

# Nonparametric Estimation of Expected Shortfall<sup>1</sup>

SONG XI CHEN

*Io St te niversity*

**ABSTRACT.** The expected shortfall is an increasingly popular risk measure in financial risk management and possesses the desired sub-additivity property, which is lacking for the Value at Risk (VaR). We consider two nonparametric expected shortfall estimators for dependent financial losses. One is a sample average of excessive losses larger than a VaR. The other is a kernel smoothed version of the first estimator (Scaillet, 2004 *M the tic Finance*), hoping that more accurate estimation can be achieved by smoothing. Our analysis reveals that the extra kernel smoothing does not produce more accurate estimation of the shortfall. This is different from the expectation of the shortfall.

## 1. INTRODUCTION

The expected shortfall (ES) and the value at risk (VaR) are popular measures of financial risks for an asset or a portfolio of assets. Artzner, Delbaen, Eber and Heath (1999) show that VaR lacks the sub-additivity property in general and hence is not a coherent risk measure. In contrast, ES is coherent (Föllmer and Schied, 2001) and has become a more attractive alternative in financial risk management.

Let  $\{X_t\}_{t=1}^n$  be the market values of an asset or a portfolio of assets over  $n$  periods of a time unit. Let  $Y_t = -\log(X_{it}/X_{it-1})$  be the negative log return (log loss) over the  $t$ -th period. Suppose  $\{Y_t\}_{j=1}^n$  is a dependent stationary process with the marginal distribution function  $F$ . Given a positive value  $p$  close to zero, the VaR at a confidence level  $1 - p$  is

$$\nu_p = \inf\{\boldsymbol{y} : F(\boldsymbol{y}) \geq 1 - p\} \quad (1)$$

which is the  $(1-p)$ -th quantile of the lose distribution  $F$ . The VaR specifies a level of

a nonparametric kernel estimator and applies it for sensitivity analysis in the context of portfolio allocation.

An advantage of the nonparametric method is its being model-free and hence is model robust and avoids bias caused by using a mis-specified loss distribution. Financial risk management is primarily concerned with characteristics of a tail part of the loss distribution. However, data are generally sparse in the tail and hence proposing a proper parametric loss model which is adequate for the tail part is not trivial. This is where the nonparametric method can play a significant role. Another advantage of the nonparametric approach is that it allows a wide range of data dependence, which makes it adaptable in the context of financial losses. The nonparametric estimators considered in this paper can accommodate data dependence explicitly since the effect of the dependence on the variance of the ES estimation can be clearly spelt out in the variance formula. This is different from the extreme value approach as the latter effectively treats high exceedances as independent and identically distribution observations which is true asymptotically under the so-called  $D$  and  $D'$  conditions (Leadbetter, Lindgren and Rootzen, 1983). An empirical study by Bellini and Figa-Talamanca (2002), by carrying out a nonparametric runs test, has shown that financial returns can exhibit strong tail dependence even for large threshold levels. This indicates the need for considering the dependence in financial returns directly, which is the approach taken by the nonparametric estimators considered in this paper.

In this paper, we evaluate two nonparametric ES estimators. One is based on a weighted sample average of excessive losses defined by a VaR estimator  $\hat{v}_p$  based on an order statistic. Another is the kernel estimator proposed in Scaillet (2004) which employs kernel smoothing in both the initial VaR estimation and the final averaging of the excessive losses. We hope that the kernel smoothing would make the estimation more accurate as observed for the case of VaR estimation by Chen and Tang (2005). However, our analysis reveals that the variance and the mean square error of the kernel estimator is not necessarily smaller than that of a sample average estimator. As a result, for estimation of the ES, the simpler sample average of excessive losses is good enough and there is no need to carry out kernel smoothing. This may be surprising considering that kernel smoothing leads to smaller variance in quantile estimation for both independent (Sheather and Marron, 1990) and dependent (Chen and

Tang, 2005) observations. The underlying reason for obtaining these different effects of kernel smoothing lays on the fact that the ES is effectively a mean parameter, which can be estimated rather accurately by simple averaging.

