mathematics

Article

Nonparametric Estimation of the Expected Shortfall Regression for Quasi-Associated Functional Data

Larbi Ait-Hennani [1], Zoulikha Kaid [2], Ali Laksaci [2] and Mustapha Rachdi [3,*]

1 Department of Statistic and Informatics, IUT, Lille 2 University, Rond-point de l'Europe, BP. 557, F 59060 Roubaix, France
2 Department of Mathematics, College of Science, King Khalid University, Abha 62529, Saudi Arabia
3 Laboratoire AGEIS EA 7407, Université Grenoble Alpes (France), UFR SHS, BP. 47, CEDEX 09, F 38040 Grenoble, France
* Correspondence: mustapha.rachdi@univ-grenoble-alpes.fr

Abstract: In this paper, we study the nonparametric estimation of the expected shortfall regression when the exogenous observation is functional. The constructed estimator is obtained by combining the double kernels estimator of both conditional value at risk and conditional density function. The asymptotic proprieties of this estimator are established under weak dependency condition. Precisely, we assume that the observations are generated from quasi-associated functional time series and we prove the almost complete convergence of the constructed estimator. This asymptotic result is obtained under a standard condition of functional time series analysis. The finite sample performance of this estimator is evaluated using artificial data.

Keywords: functional data; complete convergence (a.co.); risk analysis; expected shortfall regression; kernel method; bandwidth parameter; financial time series; quasi-associated process

MSC: 62G05; 62R07; 62G08; 62M10; 91B05; 37M10

check for updates

Citation: Ait-Hennani, L.; Kaid, Z.; Laksaci, A.; Rachdi, M. Nonparametric Estimation of the Expected Shortfall Regression for Quasi-Associated Functional Data. *Mathematics* **2022**, *10*, 4508. https://doi.org/10.3390/math10234508

Academic Editors: Maria C. Mariani and Manuel Alberto M. Ferreira

Received: 22 September 2022
Accepted: 22 November 2022
Published: 29 November 2022

1. Introduction

The financial risk management is a challenging issue in stock investments. In applied statistics, the Value at Risk (VaR) is a common model to fit this kind of risk. Furthermore, the financial institutions have officialised the use of the VaR-function by Basel Accords (1996, 2006). However, it is demonstrated in the literature that this model has some drawbacks. In particular, the main concern is its insensitivity to the magnitude of the extreme risk (see Acerbi and Tasche [1]). For this reason, the financial institutions have enhanced the reliability of their risk management tools by adding the Expected Shortfall (ES) model (Basel III in 2014, [2]). It is well documented that this last alleviates the defects of the VaR-model. Motivated by this superiority over the VaR model, the ES-model has received growing consideration in multivariate statistics, namely, as a parametric model. Alternatively, in this contribution, we focus on the functional statistics situation using the nonparametric ideas.

Historically, the use of the expected shortfall model as a risk model was started by Artzner et al. [3]. They proved the coherency property of this risk model. Yamai and Yoshiba [4] conducted a comparative analysis between VaR and ES as financial risk models. They conclude that the VaR has serious problems when the profit-loss distribution is not Gaussian. They stated that the use of the ES-model is more appropriate in this kind of situation. Furthermore, the feasibility of the ES as a financial risk model has been popularised by Acerbi [5]. Since this precursor work, the ES-model is considerably used in finance and/or insurance as a score function or as a constraint to optimise the asset allocation (see, for instance, Krokhmal et al. [6] or Alexander et al. [7]). Of course, the efficiency of this model in practice is linked to its estimation quality. In fact, the estimation of the ES has been investigated by many authors in the last decade using different statistical approaches. The parametric approach based on the GARCH model seems more popular than the other algorithms (see, for example Wong [8] and the reference therein). As recent advances in the ES-estimation, we cite Acereda et al. [9], which compares various innovation distributions in GARCH-models to approximate ES. Nevertheless, some authors have employed an alternative parametric approach based on the extreme values method to estimate the expected-shortfall (see, for instance, Novales et al. [10], for a list of reference). Concerning the nonparametric approach, the first results were obtained and Scaillet [11]. He established the asymptotic properties of an estimator constructed by the kernel smoothing. Cai and Wang [12] have proved the asymptotic normality and the strong consistency of a kernel-based estimator. Yu et al. [13] focus on the bias correction of the ES kernel estimation using the jackknife rule. Recently, an estimator based on the Bahadur-representation was constructed and used by Wu et al. [14]. They obtained the Berry–Esséen bound for the proposed nonparametric expected shortfall estimator. For an overview on the recent advances and motivations on the expected shortfall model, we refer to Mohammed et al. [15], Marri and Moutanabbir [16] or Jiang et al. [17], among others.

All these cited works construct their estimators using finite dimensional observations. In this work, we consider the infinite dimensional situation. Such a case of so-called functional statistics is actually in continuous progress. At this stage, the financial area constitutes an attractive field for functional data modeling. The necessity to provide a mathematical models allowing the real-time control of volatility and exploring the micro-structure of the financial data are the principal motivations of the use of functional statistics for financial time series data. However, most works use the parametric ideas to fit the financial data as functional observations. For previous studies on this topic, we cite Muller et al. [18], Kokoszka et al. [19] or Shang et al. [20]. Additional tools for financial risk management based on functional data analysis can be found in Cai [21]. He provided a sophistical algorithm to analyze the risk measurement of China's financial market. As a recent development in financial time series analysis by the functional date, we mention Saart et al. [22]. They introduced a new approach-based functional principal components analysis for modeling the time varying behavior of asset returns co-movements. Wang et al. [23] combine the kernel estimation to approximate the long-term covariance function and employ the functional principal components analysis to fit and reconstruct the volatility curve. They applied their study to China's CSI 300 stock index. An alternative approach to estimate the functional volatility was recently developed by Liang et al. [24]. Their estimation is based on the approximation of the total curvature of the functional observation smoothed by some adequate basis functions. The framework of the present work is the nonparametric functional modeling of the financial time series data. In this context, the first work was investigated by Ferraty et al. [25]. They established the almost complete consistency of the kernel estimator of the functional regression. Considering the same functional model, Masry [26] proves the asymptotic normality under the strong mixing dependency. The ergodic functional time series data case was studied by Laib and Laouni [27]. However, the literature on the estimation of the functional expected shortfall model is still limited. There is only one paper on this topic. It was published by Ferraty et al. [28]. They proved the almost complete consistency of a kernel estimator of the expected shortfall model under the strong mixing assumption. To that end, let us refer to [29–31] for an overview on the recent tends and advances in functional statistics.

The main goal of this paper is to study the nonparametric estimation of the expected shortfall regression as a financial risk model. The estimator is constructed by using the double kernels estimator of the Conditional Distribution Function (CDF). Precisely, the latter is used to define the kernel estimator of the conditional density as a derivative of CDF and to build the kernel estimator of the VaR as the inverse of the CDF. Under weak dependence conditions based on the quasi-association assumption, we establish the almost complete convergence of the proposed estimator with a rate. It is worth noting that the main difficulty on the nonparametric estimation of the expected shortfall model is the absence of the backtesting measure. Thus, the principal advantage of the constructed kernel estimator is the possibility to provide the explicit expression of the estimator. Such features makes its implantation in practice very easy and its computational time very fast. Furthermore, we highlight the finite sample performance of the constructed estimator using an empirical analysis.

The paper is organized as follows: The next section is dedicated to the general framework of our functional time series setting. The main asymptotic properties of the constructed estimator is given in Section 3. Section 4 is devoted to discuss some computationabilities of the estimator. A real data application is presented in Section 5. Finally, the proof of the auxiliary results is given in the Appendix A.

2. Methodology

2.1. The Quasi-Associated Functional Time Series Data Framework

Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be n copies of a random vector, identically distributed as (X, Y) and valued in $\mathcal{H} \times \mathbb{R}$, $\mathcal{H}$ is a separable real Hilbert space. We denote by $\langle \cdot, \cdot \rangle$ the inner product on H and by $\| \cdot \|$ its associated norm. We suppose that $\mathcal{H}$ has complete orthonormal basis $(e_k)_{k \geq 1}$. The functional time series data considered in this work is characterised by the quasi-association correlation (see, Bulinski and Suquet [32] for the real case and Douge [33] for its definition in the Hilbert space). All along this work, we assume that the random pair $Z_i = \{(X_i, Y_i),\ i \in \mathbb{N}\}$ is the stationary quasi-associated processes, and we put λ_k as its covariance coefficient, as defined by:

$$\lambda_k := \sup_{s \geq k} \sum_{|i-j| \geq s} \lambda_{i,j},$$

where

$$\lambda_{i,j} = \sum_{k=1}^{\infty} \sum_{l=1}^{\infty} |Cov(X_i^k, X_j^k)| + \sum_{k=1}^{\infty} |Cov(X_i^k, Y_j)| + \sum_{l=1}^{\infty} |Cov(Y_i, X_j^l)| + |Cov(Y_i, Y_j)|.$$

X_i^k denotes the kth component of X_i defined as $X_i^k := < X_i, e_k >$. Furthermore, the quasi-associated functional time series of this contribution is carried out by the following assumptions

$$\text{The covariance coefficient} \quad (\lambda_k)_{k \in \mathbb{N}} \quad \text{such that } \lambda_k \leq Ce^{-ak}, \quad a > 0, \quad C > 0. \tag{1}$$

and

$$\exists C,\, C' > 0 \text{ such that } \mathbb{E}[\exp(|Y|)] < C \,\forall i \neq j \mathbb{E}\big(|Y_i Y_j| \mid X_i, X_j\big) \leq C' < \infty. \tag{2}$$

2.2. Model and Estimator

It is well known that the expected shortfall function is an alternative risk model defined without a backtesting measure. Its definition is closely linked to the CVaR function as the inverse of the conditional cumulative distribution. Therefore, to study the estimation of the conditional expected shortfall, we assume that the regular version of the conditional probability of Y given X exists. Moreover, we suppose that the conditional distribution of Y given X is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}$ and has a bounded density. Specifically, for $x \in \mathcal{H}$, we will denote by F^x the conditional cumulative distribution function, (*cdf.*) of Y given $X = x$ and by f^x the conditional density of Y given $X = x$.

