CENTRE FOR ECONOMETRIC ANALYSIS
CEA@Cass

http://www.cass.city.ac.uk/cea/index.html

Cass Business School
Faculty of Finance
106 Bunhill Row
London EC1Y 8TZ

Estimation of and Inference about the Expected Shortfall for Time Series with Infinite Variance

Oliver Linton and Zhijie Xiao

CEA@Cass Working Paper Series

WP–CEA–04-2012

Estimation of and Inference about the Expected Shortfall for Time Series with Infinite Variance*

Oliver Linton[†] Zhijie Xiao[‡]

University of Cambridge Boston College

March 21, 2012

Abstract

We study estimation and inference of Expected Shortfall for time series with Infinite variance. Both the smoothed and nonsmoothed estimators are investigated. The rate of convergence is determined by the tail thickness parameter and the limiting distribution is in the stable class with parameters depending on the tail thickness parameter of the time series and on the dependence structure, which makes inference complicated. A subsampling procedure is proposed to carry out statistical inference. We also analyze a nonparametric estimator of the conditional expected shortfall.

1 Introduction

The Expected Shortfall (ES) of a continuously distributed random variable Y is defined as the expected loss given the loss exceeds the Value at Risk threshold

$$ES_\tau = E\left[Y|Y \leq \alpha(\tau)\right] = \frac{1}{\tau}E\left[Y1(Y \leq \alpha(\tau))\right], \tag{1}$$

*We thank Michael Jansson, J. B. Hill and two referees for their very helpful comments on an earlier version of this paper.

[†]Faculty of Economics, Austin Robinson Building, Sidgwick Avenue, Cambridge, CB3 9DD. Email: obl20@cam.ac.uk

[‡]Department of Economics, Boston College, Chestnut Hill, MA 02467, USA. Tel: 617-552-1709. Email: xiaoz@bc.edu. Financial support from Boston College is gratefully acknowledged.

where the Value at Risk $\alpha(\tau)$ satisfies $\Pr[Y \leq \alpha(\tau)] = \tau$, where $\tau \in (0, 1)$ is the loss probability and $1(.)$ is the usual indicator function. The expected shortfall is defined provided $E|Y| < \infty$. It is well known that the expected shortfall is a coherent measure of risk by contrast with the Value at Risk (see, e.g. Arztner et al. (1999)). Expected shortfall is widely used in practice. For financial applications of ES and other measures based on ES, see, e.g., Acharya, Pedersen, Philippe, and Richardson (2010), Adrian and Brunnermeier (2010), Brownlees and Engle (2010), and Danielsson (2011).

This paper is concerned with estimation of ES_τ from a stationary and mixing time series. The existing literature on estimation has focused on the case where $E|Y|^k < \infty$ for $k \geq 2$, see Chen (2008), Cai and Wang (2008), Scaillet (2004), and Fermanian and Scaillet (2005). These authors proposed slightly different estimators based on empirical versions of the relation (1), and showed that their estimators were consistent and asymptotically normal at rate square root of sample size. Furthermore, the estimation of the τ-th quantile of Y does not appear to affect the limiting distribution of the estimators of ES.

This paper studies the case where the expected shortfall is defined ($E|Y| < \infty$) but the variance of Y is infinite, and so the above results do not necessarily hold. Many financial time series have heavy tails, see, e.g., Mandelbrot (1963), Fama (1965), Taleb (2010), Rachev and Mittnik (2000), Embrechts, Kluppelberg and Mikosch (1997), Ibragimov, Jaffee and Walden (2009), and Ibragimov (2009a). For example, electricity prices can be subject to large spikes due to supply/demand imbalances that cannot be temporally mediated (Weron (2008)). There is also a literature on heavy-tail behavior about the distribution of wealth and income, see, e.g. Ibragimov (2009a), Gabaix (2008), Atkinson and Piketty (2007). Klass et al (2006) and Nirei and Sonma (2007) documented that the tail exponent of wealth is around 1.5. Gopikrishnan et al. (2000) find that trading volumes for the 1,000 largest U.S. stocks have Pareto tail with exponent around 3/2. The 1987 crash delivered daily a return on the broad S&P500 index that was over 20 standard deviations below the mean. More recently, the "flash crash" of May 6th, 2010, showed how far stock prices could move in a very short period of time. At 2:42 pm, with the Dow Jones Industrial Average down more than 300 points for the day, the index began to fall rapidly, dropping more than 600 points in 5 minutes for an almost 1000 point loss (or about nine percent) on the day by 2:47 pm. Twenty minutes later, by 3:07 pm, the market had regained most of the 600 point drop. We note that a lot of Financial theory concerning diversification and the risk return trade-off does not require a finite variance. For example the CAPM, "Mean "Variance" Efficiency", and "Two fund separation" are known to hold for the more general class of elliptical distributions that are characterized by a location vector μ and a scale matrix Ω. The scale matrix $\Omega = (\omega_{ij})$ need not be a covariance matrix. Press (1982) shows that provided the expected return exists and is finite $ER - r_f = \beta_{CAPM} E(R_m - r_f), \quad \beta_{CAPM} = \frac{\omega_{im}}{\omega_{mm}}$, see also Fama and Miller (1972) and Samuelson (1967). So, we don't need a variance. Stock and Watson (2007, Chapter 2) discussed the 1987 Black Friday effect on the Dow Jones and its implications for non-normality. Yet many risk measures are constructed from variance, and so may be non robust to large movements in series. The consequences of large crashes are enormous, and it is important to have risk management tools that reflect this possibility and are robust to it.

We analyze the estimators of expected shortfall to determine their properties under heavy tailed assumptions. We derive the asymptotic distributions of two estimators of expected shortfall: the so-called smoothed estimator proposed in Scaillet (2004), and the unsmoothed estimator proposed in Chen (2004). The estimators are consistent at a rate depending on the tail thickness parameter and have the same stable limiting distribution; estimation of the Value at Risk $\alpha(\tau)$ also does not affect the limiting distribution. However, the limiting distribution is very complex and depends on the dependence properties of the data as well as on the tail thickness parameter, so that "plug-in inference" is very complicated. We propose a subsampling method to approximate the limiting distribution of the proposed estimators, and thereby to facilitate statistical inference about the parameter of interest. This method also works when the variance exists and normal asymptotics prevail. A Monte Carlo experiment is conducted to investigate the finite sample performance of the proposed estimators and subsampling methods.

We also extend the theory to cover the nonparametric conditional expected shortfall estimators. In this case also, the convergence rate is slower than that obtained in the standard nonparametric regression framework with finite variance and the limiting distribution is in the stable class. However, the limiting distribution does not depend on the time series properties of the process so that "plug-in inference" is relatively straightforward. Hall, Peng, and Yao (2002) obtain similar results for nonparametric regression.

There is a large literature studying stable limits for sums of infinite variance random variables. Classical work of Gnedenko and Kolmogorov (1954), Feller (1971) etc. show that appropriately standardized sums of an iid sequence of infinite variance random variables converges to an infinite variance α-stable random variable. Davis and Resnick (1985, 1986), Phillips and Solo (1992) studied stable limits for sums of weakly dependent infinite variance random variables in the form of linear process or functions of linear process. Davis (1983), Denker and Jakubowski(1989), etc. studied more general stationary process. There is a literature on asymptotics for sums of functionals such as the autocovariance functions (ACFs) of heavy-tailed time series. In particular, Phillips and Solo (1992) studied limiting behavior of ACFs of linear process. Davis and Mikosch (1998), and Mikosch and Starica (2000) studied the sample ACF of a nonlinear stationary process. They show that in the case with finite variances but infinite fourth moment, providing the process is ergodic, the convergence rate of the sample ACF is slower than root-n for a nonlinear stationary process such as a GARCH process. While for a linear process with finite variances but infinite fourth moment, the sample ACF is still asymptotic normal with root-n convergence. Also see Embrechts et al (1997) for discussions on this topic. Recent work of Bartkiewicz et al. (2010), which clarifies and extends earlier work, for example Davis and Hsing (1995), provides conditions that determines the parameters of limiting distributions in terms of tail characteristics of the underlying stationary sequence. For simplicity, some conditions that we use are stronger than theirs, so as usual our conditions are sufficient but not necessary.

The rest of our paper is organized as follows: The model and assumptions are introduced in Section 2. Both the smoothed estimator and the unsmoothed estimators are given in Section 3. Limiting distributions of these estimators are also developed in this section. In Section 4, we propose a subsampling procedure to approximate the limiting distribution of the proposed estimators. In Section 5 we propose and investigate a nonparametric estimator of conditional expected shortfall. In Section 6 we provide some Monte Carlo investigations and an application to some financial time series. Section 7 concludes.

2 The Model

Consider a random sample $\{Y_1, \ldots, Y_T\}$ from a stationary and mixing time series with infinite variance. Given this random sample, we consider the problem of estimating the Expected Shortfall of Y defined by (1) for $\tau \in (0, 1)$. A stationary process $\{(\xi_t, \mathcal{F}_t), -\infty < t < \infty\}$ is said to be strong mixing if the mixing coefficient $\alpha(k)$ defined by

$$\alpha(k) = \sup_{A \in \mathcal{F}_{-\infty}^{t}, B \in \mathcal{F}_{t+k}^{\infty}} |P(A \cap B) - P(A)P(B)|$$

converges to zero as $k \to \infty$. Strong mixing includes many linear and nonlinear time series models as special cases; see Doukhan (1994) for the definition and examples. To facilitate our asymptotic analysis, we make the following assumptions.

Assumption A

A1 (Y_t) are realizations from a strictly stationary sequence with regularly varying tail probabilities, that is: for every $m \geq 1$, there exists a non-null Radon measure μ_m on the Borel sigma field of $\overline{\mathbb{R}}^m/\{\mathbf{0}\}$, where $\overline{\mathbb{R}}^m = \mathbb{R}^m \cup \{\pm\infty\}$, such that

$$T \Pr\left[a_T^{-1}(Y_1, \cdots, Y_m) \in \cdot\right] \xrightarrow{v} \mu_m(\cdot),$$

where $\xrightarrow{v}$ denotes vague convergence (Kallenberg (1983), and $a_T = T^{1/\theta} \to \infty$ satisfies $T \Pr[a_T^{-1}|Y_1| > 1] \to 1$, $\theta \in (1, 2)$. The limiting measure μ_m satisfies $\mu_m(tA) = t^{-\theta}\mu_m(A)$ for all $t > 0$, for Borel sets $A \subset \mathbb{R}^m$. Here, θ is the index of regular variation of Y_t, so that $\Pr[|Y_t| > x] = x^{-\theta}L(x)$, where L is a slowly varying function at ∞ (assumed to be one wlog).

A2 The sequence Y_t is strongly mixing with geometrically declining mixing coefficient $\alpha(k)$, i.e., $|\alpha(k)| \leq c\delta^k$ for some $\delta \in (0, 1)$ and $c < \infty$.

A3 Define $A_{\pm} = \{x \in \overline{\mathbb{R}}^m : \pm(x_1 + \cdots + x_m) > 1\}$ and $B_m = \{x_1 \leq \alpha(\tau), \ldots, x_m \leq \alpha(\tau)\}$, and let for all $m \geq 1$:

$$b_{\pm}(m) = \mu_m(A_{\pm} \cap B_m).$$

The limits $\lim_{m \to \infty} (b_{\pm}(m) - b_{\pm}(m-1)) = c_{\pm}$ exist.

A4 There exists a sequence $r_T \to \infty$ with $T\alpha(r_T) \to 0$ such that

$$\lim_{n \to \infty} \limsup_{T \to \infty} T \Pr\left(\max_{t=n+1,\ldots,r_T} |Y_t| > \frac{a_T}{r_T}, |Y_1| > a_T \right) = 0.$$

A5 Let f denote the stationary density of Y_t, where $f(y) > 0$ for all $y \in (-\infty, \infty)$, and let f_s denote the joint density of Y_t, Y_{t+s}, these densities are twice continuously differentiable.

Assumptions A1 - A 4 are similar to Bartkiewicz et al. (2010). Specifically, A1 is their condition **RV** (regular variation) restricted to the case where $\theta > 1$ so that the expected shortfall exists. We replace their mixing condition **MX** by the stronger condition of geometric strong mixing, see their Lemma 3; this is satisfied by many models of interest including GARCH(1,1) models under conditions on parameters and error moments, Pan, Linton, and Wang (2010). Condition A3 is their **TB** (tail balance) condition. Condition A4 is the sufficient condition given in Bartkiewicz et al. (2010, equation (3.8)) for their anti-clustering condition **AC**. They verify A4 in the case of m-dependent random variables for which $c_+ = b_+(m+1) - b_+(m)$ and $c_- = b_-(m+1) - b_-(m)$. They also verify A4 for GARCH(1,1) process and give expressions for the quantities c_+ and c_1. Assumption A5 is needed to handle the smoothing bias terms induced by our estimator and to ensure the regular behaviour of the quantile estimator.

In terms of notation, a stable distribution $S(\mu, \sigma, \theta, \beta)$ with location, scale, shape, and skewness parameters $(\mu, \sigma, \theta, \beta)$ is defined via its characteristic function $\phi(t)$ given by

$$\phi(t) = \exp\left\{it\mu - \sigma^\theta \left|t\right|^\theta (1 - i\beta \mathrm{sgn}(t) \tan(\pi\theta/2)\right\}.$$

In the sequel the following definition is fundamental. Let S be a stable random variable with characteristic function

$$E\left[\exp(itS)\right] = \exp\left[-|t|^\theta \frac{\Gamma(2-\theta)}{1-\theta} \left\{(c_+ + c_-)\cos(\pi\theta/2) - i\mathrm{sign}(t)(c_+ - c_-)\sin(\pi\theta/2)\right\}\right]. \quad (2)$$

Corresponding to the conventional notation, the above random variable has a stable distribution $S(0, \sigma^*, \theta, \beta^*)$ with $\beta^* = \frac{c_+ - c_-}{c_+ + c_-}$, and σ^* is determined by the equation $\sigma^{*\theta} = \frac{\Gamma(2-\theta)}{1-\theta}(c_+ + c_-)\cos(\pi\theta/2)$. Let S^* be a stable random variable with characteristic function given by (2) with $c_+ = 1$ and $c_- = 0$. We can explicitly express S^* as $S^* = \sum_{j=1}^{\infty} \left(\mathrm{E}\left(Z_j 1\left(0 < Z_j \leq 1\right)\right) - Z_j\right)$, where $Z_i = \Gamma_i^{-1/\theta}$, $\Gamma_k = V_1 + \ldots + V_k$, and $\{V_i\}$ is a sequence of independent exponential random variables with unit mean (see, e.g. Davis (1983)).

3 The Estimators

We consider the case where τ is fixed. In the case where $\tau \to 0$ an alternative approach is possible based on the model for the tail probabilities, see Danielsson (2011) for references.

3.1 The Smoothed Estimator

Estimation of the ES usually requires knowledge about the corresponding quantile. We may estimate the τ-th quantile of Y, $\alpha(\tau)$, by $\widetilde{\alpha}(\tau)$ that satisfies

$$\frac{1}{T}\sum_{t=1}^{T}\mathcal{K}_h\left(Y_t-\widetilde{\alpha}(\tau)\right)=\tau+o_p(T^{-1/2}). \tag{3}$$

Here, $\mathcal{K}_h(\cdot)=\mathcal{K}(\cdot/h)$, where $\mathcal{K}(\cdot)$ is a survivor function (defined as $1-F(\cdot)$, where $F(\cdot)$ is a CDF of a continuous random variable) and h is a bandwidth satisfying the following assumptions.

Assumption B

B1 $\mathcal{K}_h(\cdot)=\mathcal{K}(\cdot/h)$, where $\mathcal{K}$ is a twice continuously differentiable survivor function with compact support satisfying $\mathcal{K}(x)=1$ for $x\leq -1$ and $\mathcal{K}(x)=0$ for $x\geq 1$, while h is a bandwidth. We define $K(u)=-\mathcal{K}'(u)$, which is a differentiable symmetric density function with compact support $[-1,1]$.