The paper is structured as follows. We introduce the two nonparametric ES estimators in Section 2. Their statistical properties are discussed in Section 3. Section 4 reports simulation results, which is followed by an empirical study on two financial series in Section 5. All the technical details are given in the appendix.

## 2. NONPARAMETRIC ESTIMATORS

A simple nonparametric estimator of the ES is a weighted average of excessive losses larger than  $\hat{d}_p$  where  $\hat{d}_p = Y_{([n(1-p)]+1)}$  is the sample VaR (quantile) estimator of  $\nu_p$  and  $Y_{(r)}$  is the  $r$ -th order statistic of  $\{Y_t\}_{t=1}^n$ . Hence, it is

$$\hat{\mu}_p = \frac{\sum_{t=1}^n Y_t I(Y_t \geq \hat{d}_p)}{\sum_{t=1}^n I(Y_t \geq \hat{d}_p)} = ([np] + 1)^{-1} \sum_{t=1}^n Y_t I(Y_t \geq \hat{d}_p) \quad (3)$$

where  $I(\cdot)$  is the indicator function.

The kernel estimator proposed by Scaillet (2004) is the following. Let  $K$  be a kernel function, which is a symmetric probability density function, and  $G(t) = \int_t^\infty K(y) dy$  and  $G_h(t) = G(t/h)$  where  $h$  is a positive smoothing bandwidth. The kernel estimator of the survival function  $S(x) = 1 - F(x)$  is

$$S_h(z) = n^{-1} \sum_{t=1}^n G_h(z - Y_t) \quad (4)$$

A kernel estimator of  $\nu_p$ , denoted as  $\hat{d}_{p,h}$ , is the solution of  $S_h(z) = p$ , as proposed in Gourioux, Laurent and Scaillet (2000).<sup>2</sup> By replacing the indicator function and  $\hat{d}_p$  with the smoother  $G_h$  and  $\hat{d}_{p,h}$  respectively in (3) Scaillet (2004) proposed the following kernel estimator

$$\hat{\mu}_{p,h} = ([np])^{-1} \sum_{t=1}^n Y_t G_h(\hat{d}_{p,h} - Y_t). \quad (5)$$

---

<sup>2</sup>Its statistical properties and how to obtain the standard errors are considered in Chen and Tang (2005). See also Cai (2002) and Fan and Gu (2003) for kernel estimators.

Based on the improvement of the kernel VaR estimator  $\hat{d}_{p,h}$  over  $\hat{d}_p$ , it is expected that the kernel ES estimator  $\hat{e}_{p,h}$  would improve the estimation accuracy of the unsmoothed estimator  $\hat{e}_p$ . Confirming this or otherwise is the focus of the next section.

### 3. MAIN RESULTS

The properties of these two nonparametric ES estimators are evaluated in this section. We start with some conditions.

Let  $\mathcal{F}_k^l$  be the  $\sigma$ -algebra of events generated by  $\{Y_t, k \leq t \leq l\}$  for  $l > k$ . The  $\alpha$ -mixing coefficient introduced by Rosenblatt (1956) is

$$\alpha(k) = \sup_{A \in \mathcal{F}_1^k, B \in \mathcal{F}_{k+1}^\infty} |P(AB) - P(A)P(B)|.$$

The series is said to be  $\alpha$ -mixing if  $\lim_{k \rightarrow \infty} \alpha(k) = 0$ . The dependence described by the  $\alpha$ -mixing is the weakest as it is implied by other types of mixing; see Doukhan (1994) for comprehensive discussions. The following conditions are assumed in our study.

