition. The second lemma gives intermediate results allowing to state Lemma 3.

Lemma 1 *Under assumptions of Theorem 3.2, one has:*

$$p\widehat{\mu}_p(\chi) = Q_1 + Q_2 + O\left((\widehat{\nu}_p(\chi) - \nu_p(\chi))^2\right), \quad (44)$$

$$\text{with } Q_1 = g \sum_{i=1}^n W_h(\chi, \mathcal{X}_i) G\left(\frac{\nu_p(\chi) - Y_i}{g}\right) \text{ and } Q_2 = \sum_{i=1}^n W_h(\chi, \mathcal{X}_i) \xi_i,$$

$$\text{where } \xi_i = (Y_i - \nu_p(\chi)) \left(1 - H\left(\frac{\nu_p(\chi) - Y_i}{g}\right)\right) + p\nu_p(\chi) \text{ and } G(u) = \int_u^{+\infty} v K_s(v) dv.$$

Lemma 2 (i) $\mathbb{E}\left(K_{a,1}^p | G_1|^q\right) = O(g\phi_\chi(h)), \forall p > 0, \forall q > 0,$

(ii) $\mathbb{E}(K_{a,1} | Y|^p H_1^q) = O(\phi_\chi(h)), \forall p \in \{1, 2\}, \forall q > 0.$

Lemma 3 *Under the assumptions of Theorem 3.2, one gets:*

$$(i) Q_1 = O(g^2) + O_{a.co.}\left(g\sqrt{\frac{g \log n}{n \phi_\chi(h)}}\right),$$

$$(ii) Q_2 = p\mu_p(\chi) + O(g^2) + O_{a.co.}\left(\sqrt{\frac{\log n}{n \phi_\chi(h)}}\right).$$

Proof of Lemma 1. From (7), using standard Taylor expansions and according to the links between $G(\cdot)$ and $H(\cdot)$ with $K_s(\cdot)$ (i.e., $G'(v) = -v K_s(v)$, $G''(v) = -K_s(v) - vK_s'(v)$, $H'(v) = K_s(v)$), it exists $\theta \in (\min\{\nu_p(\chi), \widehat{\nu}_p(\chi)\}, \max\{\nu_p(\chi), \widehat{\nu}_p(\chi)\})$ such that:

$$\begin{aligned} G\left(\frac{\widehat{\nu}_p(\chi) - Y_i}{g}\right) &= G_i + \frac{\widehat{\nu}_p(\chi) - \nu_p(\chi)}{g} G'_i + \frac{1}{2} \left(\frac{\widehat{\nu}_p(\chi) - \nu_p(\chi)}{g}\right)^2 G''_i + \left(\frac{\theta - Y_i}{g}\right) \\ &= G_i - \frac{\nu_p(\chi)}{g} \left(\frac{\widehat{\nu}_p(\chi) - \nu_p(\chi)}{g}\right) K_s\left(\frac{\nu_p(\chi) - Y_i}{g}\right) \\ &\quad + \frac{Y_i}{g} \left(\frac{\widehat{\nu}_p(\chi) - \nu_p(\chi)}{g}\right) K_s\left(\frac{\nu_p(\chi) - Y_i}{g}\right) \\ &\quad - \frac{1}{2} \left(\frac{\widehat{\nu}_p(\chi) - \nu_p(\chi)}{g}\right)^2 K_s\left(\frac{\theta - Y_i}{g}\right) \\ &\quad + \frac{Y_i}{2g} \left(\frac{\widehat{\nu}_p(\chi) - \nu_p(\chi)}{g}\right)^2 K'_s\left(\frac{\theta - Y_i}{g}\right) \\ &\quad - \frac{\theta}{2g} \left(\frac{\widehat{\nu}_p(\chi) - \nu_p(\chi)}{g}\right)^2 K'_s\left(\frac{\theta - Y_i}{g}\right), \end{aligned}$$

and

$$\begin{aligned}
H\left(\frac{\widehat{\nu}_p(\chi) - Y_i}{g}\right) &= H_i + \frac{\widehat{\nu}_p(\chi) - \nu_p(\chi)}{g} H'_i + \frac{1}{2} \left(\frac{\widehat{\nu}_p(\chi) - \nu_p(\chi)}{g}\right)^2 H''\left(\frac{\theta - Y_i}{g}\right) \\
&= H\left(\frac{\nu_p - Y_i}{g}\right) + \frac{\widehat{\nu}_p(\chi) - \nu_p(\chi)}{g} K_s\left(\frac{\nu_p(\chi) - Y_i}{g}\right) \\
&\quad + \frac{1}{2} \left(\frac{\widehat{\nu}_p(\chi) - \nu_p(\chi)}{g}\right)^2 K'_s\left(\frac{\theta - Y_i}{g}\right).
\end{aligned}$$

This leads us to:

$$p \widehat{\mu}_p(\chi) = \sum_{i=1}^n W_h(\chi, \mathcal{X}_i) \left\{ g G_i + Y_i (1 - H_i) - \nu_p(\chi) \left(\frac{\widehat{\nu}_p(\chi) - \nu_p(\chi)}{g}\right) K_s\left(\frac{\nu_p(\chi) - Y_i}{g}\right) \right. \quad (45)$$

$$\left. - \frac{g}{2} \left(\frac{\widehat{\nu}_p(\chi) - \nu_p(\chi)}{g}\right)^2 K_s\left(\frac{\theta - Y_i}{g}\right) - \frac{\theta}{2} \left(\frac{\widehat{\nu}_p(\chi) - \nu_p(\chi)}{g}\right)^2 K'_s\left(\frac{\theta - Y_i}{g}\right) \right\} \quad (46)$$

$$= \sum_{i=1}^n W_h(\chi, \mathcal{X}_i) Y_i (1 - H_i) - \nu_p(\chi) (\widehat{\nu}_p(\chi) - \nu_p(\chi)) \widehat{f}(\nu_p(\chi)|\chi) + Q_1 \quad (47)$$

$$- \frac{1}{2} (\widehat{\nu}_p(\chi) - \nu_p(\chi))^2 \left\{ \widehat{f}(\theta|\chi) + \theta \widehat{f}'(\theta|\chi) \right\}. \quad (48)$$

By using Theorem 3.1-(ii), it is clear that the last term on the right hand side of (48) is $O\left((\widehat{\nu}_p(\chi) - \nu_p(\chi))^2\right)$.