B2 As $T\to\infty$, $h\to 0$, $Th\to\infty$, and $Th^{2\theta}\to 0$.

We consider the following estimator of the Expected Shortfall

$$\widetilde{ES}_\tau=\frac{1}{\tau T}\sum_{t=1}^{T}Y_t\mathcal{K}_h\left(Y_t-\widetilde{\alpha}(\tau)\right), \tag{4}$$

This is essentially the smoothed estimator of Scaillet (2004). The Value at Risk is asymptotically normal under our conditions, Scaillet (2004). We summarize the asymptotic distribution of $\widetilde{ES}_\tau$ in Theorem 1.

THEOREM 1. *Suppose that Assumptions A1 - A5 and B1 -B2 hold, Then*

$$T^{(\theta-1)/\theta}(\widetilde{ES}_\tau-ES_\tau)\Longrightarrow\frac{1}{\tau}S. \tag{5}$$

REMARK. The rate of convergence is strictly slower than $T^{1/2}$ and the limiting distribution is quite complicated. When the data are i.i.d., the limiting distribution is a standard asymmetric stable law determined only by the tail thickness parameter θ. However, in general the distribution of S depends on the dependence of the process through the quantities c_+,c_-.

REMARK. Theorem 1 shows that the preliminary estimation of the Value at Risk, $\widetilde{\alpha}(\tau)$, has no effect on the limiting behavior of the expected shortfall estimator, i.e., $T^{(\theta-1)/\theta}(\widetilde{ES}_\tau - \widetilde{ES}_{0\tau}) = o_p(1)$, where $\widetilde{ES}_{0\tau}$ is the infeasible procedure that relies on knowledge of $\alpha(\tau)$

$$\widetilde{ES}_{0\tau} = \frac{1}{T\tau}\sum_{t=1}^{T} Y_t \mathcal{K}_h(Y_t - \alpha(\tau)).$$

Furthermore, the joint asymptotic distribution of $\widetilde{ES}_\tau$ and $\widetilde{\alpha}(\tau)$, i.e., the expected shortfall and Value at Risk are asymptotically independent. This means that the asymptotic behaviour of any smooth function of the two risk measures will have asymptotic properties dominated by those of the expected shortfall estimator.

REMARK. From the proof of Theorem 1 we can see that the smoothed estimator contains a bias term that is of order h^2.

3.2 The Non-Smoothed Estimator

The previous estimators of the Value at Risk and Expected Shortfall are smoothed estimators. Alternatively, the τ-th quantile of Y, can be estimated simply by the τ-th sample quantile of $\{Y_1, \ldots, Y_T\}$, say, $\widehat{\alpha}(\tau)$, and we may construct the following nonsmoothed estimator of ES_τ:

$$\widehat{ES}_\tau = \frac{1}{\tau T}\sum_{t=1}^{T} Y_t 1\left(Y_t \leq \widehat{\alpha}(\tau)\right).$$

This is the unsmoothed estimator of Chen (2008).

THEOREM 2. *Suppose that Assumptions A1 - A5 hold. Then, as* $T \to \infty$*,*

$$T^{(\theta-1)/\theta}\left(\widehat{ES}_\tau - ES_\tau\right) \Rightarrow \frac{1}{\tau}S,$$

where S *is the same stable law as in Theorem 1.*

The two estimators share the same rate of convergence and limiting distribution, just as in Chen (2008).

4 Consistent Inference

We propose a method for conducting inference about the expected shortfall that is robust to heavy tails. One may be interested in a confidence interval to give some idea of the precision of the estimation. One can also use this interval to construct modified estimators that are robust to estimation error, by taking a lower α quantile of the estimated distribution. In some cases this can result in a considerably more conservative estimator.

In order to conduct statistical inference based on the proposed estimators, we need to estimate the asymptotic distributions somehow. In the case of finite variance time series, the ES estimator is root-T consistent and asymptotic normal with a variance that can be consistently estimated. In the case with infinite variance, to conduct inference about ES_τ, we need to estimate consistently the parameters (θ, c_+, c_-) under the weak conditions we have imposed, which is a difficult task. (In the special case with $c_+ = c_-$, alternative methods may be preferred, see our discussion in the end of this section on the t-statistic based approach of Ibragimov and Müller (2010). We focus on the general case in this section.) The parameter θ can be estimated consistently under weak dependence conditions by many methods, see for example Hill (2010), Embrechts et al (1997). Estimation of the spectral measure has been investigated for i.i.d data, see for example Einmahl, de Haan, and Piterbarg (2001), but these results do not cover estimation of c_+, c_- under weak dependence conditions. In this paper, we propose a general method based on subsampling (Politis and Romano (1999)). This method is also consistent when the variance of the series exists and so is robust with regard to the tail thickness parameter.

Let $\widehat{\theta}$ be a consistent estimator of θ, given the random sample $\{Y_t, t = 1, \ldots, T\}$, we consider subsamples of size M

$$\{Y_t, \ldots, Y_{t+M-1}\},\ t = 1, \ldots, T - M + 1,$$

and estimate the expected shortfall based on subsamples. Thus the nonsmoothed estimators are

$$\widehat{ES}_\tau(M, t) = \frac{1}{\tau M} \sum_{s=0}^{M-1} Y_{t+s} 1\left(Y_{t+s} \leq \widehat{\alpha}_t(\tau)\right),$$

and the smoothed estimators are

$$\widetilde{ES}_\tau(M, t) = \frac{1}{\tau M} \sum_{s=0}^{M-1} Y_{t+s} \mathcal{K}_h\left(Y_{t+s} - \widetilde{\alpha}_t(\tau)\right).$$

where $\widehat{\alpha}_t(\tau)$ and $\widetilde{\alpha}_t(\tau)$ were the corresponding nonsmoothed and smoothed estimators based on the subsample $\{Y_t, \ldots, Y_{t+M-1}\}$.[1]

[1] Since that the preliminary estimation doesn't affect the limiting distribution of estimators of the ES, one can also use $\widehat{\alpha}(\tau)$ or $\widetilde{\alpha}(\tau)$ that were estimated based on the whole sample.

We approximate the sampling distribution of $T^{(\theta-1)/\theta}\left(\widehat{ES}_\tau - ES_\tau\right)$, denoted by $\widehat{F}_T(y)$, by

$$\widehat{F}_{T,M}(y) = \frac{1}{T-M+1}\sum_{t=1}^{T-M+1} 1\left(M^{(\widehat{\theta}-1)/\widehat{\theta}}\left[\widehat{ES}_\tau(M,t) - \widehat{ES}_\tau\right] \leq y\right),$$

and approximate the sampling distribution of $T^{(\theta-1)/\theta}\left(\widetilde{ES}_\tau - ES_\tau\right)$, say $\widetilde{F}_T(y)$, by

$$\widetilde{F}_{T,M}(y) = \frac{1}{T-M+1}\sum_{t=1}^{T-M+1} 1\left(M^{(\widehat{\theta}-1)/\widehat{\theta}}\left[\widetilde{ES}_\tau(M,t) - \widetilde{ES}_\tau\right] \leq y\right).$$

Assumption C

C1 Assume that The tail index θ is estimated by $\widehat{\theta}$ at rate faster than $\log T$ in the sense that $\log(T)\,(\widehat{\theta}-\theta) \xrightarrow{P} 0$.

C2 $M \to \infty$, and $M/T \to 0$.

Assumption C1 requires that the tail index is estimated faster than rate $\log T$ so that $M^{(\widehat{\theta}-1)/\widehat{\theta}}/M^{(\theta-1)/\theta} \xrightarrow{P} 1$. There is a literature on estimation of tail index where θ is estimated by $\widehat{\theta}$ at rate T^a, for some $0 < a < (\theta-1)/\theta$. Such estimators satisfy our assumption C1 (see, e.g., Hill (2010), Embrechts et al (1997)). Assumption C2 is a basic requirement for subsampling.

Let $F(y)$ be the limiting distribution function of $T^{(\theta-1)/\theta}(\widehat{ES}_\tau - ES_\tau)$, we have the following result.

THEOREM 3. *Let y be any continuous point of $F(\cdot)$.* (1) *Suppose that Assumptions A1 -A5, B1-B2, and C1, C2 hold, then, as $T \to \infty$,*

$$\widetilde{F}_{T,M}(y) \xrightarrow{P} F(y).$$

(2) *Suppose that Assumptions A1 -A5 and C1, C2 hold, then as $T \to \infty$,*

$$\widehat{F}_{T,M}(y) \xrightarrow{P} F(y).$$

REMARK. From the proof of Theorem 3 we can see that the proposed subsampling method is "robust" in the sense that it is also consistent even in the case of finite variance, where normal asymptotics prevail.

REMARK. Ibragimov and Müller (2010) developed t-statistic based approach for inference on a parameter based on its estimator with a limit in the form of mixture of normals. Monte Carlo evidence indicates that the Ibragimov and Müller approach usually has better finite sample performance than subsampling. This approach can be applied to inference on the expected shortfall in the special case with symmetric stable limits ($c_+ = c_-$). In this special case, we suggest using the t-statistic based inference of Ibragimov and Müller (2010).

5 Conditional Expected Shortfall

In many financial applications, investors look at the *conditional* distribution of returns given current information. We consider the *Conditional* Expected Shortfall. We first remark on an obvious parametric extension of the first part of the paper. Consider the time series model

$$Y_t = \mu_t(\theta_0) + \sigma_t(\theta_0)\varepsilon_t,$$

where Y_t is the return of an asset at time t, and $\mu_t(\theta), \sigma_t(\theta) \in \mathcal{F}_{t-1}$ for each $\theta \in \Theta \subset \mathbb{R}^p$, where $\mathcal{F}_t = \sigma\{Y_{t-1}, Y_{t-2}, \cdots\}$ is the information set at time t including past values of returns and possibly the value of some covariates X_t. The random variables ε_t and $\varepsilon_t^2 - 1$ are martingale difference sequences[2]. This class of models includes many popular time series models used in practice such as the GARCH and EGARCH models as well as seasonal regression models that may be used for electricity prices (see Campbell, Lo, and Mackinlay (1997, Ch 12) for a review of GARCH and examples. The Conditional Value at Risk of Y_t given $\mathcal{F}_{t-1}$ and the Conditional Expected Shortfall of Y_t given $\mathcal{F}_{t-1}$ are:

$$\begin{aligned} VaR_{t-1}(\tau) &= \mu_t(\theta_0) + \sigma_t(\theta_0)q_\tau(\varepsilon) \\ ES_{t-1}(\tau) &= \mu_t(\theta_0) + \sigma_t(\theta_0)ES_\tau(\varepsilon), \end{aligned}$$

where $q_\tau(\varepsilon)$ and $ES_\tau(\varepsilon)$ denote the unconditional Value at Risk and expected shortfall of the error term ε_t respectively. Let $\widehat{\theta}$ be square root sample size consistent estimators of θ_0 (which may require robust methods) , and let $\widehat{\varepsilon}_t = (Y_t - \mu_t(\widehat{\theta}))/\sigma_t(\widehat{\theta})$. We can then apply the methods and results of the previous section to give the properties of the estimated $ES_\tau(\varepsilon)$ and hence the

[2]This is, evidently, satisfied for the case when ε_t are iid mean zero and variance one. For examples when ε_t are m.d.s., see de la Pena, Ibragimov and Sharakhmetov (2003).

estimated $ES_{t-1}(\tau)$. The results parallel closely those given above, except that the conditions A1-A6 must be imposed on ε_t rather than Y_t. We discuss in detail the nonparametric case, where the results are somewhat different.

Let X_t denote the d-dimensional vector that contains information available at time t, including possibly lagged dependent variables, and denote the τ-th conditional quantile of Y_t by $Q_{Y_t}(\tau|X_t)$, i.e., $Q_{Y_t}(\tau|X_t)$ satisfies $\Pr(Y_t < Q_{Y_t}(\tau|X_t)|X_t) = \tau$. The Conditional ES of Y_t is defined as

$$ES_\tau(x) = E[Y_t|Y_t \leq Q_{Y_t}(\tau|X_t), X_t = x]. \tag{6}$$

The τ-th conditional quantile of Y_t, $Q_{Y_t}(\tau|X_t)$, can be estimated from a quantile regression of Y_t on X_t. In practice, the functional form of the conditional distribution is usually unknown. For this reason, we consider in this paper nonparametric estimation of the conditional quantiles and the expected shortfall. Our work builds on that of Hall, Peng, and Yao (2002) and Peng and Yao (2004).

Denoting the conditional quantile function of Y_t as $m(\tau, X_t)$, and let $u_{t\tau} = Y_t - m(\tau, X_t)$, then we may write the model as

$$Y_t = m(\tau, X_t) + u_{t\tau}.$$

If $m(\tau, x)$ is a smooth function of x, then for any X_t in a neighborhood of x, we have

$$\begin{aligned} m(\tau, X_t) &\simeq m(\tau, x) + \sum_{1\leq|\mathbf{j}|\leq p} \frac{1}{\mathbf{j}!} D^{|\mathbf{j}|} m(\tau, x)(X_t - x)^{\mathbf{j}} + o(\|X_t - x\|^p) \\ &\equiv \sum_{0\leq|\mathbf{j}|\leq p} \beta_{\mathbf{j}}(\tau, x; h)((X_t - x)/h_1)^{\mathbf{j}} + o(\|X_t - x\|^p). \end{aligned}$$

Here, we use the notation of Masry (1996): $\mathbf{j} = (j_1, ..., j_d)$, $|\mathbf{j}| = \sum_{i=1}^d j_i$, $x^{\mathbf{j}} = \mathbf{\Pi}_{i=1}^d x_i^{j_i}$, $\sum_{0\leq|\mathbf{j}|\leq p} = \sum_{k=0}^p \sum_{j_1=0}^k \cdots \sum_{j_d=0}^k{}_{j_1+...+j_d=k}$, $D^{|\mathbf{j}|} m(\tau, x) = \frac{\partial^{|\mathbf{j}|} m(\tau, x)}{\partial^{j_1} x_1 ... \partial^{j_d} x_d}$, $\beta_{\mathbf{j}}(\tau, x; h) = \frac{h^{|\mathbf{j}|}}{\mathbf{j}!} D^{|\mathbf{j}|} m(\tau, x)$, where $\mathbf{j}! \equiv \mathbf{\Pi}_{i=1}^d j_i!$ and $h_1 = h_1(T)$ is a bandwidth parameter that controls how "close" X_t is from x in the quantile estimation. Given observations $\{(Y_t, X_t)\}_{t=1}^n$, we consider the following local-polynomial[3] quantile regression that minimizes the following objective function

$$Q_T(\tau, x; \boldsymbol{\theta}) \equiv \sum_{t=1}^{T} \rho_\tau \left(Y_t - \sum_{0 \leq |\mathbf{j}| \leq p} \beta_{\mathbf{j}} ((X_t - x)/h_1)^{\mathbf{j}} \right) K\left(\frac{x - X_t}{h_1}\right), \tag{7}$$

where $\rho_\tau(z)$ be the "check" function defined by $\rho_\tau(z) = z(\tau - 1(z \leq 0))$ with $1(\cdot)$ being the usual indicator function, K is a nonnegative kernel function on $\mathbb{R}^q$ defined as $K(u) = \prod_{r=1}^{d} k(u_r)$ where $k(u)$ is the univariate kernel, and $\boldsymbol{\theta}$ *is a stack of* $\beta_{\mathbf{j}}$ *in the lexicographical order*. Minimizing (7) with respect to $\beta_{\mathbf{j}}$, $0 \leq |\mathbf{j}| \leq p$, delivers an estimate $\widehat{\beta}_{\mathbf{j}}(\tau, x; h_1)$ of $\beta_{\mathbf{j}}(\tau, x; h_1)$. The conditional quantile function $m(\tau, x)$ and its derivatives up to p-th order are then estimated respectively by

$$\widehat{m}(\tau, x) = \widehat{\beta}_{\mathbf{0}}(\tau, x; h_1) \text{ and } \widehat{D}^{|\mathbf{j}|} m(\tau, x) = (\mathbf{j}!/h_1^{|\mathbf{j}|}) \widehat{\beta}_{\mathbf{j}}(\tau, x; h_1) \text{ for } 1 \leq |\mathbf{j}| \leq p.$$

In the special case when $p = 1$, these are the widely used local linear estimators.