Now, for $p \in (0, 1)$, the Conditional Value at Risk of order p, denoted by $CVaR_p(x)$, is defined by

$$F^x(CVaR_p(x)) = 1 - p. \tag{3}$$

Of course the existence and unicity of $CVaR_p(x)$ can be insured by assuming the strict monotony of the function $F^x(\cdot)$. Such a conditional model constitutes a preliminary step to estimate the conditional expected shortfall. Indeed, the conditional expected shortfall of order p denoted by $CES_p(\cdot)$ is defined by

$$\forall p \in (0,1) \qquad CES_p(x) = \mathbb{E}\big[Y|Y > CVaR_p(x),\, X = x\big] = \frac{1}{p}\int_{CVaR_p(x)}^{\infty} y f^x(y)dy.$$

Thus, the kernel estimator of $CES_p(\cdot)$ is constructed from the kernel estimation of both conditional models f^x and $CVaR_p(x)$. Both estimators are linked to the functional version of the kernel estimator of the conditional cumulative distribution function F^x introduced by [34], as follows

$$\widehat{F}^x(y) = \frac{\sum_{i=1}^n K(a_n^{-1}\|x - X_i\|)H(b_n^{-1}(y - Y_i))}{\sum_{i=1}^n K(a_n^{-1}\|x - X_i\|)} \tag{4}$$

where K is a kernel, H is a *cdf* and a_n (*resp.* b_n) is a sequence of positive real numbers.

Now, from Equation (4), we deduce that the natural estimator of $CVaR_p(x)$ is :

$$\widehat{F}^x(\widehat{CVaR}_p(x)) = p. \tag{5}$$

and of the conditional density f^x is

$$\widehat{f}^x(y) = \frac{\sum_{i=1}^n K(a_n^{-1}\|x - X_i\|)H'(b_n^{-1}(y - Y_i))}{b_n \sum_{i=1}^n K(a_n^{-1}\|x - X_i\|)}. \tag{6}$$

Thus, by a simple algebra, we express the kernel estimator of $CES_p(x)$, by

$$\widehat{CES}_p(x) = \frac{\sum_{i=1}^n K(a_n^{-1}\|x - X_i\|)\Big(b_n G(b_n^{-1}(\widehat{V}_p(x) - Y_i)) + Y_i\Big(1 - H(b_n^{-1}\Big(\widehat{V}_p(x) - Y_i\Big)\Big)\Big)\Big)}{p\sum_{i=1}^n K\Big(a_n^{-1}\|x - X_i\|\Big)} \tag{7}$$

where

$$G(s) = \int_s^{\infty} uH'(u).$$

3. Main Results

Throughout this paper, we set by c or c' some strictly positive generic constants, and N_x a given neighborhood of x and by $S_p(x)$ a given compact subset contains $CVaR_p(x)$. In addition, we assume the operators $\mathbb{E}[Y|X = x]$ and $CES_p(x)$ are a Holderian function of exponent $\beta_0 > 0$ and we consider the following assumptions:

(H1) $P(X \in B(x,r)) = \phi_x(r) > 0$ where $B(x,h) = \{x' \in \mathcal{F} \ : d(x',x) < h\}$.

(H2) $\exists\delta > 0, \forall(t_1,t_2) \in S_p(x)^2, \forall(x_1,x_2) \in N_x^2$,

$$|F^{x_1}(t_1) - F^{x_2}(t_2)| \le C_2\Big(d(x_1,x_2)^{b_1} + |t_1 - t_2|^{b_2}\Big), \quad \beta_1 > 0,\, \beta_2 > 0.$$

(H3) $\forall i \neq j$,

$$0 < \sup_{i \neq j} \mathbb{P}\big[(X_i, X_j) \in B(x,h) \times B(x,h)\big] = O(\phi_x^{1+a/(a+1))}(h)).$$

(H4) The kernel function $K(\cdot)$ is a bounded continuous Lipschitz function on $(0,1)$ such that

$$C_1 1\!\!1_{[0,1]}(t) < K(t) < C_2 1\!\!1_{[0,1]}(t).$$

(H5) The function H is of class $\mathcal{C}^2$ and its derivative H' is a symmetric kernel function and has compact support $[-1,1]$

(H6) There exist $\xi \in (0,1)$ and ξ_1, ξ_2, $\gamma > 0$ such that
$\frac{\log n^5}{n^{1-\xi-\xi_1}} \leq \phi_x(a_n) \leq \frac{1}{\log n^{1+\xi_2}} \leq (a_n b_n)^{2a/(a+1)}$ and $n^\gamma b_n \to 0$.

Comments on the hypotheses.

Clearly, all these assumptions are classical in this context of functional time series analysis. They are comparable similar to those considered by Douge (2010). Recall that quasi-associated correlation constitutes a very weak dependence condition allowing one to include many functional time series. Such consideration permits to increase the scope of the application of the proposed functional model. We refer to Ango Nze et al. (2002) for a list of the quasi-associated process including Bernoulli shifts class, Markov processes driven by discrete innovations and the AR(1) process with $\rho < 1/2$ and Bernoulli innovation, among others. They mentioned the GARCH process, which is common in the financial area and satisfies the weak dependency assumption. Moreover, it is demonstrated in Chernick (1981) that the autoregressive $\rho = 0.1$ and the innovation random variable as $Binom(10, 0.25)$ is not the $\alpha-$mixing assumption but is quasi-associated because it can be treated as a linear process with positive coefficients. On the other hand, the additional assumptions are (H5) and (H6) and are closely related to the additional parameters that are the smoothing parameter b_n and the function H. All in all, they are technical assumptions necessary for the simplicity of the proof. They are sufficient but not necessary. In particular, Condition (H4) can be replaced by taking K, as the Rosenblatt kernel has a continuous derivative function such that

$$\left| \int (K^j)'(s)\beta_x(s)ds \right| \leq C \quad \text{and} \quad j = 1, 2$$

where $\beta_x(s) = \lim_{r\to 0} \frac{\phi_x(sr)}{\phi_x(r)}$. Clearly such consideration can regroup more kernels not necessary with compact support.

Theorem 1. *Under the Hypotheses (H1)–(H6), we have,*

$$\left| \widehat{CES}_p(x) - CES_p(x) \right| = O\left(a_n^{\beta} + b_n^{\beta_2} + \sqrt{\frac{\log n}{n^{1-\xi}\phi_x(a_n)}} \right) \qquad a.co. \tag{8}$$

where $\beta = \min(\beta_0, \beta_1)$

Proof of the Theorem 1. By a simple mathematical analysis (see [28]) we prove that

$$\widehat{CES}_p(x) - CES_p(x) = \frac{1}{p}(\widehat{T}_n + b_n\widehat{S}_n + CVaR_p(x)\widehat{G}_n + \widehat{Q}_n + pCVaR_p(x) - pCES_p(x))$$

where

$$\widehat{T}_n = O(\widehat{CVaR}_p(x) - CVaR_p(x))^2,$$

$$\widehat{S}_n = \frac{\sum_{i=1}^n K(a_n^{-1}\|x - X_i\|)G(b_n^{-1}(CVaR_p(x) - Y_i))}{\sum_{i=1}^n K(a_n^{-1}\|x - X_i\|)},$$

$$\widehat{Q}_n = \frac{\sum_{i=1}^n K(a_n^{-1}\|x - X_i\|)Y_i(1 - H(b_n^{-1}(CVaR_p(x) - Y_i)))}{\sum_{i=1}^n K(a_n^{-1}\|x - X_i\|)}$$

and

$$\widehat{G}_n = \frac{\sum_{i=1}^n K(a_n^{-1}\|x - X_i\|)(1 - H(b_n^{-1}(CVaR_p(x) - Y_i)))}{\sum_{i=1}^n K(a_n^{-1}\|x - X_i\|)}$$

Firstly, we use the monotony of F^x and $\widehat{F}^x$

$$O(\widehat{CVaR}_p(x) - CVaR_p(x)) = O\left(\sup_{t \in S_p(x)} |\widehat{F}^x(t) - F^x(t)|\right).$$

Concerning the right term, we use the usual decomposition to write

$$\widehat{F}^x(t) - F^x(t) = \frac{1}{\widehat{F}_D^x}\left[\left(\widehat{F}_N^x(t) - \mathbb{E}\left[\widehat{F}_N^x(t)\right]\right) - \left(F^x(t)) - \mathbb{E}\left[\widehat{F}_N^x(t)\right]\right)\right] - \frac{F^x(t)}{\widehat{F}_D^x}\left[\widehat{F}_D^x - \mathbb{E}\left[\widehat{F}_D^x\right]\right]$$

where

$$\widehat{F}_N^x(t) := \frac{1}{n\,\mathbb{E}[K_1(x)]}\sum_{i=1}^n K_i(x)H(b_n^{-1}(t - Y_i)) \qquad \text{and} \qquad \widehat{F}_D^x := \frac{1}{n\,\mathbb{E}[K_1(x)]}\sum_{i=1}^n K_i(x).$$

with $K_i(x) = K(a_n^{-1}\|x - X_i\|)$.

Secondly, we use a similar decomposition to evaluate the quantities $\widehat{S}_n$, $\widehat{G}_n$ and $\widehat{Q}_n$. The main difference is only in the numerators. The latter is denoted, respectively,

$$\widehat{S}_N^x = \frac{1}{n\,\mathbb{E}[K_1(x)]}\sum_{i=1}^n K_i(x)G(b_n^{-1}(CVaR_p(x) - Y_i)),$$

$$\widehat{G}_N^x = \widehat{F}_D^x - \widehat{F}_N^x(CVaR_p(x))$$

$$\widehat{Q}_N^x = \frac{1}{n\,\mathbb{E}[K_1(x)]}\sum_{i=1}^n K_i(x)Y_i(1 - H(b_n^{-1}(CVaR_p(x) - Y_i)))$$

Thus, the Theorem 1 is a consequence of the following intermediate results, of which the proofs are given in the Appendix A. □

Lemma 1. *Under the hypotheses (H1), (H3), (H4) and (H6), we have,*

$$\sup_{t \in S_p(x)}\left|\widehat{F}_N^x(t) - \mathbb{E}\left[\widehat{F}_N^x(t)\right]\right| = O\left(\left(\frac{\log n}{n\,\phi_x(h)}\right)^{1/2}\right), \qquad a.co.$$

Lemma 2. *(see [34]) Under Hypotheses (H1), (H2) and (H4)–(H6), we have,*

$$\sup_{t \in S_p(x)}\left|F^x(t) - \mathbb{E}\left[\widehat{F}_N^x(t)\right]\right| = O\left(a_n^{\beta_1} + b_n^{\beta_2}\right).$$

Lemma 3. *Under the hypotheses (H1), (H3), (H4) and (H6), we have,*

$$\left|\widehat{S}_N^x - \mathbb{E}\left[\widehat{S}_N^x\right]\right| = O\left(\left(\frac{\log n}{n\,\phi_x(h)}\right)^{1/2}\right) \tag{9}$$

and

$$\widehat{F}_D^x - \mathbb{E}\left[\widehat{F}_D^x\right] = O\left(\left(\frac{\log n}{n\,\phi_x(h)}\right)^{1/2}\right) \qquad a.co. \tag{10}$$