A standard Taylor expansion of $\widehat{F}(\cdot|\chi)$ combined with Theorem 3.1-(ii) leads to

$$p = 1 - \widehat{F}(\nu_p(\chi)|\chi) - (\widehat{\nu}_p(\chi) - \nu_p(\chi)) \widehat{f}(\nu_p(\chi)|\chi) + O\left((\widehat{\nu}_p(\chi) - \nu_p(\chi))^2\right), \quad (49)$$

which implies that

$$-\nu_p(\chi) (\widehat{\nu}_p(\chi) - \nu_p(\chi)) \widehat{f}(\nu_p(\chi)|\chi) = p \nu_p(\chi) - \nu_p(\chi) \sum_{i=1}^n W_h(\chi, \mathcal{X}_i) (1 - H_i) \quad (50)$$

$$+ O\left((\widehat{\nu}_p(\chi) - \nu_p(\chi))^2\right). \quad (51)$$

Now, by combining (48) with (51) one gets

$$p \widehat{\mu}_p(\chi) = Q_1 + Q_2 + O\left((\widehat{\nu}_p(\chi) - \nu_p(\chi))^2\right), \quad (52)$$

which ends the proof of Lemma 1.

Proof of Lemma 2.

(i). We have

$$\begin{aligned}\mathbb{E}\left(K_{a,1}^p |G_1|^q\right) &\leq \mathbb{E}\left\{K_{a,1}^p \mathbb{E}(|G_1|^q | \mathcal{X}_1)\right\} \\ &= \int_{B(\chi, h)} K_a^p\left(\frac{d(\chi, z)}{h}\right) \left\{\int_{\mathbb{R}} \left|G\left(\frac{\nu_p(\chi) - y}{g}\right)\right|^q f(y|z)\right\} dP(z).\end{aligned}$$

Now, because of **(H4)** and **(H7)**, it is easy to get

$$\int_{\mathbb{R}} \left|G\left(\frac{\nu_p(\chi) - y}{g}\right)\right|^q f(y|z) dy \leq C g f(\nu_p(\chi)|z) + O(g^2), \quad (53)$$

which implies that

$$\mathbb{E}\left(K_{a,1}^p |G_1|^q\right) \leq C g \int_{B(\chi, h)} K_a^p\left(\frac{d(\chi, z)}{h}\right) f(\nu_p(\chi)|z) dP(z) \quad (54)$$

$$\leq C' g f(\nu_p(\chi)|\chi) \mathbb{E}\left(K_{1,a}^p\right) \left\{1 + O(h^{\beta_1})\right\} + \mathbb{E}\left(K_{1,a}^p\right) O(g^2), \quad (55)$$

the last inequality coming from **(H5)**. The combination of (32) with (55) leads us to:

$$\mathbb{E}\left(K_{a,1}^p |G_1|^q\right) = O(g \phi_\chi(h)). \quad (56)$$

(ii). Since $H_1^q \leq 1$, one has:

$$\begin{aligned}\mathbb{E}(K_{a,1}|Y_1|^p H_1^q) &\leq \mathbb{E}K_{a,1} \mathbb{E}(|Y_1|^p | \mathcal{X}_1) \\ &\leq \mathbb{E}K_{a,1} |\sigma_p(\mathcal{X}_1) - \sigma_p(\chi)| + \sigma_p(\chi) \mathbb{E}K_{a,1}.\end{aligned}$$

We get $\mathbb{E}(K_{a,1}|Y_1|^p H_1^q) = O(\phi_\chi(h))$, which ends the proof of Lemma 2-(ii).

Proof of Lemma 3.

- Lemma 3-(i). First of all, (39) implies that

$$Q_1 = (Q_{11} + Q_{12}) \left\{1 + O_{a.co.} \left(\sqrt{\frac{\log n}{n \phi_\chi(h)}}\right)\right\}, \quad (57)$$

with $Q_{11} = \frac{g}{\mathbb{E}K_{a,1}} \sum_{i=1}^n \left(\underbrace{K_{a,i} G_i - \mathbb{E} K_{a,1} G_1}_{Z_i} \right)$ and $Q_{12} = \frac{g}{\mathbb{E}K_{a,1}} \mathbb{E} K_{a,1} G_1$. On the one hand, (32) and Lemma 2-(i) imply that $|\text{Cov}(Z_i, Z_j)| = O(g^2 \phi_\chi(h)^{2-1/(\kappa+1)})$ for $i \neq j$ and $\text{Var}(Z_i) = O(g \phi_\chi(h))$. On the other hand, the Davydov covariance inequality (see Rio 2000, formula 1.12a) allows us to get $|\text{Cov}(Z_i, Z_j)| = O(|i-j|^{-\kappa})$. Then, by using a similar decomposition as in (37), it comes:

$$s_n^2(Z_1, \dots, Z_n; u_n) = O(g n \phi_\chi(h)), \quad (58)$$

where one sets $u_n = g^{-1} \phi_\chi(h)^{-1+1/(\kappa+1)}$. By remarking that

$$P(g^{-1} |Q_{11}| > \epsilon) \leq P\left(\left|\sum_{i=1}^n Z_i\right| > n \phi_\chi(h) \epsilon\right), \quad (59)$$

the use of the Fuk-Nagev inequality (31) with $\epsilon = \epsilon_0 \frac{\sqrt{g n \phi_\chi(h) \log n}}{n \phi_\chi(h)}$ and $r = C (\log n)^2$ leads to:

$$Q_{11} = O_{a.co.} \left(g \sqrt{\frac{g \log n}{n \phi_\chi(h)}} \right). \quad (60)$$

In addition, (32) and Lemma 2-(i) allow us to get $Q_{12} = O(g^2)$, which combined with (60) implies that:

$$Q_1 = O(g^2) + O_{a.co.} \left(g \sqrt{\frac{g \log n}{n \phi_\chi(h)}} \right). \quad (61)$$

- Lemma 3-(ii).

$$Q_2 = \underbrace{\sum_{i=1}^n W_h(\chi, \mathcal{X}_i) (\xi_i - \mathbb{E}(\xi_i | \mathcal{X}_i))}_{Q_{21}} + \underbrace{\sum_{i=1}^n W_h(\chi, \mathcal{X}_i) \mathbb{E}(\xi_i | \mathcal{X}_i)}_{Q_{22}}. \quad (62)$$

On the one hand,

$$\mathbb{E}(Y_i(1-H_i) | \mathcal{X}_i = \chi) = p \mu_p(\chi) - \frac{g^2}{2} (f(\nu_p(\chi)|\chi) + \nu_p(\chi) f'(\nu_p(\chi)|\chi)) m_2(K_s) + o(g^2), \quad (63)$$

and on the other hand,

$$\mathbb{E}((1-H_i) | \mathcal{X}_i = \chi) = F(\nu_p(\chi)|\chi) - \frac{g^2}{2} f'(\nu_p(\chi)|\chi) m_2(K_s) + o(g^2), \quad (64)$$

which implies that

$$\mathbb{E}(\xi_i | \mathcal{X}_i = \chi) = p \mu_p(\chi) - \frac{g^2}{2} f(\nu_p(\chi) | \chi) m_2(K_s) + o(g^2). \quad (65)$$