Notice that

$$ES_\tau(x) = \mathrm{E}\left[Y_t 1\left(Y_t < Q_{Y_t}(\tau | X_t)\right) \middle| X_t = x\right] = \mathrm{E}(Y | X = x) - \frac{1}{\tau} \mathrm{E}\left[\rho_\tau\left(Y - m(\tau, X)\right) \middle| X = x\right],$$

we can estimate the conditional Expected Shortfall $ES_\tau(x)$ by replacing the conditional expectations with corresponding local sample averages. We propose the following estimator for $ES_\tau(x)$:

$$\widehat{ES_\tau(x)} = \frac{1}{\widehat{f}_X(x)} \frac{1}{Th^d} \sum_{t=1}^{T} K\left(\frac{X_t - x}{h}\right) \left[Y_t - \frac{1}{\tau} \rho_\tau\left(Y_t - \widehat{m}(\tau, X_t)\right)\right], \tag{8}$$

where $\widehat{f}_X(x) = T^{-1} h^{-d} \sum_{t=1}^{T} K\left(\frac{X_t - x}{h}\right)$, $h = h(T)$ is the bandwidth parameter that we use for estimating ES. An alternative more complicated estimator can be constructed by fitting a local polynomial estimator to $\widehat{Z}_t = Y_t 1(Y_t < \widehat{m}(\tau, X_t))$. Also see Cai and Wang (2008) and Scaillet (2004), Scaillet (2005) for estimation of conditional ES in the case with finite variances.

Again, to facilitate asymptotic analysis, we assume that the following regularity conditions hold.

Assumption D

[3] Different nonparametric estimators, including both the simple kernel smoother and the local polynomial estimator, can be used. In this paper, we consider the local polynomial procedures. For additional information about local polynomial estimation, see Fan (1992), and Fan and Gijbels (1996) for discussions on the attractive properties of this approach.

D1 Let $\mathcal{F}_{t-1}$ be the σ-field generated by past information $(Y_{t-j}, j \geq 1; X_{t-i}, i \geq 0)$, and denote the conditional quantile function of Y_t as

$$Q_{Y_t}(\tau|\mathcal{F}_{t-1}) = m(\tau, X_t),$$

where $m(\tau, x)$ is continuously differentiable at x to the order p, and $ES_\tau(x)$ defined by (6) is continuously differentiable at x to the second order.

D2 $\left(Y_t, X_t^\top\right)$ are realizations from a strictly stationary sequence with regularly varying tail probabilities satisfying Assumption A1. The conditional CDF of Y_t given X_t, $F_{Y|X}(y|x) = \Pr(Y_t < y|X_t = x)$, has derivative $f_{Y|X}(y|x) > 0$ with $\left|f_{Y|X}(y_n|x)\right|$ uniformly integrable for any sequence $y_n \to m(\tau, x)$. In addition, the sequence $\left(Y_t, X_t^\top\right)$ is ρ-mixing with geometrically declining mixing coefficient $\alpha(k)$, i.e., $|\alpha(k)| \leq c\delta^k$ for some $\delta \in (0,1)$ and $c < \infty$.

D3 $u_{t\tau} = Y_t - m(\tau, X_t)$ also has regularly varying tail probabilities (in the conditional distribution given $X_t = x$) satisfying Assumption A1.

D4 The pdf of X_t, $f_X(x)$, is continuously differentiable at x and $f_X(x) > 0$.

D5 The kernel function $K(\cdot)$ is a product kernel of $k(\cdot)$, which is a symmetric density function with compact support $[-1,1]$, and $|k(a) - k(a')| \leq \bar{c}_2|a - a'|$ for any $a, a' \in \mathbb{R}$ and some $\bar{c}_2 < \infty$. The functions $H_{\mathbf{j}}(x) = x^{\mathbf{j}}K(x)$ for all $\mathbf{j}$ with $0 \leq |\mathbf{j}| \leq 2p+1$ are Lipschitz continuous. Let $\mu_2(K) = \int u^2 K(u)\, du$

D6 $h \to 0$, $h_1^{p+1}/h^2 \to 0$, $T^{1/2-1/\theta}h^{d(1-1/\theta)}h_1^{-1/2} \to 0$, *and* $Th^d \to \infty$ *as* $T \to \infty$.

Let $ES_{\tau,X}^{(r)}(x)$ and $f_X^{(r)}(x)$ be column vectors of

$$\frac{1}{r_1! \cdots r_d!}\frac{\partial^r ES_\tau(x)}{\partial x_1^{r_1} \cdots \partial x_d^{r_d}}, \frac{1}{r_1! \cdots r_d!}\frac{\partial^r f_X(x)}{\partial x_1^{r_1} \cdots \partial x_d^{r_d}}, \text{ for } r_1 + \cdots + r_d = r,$$

and define the bias function

$$B(\tau, x) = \frac{1}{\tau}\mu_2(K)\left\{\frac{f_X^{(1)}(x)}{f_X(x)}ES_{\tau,X}^{(1)}(x) + ES_{\tau,X}^{(2)}(x)\right\}. \tag{9}$$

Let

$$\eta_t = u_{t\tau} 1 (u_{t\tau} < 0) - \mathrm{E}\left[u_{t\tau} 1 (u_{t\tau} < 0) | X_t\right],$$

and define $W_t = K\left(\frac{X_t - x}{h}\right)\eta_t$, as shown in the proof, the leading component of the estimator (8) is determined by sums of W_t. Notice that $u_{t\tau} 1 (u_{t\tau} < 0) \leq 0$, and the kernel function $K(\cdot)$ has a bounded support, thus W_t has a positive upper bound but no lower bound. It follows that

$$\lim_{x\to\infty} \frac{\Pr\left[W_t > x\right]}{\Pr\left[|W_t| > x\right]} = p_1 = 0 \quad ; \quad \lim_{x\to\infty} \frac{\Pr\left[W_t < -x\right]}{\Pr\left[|W_t| > x\right]} = q_1 = 1.$$

Let c_T be the infimum of value satisfying $\Pr(|W_t| \geq c_T) \leq 1/T$, and $\mathrm{E}[W_t 1 (|W_t| \leq c_T)] = b_T$, under Assumption D3, $T \Pr[a_T^{-1}|\eta_t| > 1] \sim 1$, where $a_T = T^{1/\theta}$. By Breiman (1965) (also see Davis and Mikosch 2001),

$$\begin{aligned}
\Pr\left(|W_t| \geq c_T\right) &= \Pr\left(\left|K\left(\frac{X_t - x}{h}\right)\eta_t\right| \geq c_T\right) \\
&= \Pr\left(\frac{a_T}{c_T}\left|K\left(\frac{X_t - x}{h}\right)\eta_t\right| \geq a_T\right) \\
&= \left[\frac{a_T}{c_T}\right]^\theta \mathrm{E}\left[K\left(\frac{X_t - x}{h}\right)\right]^\theta \Pr\left(|\eta_t| \geq a_T\right) \\
&= \frac{h^d}{c_T^\theta} f_X(x) \int K(u)^\theta du
\end{aligned}$$

thus $c_T \sim \left(Th^d\right)^{1/\theta}$. By definition, η_t are recentered and $\mathrm{E}(W_t) = 0$. However, due to skewness of the distribution of W_t, b_T is non-zero and (as shown in Theorem 4) contributes as a second component of bias term in the estimation of conditional ES. Under Assumption D1, $b_+(m) = p_1 m = 0$, $b_-(m) = q_1 m = m$, hence the limits $\lim_{m\to\infty}\left(b_\pm(m) - b_\pm(m-1)\right) = c_\pm$ exist and $c_+ = 0$, $c_- = 1$.

We summarize the asymptotic behavior of the estimator of the conditional expected shortfall in Theorem 4.

Theorem 4. *Suppose that Assumptions D1 - D6 hold. Then, as $T \to \infty$,*

$$\left(Th^d\right)^{(\theta-1)/\theta}\left(\widehat{ES_\tau(x)} - ES_\tau(x) - h^2 B(\tau, x)\right) - \frac{T}{\left(Th^d\right)^{1/\theta}} \frac{b_T}{\tau f_X(x)} \Rightarrow \frac{1}{\tau f_X(x)} S^*.$$

Furthermore, for $x \neq x'$, $\widehat{ES_\tau(x)}$ and $\widehat{ES_\tau(x')}$ are asymptotically independent.

Remark. The "variance" effect in the estimated conditional ES is of order $\left(Th^d\right)^{-(\theta-1)/\theta}$. The bias terms in the estimated conditional ES contains two parts: the first term $h^2 B(\tau, x)$ is the standard bias term in nonparametric estimation, which is of order h^2; in addition, there is a second part, $\frac{b_T}{h^d T f_X(x)}$. If we denote the conditional density of η_t as $f_{\eta|X}(\eta|x)$, notice that

$$\begin{aligned} b_T &= \mathrm{E}\left[K\left(\frac{X_t-x}{h}\right)\eta_t 1\left(\left|K\left(\frac{X_t-x}{h}\right)\eta_t\right| \le c_T\right)\right] \\ &= \mathrm{E}\left\{\mathrm{E}\left[K\left(\frac{X_t-x}{h}\right)\eta_t 1\left(\left|K\left(\frac{X_t-x}{h}\right)\eta_t\right| \le c_T\right)\middle| X_t\right]\right\} \\ &= \int\int K\left(\frac{x'-x}{h}\right)\eta 1\left(K\left(\frac{x'-x}{h}\right)|\eta| \le c_T\right) f_{\eta|X}(\eta|x') f_X(x') d\eta dx' \\ &= h^d \int\int K(u)\,\eta 1\left(K(u)\,|\eta| \le c_T\right) f_{\eta|X}(\eta|x+uh) f_X(x+uh) d\eta du \\ &\simeq h^d f_X(x) \int\int K(u)\,\eta 1\left(K(u)\,|\eta| \le c_T\right) f_{\eta|X}(\eta|x) d\eta du \end{aligned}$$

which is of order h^d times the magnitude of $E[\eta_t 1(|\eta_t| \le c_T)|X_t = x]$. Since $\mathrm{E}(\eta_t|X_t = x) = 0$, the order of magnitude of $E[\eta_t 1(|\eta_t| \le c_T)|X_t = x]$ is the same as that of $E[\eta_t 1(|\eta_t| > c_T)|X_t = x]$. Notice that $c_T \to \infty$ as $T \to \infty$, and η_t is bounded from above but has no lower bound, $E[\eta_t 1(|\eta_t| > c_T)|X_t = x] = E[\eta_t 1(\eta_t < -c_T)|X_t = x]$ for large enough T. Under assumption D3, $\Pr[\eta_t < -c|X_t = x] \sim c^{-\theta}$ for large c, thus $\mathrm{E}[\eta_t 1(\eta_t < -c_T)|X_t = x]$ is of order $\left(Th^d\right)^{-(\theta-1)/\theta}$. Thus, the second part of bias is of order $\left(Th^d\right)^{-(\theta-1)/\theta}$. Under the assumption that $\theta \in (1,2)$, both parts in the bias term converge to zero and the estimator is consistent. The bandwidth that matches the smoothing bias and the stochastic term is of order $h \sim T^{-(\theta-1)/((2+d)\theta-1)}$, which results in the rate of convergence being $T^{-2(\theta-1)/((2+d)\theta-1)}$.

Remark. Notice that in the presence of infinite variance, the variance effect of the preliminary quantile regression is of order $\left(Th_1^d\right)^{-1/2}$, which is smaller than the variance effect of a local least-squares estimation (which would be $\left(Th_1^d\right)^{-(\theta-1)/\theta}$). If $p > 1$, we may simply choose $h = h_1$ in the estimator (8), and the preliminary nonparametric estimation does not affect the final estimation of expected shortfall.

Remark. The asymptotic distribution in Theorem 4 does not depend on the time series properties, meaning that the asymptotic distribution is the same as if the data were iid with the same marginal distributions. This result is typical in nonparametric problems, see Robinson (1983). The localization by the kernel shuffles the time series so that their dependence is no longer so important. It follows that one can construct consistent confidence intervals from an estimator of θ alone (provided one undersmooths to make the smoothing bias of smaller order), see McCulloch (1986) and Rachev and Mittnik (2000). However, if $X_t = t/T$, then the above result does not apply. In this case, the limiting distribution will be more like that from Theorem 1, although the rate of convergence will be nonparametric, see Hart (1991) and Peng and Yao (2004).

The last remark suggests an alternative way of estimating and conducting inference about the unconditional expected shortfall. This approach sacrifices some efficiency (rate of convergence) with the benefit of simpler inference. Specifically, let $X_1, \ldots, X_T$ be i.i.d. uniform $[0,1]$ random variables independent of $Y_1, \ldots, Y_T$. Then compute $\widehat{ES_\tau(x)}$ from (8) for a set of points $x \in \{x_1, \ldots, x_n\}$, where the grid is chosen such that $\widehat{ES_\tau(x_j)}$ and $\widehat{ES_\tau(x_k)}$ are asymptotically independent (for example, for $n = 1$ this is trivially satisfied). Then let

$$\widehat{ES_\tau} = \sum_{j=1}^{n} w_{nj} \widehat{ES_\tau(x_j)},$$

where w_{nj} are weights with $\sum_{j=1}^{n} w_{nj} = 1$. We can take n to be almost as but not quite as large as $n = O(Th)$ when the locations x_j are chosen correctly in $[0,1]$, i.e., so that $\{\widehat{ES_\tau(x_j)}, j = 1, \ldots, n\}$ are mutually independent. For an appropriate choice of weights we will obtain an improved rate of convergence over that in Theorem 4, but not quite as good as in Theorem 2; heuristically, the leading stochastic term of $\widehat{ES_\tau}$ behaves like a sample average of the form $\sum_{j=1}^{n} w_{nj} S_j^*$, where S_j^* are i.i.d. stable random variables with mean zero, and this will converge in distribution at rate $n^{(\theta-1)/\theta}$ to a stable random variable. The limiting distribution will be stable but will not depend on the dependence structure of the data.

6 Numerical Results

6.1 Monte Carlo

In this section, we report some results based on a limited Monte Carlo experiment designed to examine the sampling performance of the estimators and the subsampling procedure. The data is generated from the following AR(1) process with different choices of parameter values and error distributions:

$$Y_t = \alpha Y_{t-1} + u_t,$$

where u_t are iid stable random variables with regularly varying tail probabilities satisfying $T \Pr[a_T^{-1}|u_1| > 1] \to 1$, and $a_T = T^{1/\theta}$, $\theta \in (1,2)$. Using the conventional notation, we denoted the stable distribution as $S(\mu, \sigma, \theta, \beta)$, whose characteristic function of u, $\phi(t) = E\exp(itu)$, is given by

$$\phi(t) = \exp\left\{it\mu - \sigma^{\theta} |t|^{\theta} (1 - i\beta \text{sgn}(t) \tan(\pi\theta/2)\right\},$$

where the four parameters μ, $\sigma \geq 0$, $\theta \in (0, 2]$, $\beta \in [-1, 1]$, are location, scale, shape, and skewness parameters. Two sets of parameter values, DGP1: $\mu = 0$, $\sigma = 1$, $\theta = 1.5$, $\beta = 0$, and DGP2: $\mu = 0$, $\sigma = 1$, $\theta = 1.5$, $\beta = 0.6$, are investigated in the simulation. The first DGP correspond to symmetric innovations, and the second DGP correspond to skewed u_t. Giving the true vales of the parameters, the true vale of $ES(\tau)$ can be calculated numerically. These random variables are generated using the stable random variable generator STABRND.M of McCulloch (1996).