Moreover

$$\sum_n \mathbb{P}\left(\widehat{F}_D^x < \frac{1}{2}\right) < \infty.$$

Lemma 4. *Under the hypotheses (H1), (H3), (H4) and (H6), we have,*

$$\left|\widehat{Q}_N^x - \mathbb{E}\left[\widehat{Q}_N^x\right]\right| = O\left(\left(\frac{\log n}{n^{1-\xi}\,\phi_x(h)}\right)^{1/2}\right), \qquad a.co.$$

Lemma 5. *Under Hypotheses (H1), (H2) and (H4)–(H6), we have,*

$$\left|\mathbb{E}\left[\widehat{S}_N^x\right]\right| = O\left(a_n^{\beta_1} + b_n^{\beta_2}\right),$$

and

$$\left|pCES_p(x) - \mathbb{E}\left[\widehat{Q}_N^x\right]\right| = O\left(a_n^{\beta_1} + b_n^{\beta_2}\right).$$

4. A Simulation Study

The main purpose of this empirical analysis is to examine the behavior of the proposed kernel estimator. More precisely, we aim to evaluate the impact of the different components of the estimator. We divide the set of these parameters into components : structural parameters and technical ones. The structural parameters concerns the nonparametric model used to generate the observations $(X_i, Y_i)_i$ and the technical parameters are the kernels, the metric and the bandwidths (a_n, b_n). So in the first illustration, we examine the effect of the structural parameters while the technical ones are postponed in the second part. Therefore, in the first part, we compare the performance of this estimator using different conditional distributions and different sample size. To do this, we generate an artificial data using the nonparametric regression relationship between the input and the output variables, as follows

$$Y_i = r(X_i) + \epsilon_i \text{ for } i = 1, \ldots, n \tag{11}$$

where the functional co-variates X are drawn by the following process

$$X(t) = cos(W^3 t^3) + sin(W^2 t^2) + Wt, \quad W \rightsquigarrow ARMA(2,2), \text{and } t \in [-\pi, \pi].$$

The coefficients of W are arbitrarily chosen equal to $ar = c(0.8897, -0.4858)$, $ma = c(-0.2279, 0.2488)$ and $\sigma^2 = 0.1796$. A sample of the regressors curves $X(t)$ are plotted in Figure 1.

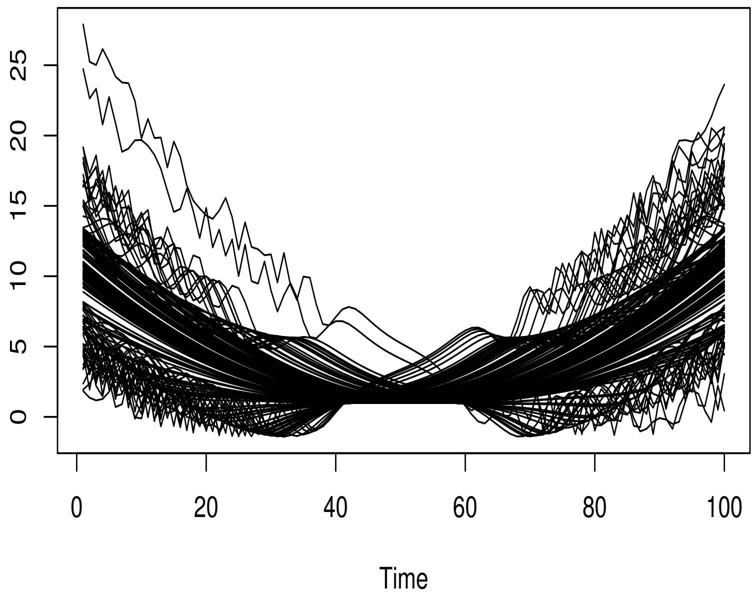

Figure 1. A sample of the functional curves.

The interest variable Y is obtained by taking as the regression operator

$$r(x) = 2\int_0^{\pi} \frac{x^2(t)}{1+x^2(t)}dt.$$

Recall that, in this kind of mechanism to generate the simulated data, the conditional distribution of Y given $X = x$ is obtained by shifting the law of the white noise of ϵ_i by the quantity $r(x)$. Thus, the theoretical expected shortfall as well as the true VaR function can be is explicitly calculated. Thus, in this experiment, we simulate with two distribution white noise ϵ_i. The first one is normal distribution ($N(0,1)$) and the second one is the Laplace distribution $Laplace(0,1)$.

Now, we describe the determination of the parameters involved in the estimator $\widehat{CES_p}$. Firstly, the optimal bandwidth a_n, b_n is chosen locally by using the following cross-validation rule:

$$(a_{CV_{opt}}(x), b_{CV_{opt}}(y)) = \arg \min_{a,b \in H_n(x,y)} \sum_{i=1}^{n} \left| Y_i - \widehat{CVaR_{0.5}}(X_i) \right|, \tag{12}$$

where $H_n(x,y)$ is the set of the positive real numbers (a,b), such that the ball centred at x with radius a and (respectively, the interval $(y-b, y+b)$) contains exactly k observations of X_i (respectively, of Y_i), $k \in \{5, 10, 15, \ldots n/2\}$. Finally, we propose that the estimator was computed by quadratic kernel on $(0,1)$ and we have used the L^2 metric associated to the PCA definition with $m = 3$ (see, Ferraty and Vieu [34])).

The efficiency of the estimation procedures is evaluated using the average absolute error:

$$ASE = \frac{1}{n}\sum_{i=1}^{n} \left| \widehat{E_p}(X_i) - E_p(X_i) \right|. \tag{13}$$

The simulation study performed over 100 replications. The obtained errors are displayed in the following box-plot Figure 2. It concerns the ASE of the two considered conditional distribution, of three values of sample size $n = 50, 150, 300$ and for $p = 0.01$.

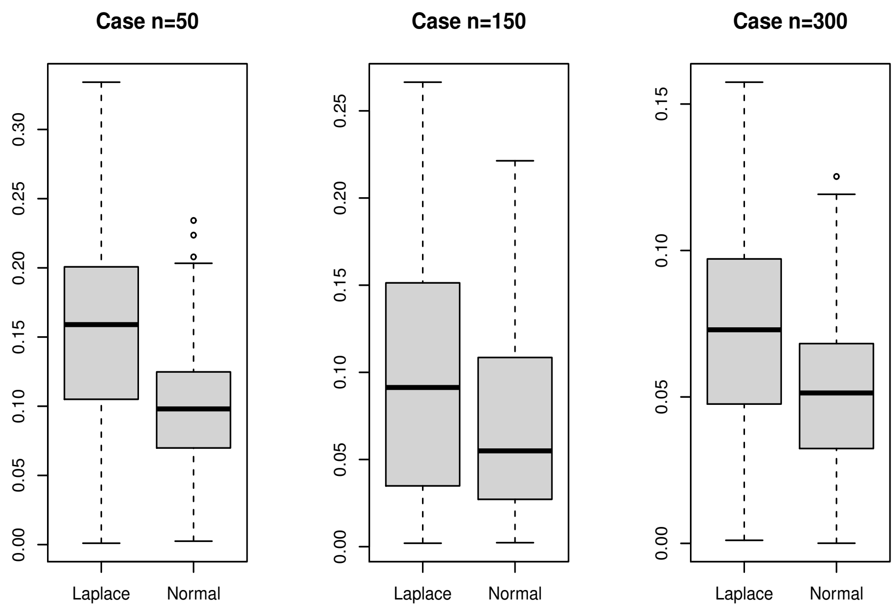

Figure 2. ASE with respect the sample size.

It clearly appears that the nonparametric estimation of the expected shortfall regression depends on the conditional distribution of Y given X. In that sense, there are significative difference in the quality estimation between the normality case and the non-normality situation. Secondly, it is clear that the performance of the estimator increases with the values of the sample size n. However, all in all the estimator $\widehat{CES_p}$ has a satisfactory performance for the small size in the case $n = 50$, and the execution time for the large sample size $n = 300$ is fast. Even if the execution time of any routine code is related to the characteristics of the microcomputer, for IdeaPad S340 with processor N4020, the function *Sys.time* in R computes the execution time by o 0.34801 secs for $n = 300$.

The second illustration concerns the examination of the impact of the technical parameters in the estimator. We focus on the choice of the kernel K and the smoothing parameters (a_n, b_n). For this aim, we compare three kernels

$$\begin{cases} K(t) = \frac{3}{2}(1-t^2)\mathbb{1}_{[0,1]} & \text{Quadratic on } [0,1], \\ K(t) = \frac{1}{\beta(2,3)}t(1-t)^2\mathbb{1}_{[0,1]} & \beta-\text{kernel on } [0,1], \\ K(t) = \mathbb{1}_{[0,1]} & \text{Uniform on } [0,1], \end{cases}$$

and three selectors of (a_n, b_n). The first one is defined as Equation (12) and the two others' rule defined by

$$(a_{CV_{opt}}(x), b_{CV_{opt}}(y)) = \arg\min_{a,b\in H_n(x,y)} \sum_{i=1}^{n} L_p(Y_i - \widehat{CVaR_p}(X_i)), \tag{14}$$

where $L_p(t) = (2p-1)t + |t|$ and

$$(a_{CV_{opt}}(x), b_{CV_{opt}}(y)) = \arg\min_{a,b\in H_n(x,y)} \sum_{i=1}^{n} |\widehat{CES_p}(X_i) - \overline{CES_p}(X_i)| \tag{15}$$

where $\overline{CES_p}(X_i)$ is the local empirical expected shortfall obtained from the fixed neighbourhood of X_i that contains J closest curves to X_i. It is defined by

$$\overline{CES_p}(X_i) = \frac{1}{J}\sum_{i=1}^{J} Y_i \mathbb{1}_{Y_i > \widehat{CVaR_p}(X_i)},$$

J is optimised over the subset $\{10, 15, 20, 30\}$. Next, we keep the same metric and the same schema of the data generating of the first illustration; we change only the conditional distribution which is related to the distribution of the white noise. Specifically, instead of the normal and laplace distributions of the first part, we consider the student distribution, which is more popular in financial time series data. The impact of the technical parameters is checked using the ASE-error Equation (13) for $n = 100$ and three values of $p = 0.01, 0.05$, and$p = 0.1$. An average of 100 replications of this error is given in the following table.

The results of Table 1 proves that the technical parameters also have an important effect on the estimation quality. The variability of the errors between different selectors or different kernels is relatively significative. In particular, it seems that the selection of the bandwidth parameter has more impact than the kernel function, in the sense that the variability of ASE with respect to the three selectors of (a_n, b_n) is more significative than the kernels.