Moreover,

$$\mathcal{X}_i \in B(\chi, h) \Rightarrow \sum_{i=1}^n W_h(\chi, \mathcal{X}_i) \mathbb{E}(\xi_i | \mathcal{X}_i) = \sum_{i=1}^n W_h(\chi, \mathcal{X}_i) \mathbb{E}(\xi_i | \mathcal{X}_i = \chi) + o(g^2). \quad (66)$$

This last result combined with (65) leads to

$$Q_{22} = p \mu_p(\chi) + O(g^2). \quad (67)$$

Now, according to (39), one has

$$Q_{21} = \underbrace{\frac{1}{\mathbb{E}K_{a,1}} \sum_{i=1}^n K_{a,i} (\xi_i - \mathbb{E}(\xi_i | \mathcal{X}_i))}_T \left\{ 1 + O_{a.co.} \left(\sqrt{\frac{\log n}{n \phi_\chi(h)}} \right) \right\}, \quad (68)$$

where $n \mathbb{E}K_{a,i} T = \sum_{i=1}^n W_{i,1,1} - \nu_p(\chi) \sum_{i=1}^n W_{i,0,1} + p \nu_p(\chi) \sum_{i=1}^n W_{i,0,0}$ with

$W_{i,k,l} = K_{a,i} Y_i^k \bar{H}_i^l - K_{a,i} \mathbb{E}(Y_i^k \bar{H}_i^l | \mathcal{X}_i)$ and $\bar{H}_i = 1 - H_i$. Use again (32) with Lemma 2-(ii)

to derive that, for any $(k, l) \in \{0, 1\}^2$, the decomposition of the covariances between the $W_{i,k,l}$'s satisfies:

$$s_n^2(W_{1,k,l}, \dots, W_{n,k,l}; u_n) = O(n \phi_\chi(h)), \quad (69)$$

where $u_n = \phi_\chi(h)^{-1+1/(\kappa+1)}$. Now, the Fuk-Nagaev inequality with (69),

$\epsilon = \epsilon_0 \frac{\sqrt{n \phi_\chi(h) \log n}}{n \phi_\chi(h)}$ and $r = C (\log n)^2$ allows us to write:

$$P \left(\left| \sum_{i=1}^n W_{i,k,l} \right| > n \phi_\chi(h) \epsilon \right) = O_{a.co.} \left(\sqrt{n \phi_\chi(h) \log n} \right). \quad (70)$$

Now, according to definition of T , with (32) and (70) it comes:

$$T = O_{a.co.} \left(\sqrt{\frac{\log n}{n \phi_\chi(h)}} \right). \quad (71)$$

This last result combined with (68) allows us to get

$$Q_{21} = O_{a.co.} \left(\sqrt{\frac{\log n}{n \phi_{\chi}(h)}} \right). \quad (72)$$

The proof of Lemma 3-(ii) is achieved by gathering (62), (67) and (72).

References

- Acerbi, C. (2002). Spectral measures of risk: a coherent representation of subjective risk aversion. *J. Bank Financ.* 26: 1505-1518.
- Bao, Y., Lee, T.H., Saltoglu, B. (2006). Evaluating predictive performance of Value-at-Risk models in emerging markets: a reality check. *J. Forecast.* 25: 101-128.
- Basel Committee on Banking Supervision (1996). Amendment to the capital accord to incorporate market risks, Bank for International Settlements.
- Basel Committee on Banking Supervision (2006). Basel II: International convergence of capital measurement and capital standards: a revised framework. *Report # 128*. Bank for International Settlements.
- Beirlant, J., de Wet, T., Goegebeur, Y. (2004). Nonparametric estimation of extreme conditional quantiles. *J. Stat. Comput. Sim.* 74: 567-580.
- Berkowitz, J., O'Brien, J. (2002). How accurate are the value-at-risk models at commercial banks? *J. Financ.* 57: 1093-1111.
- Berlinet, A., Gannoun, A., Matzner-Løber, E. (2001). Asymptotic normality of convergent estimates of conditional quantiles. *Statistics* 35: 139-169.

- Bertail, P., Haefke, C., Politis, D.N., White, H. (2004). A subsampling approach to estimating the distribution of diverging statistics with applications to assessing financial market risks. *J. Econometrics* 120: 295-326.
- Bollerslev, T (1986). Generalized Autoregressive Conditional Heteroskedasticity. *J. Econometrics* 31: 307-327.
- Bosq, D. (2000). *Linear processes in functional spaces. Theory and Applications*. Lecture Notes in Statistics 149. New York: Springer-Verlag.
- Cai, Z., Wang, X. (2008). Nonparametric estimation of conditional VaR and expected shortfall. *J. Econometrics* 147: 120-130.
- Chen, S.X. (2007). Nonparametric Estimation of Expected Shortfall. *J. Financ. Econometrics* 6, 87-107.
- Chen, S.X., Tang, C.Y. (2005). Nonparametric inference of value at risk for dependent financial returns. *J. Financ. Econometrics* 3: 227-255.
- Cosma, A., Scaillet, O., von Sachs, R. (2007). Multivariate wavelet-based shape preserving estimation for dependent observations. *Bernoulli* 13: 301-329.
- Danielsson, J., De Vries, C.G. (2000). Value-at-Risk and Extreme Returns. *Annales d'Economie et de Statistique, ADRES*, issue 60.
- Dettling, M., Bühlmann, P. (2004). Risk and Volatility Estimation with Linear and Nonlinear Methods Based on High-Frequency-Data. *Appl. Financ. Econ.* 14: 717-729.
- Embrechts, P., Kluppelberg, C., Mikosch, T. (1997). *Modeling extremal events for finance and insurance*. New York: Springer-Verlag.

- Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflation. *Econometrica* 50: 987-1008.
- Engle, R.F., Patton, A.J. (2001). What good is a volatility model? *Quantitative Finance*. Institute of Physics Publishing, 1, pp. 237-245.
- Fermanian, J.D., Scaillet, O. (2005). Sensitivity analysis of VaR and Expected Shortfall for portfolios under netting agreements. *J. Bank Financ.* 29: 927-958.
- Ferraty, F., Laksaci, A., Vieu, P. (2006). Estimating some characteristics of the conditional distribution in nonparametric functional models. *Stat. Inference Stoch. Process* 9: 47-76.
- Ferraty, F., Laksaci, A., Tadj, A., Vieu, P. (2010). Rate of uniform consistency for nonparametric estimates with functional variables. *J. Statist. Plann. Infer.* 140: 335-352.
- Ferraty, F., Quintela-del-Río, A., Vieu, P. (2012). Specification test for conditional distribution with functional variables. *Economet. Theor.* 28: 363-383.
- Ferraty, F., Rabhi, A., Vieu, P. (2005). Conditional quantiles for functional dependent data with application to the climatic El Niño phenomenon. *Sankhya* 67: 378-398.
- Ferraty, F., Romain, Y. (2011). *The Oxford Handbook of Functional Data Analysis* (Eds.). Oxford University Press.
- Ferraty, F., Van Keilegom, I., Vieu, P. (2010b). On the validity of the bootstrap in nonparametric functional regression. *Scand. J. Statist.* 37: 286-306.
- Ferraty, F., Vieu, P. (2006). *Nonparametric functional data analysis*. Springer Series in Statistics. New York: Springer-Verlag.