We consider estimation of the expected shortfall at both the $\tau = 0.05$ and $\tau = 0.1$ quantiles. Two different choices of AR parameter value are considered: (i) $\alpha = 0.5$; (ii) $\alpha = 0.8$. We examine the sampling performance for a range of sample sizes $T = 100$, $T = 200$, $T = 500$, $T = 1000$. Number of repetition is 5000. If we denote the unsmoothed estimator in the i-th iteration as $\widehat{ES}_{\tau,i}$, we evaluate the sampling performance of the estimate based on the average bias of the estimate defined as

$$BIAS = \frac{1}{N} \sum_{i=1}^{N} \widehat{ES}_{\tau,i} - ES_{\tau},$$

and the Absolute Deviation Error

$$ADE = \frac{1}{N} \sum_{i=1}^{N} \left| \widehat{ES}_{\tau,i} - ES_{\tau} \right|.$$

For the smoothed estimator $\widetilde{ES}_{\tau}$, we consider two choices of bandwidth $h_1 = T^{-1/4}$ and $h_2 = 0.5T^{-1/4}$.

Tables 1 and 2 report result of average bias $BIAS$ and mean absolute error ADE for different estimators of the expected shortfall at the 5% ($\tau = 0.05$) quantile. In particular, Tale 1 reports the result for the case of DGP1 and Table 2 reports the results corresponding to DGP2. The Monte Carlo results in Tables 1 and 2 indicates that the difference of the smoothed and unsmoothed estimators is small. As the sample size T increases, both the bias and absolute deviation errors are decreasing, corroborating the asymptotic theory. As we move from $\alpha = 0.5$ to $\alpha = 0.8$, estimation error increases when serial correlation increases.

We next look at the 10% ($\tau = 0.1$) quantile. Tables 3 and 4 report the corresponding results for the case of DGP1 and DGP2. The results in Tables 3 and 4 are qualitatively similar to those of Tables 1 and 2, but we can see that estimation of ES improves when we move from tail to relatively central quantile. The closer to the extremal points, the poor estimates.

Table 1: BIAS and ADE ($ES(5\%)$, DGP1)

T		Unsmoothed	Smoothed h_1	Smoothed h_2	Unsmoothed	Smoothed h_1	Smoothed h_2
			$\alpha = 0.8$			$\alpha = 0.5$	
100	BIAS	3.7161	3.7191	3.7189	0.4890	0.4924	0.4927
	ADE	13.9550	13.9581	13.9578	7.2158	7.2192	7.2194
200	BIAS	1.6336	1.6323	1.6328	0.4648	0.4643	0.4642
	ADE	12.2525	12.2527	12.2524	5.9460	5.9475	5.9460
500	BIAS	0.7587	0.7544	0.7570	-0.1131	-0.1171	-0.1153
	ADE	10.8352	10.8330	10.8343	5.0509	5.0499	5.0500
1000	BIAS	0.6221	0.6162	0.6203	0.2273	0.2266	0.2304
	ADE	9.1616	9.1614	9.1638	3.7967	3.7959	3.7974

Table 2: BIAS and MAE ($ES(5\%)$, DGP2)

T		Unsmoothed	Smoothed h_1	Smoothed h_2	Unsmoothed	Smoothed h_1	Smoothed h_2
			$\alpha = 0.8$			$\alpha = 0.5$	
100	BIAS	1.7966	1.7908	1.8004	0.2614	0.2521	0.2660
	ADE	7.1033	7.0988	7.1070	3.7384	3.7353	3.7437
200	BIAS	0.7400	0.7358	0.7381	0.2590	0.2584	0.2587
	ADE	6.7778	6.7770	6.7775	3.0945	3.0990	3.0957
500	BIAS	0.3123	0.3128	0.3176	-0.0442	-0.0482	-0.0484
	ADE	5.5750	5.5730	5.5774	2.6568	2.6587	2.6564
1000	BIAS	0.2587	0.2464	0.2535	0.1435	0.1336	0.1379
	ADE	4.7402	4.7379	4.7414	2.0066	2.0055	2.0085

Table 3: BIAS and ADE ($ES(10\%)$, DGP1)

T		Unsmoothed	Smoothed h_1	Smoothed h_2	Unsmoothed	Smoothed h_1	Smoothed h_2
		$\alpha = 0.8$			$\alpha = 0.5$		
100	BIAS	1.4851	1.4981	1.4902	0.3274	0.3404	0.3489
	ADE	7.7923	7.8736	7.8663	3.6355	3.6483	3.6521
200	BIAS	1.1033	1.1045	1.1056	0.0343	0.0367	0.0418
	ADE	6.6707	6.6722	6.6725	3.2760	3.2848	3.2884
500	BIAS	0.5177	0.5144	0.5178	0.0469	0.0581	0.0636
	ADE	5.4938	5.4927	5.4942	2.4974	2.5031	2.5066
1000	BIAS	-0.000608	-0.0041	-0.000989	0.0126	0.0202	0.0230
	ADE	4.9667	4.9655	4.9667	2.0419	2.0450	2.0464

Table 4: BIAS and MAE ($ES(10\%)$, DGP2)

T		Unsmoothed	Smoothed h_1	Smoothed h_2	Unsmoothed	Smoothed h_1	Smoothed h_2
		$\alpha = 0.8$			$\alpha = 0.5$		
100	BIAS	0.5690	0.5751	0.5830	0.2039	0.2266	0.2146
	ADE	4.3807	4.3871	4.3962	1.9104	1.9306	1.9187
200	BIAS	0.5982	0.6005	0.5951	0.0828	0.1049	0.0892
	ADE	3.5007	3.5017	3.5028	1.6389	1.6550	1.6433
500	BIAS	0.2175	0.2182	0.2224	0.0628	0.0767	0.0657
	ADE	2.9594	2.9606	2.9643	1.2808	1.2891	1.2827
1000	BIAS	0.1984	0.1986	0.1972	0.0225	0.0298	0.0234
	ADE	2.4301	2.4311	2.4297	1.0810	1.0854	1.0818

The second part of Monte Carlo examines the performance of the subsampling procedure. We focus on the unsmoothed estimator since the two versions estimators have qualitatively similar performance. In each iteration, the empirical quantiles of

$$\left\{M^{(\widehat{\theta}-1)/\widehat{\theta}}\left[\widehat{ES}_{\tau}(M,t)-\widehat{ES}_{\tau}\right], t=1,\ldots,T-M+1\right\} \tag{10}$$

can be calculated. If we denote, say, the p-th quantile of (10) in the i-th iteration by

$Q_{\widehat{ES}_\tau(M,t),i}(p)$, we expect that

$$\frac{1}{N}\sum_{i=1}^{N} 1\left[T^{(\theta-1)/\theta}\left(\widehat{ES}_\tau - ES_\tau\right) < Q_{\widehat{ES}_\tau(M,t),i}(p)\right] \to p.$$

To evaluate the approximation of the sampling distribution of $T^{(\theta-1)/\theta}\left(\widehat{ES}_\tau - ES_\tau\right)$ by the subsampling distribution, we report the empirical value of

$$\frac{1}{N}\sum_{i=1}^{N} 1\left[T^{(\theta-1)/\theta}\left(\widehat{ES}_\tau - ES_\tau\right) < Q_{\widehat{ES}_\tau(M,t),i}(p)\right]$$

for different sample sizes and data generating processes DGP1 and DGP2 that we used in the first part simulation. For the size of subsample, we consider two choices: $M_1 = \sqrt{T}$ and $M_2 = 1.5\sqrt{T}$. We report the results corresponding to $p = 0.05$ and $p = 0.1$ in Tables 5 and 6 respectively. Again, as we move closer to the extremal points, the poor sampling performance is observed.

Table 5 ($p = 0.05$)

T	DGP1				DGP2			
	$\alpha = 0.8$		$\alpha = 0.5$		$\alpha = 0.8$		$\alpha = 0.5$	
	M_1	M_2	M_1	M_2	M_1	M_2	M_1	M_2
100	0.0196	0.0654	0.0120	0.0142	0.0210	0.0574	0.0098	0.0116
200	0.0286	0.0928	0.0124	0.0190	0.0244	0.0814	0.0096	0.0182
500	0.0528	0.0122	0.0326	0.0272	0.0466	0.0224	0.0146	0.0138
1000	0.0754	0.0316	0.0656	0.0276	0.0684	0.0306	0.0524	0.0234

Table 6 ($p = 0.1$)

T	DGP1				DGP2			
	$\alpha = 0.8$		$\alpha = 0.5$		$\alpha = 0.8$		$\alpha = 0.5$	
	M_1	M_2	M_1	M_2	M_1	M_2	M_1	M_2
100	0.1028	0.0464	0.0382	0.0290	0.0932	0.0448	0.0366	0.0236
200	0.2246	0.0438	0.1760	0.0302	0.2328	0.0486	0.2186	0.0294
500	0.1828	0.1690	0.1470	0.1172	0.1832	0.1710	0.1494	0.1362
1000	0.1680	0.1074	0.1304	0.0868	0.1720	0.1090	0.1322	0.0852

6.2 Application

Ibragimov, Ibragimov and Kattuman (2010) have documented heavy tails in emerging market exchange rates, including the Russian Ruble, the Thai Baht, and the Phillipine PP. They looked at the period from October 1997 to October 2008 (the Russian Ruble for example is from January 1st 1999 to October 29th 2008.). They found these three currencies (and others) had tail index rate less than two. We will examine again the Ruble over the time period 1993-2011, which contains daily 4649 observations (we find similar results, not shown, for the two other currencies mentioned above). We show below the standardized time series (that is, normalized to have sample mean and variance one), which shows the extreme movements within the sample, especially during the default related crisis period of 1999.

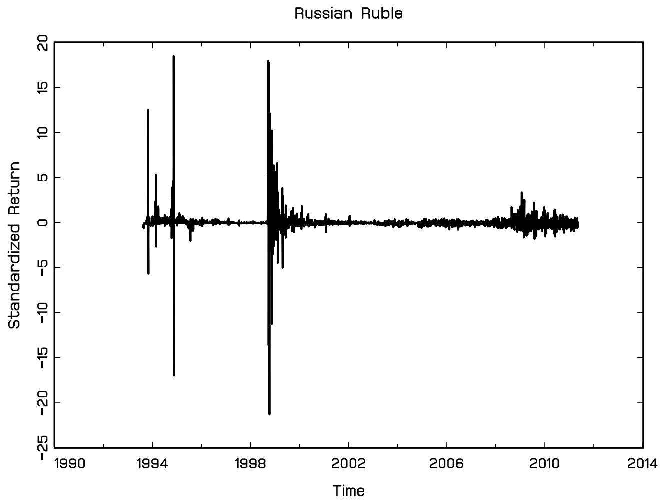

Figure 1. Russian Ruble 1993-2011 Daily standardized return.

We first report the estimated tail indexes of the series using the Gabaix and Ibragimov (2011) rank-1/2 method. We report this a function of the threshold parameter (or number of observations used in the regression) upto and including the value of 500, which is slightly more than 10% of the data. The estimated parameters are consistently below 2 except for the upper tail for the very small threshold. In fact, the estimates are remarkably stable, and more so than the corresponding Hill estimates that we also computed. We find some difference between the upper and lower tails of the data, but a value of 1.3 for both would not be far wrong; this is consistent with the findings of Gabaix and Ibragimov (2011).

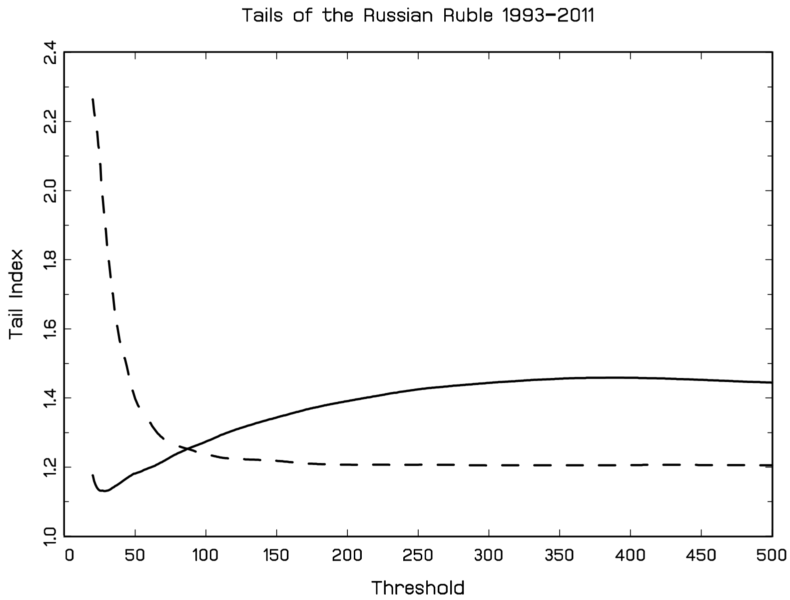

Figure 2. Estimated tail thickness parameters for different values of threshold. Solid line is the lower tail, dashed line is the upper tail.

We then estimated the 5% annual Value at Risk and Expected Shortfall of the time series using our methods described above. The 5% VaR and ES for the whole sample were -0.00247 and -0.00706 respectively. We next estimated both quantities using a rolling window, backward looking, subsample to show the variation of the risk measure, as we may expect it not to be constant over the whole period. We do this for an annual window (250 observations), which is consistent with much recent practice. Similar results are obtained for longer time periods. The results are shown below.

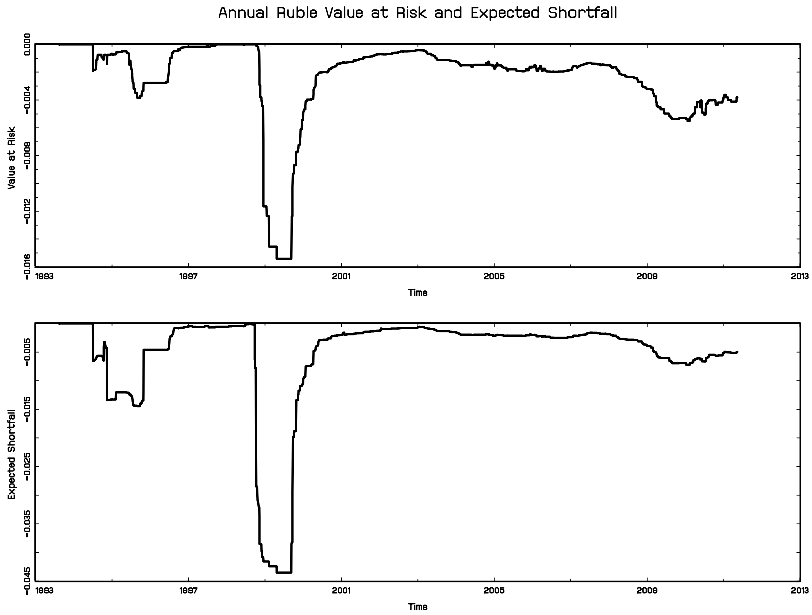

Figure 3. Estimated annual 5% Value at Risk and Expected shortfall, rolling window.

Both series vary a lot over the sample period with the crisis period at the turn of the millennium dominating numerically. More recently, both risk measures increased during the financial crisis but seem to have recovered. We note that by taking account of the estimation error[4] we obtain a full sample ES of -0.0179, which is considerably smaller than the original estimator.