Table 1. Comparison of ASE-error using different kernels and smoothing parameters.

Selector	Kernel	$p = 0.01$	$p = 0.05$	$p = 0.1$
Rule (12)	Quadratic	0.094	0.083	0.088
	β-kernel	0.089	0.078	0.084
	Uniform	0.109	0.112	0.142
Rule (14)	Quadratic	0.052	0.062	0.071
	β-kernel	0.048	0.057	0.067
	Uniform	0.074	0.079	0.088
Rule (15)	Quadratic	0.041	0.048	0.061
	β-kernel	0.039	0.043	0.059
	Uniform	0.055	0.050	0.054,

5. Real Data Example

The object of this section is to evaluate the performance of the kernel estimator of the expected shortfall regression using a real financial time series data. In the previous section, we proved that the easy implantation of the constructed estimator for different kernels, different bandwidths as well as different conditional models. We have observed, without surprising, that its efficiency is strongly influenced by these parameters. Recall that the main motivation of this kernel smoothing is to overcome the problem of the non-elicitability of the expected shortfall model. However, in the multivariate statistics there exists some alternative solutions based on jointly estimation the $CVaR_p(\cdots)$ and $CES_p(\cdot)$ using some adapted parametric scoring functions. Such solution has been studied by [35–37]. Thus the main purpose of this section is to conduct a comparative empirical analysis between our nonparametric functional approach based on the kernel smoothing and the cited studies based on the semi-parametric multivariate techniques. It's worth remembering that the nonparametric functional tools is more appropriate that the parametric models to fit the high-frequency data (see Ferraty and View [34]). So, we want in this paragraph to highlight this feature of nonparametric functional data analysis. For this aim, we consider as financial time series the interday return of the stock index Nikkei during the period 11 October 1983 to 11 October 2022. The studied data was obtained from the website https://fred.stlouisfed.org/series/NIKKEI225 (accessed on 14 September 2022). Undoubtedly, this kind of financial times series data exhibits the principal characteristics of financial data: volatility, non-normality, heavy tailed distribution, excess kurtosis, ... In Figure 3, we plot the process of the return index $r(t)$.

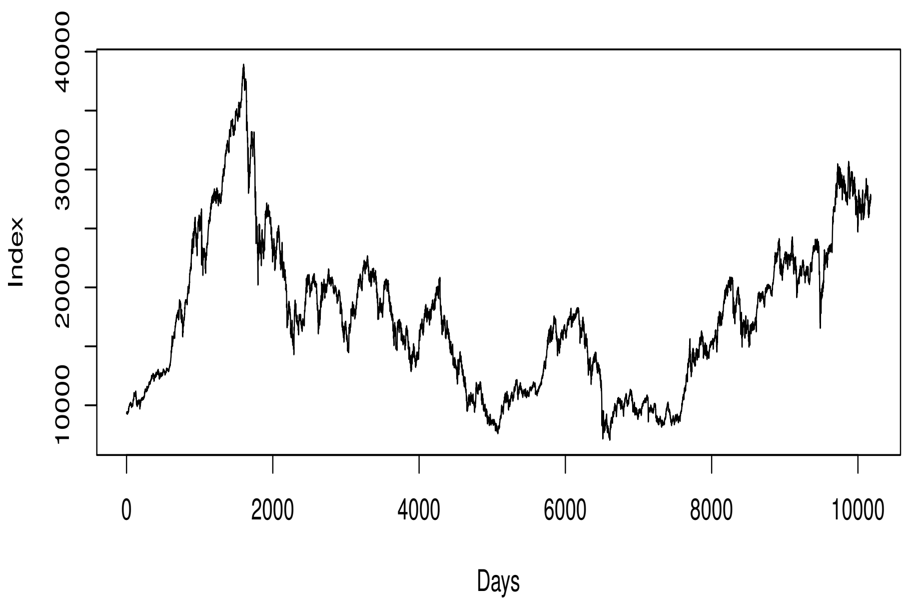

Figure 3. The daily return for the stock index.

Often in financial times series analysis we aim to forecast a future values of the process $Z(t) = -100 \log\left(\frac{r(t)}{r(t-1)}\right)$ index knowing its historical trajectory. In our functional context, we formulate this forecasting problem by putting $Y_i = Y = Z(d_i)$ (the value of $Z(.)$ at day d_i) and the functional covariate $X_i(t) = Z(t)_{t\in[d_i-R,d_i-1]}$ that the historical values of the index of the last R-days before d_i. In this empirical study, we conduct comparative study using different values of $R = 3, 6, 10, 30, 90$. We point out that for the Multivariate Semi-Parametric Model (MSPM), we use the code routine *esreg* from the R-package *esrreg* using two types of specification functions that defined the scoring functions (see Dimitriadis and Bayer [37] for more details in this approach). Recall that this kind of approach requires a limited dimension of regressors. Therefore, we have combined this function with the principal component analysis to reduce the dimension of the regressors when the R is large. Next, for the Non-Parametric Functional Model (NPFM), we use the β-kernel and we select the smoothing parameters by the rule (14). Moreover, we explore the functional structure of the data by considering two situations: smoothing and non-smoothing cases. In the smoothing case, we use the spline metric, whereas in the non-smoothing case, we use the PCA-metric. We return to Ferraty and Vieu [34] for more details on the mathematical formulation of the two metrics. In the following Figures 4 and 5 we display the initial curves and their smoothed version one.

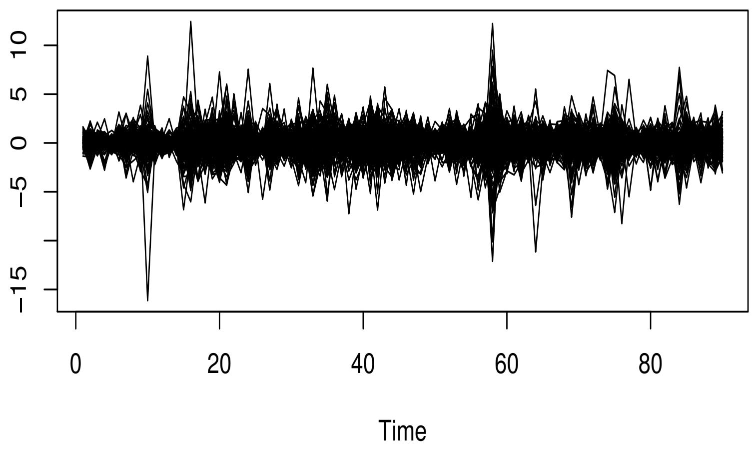

Figure 4. The initial curves; the case when $R = 90$.

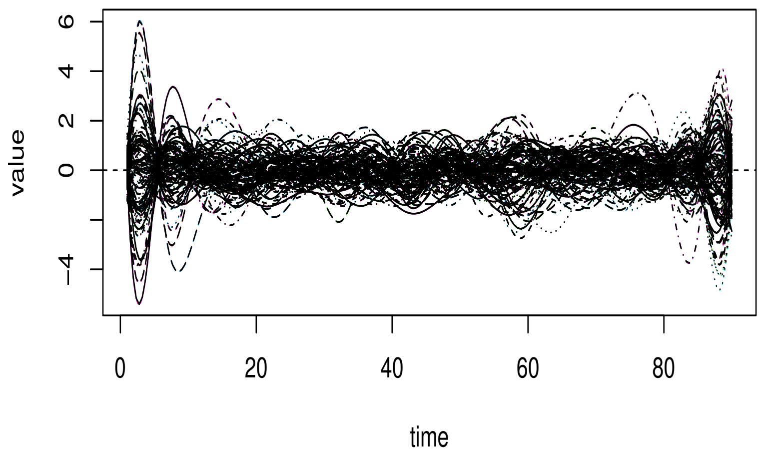

Figure 5. The smoothed curves; the case when $R = 90$.

Furthermore, the performance of both models are examined for three values of $p = 0.01, 0.05, 0.1$ using the error defined by the mean absolute value, as follows

$$MAE(p) = \frac{1}{N}\sum_{j=1}^{n} |\widehat{CES}_p(X_i) - \widetilde{CES}_p(X_i)|$$

where $\widetilde{CES}_p(X_i) = \frac{1}{N}\sum_{j=1}^{n} Y_i \mathbb{1}_{Y_i > \widehat{CVaR}_p(X_i)}$ and $N = \sum_{j=1}^{n} \mathbb{1}_{Y_i > \widehat{CVaR}_p(X_i)}$. The comparison results are reported in the Table 2.

Table 2. Comparison of the MAE- error between Nonparametric functional model and semiparametric multivariate model.

Model	p	$R = 3$	$R = 6$	$R = 10$	$R = 30$	$R = 90$
MSPM with	0.01	1.92	2.05	2.29	2.56	2.98
specification functions	0.05	1.75	1.96	2.16	2.29	2.81
$G_1(t) = 0$ and $G_2(t) = e^t$	0.1	1.67	1.84	1.93	2.07	2.28
MSPM with	0.01	1.81	1.96	1.07	1.43	1.78
specification functions	0.05	1.62	1.81	1.08	1.17	1.74
$G_1(t) = t$ and $G_2 = \frac{e^t}{1+e^t}$	0.1	1.43	1.63	1.71	1.94	2.19
NPFM with	0.01	1.87	1.98	2.26	1.52	2.91
non-smoothing curves	0.05	1.71	1.89	2.09	2.22	2.76
	0.1	0.65	1.79	1.88	2.01	2.23
NPFM with	0.01	2.35	1.04	1.13	2.32	2.09
smoothing curves	0.05	2.26	0.97	1.06	1.28	1.78
	0.1	1.95	0.84	0.83	0.76	0.99

Unsurprisingly, we can observe that the semi-parametric approach has a good performance for small $R = 3$, while the non-smoothing functional approach is equivalent to MSPM for small $R = 3$, and its smoothing version has a substantial efficiency when it is large $R \geq 30$. However, the smoothing functional approach is not applicable of $R \leq 10$. Next, it appears clearly that the two approaches are impacted by the choice of the parameters used in the estimation processing. In particular, the parametric approach is affected by the choice of the specification functions associated to the scoring function. This is the case when $G_1(t) = t$ and $G_2(t) = \frac{e^t}{1+e^t}$ seems more appropriate than the first case. It is he same thing for the functional case: the smoothing case is better than the non-smoothing case when R is large.