- Gaglianone, W.P., Lima, L.R., Linton, O., Smith, D. (2011). Evaluating value-at-risk models via quantile regression. *J. Bus. Econ. Stat.* 29: 150-160.
- Gilli, M., K ellezi, E. (2006). Application of Extreme Value Theory for Measuring Financial Risk. *Comput. Econ.* 27: 207-228.
- Giot, P., Laurent, S. (2003). Value-at-Risk for long and short trading positions. *J. Appl. Econom.* 18: 641-664.
- Jorion, P. (2000). *Value at Risk (2nd edition)*. New York: McGraw Hill.
- Koenker, R., Bassett, G. (1978). Regression quantiles. *Econometrica* 46: 33-50.
- Lemdani, M., Ould-Said, E., Poulin, N. (2009). Asymptotic properties of a conditional quantile estimator with randomly truncated data. *J. Multivariate Anal.* 100: 546-559.
- Linton, O., Sancetta, A. (2009). Consistent estimation of a general nonparametric regression function in time series. *J. Econometrics* 152: 70-78.
- Linton, O., Xiao, Z. (2011). Estimation of and inference about the Expected Shortfall for time series with infinite variance. Available at SSRN: <http://ssrn.com/abstract=1954618>.
- McNeil, A.J., Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *J. Empir. Finance* 7: 271-300.
- Masry, E. (2005). Nonparametric regression estimation for dependent functional data: asymptotic normality. *Stochastic Process. Appl.* 115: 155-177.
- Perignon, C., Smith, D.R. (2008). The level and quality of Value-at-Risk disclosure by commercial banks. AFA 2008 New Orleans Meetings Paper. Available at SSRN: <http://ssrn.com/abstract=952595>.

- Quintela-del-Río, A. (2008). Hazard function given a functional variable: Non-parametric estimation under strong mixing conditions. *J. Nonparametr. Stat.* 20: 413-430.
- Quintela-del-Río, A., Ferraty, F., Vieu, P. (2011). Analysis of time of occurrence of earthquakes: a functional data approach. *Math.Geosci.* 43: 695-719.
- Quintela-del-Río, A., Francisco-Fernández, M. (2011). Nonparametric functional data estimation applied to ozone data: prediction and extreme value analysis. *Chemosphere* 82: 800-808.
- Raggi, D., Bordignon, S. (2006). Comparing stochastic volatility models through Monte Carlo simulations. *Comput. Stat. Data An.* 50: 1678-1699.
- Rio, E. (2000). *Théorie Asymptotique des Processus Aléatoires Dépendants* (in French). Mathématiques et Applications 31. Berlin: Springer-Verlag.
- Riskmetrics (1995). RiskMetrics Technical Document. J.P. Morgan.
- Scaillet, O. (2004). Nonparametric estimation and sensitivity analysis of expected shortfall. *Math. Financ.* 14: 115-129.
- Scaillet, O. (2005). Nonparametric estimation of conditional expected shortfall. *Revue Assurances et Gestion des Risques/Insurance and Risk Management Journal*, 74, pp. 639-660.
- Zhu, D., Galbraith, J.W. (2011). Modeling and forecasting expected shortfall with the generalized asymmetric Student- t and asymmetric exponential power distributions. *J. Empirical Finance* 18: 765-778.

Table Captions:

Table 1: Average and standard deviation of the absolute errors for the three considered CVaR estimates, for 1000 replications of the simulated GARCH(1,1) model.

Table 2: Average and standard deviation of the absolute errors for the CES estimates, for 1000 replications of the simulated GARCH(1,1) model.

Table 3: Expected loss (29) for the three considered CVaR estimates, for the S&P 500 Index time series. Values of $p = 0.05$ and 0.01 were used.

Table 4: Mean absolute error (30) for the parametric and nonparametric CES estimates, for values of $p = 0.05$ and 0.01 , for the S&P500 Index time series.

Figure Captions:

Figure 1: (a) The generated time series, according to Model (14). (b) Five particular functional data χ_i .

Figure 2: Boxplot of the 500 absolute errors between the true values and their nonparametric estimates, for three particular sample sizes, $n = 100, 250$ and 500 . Left panel: CVaR estimates. Right panel: CES estimates.

Figure 3: Boxplot of the 100 absolute errors, using the optimal and the estimated bootstrap bandwidths.

Figure 4: Estimation of the CVaR for a simulated GARCH(1,1) model, for the three considered methods: Historical simulation (top), GARCH fitting (center) and NFDE (bottom). The quantile estimation is displayed as a dotted line.

Figure 5: Estimation of the CES for a simulated GARCH(1,1) model, by means of a GARCH fitting (top) and NFDE (bottom). The dotted line corresponds to the CES estimates.

Figure 6: Plot of the S&P500 Index time series data, from 07/April/1986 to 31/August/2010 (6156

data).

Figure 7: Estimation of the CVaR (dotted line) for the S&P500 Index time series data, using the Historical Simulation (HS) and the GARCH fitting methods.

Figure 8: Nonparametric estimates for the S&P500 Index time series data. Top: CVaR estimates (dotted line). Bottom: CES estimates (dotted line).

HS	GARCH	NFDE
0.156 (0.069)	0.0062 (0.005)	0.069 (0.063)

Table 1:

GARCH	NFDE
0.009 (0.008)	0.116 (0.121)