The subsample VaR and ES take different numerical values but they move very closely together. One might think that the expected shortfall is just a constant multiple of the value at risk, but this is not true at all. The next graph shows the relationship between the expected shortfall and the Value at Risk as a scatter plot with a regression line included. The relationship between these two risk measures seems quite nonlinear and complex and perhaps characterized by a number of different regimes. Notably, when risk is low there seems to be a strong linear relationship, which would be consistent with a normal type innovation with time varying volatility but when risk is higher this relationship disappears. Perhaps it is necessary to allow a changing tail thickness parameter to explain this.

[4] We computed the 0.05 quantile of the standardized subsample estimates (using $\theta = 1.4$) denoted by $\widehat{c}_\alpha$, and then computed $ES + \widehat{c}_\alpha / T^{(\theta-1)/\theta}$.

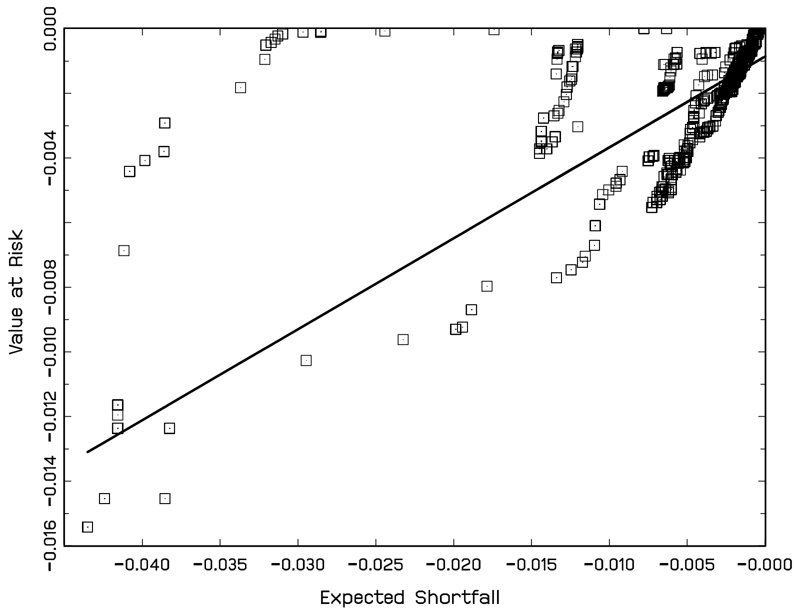

Figure 4. Scatter plot of annual Value at Risk versus Expected Shortfall.

7 Conclusions

We have shown that the usual asymptotic inference procedures for expected shortfall estimation break down when the series has heavy tails. In some sense this shows a weakness of expected shortfall as a risk measure, since by contrast the Value at Risk has the same asymptotic behaviour regardless of the thickness of the tails. Nevertheless, the expected shortfall is widely used. We have shown how to conduct correct inference in the heavy tail case. The results complement existing results for the light tail case with finite variances.

The analysis in this paper can be extended to other similar concepts. For example, our analysis in this paper can be applied to similar risk measures such as median absolute deviation $\mathrm{E}|Y-\mathrm{median}(Y)|$ or the medianhalf deviation $\mathrm{E}|Y|\,1(Y \leq \mathrm{median}(Y))$. The results can also be extended to higher moment coherent risk measures studied by Krokhmal (2007). In addition, the method can also be applied to Lorenz curve estimation and inference. The definition of Lorenz curve for a given income distribution is similar to the expected shortfall: suppose that income Y is a random variable with distribution function F, the corresponding Lorenz curve is defined as $L(\tau) = \mu^{-1}\int_0^\tau F^{-1}(t)dt$. Cumulated evidence indicates heavy-tail behavior of income or wealth distributions. Thus the proposed estimators and inference procedures may be extended to the study of Lorenz curves.

8 Appendix

8.1 Proof of Theorem 1

First note that $\widetilde{\alpha}(\tau) - \alpha(\tau) = O_p(T^{-1/2})$, Scaillet (2004). Using the definition of $\widetilde{\alpha}(\tau)$ we have

$$\begin{aligned}
\widetilde{ES}_\tau - ES_\tau &= \alpha(\tau)\frac{1}{T\tau}\sum_{t=1}^{T}\mathcal{K}_h(Y_t - \widetilde{\alpha}(\tau)) + \frac{1}{T\tau}\sum_{t=1}^{T}(Y_t - \alpha(\tau))\mathcal{K}_h(Y_t - \widetilde{\alpha}(\tau)) - ES_\tau \\
&= \alpha(\tau) + \frac{1}{T\tau}\sum_{t=1}^{T}(Y_t - \alpha(\tau))\mathcal{K}_h(Y_t - \alpha(\tau)) - ES_\tau \\
&\quad + \frac{1}{T\tau}\sum_{t=1}^{T}(Y_t - \alpha(\tau))\left[\mathcal{K}_h(Y_t - \widetilde{\alpha}(\tau)) - \mathcal{K}_h(Y_t - \alpha(\tau))\right] \\
&= \frac{1}{T\tau}\sum_{t=1}^{T}\{(Y_t - \alpha(\tau))1(Y_t \leq \alpha(\tau)) - \alpha(ES_\tau - \alpha(\tau))\} \\
&\quad + \frac{1}{T\tau}\sum_{t=1}^{T}(Y_t - \alpha(\tau))\left[\mathcal{K}_h(Y_t - \alpha(\tau)) - 1(Y_t \leq \alpha(\tau))\right] \\
&\quad + \frac{1}{T\tau}\sum_{t=1}^{T}(Y_t - \alpha(\tau))\left[\mathcal{K}_h(Y_t - \widetilde{\alpha}(\tau)) - \mathcal{K}_h(Y_t - \alpha(\tau))\right] \\
&= W_T + R_{T1} + R_{T2}.
\end{aligned}$$

Under assumption B1, by Taylor expansion,

$$\begin{aligned}
R_{T2} &= \frac{1}{T\tau}\sum_{t=1}^{T}(Y_t - \alpha(\tau))\left[\mathcal{K}_h(Y_t - \widetilde{\alpha}(\tau)) - \mathcal{K}_h(Y_t - \alpha(\tau))\right] \\
&= (\widetilde{\alpha}(\tau) - \alpha(\tau))\frac{1}{T\tau}\sum_{t=1}^{T}(Y_t - \alpha(\tau))K_h(Y_t - \alpha(\tau)) \\
&\quad - \frac{1}{2\alpha}(\widetilde{\alpha}(\tau) - \alpha(\tau))^2\frac{1}{Th^2}\sum_{t=1}^{T}(Y_t - \alpha(\tau))K'\left(\frac{Y_t - \overline{\nu}_\tau}{h}\right) \\
&= R_{T21} + R_{T22},
\end{aligned}$$

where $\overline{\nu}_\tau$ are intermediate values and $K_h(.) = K(./h)/h$ for a kernel density K that is bounded and has compact support. Therefore, $|(Y_t - \alpha(\tau))K_h(Y_t - \alpha(\tau))| \leq C(h)$ for some constant $C(h) < \infty$. We can apply the central limit theorem for triangular arrays of stationary Geometrically mixing processes (e.g. Withers 1981), so that

$$\frac{1}{T}\sum_{t=1}^{T}(Y_t - \alpha(\tau))K_h(Y_t - \alpha(\tau)) - E[(Y_t - \alpha(\tau))K_h(Y_t - \alpha(\tau))] = O_p(T^{-1/2}h^{-1/2}).$$

By a change of variable $x \mapsto u = (x - \alpha(\tau))/h$ and Taylor expansion we have

$$\begin{aligned} E\left[(Y_t - \alpha(\tau))K_h(Y_t - \alpha(\tau))\right] &= \int (x - \alpha(\tau))K_h(x - \alpha(\tau))f(x)dx \\ &= h \int uK(u)f(\alpha(\tau) + uh)du \\ &\simeq \frac{h^2}{2} f'(\alpha(\tau)) \int u^2 K(u)du. \end{aligned}$$

Therefore, $R_{T21} = O_p(T^{-1}h^{-1/2}) + O_p(T^{-1/2}h^2) = o_p(T^{-1/2})$.

Similar arguments apply to R_{T22}. By the square-root T consistency of $\widetilde{\alpha}(\tau)$, there exists a sequence $\epsilon_T \to 0$ such that $\Pr[\sqrt{T}\,|\widetilde{\alpha}(\tau) - \alpha(\tau)| > \epsilon_T] \to 0$. Therefore, we can restrict attention to deterministic sequences $\alpha(\tau) + \delta/\sqrt{T}$ for δ with $|\delta| \leq \Delta < \infty$. Furthermore,

$$\sup_{|\delta| \leq \Delta} \left| \frac{Y_t - \alpha(\tau)}{h} K' \left(\frac{Y_t - \alpha(\tau) - \delta T^{-1/2}}{h} \right) \right| \leq C'$$

for some $C' < \infty$. Therefore, we can apply the law of large numbers for triangular arrays of stationary Geometrically mixing processes (e.g. Davidson 1994), so that

$$\frac{1}{Th} \sum_{t=1}^{T} (Y_t - \alpha(\tau))K' \left(\frac{Y_t - \overline{\nu}_\tau}{h} \right) = O_p(1),$$

so that $R_{T22} = O_p(T^{-1}h^{-1}) = o_p(T^{-1/2})$.

Furthermore, consider

$$R_{T1} = \frac{1}{T} \sum_{t=1}^{T} (Y_t - \alpha(\tau)) \left\{ \mathcal{K}_h(Y_t - \alpha(\tau)) - 1(Y_t \leq \alpha(\tau)) \right\}.$$

Under assumption B1, we have $|R_{T1}| \leq Ch$ for some positive constant C. Therefore,

$$\frac{R_{T1} - ER_{T1}}{h} = O_p(T^{-1/2}).$$

By Taylor expansion and change of variables we have

$$\begin{aligned} ER_{T1} &= \int (x - \alpha(\tau)) \left\{ \mathcal{K}\left(\frac{x - \alpha(\tau)}{h} \right) - 1(x \leq \alpha(\tau)) \right\} f(x)dx \\ &= \frac{h^2}{2} \int_{-1}^{1} u \left\{ \mathcal{K}(u) - 1(u \leq 0) \right\} f(\alpha(\tau) + uh)du \\ &\simeq \frac{h^2}{2} f(\alpha(\tau)) \int_{-1}^{1} u \left\{ \mathcal{K}(u) - 1(u \leq 0) \right\} du. \end{aligned}$$

Therefore, $R_{T1} = O_p(h^2) + O_p(T^{-1/2}h) = o_p(T^{-1/2})$.

Finally, write

$$W_T = \frac{1}{\tau}\frac{1}{T}\sum_{t=1}^{T} Z_t,$$
$$Z_t = (Y_t - \alpha(\tau))1(Y_t \le \alpha(\tau)) - \tau(ES_\tau - \alpha(\tau)).$$

Note that Z_t are mean zero and strictly stationary. We have $Z_t \le -\tau(ES_\tau - \alpha(\tau))$, where $ES_\tau - \alpha(\tau) \in (-\infty, 0)$, so that Z_t has a positive upper bound but no lower bound. It follows that

$$\lim_{x\to\infty} \frac{\Pr[Z_t > x]}{\Pr[|Z_t| > x]} = p = 0 \quad ; \quad \lim_{x\to\infty} \frac{\Pr[Z_t < -x]}{\Pr[|Z_t| > x]} = q = 1.$$

Furthermore, for $x > 0$,

$$\begin{aligned}
\Pr[|Z_t| > x] &= \Pr[|\{(Y_t - \alpha(\tau))1(Y_t \le \alpha(\tau)) - \alpha(ES_\tau - \alpha(\tau))\}| > x] \\
&= \Pr[\{(Y_t - \alpha(\tau))1(Y_t \le \alpha(\tau)) - \alpha(ES_\tau - \alpha(\tau))\} < -x] \\
&= \Pr[Y_t < x'] \\
&= |x'|^{-\theta} L'(x') \\
&\simeq x^{-\theta} L'(-x),
\end{aligned}$$

where $x' = -x + \tau(ES_\tau - \alpha(\tau)) + \alpha(\tau) < 0$ and L' is a slowly varying function at $-\infty$. The first four lines are exact for large enough positive x, and the approximation is valid as $x \to \infty$. Furthermore, Z_t is regularly varying with limiting measure $\overline{\mu}_m(\cdot)$ such that $\overline{\mu}_m(A) = \mu_m(A \cap B_m)$, where $B_m = \cap_{t=1}^{m}\{Y_t \le \alpha(\tau)\}$.

We use Bartkiewicz et al. (2010, Theorem 3.1) applied to $\sum_{t=1}^{T} Z_t$. Thus

$$a_T^{-1}\sum_{t=1}^{T} Z_t \Longrightarrow S.$$

The result follows. ■

8.2 Proof of Theorem 2

Let $u_{t\tau} = Y_t - \alpha(\tau)$, and $g(u) = 1(u \le 0)$, since $g(\cdot)$ is not everywhere differentiable, we treat the function $g(\cdot)$ as a generalized function (as in Phillips (1995)) with a smooth regular sequence

$$g_m(u) = \int_{-\infty}^{\infty} g(v) H[m(v-u)]\, m e^{-v^2/m^2} dv$$

where $H(\cdot)$ is a smudge function whose role in $g_m(u)$ is to smudge out $g(v)$ when v is outside the interval $(u-m^{-1}, u+m^{-1})$ (see Phillips (1995) for more discussions on smudge function and related literature). Notice that $g(u)$ has first derivative everywhere except $u=0$ and

$$\dot{g}(u) = \frac{dg(u)}{du} = \delta(u) = \text{ Dirac delta function,}$$

has a regular sequence $\dot{g}_m(\cdot) = (m/\pi)^{1/2} e^{-mu^2} = \delta_m(u)$.

If we denote $\Psi_n(g) = \frac{1}{\tau T}\sum_{t=1}^{T} Y_t g(\widehat{u}_{t\tau})$, then $\Psi_n(g)$ is a generalized process defined by the following regular sequence of processes

$$\Psi_{n,m}(g) = \frac{1}{\tau T}\sum_{t=1}^{T} Y_t g_m(\widehat{u}_{t\tau}).$$

By a Taylor expansion of $g_m(u)$ around $u = u_{t\tau}$ gives

$$\begin{aligned}\Psi_{n,m}(g) &= \frac{1}{\tau T}\sum_{t=1}^{T} Y_t g_m(u_{t\tau}) - \frac{1}{\tau T}\sum_{t=1}^{T} Y_t \dot{g}_m(u_{t\tau})(\widehat{\alpha}(\tau)-\alpha(\tau)) \\ &\quad + \frac{1}{\tau T}\sum_{t=1}^{T} Y_t\left[\dot{g}_m(u_{t\tau}) - \dot{g}_m(u_{t\tau}^*)\right](\widehat{\alpha}(\tau)-\alpha(\tau)),\end{aligned} \tag{11}$$

where $u_{t\tau}^*$ is a middle value.