An additional check is performed using another backtest based on the cover test developed by Bayer and Dimitriadis [38]. We apply the version, the so-called one-side intercept expected shortfall regression backtest. Precisely, we compare our functional approach to the GARCH specifications with Student-t using this cover-test. The latter is available by the routine code *esr-backtest* from the R-package *esrback*. Here, we choose $\alpha = 0.05$. For our aim we split, randomly, many times (exactly 100 times), the data into subsets (70% learning sample and 30% testing sample), and for each time we compute the empirical p-values of both approaches when $R = 90$. In the following Figure 6, we display the obtained empirical values versus the threshold $\alpha = 0.025$ in the horizontal line.

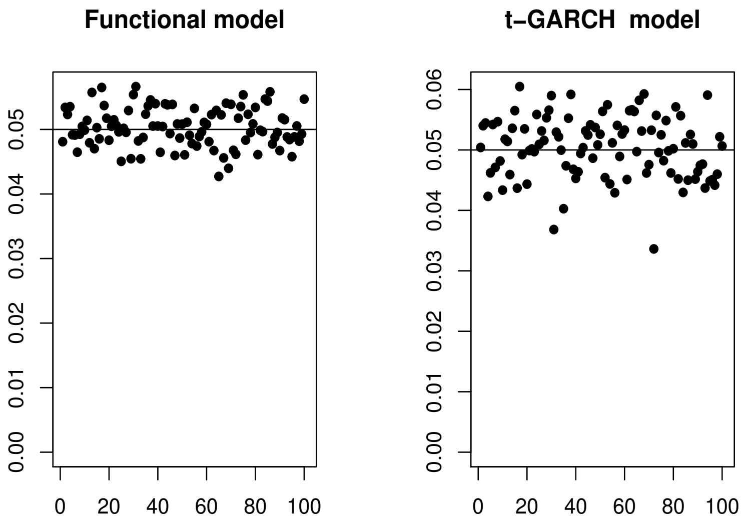

Figure 6. Comparison of the p-values between t-GARCH nonparametric functional approach.

The obtained results confirms the statement concluded by the table. In sense that the nonparametric functional procedure is good solution of the high-frequency data associated to large R. It's worth noting that also of small large R the functional approach with the PCA metric constitutes a competitive algorithm to the semi-parametric approach.

6. Conclusions

This contribution focuses on the nonparametric estimation of the expected shortfall regression. An estimator based on the kernel method was constructed and its asymptotic property was proved under a weak dependence condition. The easy implementation of this estimator was checked using a simulated data. The empirical analysis incorporates the theoretical part, where we observe that the efficiency of the estimator is linked to various aspects including the regularity of the nonparametric models, the smoothness of the functional regressor as well as the choice of the bandwidth parameters a_n and b_n. One of the main feature of the our estimator is the possibility to express its form explicitly. Such a statement allows one to overcome the problem of the nonexistence of the backtesting measure in this model. The present contribution offers also an important number of prospects in the future. The first natural prospect is the establishment of the asymptotic normality of the constructed estimator under the quasi-association assumption. The generalisation of the present result to the spatial case constitutes an important open question. In addition to the treatment of an alternative functional time series structure, there is the use of other smoothing methods such as the kNN method, local linear approach or the recursive algorithm, among others.

Author Contributions: The authors contributed approximately equally to this work. Formal analysis, L.A.-H.; Validation, Z.K.; Writing—review and editing, A.L. and M.R. All authors have read and agreed to the final version of the manuscript.

Funding: This research was funded by the Deanship of Scientific Research at King Khalid University through the Research Groups Program under grant number R.G.P. 1/177/43.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this study are available through https://fred.stlouisfed.org/series/DJIA (accessed on 14 September 2022).

Acknowledgments: The authors are indebted to the Editor-in-Chief, Associate Editor and the three referees for their very generous comments and suggestions on the first version of our article, which helped us to improve content, presentation, and layout of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

In what follows, when there is no confusion, we will denote, for all $i = 1, \ldots, n$, by:

$$H_i(y) = H(b_n^{-1}(y - Y_i)) \quad \text{and} \quad G_i(y) = G(b_n^{-1}(y - Y_i)).$$

Now, we state the following lemmas, which are needed to establish our asymptotic results

Lemma A1. *(See, Kallabis and Neumann [39])*
Let $X_1, \ldots, X_n$ the real random variables such that $\mathbb{E}X_j = 0$ and $\mathbb{P}(|X_j| \le M) = 1$, for all $j = 1, \ldots, n$ and some $M < \infty$. Let $\sigma_n^2 = Var\left(\sum_{j=1}^n X_j\right)$. Assume, furthermore, that there exist $K < \infty$ and $\beta > 0$ such that, for all u-tuplets $(s_1, \ldots, s_u)$ and all v-tuplets $(t_1, \ldots, t_v)$ with $1 \le s_1 \le \ldots \le s_u \le t_1 \le \ldots \le t_v \le n$, the following inequality is fulfilled :

$$\left| Cov(X_{s_1} \ldots X_{s_u}, X_{t_1} \ldots X_{t_v}) \right| \le K^2 M^{u+v-2} v e^{-\beta(t_1 - s_u)}.$$

Then,

$$\mathbb{P}\left(\left| \sum_{j=1}^n X_j \right| > t \right) \le \exp\left\{ -\frac{t^2/2}{A_n + B_n^{\frac{1}{3}} t^{\frac{5}{2}}} \right\}$$

for some $A_n \le \sigma_n^2$ *and* $B_n = \left(\dfrac{16nK^2}{9A_n(1 - e^{-\beta})} \vee 1 \right) \dfrac{2(K \vee M)}{1 - e^{-\beta}}$.

Proof of Lemma 3. The proof of this lemma is based on the compactness of $S_p(x)$. Indeed, we use the fact that $S_p(x) \subset \bigcup_{k=1}^{d_n} S_k$ where $S_k = (t_k - l_n,\ t_k + l_n)$, $\quad d_n = O(l_n^{-1}) = O(n^{(2\gamma+1)/2})$. We define $t_y = \arg\min_{t \in \{t_1, \ldots, t_{s_n}\}} |y - t|$, we have

$$\begin{aligned}
\frac{1}{\widehat{F}_D^x} \sup_{y \in S_p(x)} \left| \widehat{F}_N^x(y) - \mathbb{E}\widehat{F}_N^x(y) \right| \quad &\le \quad \underbrace{\frac{1}{\widehat{F}_D^x} \sup_{y \in S_p(x)} \left| \widehat{F}_N^x(y) - \widehat{F}_N^x(t_y) \right|}_{I_1} + \\
\underbrace{\frac{1}{\widehat{F}_D^x} \sup_{y \in S_p(x)} \left| \widehat{F}_N^x(t_y) - E\widehat{F}_N^x(t_y) \right|}_{I_2} + \quad & \quad \underbrace{\frac{1}{\widehat{F}_D^x} \sup_{y \in S_p(x)} \left| \mathbb{E}\widehat{F}_N^x(t_y) - \mathbb{E}\widehat{F}_N^x(y) \right|}_{I_3}. \qquad \text{(A1)}
\end{aligned}$$

- Concerning (I_1): We use the Lipschitzian condition on H, we obtain

$$\begin{aligned}
\frac{1}{\widehat{F}_D^x} \sup_{y \in S_p(x)} \left| \widehat{F}_N^x(y) - \widehat{F}_N^x(t_y) \right| \quad &\le \quad \frac{1}{\widehat{F}_D^x} \sup_{y \in S_p(x)} \frac{1}{n\, \mathbb{E}[K_1(x)]} \sum_{i=1}^n \left| H_i(y) - H_i(t_y) \right| K_i(x), \\
&\le \quad \frac{1}{\widehat{F}_D^x} \sup_{y \in S_p(x)} \frac{C|y - t_y|}{b_n} \left(\frac{1}{n\, \mathbb{E}[K_1(X)]} \sum_{i=1}^n K_i(x) \right), \\
&\le \quad C \frac{l_n}{b_n}. \qquad \text{(A2)}
\end{aligned}$$

Because of (H6), we have

$$\frac{l_n}{b_n} = o\left(\sqrt{\frac{\log n}{n\, \phi_x(a_n)}} \right).$$

Thus, for n large enough, we can write

$$\mathbb{P}\left(\frac{1}{\widehat{F}_D^x}\sup_{y\in S_p(x)}\left|\widehat{F}_N^x(y)-\widehat{F}_N^x(t_y)\right| > \frac{\eta}{3}\sqrt{\frac{\log n}{n\,\phi_x(a_n)}}\right) \;=\; 0.$$

It follows that

$$\frac{1}{\widehat{F}_D^x}\sup_{y\in S_p(x)}\left|\widehat{F}_N^x(y)-\widehat{F}_N^x(t_y)\right| = o\left(\sqrt{\frac{\log n}{n\,\phi_x(a_n)}}\right) \tag{A3}$$

and

$$\frac{1}{\widehat{F}_D^x}\sup_{y\in S_p(x)}\left|\mathbb{E}\widehat{F}_N^x(y)-\mathbb{E}\widehat{F}_N^x(t_y)\right| = o\left(\sqrt{\frac{\log n}{n\,\phi_x(a_n)}}\right). \tag{A4}$$

- Concerning (I_2): It is clear that

$$\begin{aligned}
&\mathbb{P}\left(\sup_{y\in S_p(x)}\left|\widehat{F}_N^x(t_y)-E\widehat{F}_N^x(t_y)\right| > \frac{\eta}{3}\sqrt{\frac{\log n}{n\,\phi_x(a_n)}}\right)\\
&= \mathbb{P}\left(\max_{t_y\in\{t_1,\dots,t_{s_n}\}}\left|\widehat{F}_N^x(t_y)-E\widehat{F}_N^x(t_y)\right| > \frac{\eta}{3}\sqrt{\frac{\log n}{n\,\phi_x(a_n)}}\right)\\
&\leq\; s_n \max_{t_y\in\{t_1,\dots,t_{s_n}\}}\mathbb{P}\left(\left|\widehat{F}_N^x(t_y)-E\widehat{F}_N^x(t_y)\right| > \frac{\eta}{3}\sqrt{\frac{\log n}{n\,\phi_x(a_n)}}\right).
\end{aligned}$$

We write

$$\widehat{F}_N^x(t_y)-E\widehat{F}_N^x(t_y)=\sum_{i=1}^n \Delta_i.$$

with $\forall 1\leq i\leq n$,

$$\Delta_i=\frac{1}{n\mathbb{E}[K_1(x)]}\left(K_i(x)H_i(t_y)-\mathbb{E}\left[K_1(x)H_1(t_y)\right]\right)=\frac{1}{n\mathbb{E}[K_1(x)]}\chi(X_i,Y_i),$$

with

$$\chi(z,w)=H(b_n^{-1}(t_y-w))K(a_n^{-1}\|x-z\|))-\mathbb{E}\left[H_1(t_y)K_1(x)\right],\; z\in\mathcal{H},\; w\in\mathbb{R}.$$