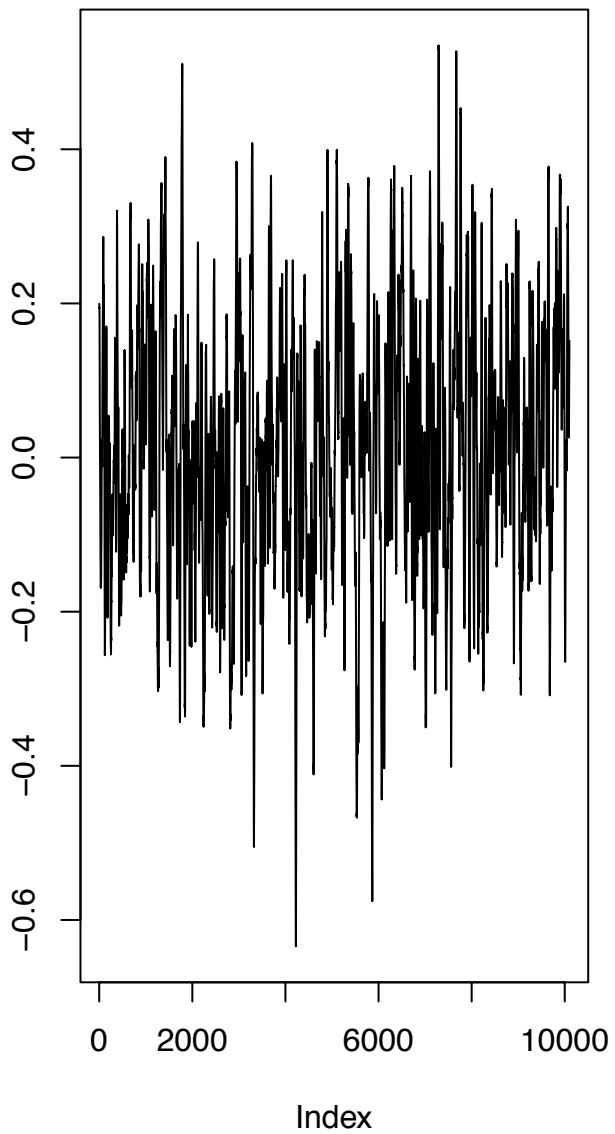
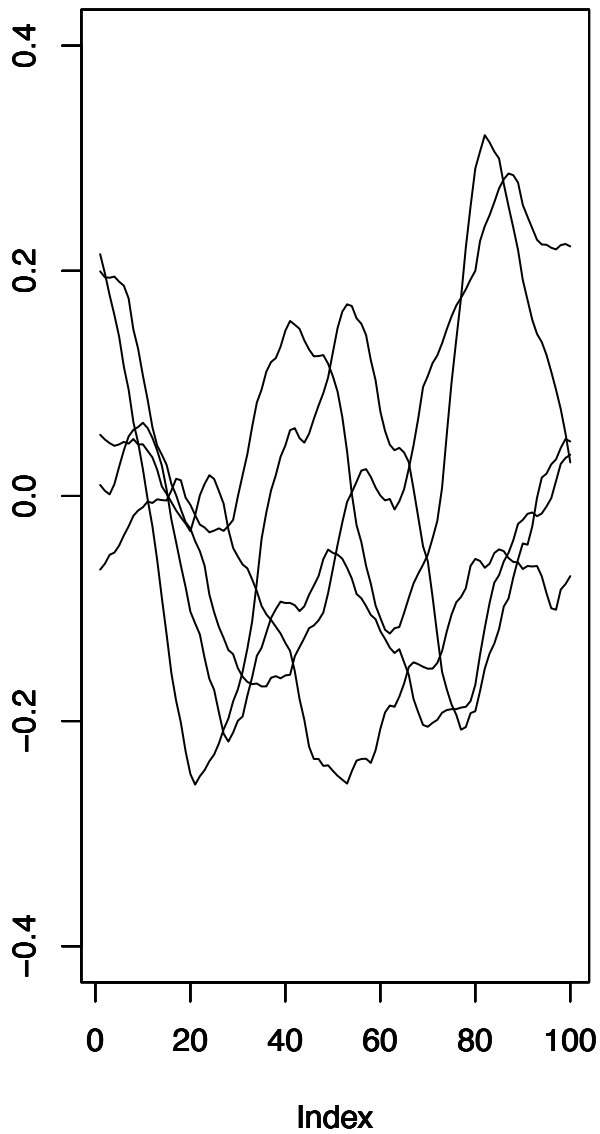
Table 2:

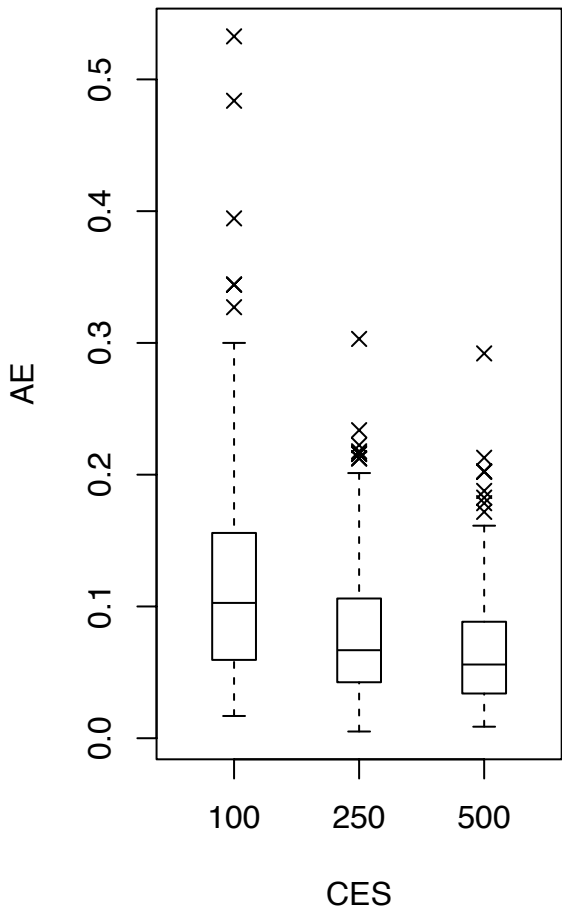
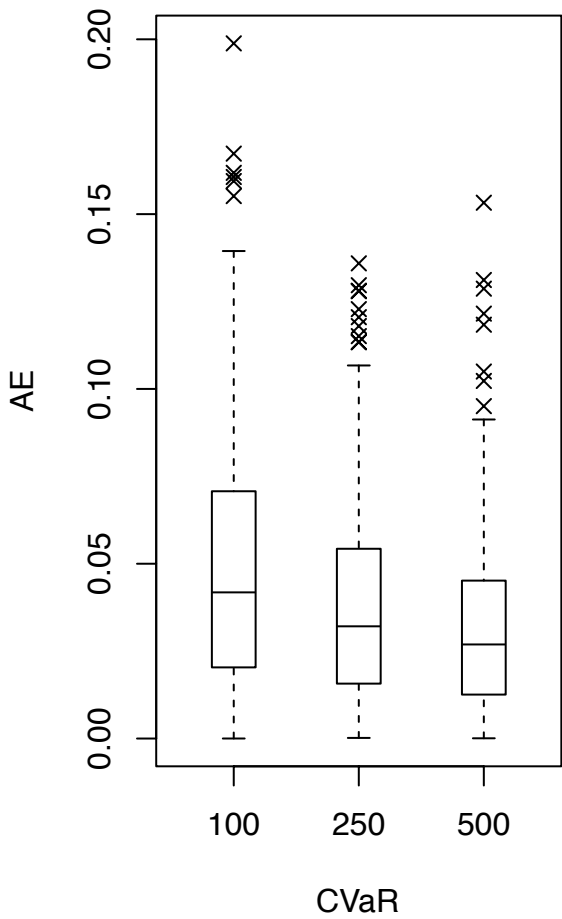
p	$Q(HS)$	$Q(GARCH)$	$Q(NFDE)$
0.1	0.20	0.22	0.28
0.05	0.13	0.16	0.14