We next show that the last two terms $\xrightarrow{P} 0$, and derive the limit of the first term in the above expansion. First, notice that the regular sequence $\dot{g}_m$ is differentiable and has bounded derivative (with a bound that depends on m) for all m, and $\widehat{\alpha}(\tau)-\alpha(\tau) = O_p(T^{-1/2})$, thus as $T\to\infty$,

$$\left|\frac{1}{\tau T}\sum_{t=1}^{T} Y_t\left[\dot{g}_m(u_{t\tau}) - \dot{g}_m(u_{t\tau}^*)\right](\widehat{\alpha}(\tau)-\alpha(\tau))\right| \xrightarrow{P} 0,$$

and

$$\left|\frac{1}{\tau T}\sum_{t=1}^{T} Y_t \dot{g}_m(u_{t\tau})(\widehat{\alpha}(\tau)-\alpha(\tau))\right| \xrightarrow{P} 0.$$

For the first term on the right hand side of (11), notice that $H[m(v-u_{t\tau})] = H[m(v+\alpha(\tau)-Y_t)]$ is a measurable and integrable function of $u_{t\tau}$, and thus also a mixing process satisfying the same dependence assumption, thus, as $T \to \infty$,

$$\begin{aligned}
&\frac{1}{\tau T}\sum_{t=1}^{T} Y_t g_m\left(u_{t\tau}\right) \\
=\ & \frac{1}{\tau}\int_{-\infty}^{\infty} g(v)\left\{\frac{1}{T}\sum_{t=1}^{T} Y_t H\left[m(v-u_{t\tau})\right]\right\} m e^{-v^2/m^2} dv \\
\to\ & \frac{1}{\tau}\int_{-\infty}^{\infty} g(v)\mathrm{E}\left\{Y_t H\left[m(v-u_{t\tau})\right]\right\} m e^{-v^2/m^2} dv \\
=\ & \frac{1}{\tau}\mathrm{E}\left(Y_t g_m\left(u_{t\tau}\right)\right) = \mu_{m\tau}
\end{aligned}$$

and, as $m \to \infty$,

$$\mu_{m\tau} \to \frac{1}{\tau}\mathrm{E}\left(Y_t g\left(u_{t\tau}\right)\right) = ES_\tau.$$

Thus,

$$\widehat{ES}_\tau \to ES_\tau.$$

Limiting distribution:

$$\begin{aligned}
& T^{(\theta-1)/\theta}\left[\widehat{ES}_\tau - ES_\tau\right] \\
=\ & T^{(\theta-1)/\theta}\left[\frac{1}{\tau T}\sum_{t=1}^{T} Y_t 1\left(Y_t - \widehat{\alpha}(\tau) \le 0\right) - \frac{1}{\tau}\mathrm{E}Y_t 1\left(Y_t - \alpha(\tau) \le 0\right)\right] \\
=\ & \frac{1}{T^{(1-\theta)/\theta}}\frac{1}{\tau T}\sum_{t=1}^{T}\left[Y_t 1\left(Y_t - \alpha(\tau) \le 0\right) - \mathrm{E}Y_t 1\left(Y_t - \alpha(\tau) \le 0\right)\right] \\
& + \frac{1}{\tau T}\sum_{t=1}^{T}\left[Y_t 1\left(Y_t - \widehat{\alpha}(\tau) \le 0\right) - Y_t 1\left(Y_t - \alpha(\tau) \le 0\right)\right] \\
=\ & \frac{1}{\tau T^{1/\theta}}\sum_{t=1}^{T}\left[\left(Y_t - \alpha(\tau)\right) 1\left(Y_t \le \alpha(\tau)\right) - \mathrm{E}\left(Y_t - \alpha(\tau)\right) 1\left(Y_t \le \alpha(\tau)\right)\right] \\
& + \frac{1}{\tau T^{1/\theta}}\sum_{t=1}^{T}\left[\alpha(\tau) 1\left(Y_t \le \alpha(\tau)\right) - \alpha(\tau)\mathrm{E}1\left(Y_t \le \alpha(\tau)\right)\right] \\
& + \frac{1}{\tau T^{1/\theta}}\sum_{t=1}^{T}\left[\left(Y_t - \alpha(\tau)\right) 1\left(Y_t - \widehat{\alpha}(\tau) \le 0\right) - \left(Y_t - \alpha(\tau)\right) 1\left(Y_t - \alpha(\tau) \le 0\right)\right] \\
& + \frac{1}{\tau T^{1/\theta}}\sum_{t=1}^{T}\left[\alpha(\tau) 1\left(Y_t \le \widehat{\alpha}(\tau)\right) - \alpha(\tau) 1\left(Y_t \le \alpha(\tau)\right)\right] \\
=\ & Z_{T1} + Z_{T2} + Z_{T3} + Z_{T4}
\end{aligned}$$

where

$$
\begin{aligned}
Z_{T1} &= \frac{1}{\tau T^{1/\theta}} \sum_{t=1}^{T} \left[(Y_t - \alpha(\tau)) \, 1 \, (Y_t \le \alpha(\tau)) - \mathrm{E} \, (Y_t - \alpha(\tau)) \, 1 \, (Y_t \le \alpha(\tau)) \right] \\
Z_{T2} &= \frac{1}{\tau T^{1/\theta}} \sum_{t=1}^{T} \left[(Y_t - \alpha(\tau)) \, 1 \, (Y_t \le \widehat{\alpha}(\tau)) - (Y_t - \alpha(\tau)) \, 1 \, (Y_t \le \alpha(\tau)) \right] \\
Z_{T3} &= \frac{\alpha(\tau)}{\tau T^{1/\theta}} \sum_{t=1}^{T} \left[1 \, (Y_t \le \alpha(\tau)) - \mathrm{E} 1 \, (Y_t \le \alpha(\tau)) \right] \\
Z_{T4} &= \frac{\alpha(\tau)}{\tau T^{1/\theta}} \sum_{t=1}^{T} \left[1 \, (Y_t \le \widehat{\alpha}(\tau)) - 1 \, (Y_t \le \alpha(\tau)) \right].
\end{aligned}
$$

We analyze each of these terms.

We first analyze

$$
\begin{aligned}
Z_{T2} &= \frac{1}{\tau T^{1/\theta}} \sum_{t=1}^{T} \left[(Y_t - \alpha(\tau)) \, 1 \, (Y_t \le \widehat{\alpha}(\tau)) - (Y_t - \alpha(\tau)) \, 1 \, (Y_t \le \alpha(\tau)) \right] \\
&= \frac{1}{\tau T^{1/\theta}} \sum_{t=1}^{T} \left[(Y_t - \alpha(\tau)) \, g(\widehat{u}_{t\tau}) - (Y_t - \alpha(\tau)) \, g(u_{t\tau}) \right].
\end{aligned}
$$

Again, this is a generalized process defined by the following regular sequence:

$$
Z_{T2,m} = \frac{1}{\tau T^{1/\theta}} \sum_{t=1}^{T} \left[(Y_t - \alpha(\tau)) \, g_m(\widehat{u}_{t\tau}) - (Y_t - \alpha(\tau)) \, g_m(u_{t\tau}) \right].
$$

Expanding $g_m(u)$ around $u = u_{t\tau}$ gives

$$
\begin{aligned}
Z_{T2,m} &= -\frac{1}{\tau T^{1/\theta}} \sum_{t=1}^{T} (Y_t - \alpha(\tau)) \, \dot{g}_m(u_{t\tau}) \, (\widehat{\alpha}(\tau) - \alpha(\tau)) \\
&\quad + \frac{1}{\tau T^{1/\theta}} \sum_{t=1}^{T} (Y_t - \alpha(\tau)) \left[\dot{g}_m(u_{t\tau}) - \dot{g}_m(u_{t\tau}^*) \right] (\widehat{\alpha}(\tau) - \alpha(\tau)),
\end{aligned}
$$

where u_t^* is a middle value. Notice that $\dot{g}_m$ is a regular sequence, $\theta \in (1, 2)$, and $\widehat{\alpha}(\tau) - \alpha(\tau) = O_p(T^{-1/2})$, and thus as $T \to \infty$, both of these two terms $\xrightarrow{P} 0$. Thus

$$
Z_{T2} \xrightarrow{P} 0.
$$

Similarly, $Z_{T4} \xrightarrow{P} 0$.

Next, for

$$Z_{T3} = \frac{\alpha(\tau)}{\tau T^{1/\theta}} \sum_{t=1}^{T} \left[1\left(Y_t \leq \alpha(\tau)\right) - \tau\right].$$

Notice that $\mathrm{E}(Z_{T3}) = 0$, the indicator function is bounded,

$$\begin{aligned}
\mathrm{Var}\left(Z_{T3}\right) &= \mathrm{E}\left[\frac{\alpha(\tau)}{\tau T^{1/\theta}} \sum_{t=1}^{T} \left[1\left(Y_t \leq \alpha(\tau)\right) - \tau\right]\right] \left[\frac{\alpha(\tau)}{\tau T^{1/\theta}} \sum_{s=1}^{T} \left[1\left(Y_s \leq \alpha(\tau)\right) - \tau\right]\right] \\
&= \frac{\alpha(\tau)^2}{\tau^2 T^{2/\theta}} \sum_{t=1}^{T} \sum_{s=1}^{T} \left[\mathrm{E}\left[1\left(Y_t \leq \alpha(\tau)\right) - \tau\right] \left[1\left(Y_s \leq \alpha(\tau)\right) - \tau\right]\right].
\end{aligned}$$

By the mixing condition on Y_t,

$$\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \left[\mathrm{E}\left[1\left(Y_t \leq \alpha(\tau)\right) - \tau\right] \left[1\left(Y_s \leq \alpha(\tau)\right) - \tau\right]\right] < \infty,$$

thus $\mathrm{Var}(Z_{T3}) = o(1)$.

The leading term Z_{T1} is the same as the leading term in Theorem 1, and thus

$$T^{(\theta-1)/\theta} \left[\widehat{ES}_\tau - ES_\tau\right] \Rightarrow \frac{1}{\tau} S.$$

■

8.3 Proof of Theorem 3

We show the result for the unsmoothed version. Notice that

$$\widetilde{ES}_\tau(M, t) = \frac{1}{\tau M} \sum_{s=0}^{M-1} Y_{t+s} \mathcal{K}_h \left(Y_{t+s} - \widetilde{\alpha}_t(\tau)\right).$$

and we approximate the sampling distribution of $T^{(\theta-1)/\theta}\left(\widetilde{ES}_\tau - ES_\tau\right)$, say $\widetilde{F}_T(y)$, by

$$\widetilde{F}_{T,M}(y) = \frac{1}{T - M + 1} \sum_{t=1}^{T-M+1} 1\left(M^{(\widehat{\theta}-1)/\widehat{\theta}} \left[\widetilde{ES}_\tau(M, t) - \widetilde{ES}_\tau\right] \leq y\right).$$

We follow the proof of Theorem 8.3.1 of Politis, Romano and Wolf (1999). For simplicity, we denote $\delta = (\theta - 1)/\theta$, and $\widehat{\delta} = (\widehat{\theta} - 1)/\widehat{\theta}$. First,

$$\begin{aligned} & \widetilde{F}_{T,M}(y) \\ = & \frac{1}{T-M+1} \sum_{t=1}^{T-M+1} 1\left(M^{(\widehat{\theta}-1)/\widehat{\theta}} \left[\widetilde{ES}_\tau(M,t) - \widetilde{ES}_\tau \right] \leq y \right) \\ = & \frac{1}{T-M+1} \sum_{t=1}^{T-M+1} 1\left(M^{\widehat{\delta}} \left[\widetilde{ES}_\tau(M,t) - ES_\tau \right] \leq y + M^{\widehat{\delta}} \left[\widetilde{ES}_\tau - ES_\tau \right] \right). \end{aligned}$$

Notice that, (i) $\widetilde{ES}_\tau \to ES_\tau$ at rate T^δ; (ii) $M^{\widehat{\delta}}/M^\delta \to 1$; (iii) $M/T \to 0$; thus for any $\epsilon > 0$,

$$\Pr\left(\left| M^{\widehat{\delta}} \left[\widetilde{ES}_\tau - ES_\tau \right] \right| \geq \epsilon \right) = \Pr\left(\left| \frac{M^{\widehat{\delta}}}{M^\delta} \frac{M^\delta}{T^\delta} T^\delta \left[\widetilde{ES}_\tau - ES_\tau \right] \right| \geq \epsilon \right) \to 0,$$

and thus, as $T \to \infty$, with probability tending to one,

$$\breve{F}_{T,M}(y - \epsilon) \leq \widetilde{F}_{T,M}(y) \leq \breve{F}_{T,M}(y + \epsilon),$$

where

$$\breve{F}_{T,M}(y) = \frac{1}{T-M+1} \sum_{t=1}^{T-M+1} 1\left(M^{\widehat{\delta}} \left[\widetilde{ES}_\tau(M,t) - ES_\tau \right] \leq y \right).$$

It suffice to show that

$$\breve{F}_{T,M}(y) \xrightarrow{P} F(y).$$

Let

$$F_{T,M}(y) = \frac{1}{T-M+1} \sum_{t=1}^{T-M+1} 1\left(M^{\delta} \left[\widetilde{ES}_\tau(M,t) - ES_\tau \right] \leq y \right) = \frac{1}{T-M+1} \sum_{t=1}^{T-M+1} \xi_t,$$

where

$$\xi_t = 1\left(M^{\delta} \left[\widetilde{ES}_\tau(M,t) - ES_\tau \right] \leq y \right),$$

then ξ_t is strong mixing with the same mixing rate and, notice that $M^\delta \left[\widetilde{ES}_\tau(M,t) - ES_\tau \right] \Longrightarrow \frac{1}{\tau} S$, thus

$$F_{T,M}(y) \xrightarrow{P} F(y).$$

Then, by the same argument as Politis, Romano and Wolf (1999, page 185), we have $\breve{F}_{T,M}(y) \xrightarrow{P} F(y)$. ■

8.4 Proof of Theorem 4

The asymptotic behavior of the nonparametric quantile regression estimates can be analyzed following a similar argument as Hall, Peng, and Yao (2002). Notice that the conditional expected shortfall focus on $X = x$, we only need that, under our assumption, uniformly in a shrinking neighborhood of x, $N_{h_1}(x) = \{x' : \|x' - x\| < h_1\}$

$$\sup_{u \in N_{h_1}(x)} |\widehat{m}(\tau, u) - m(\tau, u)| = O_P\left(h_1^{p+1} + \frac{\log T}{\sqrt{Th_1^d}}\right).$$

The regression quantile problem can be re-written as minimizing $Q_T(\tau, x; \boldsymbol{\theta}) - Q_T(\tau, x; \boldsymbol{\theta}_0)$. Let $Z_t = Z_t(x, h_1)$ be an $N \times 1$ vector that contains the regressors $((X_t - x)/h_1)^{\mathbf{j}}$ in the local-polynomial quantile regression (7) in the lexicographical order, and denote the vector of corresponding parameters as β, and the true parameters by $\beta_{\mathbf{j0}}$, let

$$Y_t - \sum_{0 \le |\mathbf{j}| \le p} \beta_{\mathbf{j0}}((X_t - x)/h_1)^{\mathbf{j}} = Y_{t\tau}^*$$

then

$$Y_t - \sum_{0 \le |\mathbf{j}| \le p} \beta_{\mathbf{j}}((X_t - x)/h_1)^{\mathbf{j}} = Y_{t\tau}^* - (\theta - \theta_{\mathbf{0}})^\top Z_t),$$

and the objective function of the minimization problem can be written as

$$\sum_{t=1}^{T} K\left(\frac{x - X_t}{h_1}\right)\left[\rho_\tau(Y_{t\tau}^* - (\theta - \theta_{\mathbf{0}})^\top Z_t) - \rho_\tau(Y_{t\tau}^*)\right].$$

Define $v = \sqrt{Th_1^d}(\theta - \theta_{\mathbf{0}})$, then $\widehat{v} = \sqrt{Th_1^d}\left(\widehat{\theta} - \theta_{\mathbf{0}}\right)$ is a minimizer of

$$\sum_{t=1}^{T}\left[\rho_\tau\left(Y_{t\tau}^* - \frac{1}{\sqrt{Th_1^d}} v^\top Z_t\right) - \rho_\tau(Y_{t\tau}^*)\right] K\left(\frac{x - X_t}{h_1}\right)$$

Let $\psi_\tau(u) = \tau - 1(u < 0)$, $u_{t\tau} = Y_t - m(\tau, X_t)$, and

$$G_{T,h_1}(v; x) = \sum_{t=1}^{T} K\left(\frac{x - X_t}{h_1}\right)\left[\rho_\tau(Y_{t\tau}^* - \frac{1}{\sqrt{Th_1^d}} v^\top Z_t) - \rho_\tau(Y_{t\tau}^*) + \frac{1}{\sqrt{Th_1^d}} v^\top Z_t \psi_\tau(u_{t\tau})\right]$$

we show that, for any $x' \in N_{h_1}(x) = \{x' : |x' - x| \le h_1\}$, (1) $\mathrm{E}\{G_{T,h_1}(v; x')\} - \frac{1}{2} v^\top H_T(\tau, x') v = O(\frac{1}{T^{1/2} h_1^{d/2}})$; (2) $\Pr\{|G_{T,h_1}(v; x') - \mathrm{E}\{G_{T,h_1}(v; x')\}| > \epsilon\} = O(\frac{1}{T^{1/2} h_1^{d/2}})$, where

$$H_T(\tau, x) \equiv \frac{1}{Th_1^d} \sum_{t=1}^{T} f_{Y_t|X_t}\left(\theta_{\mathbf{0}}^\top Z_t\right) Z_t Z_t' K\left(\frac{x - X_t}{h_1}\right),$$

and then use standard argument we obtain that the above results hold uniformly over $x' \in N_{h_1}(x)$.