So, the rest of the proof is based on the application of Lemma A1 on the variables Δ_i. To do that, we use the fact that (For any function f we denote by $\|f\|_\infty$ the supremun norm)

$$\|\chi\|_\infty \leq 2C\|K\|_\infty\|H\|_\infty$$
$$\mathrm{Lip}(\chi)\leq C(a_n^{-1}\|H\|_\infty\mathrm{Lip}(K)+b_n^{-1}\|K\|_\infty\mathrm{Lip}(H))$$

and we evaluate the covariance term

$$Cov(\Delta_{s_1}\dots\Delta_{s_u},\Delta_{t_1}\dots\Delta_{t_v}),\quad (s_1,\dots,s_u,t_1,\dots,t_v))\in\mathbb{N}^{u+v}.$$

We consider two cases:

– If $t_1=s_u$. By using the fact that $\mathbb{E}\left[\left|H_1^2(t_y)K_1^2(x)\right|\right]=O(\phi_x(a_n))$ and $\mathbb{E}[|K_1(x)|]=O(\phi_x(a_n))$ to get

$$\begin{aligned}
|Cov(\Delta_{s_1}\dots\Delta_{s_u},\Delta_{t_1}\dots\Delta_{t_v})| &\leq \left(\frac{C\|H\|_\infty\|K\|_\infty}{n\mathbb{E}[K_1(x)]}\right)^{u+v}\mathbb{E}\left[\left|K_1^2(x)H_1^2(t_y)\right|\right]\\
&\leq \phi_x(a_n)\left(\frac{C}{n\phi_x(a_n)}\right)^{u+v}.
\end{aligned} \tag{A5}$$

– If $t_1 > s_u$, we use the quasi-association definition to write that

$$\begin{aligned}|Cov(\Delta_{s_1}\dots\Delta_{s_u},\Delta_{t_1}\dots\Delta_{t_v})| &\leq \left(\left(a_n^{-1}\mathrm{Lip}(K)+b_n^{-1}\mathrm{Lip}(H)\right)(n\mathbb{E}[K_1(x)])^{-1}\right)^2 \\ &\quad \left(\frac{C}{n\mathbb{E}[K_1(x)]}\right)^{u+v-2}\sum_{i=1}^{u}\sum_{j=1}^{v}\lambda_{s_i,t_j} \\ &\leq \left(a_n^{-1}\mathrm{Lip}(K)+b_n^{-1}\mathrm{Lip}(H)\right)^2\left(\frac{C}{n\mathbb{E}[K_1(x)]}\right)^{u+v} v\lambda_{t_1-s_u} \\ &\leq \left(a_n^{-1}\mathrm{Lip}(K)+b_n^{-1}\mathrm{Lip}(H)\right)^2\left(\frac{C}{n\phi_x(a_n)}\right)^{u+v} ve^{-a(t_1-s_u)}. \end{aligned} \quad \text{(A6)}$$

On the other hand, we have,

$$\begin{aligned}|Cov(\Delta_{s_1}\dots\Delta_{s_u},\Delta_{t_1}\dots\Delta_{t_v})| &\leq \left(\frac{C\|H\|_\infty\|K\|_\infty}{n\mathbb{E}[K_1(x)]}\right)^{u+v-2}\times \\ &\quad (|\mathbb{E}[\Delta_{s_u}\Delta_{t_1}]|+\mathbb{E}|\Delta_{s_u}|\mathbb{E}|\Delta_{t_1}|) \\ &\leq \left(\frac{C\|H\|_\infty\|K\|_\infty}{n\mathbb{E}[K_1(x)]}\right)^{u+v-2}\left(\frac{C}{n\mathbb{E}[K_1(x)]}\right)^2\times \\ &\quad \left(\phi_x^{(a+1)/a}(a_n)+\phi_x^2(a_n)\right) \\ &\leq \left(\frac{C}{n\phi_x(a_n)}\right)^{u+v}\phi_x^{(a+1)/a}(a_n). \end{aligned} \quad \text{(A7)}$$

Furthermore, taking a $\frac{1}{2(a+1)}$-power of Equation (A6), $(\frac{2a+1}{a+1})$-power of Equation (A7), we obtain for $1 \leq s_1 \leq \ldots \leq s_u \leq t_1 \leq \ldots \leq t_v \leq n$:

$$|Cov(\Delta_{s_1}\dots\Delta_{s_u},\Delta_{t_1}\dots\Delta_{t_v})| \leq \phi_x(a_n)\left(\frac{C}{n\phi_x(a_n)}\right)^{u+v} ve^{-a(t_1-s_u)/(2(a+1))}.$$

The application of Lemma A1 requires also the evaluation of the variance term

$$\begin{aligned}Var\left(\sum_{i=1}^{n}\Delta_i\right) &= \sum_{i=1}^{n}\sum_{j=1}^{n}Cov(\Delta_i,\Delta_j) \\ &= Var(\Delta_1) \\ &+ \sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}Cov(\Delta_i,\Delta_j). \end{aligned} \quad \text{(A8)}$$

For the first term, we have

$$\mathbb{E}\left[H_1^2(t_y)K_1^2(x)\right] = \mathbb{E}\left[K_1^2(x)\mathbb{E}\left[H_1^2(t_y)|X_1\right]\right].$$

By integration on the real component y, we obtain

$$\mathbb{E}[[H_1^2(t_y)|X_1] = O(1).$$

As, for all $j \geq 1$, $\mathbb{E}\left[K_1^j(x)\right] = O(\phi_x(a_n))$, then

$$\mathbb{E}\left[H_1^2(t_y)K_1^2(x)\right] = O(\phi_x(a_n)).$$

It follows that

$$Var(\Delta_1) \quad = \quad O\left(\frac{1}{n\phi_x(a_n)}\right). \tag{A9}$$

Concerning the covariance term in Equation (A8), we use the following decomposition

$$\begin{aligned}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} Cov(\Delta_i,\Delta_j) \quad &= \quad \sum_{i=1}^{n}\sum_{\substack{j=1\\ 0<|i-j|\le m_n}}^{n} Cov(\Delta_i,\Delta_j) \\ &\quad + \sum_{i=1}^{n}\sum_{\substack{j=1\\ |i-j|>m_n}}^{n} Cov(\Delta_i,\Delta_j) \\ &=: \quad I + II.\end{aligned}$$

where (m_n) is a sequence of a positive integer, which goes to infinity as $n \to \infty$. Using the same idea as in Equation (A7), we prove that for all $i \neq j$

$$I \le Cnm_n\phi_x^{(a+1)/a}(a_n). \tag{A10}$$

Since both kernels H and K are bounded and Lipschitz, we obtain

$$\begin{aligned} II \quad &\le \quad \left(a_n^{-1}\mathrm{Lip}(K) + b_n^{-1}\mathrm{Lip}(H)\right)^2 \sum_{i=1}^{n}\sum_{\substack{j=1\\ |i-j|>m_n}}^{n} \lambda_{i,j} \\ &\le \quad C\left(a_n^{-1}\mathrm{Lip}(K) + b_n^{-1}\mathrm{Lip}(H)\right)^2 \sum_{i=1}^{n}\sum_{\substack{j=1\\ |i-j|>m_n}}^{n} \lambda_{i,j} \\ &\le \quad C\left(a_n^{-1}\mathrm{Lip}(K) + b_n^{-1}\mathrm{Lip}(H)\right)^2 \sum_{i=1}^{n}\sum_{\substack{j=1\\ |i-j|>m_n}}^{n} \lambda_{i,j} \\ &\le \quad Cn\left(a_n^{-1}\mathrm{Lip}(K) + b_n^{-1}\mathrm{Lip}(H)\right)^2 \lambda_{m_n} \\ &\le \quad Cn\left(a_n^{-1}\mathrm{Lip}(K) + b_n^{-1}\mathrm{Lip}(H)\right)^2 e^{-am_n}. \end{aligned} \tag{A11}$$

Then, by Equations (A10) and (A11), we obtain

$$\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} Cov(\Delta_i,\Delta_j) \quad \le \quad C\left(nm_n\phi_x^{(a+1)/a}(a_n) + n\left(a_n^{-1}\mathrm{Lip}(K) + b_n^{-1}\mathrm{Lip}(H)\right)^2 e^{-am_n}\right).$$

By choosing $m_n = \log\left(\frac{\left(a_n^{-1}\mathrm{Lip}(K)+b_n^{-1}\mathrm{Lip}(H)\right)^2}{a\phi_x^{(a+1)/a}(a_n)}\right)$, we obtain

$$\frac{1}{n\phi_x(a_n)}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} Cov(\Delta_i,\Delta_j) \quad \to \quad 0, \text{ as } n \to \infty. \tag{A12}$$

Finally, by combining results in Equations (A9) and (A12), we obtain

$$Var\left(\sum_{i=1}^{n}\Delta_i\right) = O\left(\frac{1}{n\phi_x(a_n)}\right). \tag{A13}$$

So, the variables Δ_i, $i = 1, \ldots, n$ satisfy the conditions of Lemma A1 for

$$K_n = \frac{C}{n\sqrt{\phi_x(a_n)}}, \ M_n = \frac{C}{n\phi_x(a_n)} \ \text{ and } \ A_n = Var\left(\sum_{i=1}^{n} \Delta_i\right).$$

Thus,

$$\begin{aligned}
&\mathbb{P}\left(\left|\widehat{F}_N^x(t_y) - E\widehat{F}_N^x(t_y)\right| > \eta\sqrt{\frac{\log n}{n\phi_x(a_n)}}\right) \\
&= \quad \mathbb{P}\left(\Big|\sum_{i=1}^{n} \Delta_i\Big| > \eta\sqrt{\frac{\log n}{n\phi_x(a_n)}}\right) \\
&\leq \quad \exp\left\{-\frac{\eta^2 \log n}{(n\phi_x(a_n))\left(Var(\sum_{i=1}^{n} \Delta_i) + \frac{\log^{5/6} n}{(n\phi_x(a_n))^{(7/6)}}\right)}\right\} \\
&\leq \quad \exp\left\{-\frac{\eta^2 \log n}{\left(C + \frac{\log^{5/6} n}{(n\phi_x(a_n))^{(1/6)}}\right)}\right\} \\
&\leq \quad C' \exp\left\{-C\eta^2 \log n\right\}
\end{aligned}$$

Finally, for a suitable choice of η allows one to obtain

$$\frac{1}{\widehat{F}_D^x} \sup_{y \in S_p(x)} \left|\widehat{F}_N^x(t_y) - E\widehat{F}_N^x(t_y)\right| = O_{a.co}\left(\sqrt{\frac{\log n}{n\,\phi_x(a_n)}}\right). \tag{A14}$$