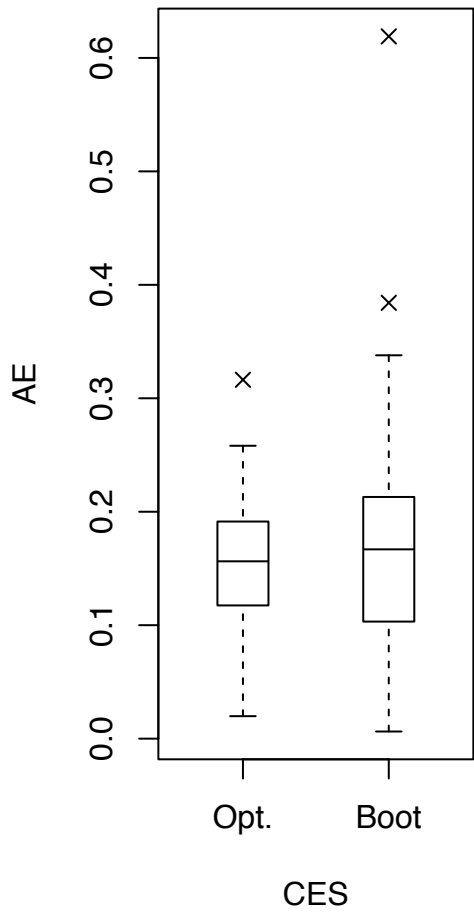
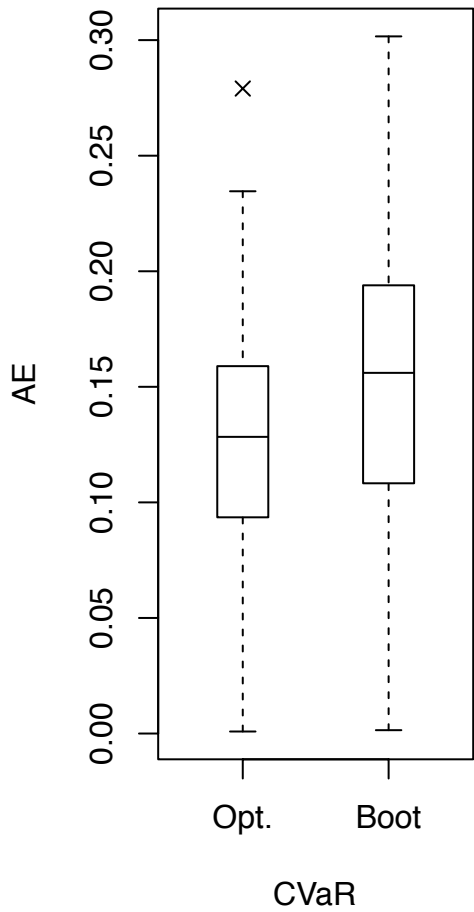
Table 3:

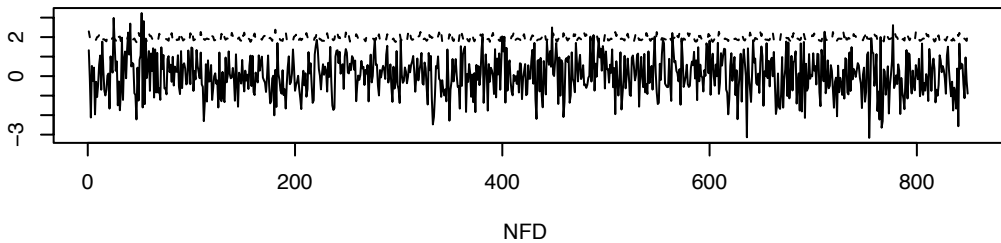
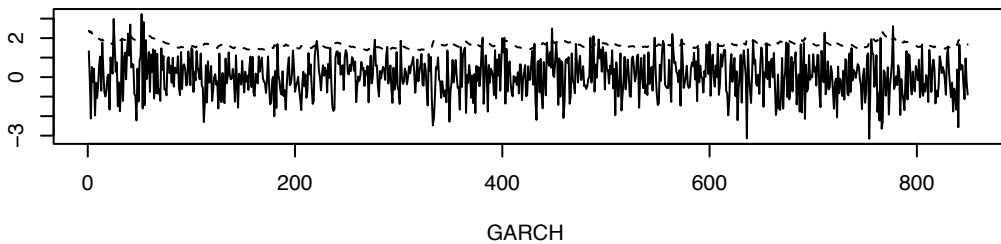
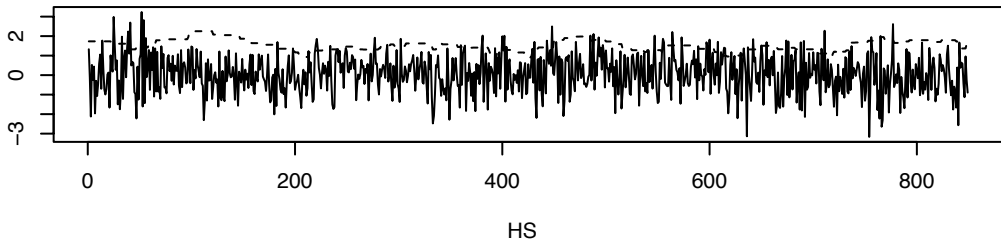
p	GARCH	NFDE
0.1	0.85	0.26
0.05	1.15	0.33