For (1), let $Z_t' = Z_t(x', h_1)$ be the vector that contains the regressors $((X_t - x')/h_1)^{\mathbf{j}}$ in the local-polynomial quantile regression around x', and denote

$$Y_t - \sum_{0 \le |\mathbf{j}| \le p} \beta_{\mathbf{j}\mathbf{0}}((X_t - x')/h_1)^{\mathbf{j}} = Y_{t\tau}'^{*},$$

notice that

$$\rho_\tau(u - v) - \rho_\tau(u) = -v\psi_\tau(u) + (u - v)\{1(0 > u > v) - 1(0 < u < v)\},$$

Under Assumption D1, by a direct moment calculation we have,

$$\begin{aligned}
& \mathrm{E}\left\{G_{T,h_1}(v; x')\right\} \\
= & \sum_{t=1}^{T} \mathrm{E}\left[K\left(\frac{x' - X_t}{h_1}\right) f_{Y_t|X_t}\left(\theta_{\mathbf{0}}^\top Z_t'\right) \frac{1}{2Th_1^d}\left\{v^\top Z_t' Z_t'^\top v + O_p\left(\frac{1}{\sqrt{Th_1^d}}\right)\right\}\right] \\
= & \frac{1}{2}\left[v^\top \mathrm{E}\left\{f_{Y_t|X_t}\left(\theta_{\mathbf{0}}^\top Z_t'\right) \frac{1}{h_1^d} K\left(\frac{x' - X_t}{h_1}\right) Z_t' Z_t'^\top\right\} v\right] + O\left(\frac{1}{\sqrt{Th_1^d}}\right) \\
\approx & \frac{1}{2} f_X(x) f_{Y|X}\left(Q_Y(\tau|x)\right) v^\top \mathbb{H} v + O\left(\frac{1}{\sqrt{Th_1^d}}\right)
\end{aligned}$$

since

$$\begin{aligned}
& \frac{1}{h_1^d}\int f_{Y_t|X_t}\left(Q_{Y_t}(\tau|x)\right) K\left(\frac{x - X_t}{h_1}\right)\left(\frac{X_t - x}{h_1}\right)^{\ell} f_X(X_t) dX_t \\
\to & f_X(x) f_{Y|X}\left(Q_Y(\tau|x)\right) \int U^\ell K(U) dU.
\end{aligned}$$

Next we show that (2) $\Pr\left\{|G_{T,h_1}(v; x') - \mathrm{E}\left\{G_{T,h_1}(v; x')\right\}| > \epsilon\right\} = O(\frac{1}{T^{1/2} h_1^{d/2}})$. Let

$$\xi_t(v; x') = K\left(\frac{x' - X_t}{h_1}\right)\left[\rho_\tau(Y_{t\tau}'^{*} - \frac{1}{\sqrt{Th_1^d}} v^\top Z_t') - \rho_\tau(Y_{t\tau}'^{*}) + \frac{1}{\sqrt{Th_1^d}} v^\top Z_t' \psi_\tau(u_{t\tau})\right],$$

First

$$
\begin{aligned}
& \mathrm{E}\left(\xi_t(v;x')^2\right) \\
= & \mathrm{E}\left|K\left(\frac{x'-X_t}{h_1}\right)\left(Y'^{*}_{t\tau}-\frac{1}{\sqrt{Th_1^d}}v^\top Z'_t\right)\right. \\
& \left.\times 1\left(0>Y'^{*}_{t\tau}>\frac{1}{\sqrt{Th_1^d}}v^\top Z'_t\right)-1\left(0<Y'^{*}_{t\tau}<\frac{1}{\sqrt{Th_1^d}}v^\top Z'_t\right)\right|^2 \\
\leq & \frac{1}{Th_1^d}\times \mathrm{E}\left\|K\left(\frac{x'-X_t}{h_1}\right)v^\top Z_t\right\|^2 1\left(|Y'^{*}_{t\tau}|<\frac{1}{\sqrt{Th_1^d}}v^\top Z'_t\right) \\
= & O\left(\frac{1}{T^{3/2}h_1^{d/2}}\right)
\end{aligned}
$$

Thus

$$
\sum_{t=1}^{T}\mathrm{E}\left(\xi_t(v;x')^2\right)=O\left(\frac{1}{T^{1/2}h_1^{d/2}}\right)
$$

and, by a similar moment calculation, using the mixing condition of $\{Y_t, X_t\}$, we have

$$
\sum_{j=1}^{T-1}(T-j)\mathrm{Cov}\left\{\xi_1(v;x'),\xi_{1+j}(v;x')\right\}=O\left(\frac{1}{T^{1/2}h_1^{d/2}}\right)
$$

Thus, by application of the Chebyshev inequality,

$$
G_{T,h_1}(v;x')-\frac{1}{2}v^\top H_T\left(\tau,x'\right)v=O_p\left(\frac{1}{\sqrt{Th_1^d}}\right).
$$

We next consider the estimator for the expected shortfall. Notice that $\rho_\tau\left(Y_t-\widehat{m}\left(\tau,X_t\right)\right)=\left(Y_t-\widehat{m}\left(\tau,X_t\right)\right)\left(\tau-1\left(Y_t-\widehat{m}\left(\tau,X_t\right)<0\right)\right)$, thus

$$
\begin{aligned}
& \widehat{ES_\tau(x)} \\
= & \frac{1}{\tau}\frac{1}{\widehat{f}_X(x)}\frac{1}{Th^d}\sum_{t=1}^{T}K\left(\frac{X_t-x}{h}\right)Y_t 1\left(Y_t-m\left(\tau,X_t\right)<0\right) \\
& +\frac{1}{\tau}\frac{1}{\widehat{f}_X(x)}\frac{1}{Th^d}\sum_{t=1}^{T}K\left(\frac{X_t-x}{h}\right)Y_t\left[1\left(Y_t-\widehat{m}\left(\tau,X_t\right)<0\right)-1\left(Y_t-m\left(\tau,X_t\right)<0\right)\right] \\
& +\frac{1}{\tau}\frac{1}{\widehat{f}_X(x)}\frac{1}{Th^d}\sum_{t=1}^{T}K\left(\frac{X_t-x}{h}\right)\widehat{m}\left(\tau,X_t\right)\left(\tau-1\left(Y_t-\widehat{m}\left(\tau,X_t\right)<0\right)\right)
\end{aligned}
$$

Notice that K has compact support and integrates to one, by a law of large number for the sum of strong mixing process,

$$\widehat{f}_X(x) = \frac{1}{T}\sum_t \frac{1}{h^d} K\left(\frac{X_t - x}{h}\right) \xrightarrow{P} f(x).$$

By an argument similar to Xiao, Linton, Carroll, and Mammen (2003) based on a geometric expansion of $\widehat{f}_X(x)^{-1}$ around $f_X(x)^{-1}$, and notice that, by definition,

$$ES_\tau(x) = E\left(Y_t | Y_t < Q_{Y_t}(\tau|X_t), X_t = x\right),$$

and $Q_{Y_t}(\tau|X_t) = m(\tau, X_t)$, let

$$\eta_t = (Y_t - m(\tau, X_t))\, 1\left(Y_t - m(\tau, X_t) < 0\right) - \mathrm{E}\left[(Y_t - m(\tau, X_t))\, 1\left(Y_t - m(\tau, X_t) < 0\right) | X_t\right],$$

we have

$$\left(Th^d\right)^{(\theta-1)/\theta}\left[\widehat{ES_\tau(x)} - ES_\tau(x)\right] = M_{1T} + M_{2T} + M_{3T} + M_{4T} + o_p(1),$$

where

$$\begin{aligned}
M_{1T} &= \frac{1}{\tau}\frac{1}{f_X(x)}\frac{1}{(Th^d)^{1/\theta}}\sum_{t=1}^{T} K\left(\frac{X_t - x}{h}\right)\eta_t \\
M_{2T} &= \frac{1}{\tau}\frac{1}{f_X(x)}\frac{1}{(Th^d)^{1/\theta}}\sum_{t=1}^{T} K\left(\frac{X_t - x}{h}\right)\left[ES_\tau(X_t) - ES_\tau(x)\right], \\
M_{3T} &= \frac{1}{\tau}\frac{1}{f_X(x)}\frac{1}{(Th^d)^{1/\theta}}\sum_{t=1}^{T} K\left(\frac{X_t - x}{h}\right)\left[Y_t - m(\tau, X_t)\right] \times \\
&\quad \left[1\left(Y_t - \widehat{m}(\tau, X_t) < 0\right) - 1\left(Y_t - m(\tau, X_t) < 0\right)\right], \\
M_{4T} &= \frac{1}{\tau}\frac{1}{f_X(x)}\frac{1}{(Th^d)^{1/\theta}}\sum_{t=1}^{T} K\left(\frac{X_t - x}{h}\right)\left[\left(\widehat{m}(\tau, X_t) - m(\tau, X_t)\right)\left(\tau - 1\left(Y_t - \widehat{m}(\tau, X_t) < 0\right)\right)\right].
\end{aligned}$$

The Theorem can be proved based on an analysis of each of the above terms. For the first term, let $W_t = K\left(\frac{X_t - x}{h}\right)\eta_t$, we analyze the asymptotic behavior of $\left(Th^d\right)^{-1/\theta}\sum_{t=1}^{T} W_t$.

Under our assumptions, $Y_t - m(\tau, X_t)$ is also a strictly stationary stable process with index θ. In addition, $(Y_t - m(\tau, X_t))\, 1\left(Y_t - m(\tau, X_t) < 0\right) \leq 0$, and notice that the kernel function $K(\cdot)$ has a bounded support, so that W_t has a positive upper bound but no lower bound. It follows that

$$\lim_{x\to\infty}\frac{\Pr\left[W_t > x\right]}{\Pr\left[|W_t| > x\right]} = p_1 = 0 \quad ; \quad \lim_{x\to\infty}\frac{\Pr\left[W_t < -x\right]}{\Pr\left[|W_t| > x\right]} = q_1 = 1.$$

Let $c_T = \left(Th^d\right)^{1/\theta}$, the limit of $c_T^{-1} \sum_{t=1}^{T} (W_t - b_T)$ can be obtained by an argument similar to those of Theorems 2 and 3 of Davis (1983) and Theorem 2 of Hall, Peng, and Yao (2002) (also see LaPage, Woodroofe and Zinn (1981)). To apply their methods, we need to verify conditions (D) and (D') of Davis (1983), i.e. we need to verify:

(D) For all $x > 0$,

$$\lim_{k\to\infty} \left[\limsup_{T\to\infty} S_{k,T}(x)\right] = 0$$

where

$$S_{k,T}(x) = T \sum_{j=2}^{[T/k]} \{p_1(j) + p_2(j) + p_3(j) + p_4(j)\}$$

and

$$\begin{aligned} p_1(j) &= \Pr(W_1 > c_T x, W_j > c_T x) \\ p_2(j) &= \Pr(W_1 > c_T x, W_j \leq -c_T x) \\ p_3(j) &= \Pr(W_1 \leq -c_T x, W_j > c_T x) \\ p_4(j) &= \Pr(W_1 \leq -c_T x, W_j \leq -c_T x) \end{aligned}$$

(D')

$$\lim_{\varepsilon\to 0} \limsup_{T\to\infty} \frac{T}{c_T^2} \sum_{j=2}^{T} \max\left(0, \operatorname{Cov}\left(W_1 1(|W_1| < \varepsilon c_T), W_j 1(|W_j| < \varepsilon c_T)\right)\right)$$

To prove (D), notice that

$$\Pr\left(\left|K\left(\frac{X_t - x}{h}\right)\eta_t\right| \geq c_T\right) \leq 1/T,$$

and η_t is stable with index θ, thus

$$\lim_{n\to\infty} T \Pr\left(\left|K\left(\frac{X_t - x}{h}\right)\eta_t\right| \geq c_T z\right) = z^{-\theta}.$$

By mixing property,

$$\Pr(W_1 > c_T x, W_j > c_T x) \leq \Pr(W_1 > c_T x) \Pr(W_j > c_T x) (1 + \alpha(j))$$

thus, there exists $C > 0$,

$$\begin{aligned}
S_{k,T}(z) &= T\sum_{j=2}^{[T/k]}\{p_1(j)+p_2(j)+p_3(j)+p_4(j)\} \\
&\leq 4T\sum_{j=2}^{[T/k]}\left[\Pr\left(W_1 > c_T z\right)\right]^2(1+\alpha(j)) \leq 4T^{-1}\sum_{j=2}^{[T/k]}\left[n\Pr\left(W_1 > c_T z\right)\right]^2(1+\alpha(j)) \\
&\leq 4T^{-1}z^{-2\theta}\sum_{j=2}^{[T/k]}(1+\alpha(j)) = O\left(\frac{T/k}{T}\right) = O\left(\frac{1}{k}\right) \to 0, \text{ as } k\to\infty.
\end{aligned}$$

For condition (D'), notice that

$$\begin{aligned}
&\operatorname{Cov}\left(W_1 1(|W_1| < \varepsilon c_T), W_j 1(|W_j| < \varepsilon c_T)\right) \\
\leq\ &\alpha(j)\mathrm{E}\left(W_1^2 1(|W_1| < \varepsilon c_T)\right)
\end{aligned}$$

and, by Hall, Peng and Yao (2002) and the regular variation property of the process, we have

$$\frac{c_T^2\Pr(|W_1| \geq \varepsilon c_T)}{\mathrm{E}\left(W_1^2 1(|W_1| < \varepsilon c_T)\right)} \to \frac{\theta}{(2-\theta)\,\varepsilon^2}$$

Thus

$$\frac{T}{c_T^2}\mathrm{E}\left(W_1^2 1(|W_1| < \varepsilon c_T)\right) \to \frac{(2-\theta)}{\theta}\varepsilon^{2-\theta}$$

and, notice that $\sum_{j=2}^{T}\alpha(j) < \infty$, we have

$$\lim_{\varepsilon\to 0}\lim\sup_{T\to\infty}\frac{T}{c_T^2}\sum_{j=2}^{T}\max\left(0, \operatorname{Cov}\left(W_1 1(|W_1| < \varepsilon c_T), W_j 1(|W_j| < \varepsilon c_T)\right)\right) = 0.$$

Then, by Davis (1983),

$$c_T^{-1}\sum_{t=1}^{T}\left[K\left(\frac{X_t - x}{h}\right)\eta_t - d_T\right] \to S^*$$

where

$$S^* = \sum_{j=1}^{\infty}\left(\delta_j Z_j - (p_1 - q_1)\mathrm{E}\left(Z_j 1\left(0 < Z_j \leq 1\right)\right)\right).$$