□

Proof of Lemma 2. The details of the proof of this lemma is omitted. It is similar to the second part of Lemma 2. It suffices to replace H in the definition of Δ_i either by G for the statement or by 1 for the Equation (10). Concerning the last required result, we have

$$\mathbb{P}\left(\left|\widehat{F}_D^x\right| \leq 1/2\right) \leq \mathbb{P}\left(\left|\widehat{F}_D^x - 1\right| > 1/2\right) \leq \mathbb{P}\left(\left|\widehat{F}_D^x - \mathbb{E}\left[\widehat{F}_D^x\right]\right| > 1/2\right)$$

and using Equation (10) to deduce that

$$\sum_{n=1}^{\infty} \mathbb{P}\left(\left|\widehat{F}_D^x\right| \leq 1/2\right) < \infty$$

which completes the proof of the last statement of this lemma. □

Proof of Lemma 4. We write

$$\widehat{Q}_N^x = \widetilde{Y}_N + \widehat{Y}_N$$

with

$$\widetilde{Y}_N = \frac{1}{n\,\mathbb{E}[K_1(x)]} \sum_{i=1}^{n} K_i(x) Y_i$$

and

$$\widehat{Y}_N = \frac{1}{n\,\mathbb{E}[K_1(x)]} \sum_{i=1}^{n} K_i(x) Y_i H(b_n^{-1}(CVaR_p(x) - Y_i))$$

So, all it remains to show that

$$\left|\widetilde{Y}_N - \mathbb{E}\left[\widetilde{Y}_N\right]\right| = O\left(\left(\frac{\log n}{n^{1-\xi}\,\phi_x(h)}\right)^{1/2}\right), \qquad a.co$$

and

$$\left|\widehat{Y}_N - \mathbb{E}\left[\widehat{Y}_N\right]\right| = O\left(\left(\frac{\log n}{n^{1-\xi}\,\phi_x(a_n)}\right)^{1/2}\right), \qquad a.co.$$

Because of the similarity between the proof of both terms, we focus only on the last one. Furthermore, its proof is based on the same arguments as in Lemma 1; the main difference is the additional variable Y_i which is is not necessary bounded. To overcome this issue we employ the truncation method and we define

$$\widehat{Y}_N^* = \frac{1}{n\,\mathbb{E}[K_1(x)]}\sum_{i=1}^{n} K\left(a_n^{-1}\|x - X_i\|\right) Y_i H(b_n^{-1}(CVaR_p(x) - Y_i))\mathbb{1}_{|Y_i|<\mu_n} \text{ with } \mu_n = n^{\xi/6}.$$

Then, the claimed result is a consequence of the following intermediates results

$$\left|\mathbb{E}[\widehat{Y}_N^*] - \mathbb{E}[\widehat{Y}_N]\right| = O\left(\sqrt{\frac{\log n}{n^{1-\xi}\phi_x(a_n)}}\right), \tag{A15}$$

$$\left|\widehat{Y}_N^* - \widehat{Y}_N\right| = O_{a.co.}\left(\sqrt{\frac{\log n}{n^{1-\xi}\phi_x(a_n)}}\right) \tag{A16}$$

and

$$\left|\widehat{Y}_N^* - \mathbb{E}[\widehat{Y}_N^*]\right| = O_{a.co.}\left(\sqrt{\frac{\log n}{n^{1-\xi}\phi_x(a_n)}}\right). \tag{A17}$$

Firstly, to prove Equation (A17) we write :

$$\widehat{Y}_N^* - \mathbb{E}\left[\widehat{Y}_N^*\right] \quad = \quad \sum_{i=1}^{n} \Lambda_i$$

where

$$\Lambda_i = \frac{1}{n\,\mathbb{E}[K_1(x)]}\chi'(X_i, Y_i),$$

with

$$\chi'(z, w) = wH(b_n^{-1}(t_y - w))K(a_n^{-1}\|x - z\|)) - \mathbb{E}\left[YH_1(t_y)K_1(x)\right],\, z \in \mathcal{H},\, w \in \mathbb{R}.$$

Clearly,

$$\|\chi'\|_\infty \leq C\mu_n\|K\|_\infty\|H\|_\infty \text{ and } \mathrm{Lip}(\chi') \leq\leq C\mu_n(a_n^{-1}\|H\|_\infty\mathrm{Lip}(K) + b_n^{-1}\|K\|_\infty\mathrm{Lip}(H)).$$

Once again, we apply the inequality of Kallabis and Newmann on Λ_i for which we must evaluate asymptotically the quantities $Var(\sum_{i=1}^n \Lambda_i)$ and $Cov(\Lambda_{s_1}\ldots\Lambda_{s_u}, \Lambda_{t_1}\ldots\Lambda_{t_v})$, for all $(s_1, \ldots, s_u) \in \mathbb{N}^u$ and $(t_1, \ldots, t_v) \in \mathbb{N}^v$. Both quantities are treated by the same techniques as in the proof of (A13) and we obtain

$$Var\left(\sum_{i=1}^{n}\Lambda_i\right) = O\left(\frac{1}{n\phi_x(a_n)}\right).$$

and

$$|Cov(\Lambda_{s_1}, \ldots \Lambda_{s_u}, \Lambda_{t_1}, \ldots \Lambda_{t_v})| \quad \leq \quad \phi_x(a_n)\left(\frac{C\mu_n}{n\phi_x(a_n)}\right)^{u+v} ve^{-a(t_1-s_u)/(2(a+1))}.$$

So, we apply Theorem 2.1 of Kallabis and Newmann (2006, p. 2) for the variables Λ_i, $i = 1, \dots, n$ where

$$K_n = \frac{C\mu_n}{n\sqrt{\phi_x(a_n)}}, \ M_n = \frac{C\mu_n}{n\phi_x(a_n)} \ \text{ and } A_n = Var\left(\sum_{i=1}^{n} \Lambda_i\right) = O\left(\frac{1}{n\phi_x(a_n)}\right).$$

It follows that

$$\begin{aligned}
&\mathbb{P}\left(\left|\widehat{Y}_N^* - \mathbb{E}\left[\widehat{Y}_N^*\right]\right| > \eta\sqrt{\frac{\log n}{n^{1-\xi}\phi_x(a_n)}}\right) \\
&\leq \mathbb{P}\left(\left|\sum_{i=1}^{n} \Lambda_i\right| > \eta\sqrt{\frac{\log n}{n^{1-\xi}\phi_x(a_n)}}\right) \\
&\leq \exp\left\{-\frac{\eta^2 \log n/(2n^{1-\xi}\phi_x(a_n))}{\left(Var(\sum_{i=1}^n \Lambda_i) + C\mu_n(n\phi_x(a_n))^{-\frac{1}{3}}\left(\frac{\log n}{n^{1-\xi}\phi_x(a_n)}\right)^{\frac{5}{6}}\right)}\right\} \\
&\leq \exp\left\{-\frac{\eta^2 \log n}{Cn^{-\xi} + \mu_n n^{-\xi/6}\left(\frac{\log^5 n}{n\phi_x(a_n)}\right)^{\frac{1}{6}}}\right\} \\
&\leq C' \exp\left\{-C\eta^2 \log n\right\}.
\end{aligned} \tag{A18}$$

The suitable choice of η allows one to finish the proof of Equation (A17).

Now, to proving Equation (A15) use the Holder's inequality to show that,

$$\begin{aligned}
\left|\mathbb{E}\left[\widehat{Y}_N\right] - \mathbb{E}\left[\widehat{Y}_N^*\right]\right| &\leq \frac{1}{n\,\mathbb{E}[K_1(x)]}\left|\mathbb{E}\left[\sum_{i=1}^{n} Y_i \mathbb{1}_{\{|Y_i|>\mu_n\}} K_i(x)\right]\right| \\
&\leq \frac{1}{\mathbb{E}[K_1(x)]}\mathbb{E}\left[|Y_i|\mathbb{1}_{\{|Y_i|>\mu_n\}} K_1(x)\right] \\
&\leq \frac{1}{\mathbb{E}[K_1(x)]}\mathbb{E}\left[\exp(|Y_1|/4)\mathbb{1}_{\{|Y_i|>\mu_n\}} K_1(x)\right] \\
&\leq \left(\mathbb{E}\left[\exp(|Y_1|/2)\mathbb{1}_{\{|Y_i|>\mu_n\}}\right]\right)^{\frac{1}{2}}\left(\mathbb{E}(K_1^2(x))\right)^{\frac{1}{2}} \\
&\leq \phi_x^{-1}(a_n)\exp(-\mu_n/4)(\mathbb{E}[\exp(|Y_1|)])^{\frac{1}{2}}\left(\mathbb{E}(K_1^2(x))\right)^{\frac{1}{2}} \\
&\leq C\phi_x^{-1/2}(a_n)\exp(-\mu_n/4).
\end{aligned}$$

Since $\mu_n = n^{\xi/6}$ then, we can write

$$\left|\mathbb{E}\left[\widehat{Y}_N\right] - \mathbb{E}\left[\widehat{Y}_N^*\right]\right| = o\left(\left(\frac{\log n}{n^{1-\xi}\phi_x(a_n)}\right)^{1/2}\right).$$

The last claimed result Equation (A16) is shown by using the Markov's inequality. Indeed, for all $\epsilon > 0$

$$\begin{aligned}
\mathbb{P}\left[\left|\widehat{Y}_N - \widehat{Y}_N^*\right| > \epsilon\right] &= \mathbb{P}\left[\frac{1}{n\phi_x(a_n)}\sum_{i=1}^{n} Y_i \mathbb{1}_{|Y_i|>\mu_n} K_i| > \epsilon\right] \\
&\leq n\mathbb{P}[|Y_1| > \mu_n] \\
&\leq n\exp(-\mu_n)\mathbb{E}(\exp(|Y|)) \\
&\leq Cn\exp(-\mu_n).
\end{aligned}$$

Then,

$$\sum_{n\geq 1}\mathbb{P}\left(\left|\widehat{\Upsilon}_N-\widehat{\Upsilon}_N^*\right|>\epsilon_0\left(\sqrt{\frac{\log n}{n^{1-\xi}\phi_x(a_n)}}\right)\right)\leq C\sum_{n\geq 1} n\exp(-\mu_n). \tag{A19}$$

The use of the definition of μ_n completes the proof of this Lemma. □

Proof of Lemma 5. For the first term we start by writing

$$\mathbb{E}\widehat{S}_N^x=\frac{1}{\mathbb{E}[K_1(x)]}\mathbb{E}\left[K_1(x)\left[\mathbb{E}\left[G_1(CVaR_p(x))|X\right]\right]\right]. \tag{A20}$$