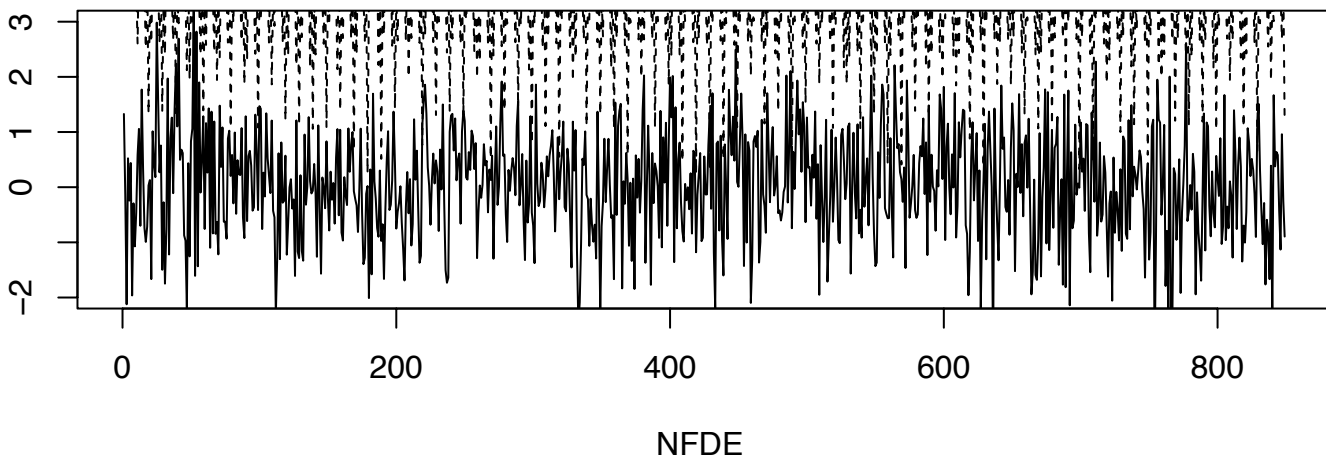
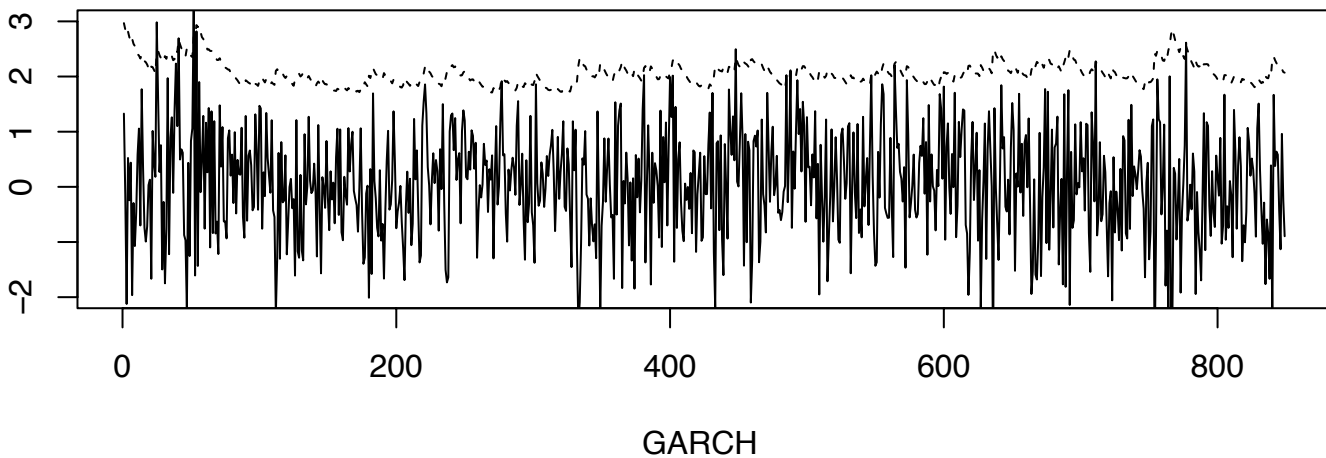
Table 4:

(a)**(b)**

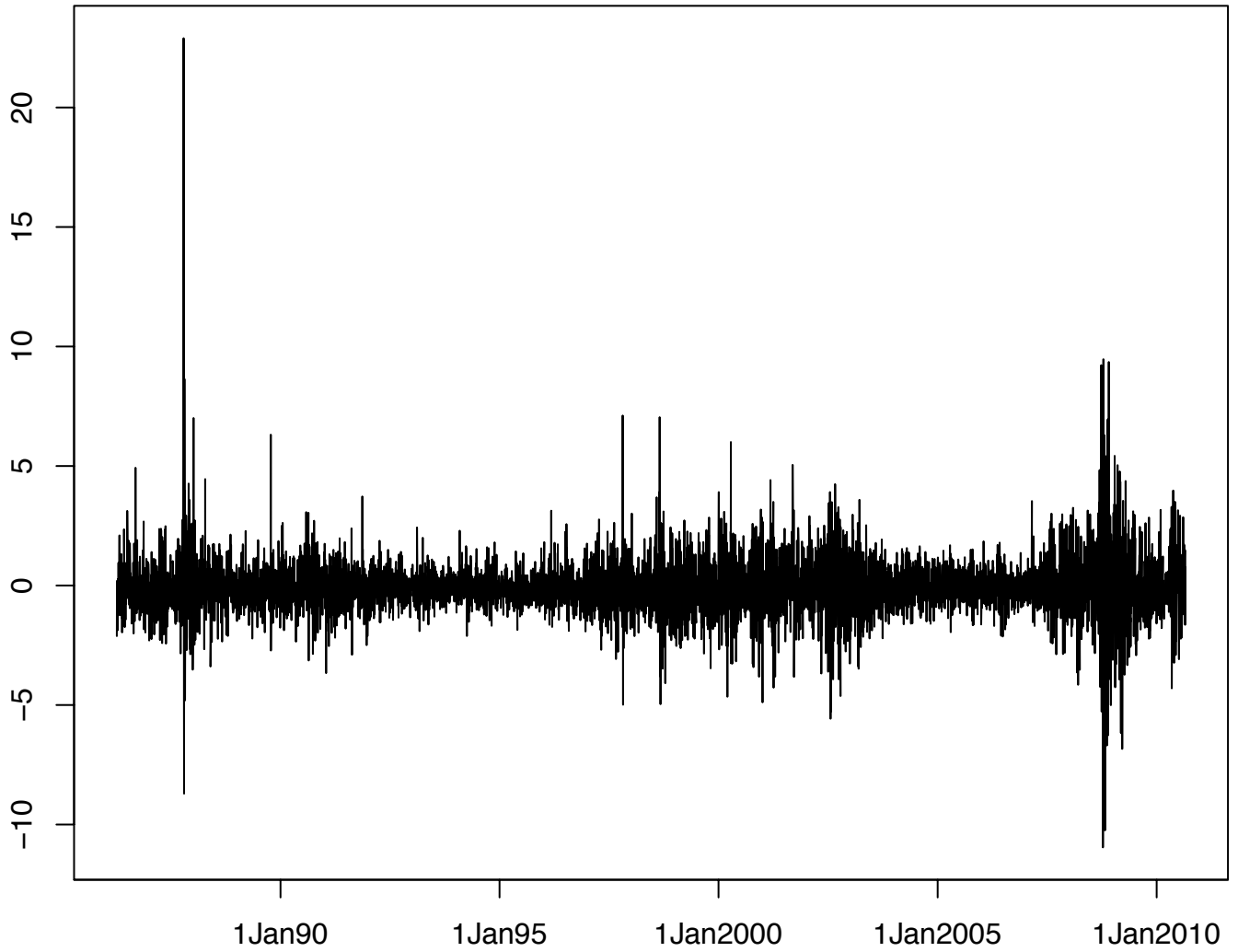


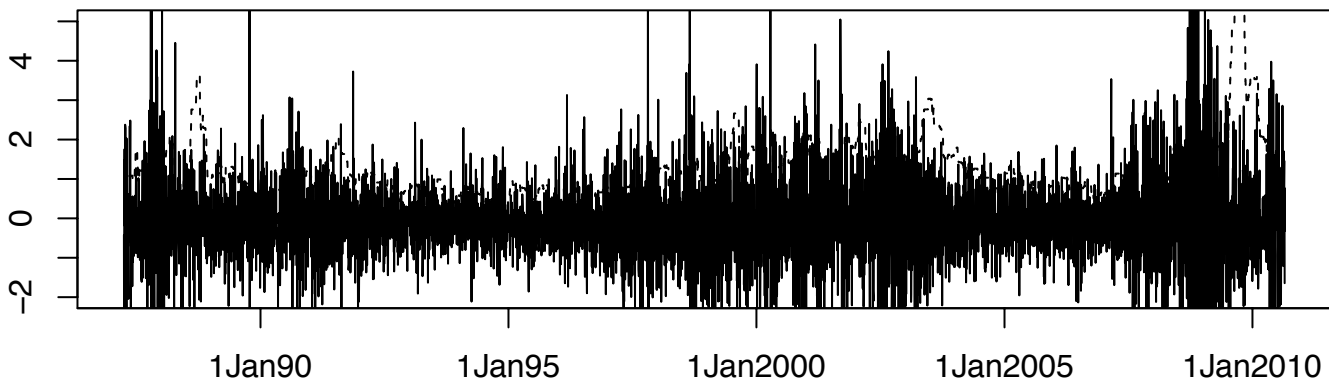




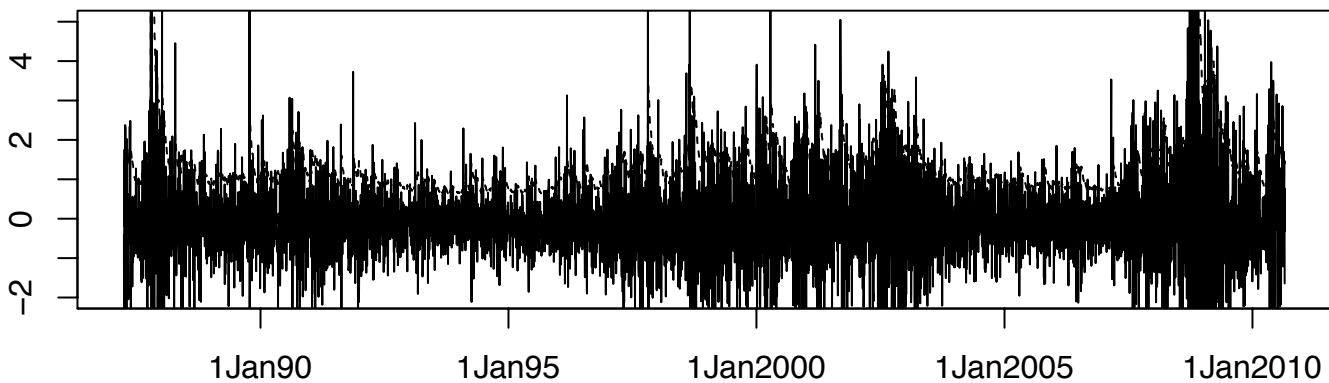


S&P 500 Index

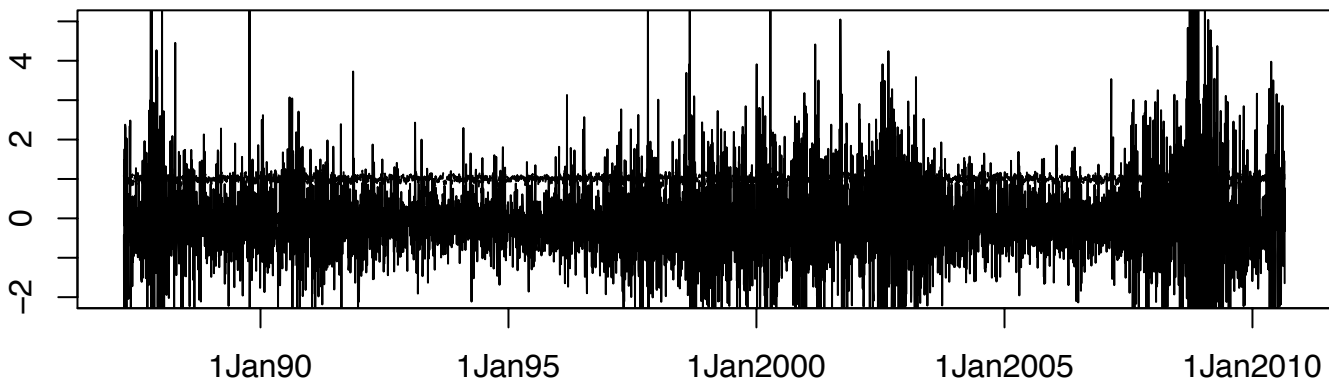




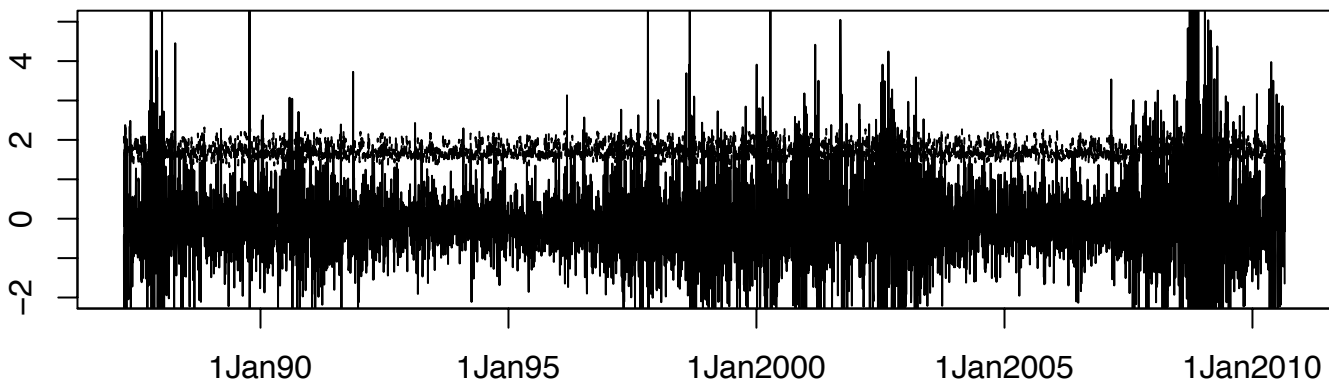
HS



GARCH



NFDE CVaR



NFDE CES