- (i) There exists a  $\rho \in (0, 1)$  such that  $\alpha(k) \leq C\rho^k$  for all  $k \geq 1$  and a positive constant  $C$ .
- (ii) The distribution  $F$  of  $Y_t$  is absolutely continuous with probability density  $f$  which has continuous second derivatives in  $\mathcal{B}(\nu_p)$ , a neighborhood of  $\nu_p$ ; the joint distribution functions of  $(Y_1, Y_{k+1})$   $F_k$  for  $k \geq 1$  have all its second partial derivatives bounded in  $\mathcal{B}(\nu_p)$ ;  $E(|Y_t|^{2+\delta}) \leq C$  for some  $\delta > 0$  and a positive constant  $C$ .
- (iii)  $K$  is a symmetric probability density satisfying the moment conditions  $\int_{-1}^1 \boldsymbol{\mu} K(\boldsymbol{\mu}) d\boldsymbol{\mu} = \mathbf{0}$  and  $\int_{-1}^1 \boldsymbol{\mu}^2 K(\boldsymbol{\mu}) d\boldsymbol{\mu} = \sigma_K^2 > 0$ , and  $K$  has bounded and Lipschitz continuous derivative.
- (iv)  $\int_{-1}^1 \boldsymbol{\mu}^3 K(\boldsymbol{\mu}) d\boldsymbol{\mu} \rightarrow 0$ ,  $n^{-3-\beta} \rightarrow \infty$  for any  $\beta > 0$  and  $n^{-4} \log^2(n) \rightarrow 0$  as  $n \rightarrow \infty$ .

Condition (i) means that the time series is geometric  $\alpha$ -mixing, which is satisfied by many commonly used financial time series which include the ARMA, ARCH, the stochastic volatility and diffusion models. For instance Masry and Tjøstheim (1995) established the  $\alpha$ -mixing for ARCH model; and Genon-Catalot, Jeantheau and Laredo (2000) for diffusion models. Condition (ii) contains standard conditions which requires underlying smoothness for the marginal and pair-wise joint densities together with finite moments for the absolute returns. Conditions (iii) and (iv) are extra ones required by the kernel estimator. While

Condition (iii) has the usual requirements on the kernel, Condition (iv) specifies a range for the bandwidth which includes  $O(n^{-1/3})$ , the optimal order for estimating VaR estimation. These conditions are comparable with conditions imposed by other authors. For instance, Conditions (i), (ii), (iii) and (iv) correspond to Assumption 3.2 (a), (b) and (f), Assumption 3.1 (a) and (c), and (b) of Scaillet (2004) respectively.

Let  $\gamma(k) = Cov\{(Y_1 - \nu_p)I(Y_1 \geq \nu_p), (Y_{k+1} - \nu_p)I(Y_{k+1} \geq \nu_p)\}$  for positive integers  $k$  and

$$\sigma_0^2(p; n) = \{Var\{(Y_1 - \nu_p)I(Y_1 \geq \nu_p)\} + 2 \sum_{k=1}^{n-1} \gamma(k)\}.$$

Assumption (i) and the Davydov inequality imply that  $\sigma_0^2(p, n)$  is finite for each  $n$  and is converging as  $n \rightarrow \infty$ .

We start with evaluating the unsmoothed estimator  $\hat{\mu}_p$  to provide a point of reference for the kernel estimators. Derivation given in the appendix shows that under conditions (i) and (ii), and for an arbitrary positive  $\kappa$ ,

$$\hat{\mu}_p - \mu_p = p^{-1} \left\{ n^{-1} \sum_{i=1}^n (Y_i - \nu_p) I(Y_i \geq \nu_p) - p(\mu_p - \nu_p) \right\} + o_p(n^{-3/4+\kappa}). \quad (6)$$

This is a Bahadur type expansion which leads to the following theorem regarding the asymptotic normality of  $\hat{\mu}_p$ .

**THEOREM 1.** Under conditions (i)- (ii), as  $n \rightarrow \infty$

$$\sqrt{pn} \sigma_0^{-1}(p; n) (\hat{\mu}_p - \mu_p) \xrightarrow{d} N(\mathbf{0}, \mathbf{1}). \quad (7)$$

The theorem indicates that the asymptotic variance of  $\hat{\mu}_p$  is  $\sigma_0^2(p; n)/(np)$ , which is the variance of  $p^{-1} \{ n^{-1} \sum_{i=1}^n (Y_i - \nu_p) I(Y_i \geq \nu_p) - p(\mu_p - \nu_p) \}$ , the leading order term in the expansion (6). The dependence in the original time series is reflected in the asymptotic variance through the covariance terms in  $\sigma_0^2(p; n)$ . This means that we need to accommodate the dependence in further statistical inference (interval estimation and hypothesis testing) for the shortfall estimation. We note also that the effective sample size for the ES estimation is only  $np$ . As  $p$  is small ranging between 1% and 5% as commonly used in financial risk management, the ES estimator is subject to high volatility which is a common challenge for statistical inference of risk measures.