Moreover, we have

$$\mathbb{E}\left[G_1(CVaR_p(x))|X\right]=\int_{-\infty}^{\infty}G\left(b_n^{-1}(y-z)\right)f^X(z)dz,$$

and integrating by parts and using the fact that H' is symmetric to deduce that

$$\mathbb{E}\left(H_1(CVaR_p(x))|X\right) = \int_{-\infty}^{\infty}tH'(t)F^X(CVaR_p(x)-b_nt)dt.$$

Thus, we have

$$\left|\mathbb{E}\left(H_1(CVaR_p(x))|X\right)-F^x(CVaR_p(x))\int_{-\infty}^{\infty}tH'(t)dt\right|$$

$$\leq \int_{-\infty}^{\infty}\left|F^X(CVaR_p(x)-b_nt)-F^x(CVaR_p(x))\right|dt.$$

Now, as $\int_{-\infty}^{\infty}tH'(t)dt=0$ then

$$\left|\mathbb{E}\left(H_1(CVaR_p(x))|X\right)\right| \leq C\int|H'(t)|\left(a_n^{\beta_1}+|t|^{\beta_2}b_n^{\beta_2}\right)dt. \tag{A21}$$

We use the same arguments for the term $\mathbb{E}\widehat{Q}_N^x$. Indeed, Observe that

$$\mathbb{E}\widehat{Q}_N^x=\mathbb{E}[\widetilde{\Upsilon}_N]+\mathbb{E}[\widehat{\Upsilon}_N].$$

It is shown in [34] that

$$\left|\mathbb{E}[\widetilde{\Upsilon}_N]-\mathbb{E}[Y|X=x]\right|\leq Ca_n^{\beta_1}.$$

So, all it remains to evaluate $\mathbb{E}[\widehat{\Upsilon}_N]$. To do that, we define $Q^x(z)=\displaystyle\int_{-\infty}^{z}tf^x(t)dt$. and once again use the integration by parts to write that

$$\mathbb{E}\left(H_1(CVaR_p(x))Y|X\right) = \int_{-\infty}^{\infty}H'(t)Q^X(CVaR_p(x)-b_nt)dt.$$

Implies that

$$\left|\mathbb{E}\left(H_1(CVaR_p(x))|X\right)-Q^x(CVaR_p(x))\right| \leq C\int|H^{'}(t)|\left(a_n^{\beta_1}+|t|^{\beta_2}b_n^{\beta_2}\right)dt.$$

Finally

$$\left|pCES_p(x)-\mathbb{E}\left[\widehat{Q}_N^x\right]\right|\leq\left|\mathbb{E}[\widetilde{\Upsilon}_N]-\mathbb{E}[Y|X=x]\right|+\left|\mathbb{E}[\widehat{\Upsilon}_N]-Q^x(CVaR_p(x))]\right|$$

$$\leq C|\left(a_n^{\beta_1}+|t|^{\beta_2}b_n^{\beta_2}\right).$$

which yields the proof of this lemma. □

References

1. Acerbi, C.; Tasche, D. On the coherence of expected shortfall. *J. Bank. Financ.* **2002**, *26*, 1487–1503. [CrossRef]
2. Basel Committee on Banking Supervision. Consultative Document. Fundamental Review of the Trading Book: A Revised Market Risk Framework. Available online: http://www.bis.org/publ/bcbs265.pdf (accessed on 10 June 2014).
3. Artzner, P.; Delbaen, F.; Eber, J.-M.; Heath, D. Coherent measures of risk. *Math. Financ.* **1999**, *9* , 203–228. [CrossRef]
4. Yamai, Y.; Yoshiba, T. On The Validity of Value-at-Risk: Comparative Analysis with Expected Shortfall. *Monet. Econ. Stud.* **2002**, *20*, 57–86.
5. Acerbi, C. Spectral measures of risk: A coherent representation of subjective risk aversion. *J. Bank Financ.* **2002**, *26*, 1505–1518. [CrossRef]
6. Krokhmal, P.; Palmquist, J.; Uryasev, S. Portfolio optimization with conditional value-at-Risk objective and constraints. *J. Risk* **2002**, *4*, 43–68. [CrossRef]
7. Alexander, S.; Coleman, T.F.; Li, Y. Minimizing cvar and var for a portfolio of derivatives. *J. Bank. Financ.* **2006**, *30*, 583–605. [CrossRef]
8. Wong, W.K. Backtesting value-at-risk based on tail losses. *J. Empir. Financ.* **2010**, *17*, 526–38. [CrossRef]
9. Acereda, B.; Leon, A.; Mora, J. Estimating the expected shortfall of cryptocurrencies: An evaluation based on backtesting. *Financ. Res. Lette* **2020**, *33*, 101181. [CrossRef]
10. Novales, A.; Garcia-Jorcano, L. Backtesting extreme value theory models of expected shortfall. *Quant. Financ.* **2019**, *19* , 799–825. [CrossRef]
11. Scaillet, O. Nonparametric estimation and sensitivity analysis of expected shortfall. *Math. Financ.* **2004**, *14*, 115–129. [CrossRef]
12. Cai, Z.; Wang, X. Nonparametric estimation of conditional VaR and expected shortfall. *J. Econom.* **2008**, *147*, 120–130. [CrossRef]
13. Yu, K.; Ally, A.K.; Yang, S.; Hand, D.J. Kernel quantile-based estimation of expected shortfall. *J. Risk* **2010**, *12*, 15.
14. Wu, Y.; Yu, W.; Balakrishnan, N.; Wang, X. Nonparametric estimation of expected shortfall via Bahadur-type representation and Berry–Esséen bounds, *J. Stat. Comput. Simul.* **2022**, *92*, 544–566. [CrossRef]
15. Mohammed, N.; Furman, E.; Su, J. Can a regulatory risk measure induce profit-maximizing risk capital allocations? the case of conditional tail expectation. *Insur. Math. Econ.* **2021**, *101*, 425–436. [CrossRef]
16. Marri, F.; Moutanabbir, K. Risk aggregation and capital allocation using a new generalized archimedean copula. *Insur. Math. Econ.* **2022**, *102*, 75–90. [CrossRef]
17. Jiang, R.; Hu, X.; Yu, K. Single-Index Expectile Models for Estimating Conditional Value at Risk and Expected Shortfall Get access Arrow. *J. Financ. Econom.* **2022**, *20*, 345–366.
18. Müller, H.; Sen, R.; Stadtmüller, U. Functional data analysis for volatility. *J. Econom.* **2011**, *165*, 233–245 [CrossRef]
19. Kokoszka, P.; Miao, H.; Zhang, X. Functional dynamic factor model for intraday price curves. *J. Financ. Econom.* **2015**, *13*, 456–477. [CrossRef]
20. Shang, H. Forecasting intraday SP 500 index returns: A functional time series approach. *J. Forecast*. **2017**, *36*, 741–755. [CrossRef]
21. Cai, T. Financial risk management based on functional data analysis *J. Discret. Math. Sci. Cryptogr.* **2018**, *21*, 1397–1400. [CrossRef]
22. Saart, P.; Xia, Y. Functional time series approach to analyzing asset returns co-movements. *J. Econom.* **2021**, *229*, 127–151. [CrossRef]
23. Wang, D.; Su, Z.; Qifang L. Intraday volatility analysis of CSI 300 index futures: A dependent functional data method. *Econ.-Res.-Ekon. Istraz.* **2022**. [CrossRef]
24. Liang, Z.; Weng, F.; Ma, Y.; Xu, Y.; Zhu, M.; Yang, C. Measurement and Analysis of High Frequency Assert Volatility Based on Functional Data Analysis. *Mathematics* **2022**, *10*, 1140 [CrossRef]
25. Ferraty, F.; Goia, A.; Vieu, P. Functional nonparametric model for time series: A fractal approach for dimension reduction. *Test* **2002**, *11*, 317–344. [CrossRef]
26. Masry, E. Nonparametric regression estimation for dependent functional data: Asymptotic normality. *Stoch. Proc. Appl.* **2005**, *115*, 155–177. [CrossRef]
27. Laïb, N.; Louani, D. Rates of strong consistencies of the regression function estimator for functional stationary ergodic data. *J. Statist. Plan. Inference* **2011**, *141*, 359–372. [CrossRef]
28. Ferraty, F.; Quintela-del-Río, A. Conditional VAR and expected shortfall: A new functional approach. *Econometric Rev.* **2016**, *35* 263–292. [CrossRef]
29. Aneiros, G.; Cao, R.; Fraiman, R.; Genest, C.; Vieu, P. Recent advances in functional data analysis and high-dimensional statistics. *J. Multivar. Anal.* **2019**, *170*, 3–9. [CrossRef]
30. Rachdi, M.; Laksaci, A.; Al-Kandari, N. Expectile regression for spatial functional data analysis (sFDA). *Metrika* **2022**, *85*, 627–655. [CrossRef]
31. Almanjahie, I.; Bouzebda, S.; Kaid, Z.; Laksaci, A. Nonparametric estimation of expectile regression in functional dependent data. *J. Nonparametr. Stat.* **2022**, *34*, 250–281. [CrossRef]
32. Bulinski, A.; Suquet, C. Normal approximation for quasi-associated random fields. *Statist. Probab. Lett.* **2001**, *54*, 215–226. [CrossRef]
33. Douge, L. *Théorèmes Limites pour des Variables Quasi-Associées Hilbertiennes*; Institut de Statistique de l'Université de Paris: Paris, France, 2010; Volume 54, pp. 51–60.
34. Ferraty, F.; Vieu, P. *Nonparametric Functional Data Analysis*; Theory and Practice; Springer: New York, NY, USA, 2006.

35. Taylor, J.W. Forecasting value at risk and expected shortfall using a semiparametric approach based on the asymmetric laplace distribution. *J. Bus. Econ. Stat.* **2019**, *37*, 121–133. [CrossRef]
36. Patton, A.J.; Ziegel, J.F.; Chen, R. Dynamic semiparametric models for expected shortfall (and value-at-risk). *Forthcom. J. Econom.* **2019**, *211*, 388–413. [CrossRef]
37. Dimitriadis, T.; Bayer, S. A joint quantile and expected shortfall regression framework. *Electron. J. Statist.* **2019**, *13*, 1823–1871. [CrossRef]
38. Bayer, S.; Dimitriadis, T. Regression-Based Expected Shortfall Backtesting. *J. Financ. Econom.* **2022**, *20*, 437–471. [CrossRef]
39. Kallabis, R ; Neumann, M. An exponential inequality under weak dependence. *Bernoulli* **2006**, *12*, 333–350.