δ_i are iid r.v. with $\Pr(\delta_i = 1) = p_1$, $\Pr(\delta_i = -1) = q_1$, and

$$Z_i = \Gamma_i^{-1/\alpha},\ \Gamma_k = V_1 + \ldots + V_k$$

where $\{V_i\}$ be a sequence of independent exponential random variables with unit mean. Notice that $p_1 = 0, q_1 = 1$, $\Pr(\delta_i = 1) = 0$, $\Pr(\delta_i = -1) = 1$,

$$S^* = \sum_{j=1}^{\infty} \left(\delta_j Z_j + \mathrm{E}\left(Z_j 1\left(0 < Z_j \leq 1\right)\right)\right) = \sum_{j=1}^{\infty} \left(\mathrm{E}\left(Z_j 1\left(0 < Z_j \leq 1\right)\right) - Z_j\right).$$

Thus

$$M_{1T} - \frac{T}{\left(Th^d\right)^{1/\theta}} \frac{b_T}{\tau f_X(x)} = \frac{1}{\tau} \frac{1}{f_X(x)} \frac{1}{\left(Th^d\right)^{1/\theta}} \sum_{t=1}^{T} (W_t - b_T) \to \frac{1}{\tau} \frac{1}{f_X(x)} S^*$$

or,

$$\tau f_X(x) M_{1T} - \frac{T}{\left(Th^d\right)^{1/\theta}} b_T \to S^*.$$

For the second term M_{2T}, under Assumption D, by a Taylor expansion of $ES_\tau(X_t)$ around x, we have

$$\begin{aligned} M_{2T} &= \frac{1}{\tau} \frac{1}{f_X(x)} \frac{1}{\left(Th^d\right)^{1/\theta}} \sum_{t=1}^{T} K\left(\frac{X_t - x}{h}\right) \left[ES_\tau(X_t) - ES_\tau(x)\right] \\ &= \left(Th^d\right)^{(\theta-1)/\theta} h^2 B(\tau, x)(1 + o_p(1)) \end{aligned}$$

where

$$B(\tau, x) = \frac{1}{\tau} \mu_2(K) \left\{ \frac{f_X^{(1)}(x)}{f_X(x)} ES_{\tau,X}^{(1)}(x) + ES_{\tau,X}^{(2)}(x) \right\}$$

We next look at the third term M_{3T}. Let

$$\xi_t = K\left(\frac{X_t - x}{h}\right) \left[Y_t - m\left(\tau, X_t\right)\right] \left[1\left(Y_t < \widehat{m}\left(\tau, X_t\right)\right) - 1\left(Y_t < m\left(\tau, X_t\right)\right)\right]$$

we verify that

$$\frac{1}{\left(Th^d\right)^{1/\theta}} \sum_{t=1}^{T} \xi_t = o_p(1).$$

Define

$$\begin{aligned} \xi_{1t} &= K\left(\frac{X_t - x}{h}\right) \left[Y_t - m\left(\tau, X_t\right)\right] \left[1\left(m\left(\tau, X_t\right) \leq Y_t < \widehat{m}\left(\tau, X_t\right)\right)\right] 1\left(m\left(\tau, X_t\right) < \widehat{m}\left(\tau, X_t\right)\right) \\ \xi_{2t} &= K\left(\frac{X_t - x}{h}\right) \left[Y_t - m\left(\tau, X_t\right)\right] \left[1\left(\widehat{m}\left(\tau, X_t\right) \leq Y_t < m\left(\tau, X_t\right)\right)\right] 1\left(m\left(\tau, X_t\right) > \widehat{m}\left(\tau, X_t\right)\right) \end{aligned}$$

then $\xi_t = \xi_{1t} - \xi_{2t}$. It suffices that we show

$$\frac{1}{\left(Th^d\right)^{1/\theta}} \sum_{t=1}^{T} \xi_{1t} = o_p(1), \quad \frac{1}{\left(Th^d\right)^{1/\theta}} \sum_{t=1}^{T} \xi_{2t} = o_p(1)$$

$$\begin{aligned}
& \left|\frac{1}{\left(Th^d\right)^{1/\theta}}\sum_{t=1}^{T}\xi_{1t}\right| \\
= & \frac{1}{\left(Th^d\right)^{1/\theta}}\sum_{t=1}^{T}K\left(\frac{X_t-x}{h}\right) \\
& \times\left|Y_t-m\left(\tau,X_t\right)\right|\left[1\left(m\left(\tau,X_t\right)\leq Y_t<\widehat{m}\left(\tau,X_t\right)\right)\right]1\left(m\left(\tau,X_t\right)<\widehat{m}\left(\tau,X_t\right)\right) \\
\leq & \frac{1}{\left(Th^d\right)^{1/\theta}}\sum_{t=1}^{T}K\left(\frac{X_t-x}{h}\right)\left|\widehat{m}\left(\tau,X_t\right)-m\left(\tau,X_t\right)\right| \\
& \times 1\left(m\left(\tau,X_t\right)\leq Y_t<\widehat{m}\left(\tau,X_t\right)\right)1\left(m\left(\tau,X_t\right)<\widehat{m}\left(\tau,X_t\right)\right) \\
\leq & \frac{1}{\left(Th^d\right)^{1/\theta}}\sum_{t=1}^{T}K\left(\frac{X_t-x}{h}\right)\left|\widehat{m}\left(\tau,X_t\right)-m\left(\tau,X_t\right)\right| \\
\leq & \sup_{X_t\in N_\delta(x)}\left|\widehat{m}\left(\tau,X_t\right)-m\left(\tau,X_t\right)\right|\frac{1}{\left(Th^d\right)^{1/\theta}}\sum_{t=1}^{T}\left|K\left(\frac{X_t-x}{h}\right)\right|
\end{aligned}$$

As $T\to\infty$, $h\to 0$,

$$\left|\frac{1}{\left(Th^d\right)^{1/\theta}}\sum_{t=1}^{T}\xi_{1t}\right|\leq\sup_{X_t\in N_\delta(x)}\left|\widehat{m}\left(\tau,X_t\right)-m\left(\tau,X_t\right)\right|\frac{1}{\left(Th^d\right)^{1/\theta}}\sum_{t=1}^{T}\left|K\left(\frac{X_t-x}{h}\right)\right|$$

notice that

$$\sup_{X_t\in N_\delta(x)}\left|\widehat{m}\left(\tau,X_t\right)-m_0\left(\tau,X_t\right)\right|=O_P\left(h_1^{p+1}+\frac{\log T}{\sqrt{Th_1^d}}\right)$$

Thus

$$\begin{aligned}
\left|\frac{1}{\left(Th^d\right)^{1/\theta}}\sum_{t=1}^{T}\xi_{1t}\right| & \leq O_P\left(h_1^{p+1}+\frac{1}{\sqrt{Th_1^d}}\right)\frac{1}{\left(Th^d\right)^{1/\theta}}\sum_{t=1}^{T}\left|K\left(\frac{X_t-x}{h}\right)\right| \\
& = O_P\left(\left(h_1^{p+1}+\frac{\log T}{\sqrt{Th_1^d}}\right)\left(Th^d\right)^{1-1/\theta}\right) \\
& = o_P(1)
\end{aligned}$$

under Assumption D6. The analysis of $\frac{1}{\left(Th^d\right)^{1/\theta}}\sum_{t=1}^{T}\xi_{2t}$ is similar.

For the last term M_{4T},

$$M_{4T}=\frac{1}{\tau}\frac{1}{f_X(x)}\frac{1}{\left(Th^d\right)^{1/\theta}}\sum_{t=1}^{T}K\left(\frac{X_t-x}{h}\right)\left[\left(\widehat{m}\left(\tau,X_t\right)-m\left(\tau,X_t\right)\right)\left(\tau-1\left(Y_t-\widehat{m}\left(\tau,X_t\right)<0\right)\right)\right]$$

notice that

$$\left|\tau-1\left(Y_t-\widehat{m}\left(\tau,X_t\right)<0\right)\right|<1$$

$$|M_{4T}| \leq \frac{1}{\tau} \frac{1}{f_X(x)} \frac{1}{(Th^d)^{1/\theta}} \sum_{t=1}^{T} \left| K\left(\frac{X_t - x}{h}\right) \right| \left| \widehat{m}(\tau, X_t) - m(\tau, X_t) \right|$$

which can be analyzed similarly to the previous terms. ■

References

[1] Acharya, V., L. Pedersen, T. Philippe, and M. Richardson (2010), Measuring systemic risk, working paper, NYU.

[2] Adrian, T., and M. Brunnermeier (2010), CoVaR, Federal Reseve Bank of New York.

[3] Artzner, P.F., J. Delbaenm, M. Eber, and D. Heath (1999), Coherent measures of risk. Mathematical Finance 9, 203-228.

[4] Bartkiewicz, K., A. Jakubowski, T. Mikosch, and O. Wintenberger (2010), Stable limits for sums of dependent infinite variance random variables. Probability Theory and Related Fields, 150, 337 - 372.

[5] Brownlees C.T. and R. Engle (2010), Volatility, Correlation and Tails for Systemic Risk Measurement, working paper, NYU.

[6] Cai, Z. and X. Wang (2008), Nonparametric estimation of conditional VaR and expected shortfall, Journal of Econometrics, V147, p120-130.

[7] Campbell, J., Lo, A., and A. C. Mackinlay, 1997, The ecetrics of financial markets, Princeton University Press, Princeton.

[8] Chen, S.X. (2008), Nonparametric estimation of expected shortfall. Journal of Financial Econometrics 1, 87-107.

[9] Danielsson, J. (2011), *Financial Risk Forecasting*. Wiley, London.

[10] Davis, R. (1983), Stable Limits for Partial sums of Dependent Random Variables., The Annals of Probability, V11, 262-269.

[11] Davis, R.A. and T. Hsing (1995), Point Process and Partial Sum Convergence for Weakly Dependent Random Variables with Infinite Variance. Annals of Probability 23, 879-917.

[12] Davis, R. and T. Mikosch (1998), The Sample Autocorrelations of Heavy-Tailed Processes with Applications to Arch, Annals of Statistics, 26, 2046-2080.

[13] Davis, R. and T. Mikosch (2001), Point Process Convergence of Stochastic Volatility Processes with Application to Sample Autocorrelation, Journal of Applied Probability, Vol. 38, Probability, Statistics and Seismology (2001),pp. 93-104.

[14] Einmahl, John H. J., Laurens de Haan, and Vladimir I. Piterbarg (2001), Nonparametric Estimation of the Spectral Measure of an Extreme Value Distribution. The Annals of Statistics, Vol. 29, 1401-1423.

[15] Embrechts, P., C. Kluppelberg and T. Mikosch (1997), Modelling extremal events for insurance and finance, Springer, New York.

[16] Fama, E. (1965) "The Behavior of Stock Market Prices". Journal of Business 38 (1): 34–105.

[17] Fama, E. (1965). Portfolio Analysis in a Stable Paretian Market. Mngmt Sci. 11, 404-419.

[18] Fermanian, JD, O Scaillet, Sensitivity analysis of VaR and Expected Shortfall for portfolios under netting agreements, Journal of Banking & Finance (2005), V29, 927-958.

[19] Gabaix, X. (2008), "Power Laws", New Palgrave Dictionary of Economics, 2nd Edition.

[20] Gabaix, X., and R. Ibragimov (2011), Rank-1/2: A simple way to improve the OLS estimation of tail exponents. Journal of Business and Economic Statistics 29, 24-39.

[21] Gopikrishnan, Parameswaran Vasiliki Plerou, Xavier Gabaix, and H. Eugene Stanley (2000). "Statistical Properties of Share Volume Traded in Financial Markets." Physical Review E, 62(4), R4493-R4496.

[22] Hall, P., L. Peng, and Q. Yao (2002), Prediction and Nonparametric Estimation for Time Series with Heavy-Tails. Journal of Time Series Analysis 23, 313-331.

[23] Hart, J.D. (1991), Kernel regression estimation with time series errors. *Journal of the Royal Statistical Society, Series B.* 53, 173-187.

[24] Hill, J.B. (2010), On tail index estimation for dependent heterogeneous data. Forthcoming in Econometric Theory.

[25] Ibragimov, R. (2009a), "Heavy-tailed densities." in The New Palgrave Dictionary of Economics. Eds. Steven N. Durlauf and Lawrence E. Blume. Palgrave Macmillan.

[26] Ibragimov, R. (2009b), "Portfolio diversification and Value at Risk under thick-tailedness. Quantitative Finance 9, 565-580.

[27] Ibragimov, R., D. Jaffee and J. Walden (2009), Nondiversification traps in catastrophe insurance markets, Review of Financial Studies, 22, 959-993.

[28] Ibragimov, M., R. Ibragimov, and P. Kattuman (2010), Emerging markets and heavy tails. Harvard university.

[29] Ibragimov, R., UK Müller (2010), t-Statistic based correlation and heterogeneity robust inference, Journal of Business and Economic Statistics, 28, 453-468.

[30] Kallenberg, O. (1983), *Random Measures*. 3rd edition. Akademie-Verlag, Berlin.

[31] LePage, R., M. Woodroofe, J. Zinn (1981), Convergence to a stable distribution via order statistics, The Annals of Probability, V9, 624-632.

[32] Li., D, Z. Lu, and O. Linton (2010), Local Linear Fitting under Near-Epoch Dependence: Uniform Consistency with Convergence Rates.

[33] Mandelbrot, B. (1963) The Variation of Certain Speculative Prices,"J . Business 3 6, 394-411

[34] Masry, E., (1996), "Multivariate regression estimation: Local polynomial fitting for time series. *Stochastic Processes and their Applications* 65, 81-101.

[35] McCulloch, H. (1986) "Simple Consistent Estimators of Stable Distribution Parameters," Communications in Statistics - Simulation and Computation 15: 1109-36.

[36] Mikosch, T., and C. Starica (2000), Limit Theory for the sample autocorrelations and extremes of a GARCH(1,1) process, Annals of Statistics, 28, 1427-1451.

[37] Mikosch, T., and D. Straumann (2006), Stable limits of martingale transforms with application to the estimation of GARCH parameters. Annals of Statistics 34, 493-522.

[38] Pan, J., O. Linton, and H. Wang (2010), Estimation for a non-stationary semi-strong GARCH(1,1) model with heavy tailed errors *Econometric Theory* 26, 1-28.

[39] Peng, L., and Q. Yao (2004), Nonparametric regression under dependent errors with infinite variance. Ann. Inst. Statist. Math 1, 73-86.

[40] Politis, D. N., J. P. Romano and M. Wolf (1999), *Subsampling.* Springer-Verlag, New York.

[41] Phillips, P., and V. Solo (1992), Asymptotics for linear processes, Annals of Statistics, 971-1001.

[42] Press, S.J. (1982). Applied Multivairate Analysis: Using Bayesian and Frequentist Methods of Inference. 2nd edition Krieger, Malabar, FL.

[43] Rachev, S. and S. Mittnik (2000), *Stable Paretian Models in Finance.* Wiley, New York.

[44] Samuelson, P.A. (1967). Efficient portfolio selection for pareto levy instruments. 2, J, Finance Quant Anal. 102-127

[45] Scaillet, O. (2004), Nonparametric estimation and sensitivity analysis of expected shortfall. Mathematical Finance 14, 115-129.

[46] Scaillet, O (2005), Nonparametric estimation of conditional expected shortfall, Insurance and Risk Management Journal.

[47] Stock, J. and M. Watson, 2007, Introduction to Econometrics, 2nd Ed., Pearson Addison Wesley.

[48] Taleb, N.N. (2010) The Black Swan, Second Edition, Penguin.

[49] Weron, Rafal, Heavy-Tails and Regime-Switching in Electricity Prices (May 1, 2008). Mathematical Methods of Operations Research, Vol. 69, No. 3, pp. 457-473. Available at SSRN: http://ssrn.com/abstract=1140166.

[50] Withers, CS., (1981), Central limit theorems for dependent variables. I, Probability Theory and Related Fields, Springer.

[51] Xiao, Z., O. Linton, R. Carroll, and E. Mammen (2003), More Efficient Local Polynomial Estimation in Nonparametric Regression with Autocorrelated Errors. *Journal of the American Statistical Association* 98, 980-992.