The following theorem summarizes the properties of the kernel estimator (5).

**THEOREM 2.** Under conditions (i)- (iv), as  $n \rightarrow \infty$

$$\sqrt{pn}\sigma_0^{-1}(p; n)(\hat{\mu}_{p,h} - \mu_p) \xrightarrow{d} N(\mathbf{0}, \mathbf{1}) \quad (8)$$

and furthermore,

$$Bias(\hat{\mu}_{p,h}) = -\frac{1}{2}p^{-1}\sigma_K^2 f(\nu_p) + o(p^{-2}) \quad \text{and} \quad (9)$$

$$Var(\hat{\mu}_{p,h}) = p^{-1}n^{-1}\sigma_0^2(p; n) + o(n^{-1}). \quad (10)$$

By comparing with Theorem 1, it is found that the kernel estimator has the same asymptotic normal distribution as the unsmoothed sample estimator  $\hat{\mu}_p$ . This is similar to the corresponding results for VaR estimation as reported in Chen and Tang (2005). We also note that both  $\hat{\mu}_p$  and  $\hat{\mu}_{p,h}$  converges to  $\mu_p$  at the rate of  $\sqrt{n}$  or more precisely at the rate of  $\sqrt{np}$ ; whereas the VaR estimators  $\hat{\nu}_p$  and  $\hat{\nu}_{p,h}$  converge to  $\nu_p$  at the rate of  $\sqrt{n}$  or more precisely at the rate of  $\sqrt{n}f(\nu_p)$  where  $f$  is the probability density of  $Y_t$ .

The second part of the theorem conveys a different story from VaR estimation. First of all, unlike the VaR estimation, the kernel estimator does not offer a variance reduction at the second order of  $n^{-1}$  as the second order term vanishes. At the same time, the smoothing brings in an bias which will lead to an overall increase in the mean square error. Therefore, for the purpose of estimating the ES, the kernel smoothing is counter-productive. The underlying reason is the fact that the ES is effectively a mean parameter, which can be estimated rather accurately without smoothing. The situation is similar to nonparametric estimation of the mean parameter, which can be well estimated by the sample mean.

It should be noted that the above statement is only applicable for point estimation of ES. For constructing confidence intervals and testing hypothesis on  $\mu_p$  in the presence of data dependence, the kernel smoothing can play a significant role in estimating  $\sigma_0^2(p; n)$  via the spectral density estimation approach.

#### 4. SIMULATION STUDY

In this section we report results from a simulation study which evaluate the performance of the nonparametric ES estimators. The main objective is to confirm our theoretical findings in the previous section.

The models chosen for the log loss  $Y_t$  in the simulation are

an

via a kernel estimation of the spectral density of

$p = n/(2q)$  and  $\nu^2(q) = \max_{0 \leq j \leq 2q-1} E \left( \sum_{l=[jp]+1}^{[(j+1)p]} X_l \right)^2$ . From an equality given in Yokoyama (1980),  $\nu^2(q) \leq Cp$ . Apply Theorem 1.3 in Bosq (1998) for  $\alpha$ -mixing sequences,

$$\begin{aligned} & P\{|F_n(\nu_p + \epsilon_n) - F(\nu_p + \epsilon_n)| > C_1 \epsilon_n\} \\ & \leq 4 \exp\left(-\frac{C_1^2 \epsilon_n^2 q}{8\sigma^2(q)}\right) + 22\left\{1 + \frac{8}{C_1 \epsilon_n}\right\}^{1/2} q \alpha\{[n/(2q)]\} \end{aligned} \quad (\text{A.2})$$

where  $\sigma^2(q) = 2p^{-2}\nu^2(q) + \epsilon_n = C\epsilon_n$ . It is obvious that

$$4 \exp\left(-\frac{C_1^2 \epsilon_n^2 q}{8\sigma^2(q)}\right) \leq 4 \exp\{-C_2 \epsilon_n q\} \quad (\text{A.3})$$

where  $C_2 > 0$ . Since  $n\epsilon_n^2 \rightarrow \infty$  means  $q\epsilon_n \rightarrow \infty$ , the first term in (A.2) converges to zero exponentially fast. On the second term of (A.2), the geometric  $\alpha$ -mixing implies that

$$22\left\{1 + \left(\frac{8}{C_1 \epsilon_n}\right)^{1/2}\right\} q \alpha\{[n/(2q)]\} \leq C \epsilon_n^{-1/2} q \rho^{[n^{1/2} \log^{-1}(n)/2]} \quad (\text{A.4})$$

which converges to zero exponentially fast too. This completes the proof of Lemma 1.  $\square$

**LEMMA 2.** Under the conditions (i)-(ii) and for any  $\kappa > 0$ ,

$$n^{-1} \sum (Y_t - \nu_p) \{I(Y_t \geq \hat{\nu}_p) - I(Y_t \geq \nu_p)\} = o_p(n^{-3/4+\kappa}).$$

**PROOF:** Let  $W_t = (Y_t - \nu_p) \{I(Y_t \geq \hat{\nu}_p) - I(Y_t \geq \nu_p)\}$ . We first evaluate  $E(W_t)$ . Note that  $E(W_t) =: -I_{t1} + I_{t2}$  where

$$\begin{aligned} I_{t1} &= E\{(Y_t - \nu_p) I(\nu_p \leq Y_t < \hat{\nu}_p) I(\hat{\nu}_p > \nu_p)\} \quad \text{and} \\ I_{t2} &= E\{(Y_t - \nu_p) I(\hat{\nu}_p \leq Y_t < \nu_p) I(\hat{\nu}_p < \nu_p)\}. \end{aligned}$$

Furthermore let  $I_{t1} = I_{t11} + I_{t12}$  and  $I_{t2} = I_{t21} + I_{t22}$  where, for  $a \in (0, 1/2)$  and  $\eta > 0$ ,

$$\begin{aligned} I_{t11} &= E\{(Y_t - \nu_p) I(\nu_p \leq Y_t < \hat{\nu}_p) I(\hat{\nu}_p \geq \nu_p + n^{-a}\eta)\}, \\ I_{t12} &= E\{(Y_t - \nu_p) I(\nu_p \leq Y_t < \hat{\nu}_p) I(\nu_p < \hat{\nu}_p < \nu_p + n^{-a}\eta)\}, \\ I_{t21} &= E\{(Y_t - \nu_p) I(\nu_p > Y_t \geq \hat{\nu}_p) I(\hat{\nu}_p \leq \nu_p - n^{-a}\eta)\} \quad \text{and} \\ I_{t22} &= E\{(Y_t - \nu_p) I(\nu_p > Y_t \geq \hat{\nu}_p) I(\nu_p > \hat{\nu}_p > \nu_p - n^{-a}\eta)\}. \end{aligned}$$

Applying the Cauchy-Swartz inequality, for  $k = 1$  and  $2$ ,

$$|I_{tk1}| \leq \sqrt{E(\hat{\nu}_p - \nu_p)^2 P(|\hat{\nu}_p - \nu_p| \geq n^{-a}\eta)}.$$

Then Lemma 1 and the fact that  $E(\hat{\nu}_p - \nu_p)^2 = O(n^{-1})$  imply

$$I_{tk1} \rightarrow 0 \quad \text{exponentially fast.} \quad (\text{A.5})$$

To evaluate  $I_{t12}$ , we note that  $|I_{t12}| \leq E\{(Y_t - \nu_p)I(\nu_p \leq Y_t < \nu_p + n^{-a}\eta)\}$ . This means

$$I_{t12} \leq \int_{\nu_p}^{\nu_p + n^{-a}\eta} d\mathbf{y}(z - \nu_p)f(z)dz = O(n^{-2a}).$$

Using the exactly same approach we can show that  $I_{t22} = O(n^{-2a})$  as well. These and (A.5) mean, by choosing  $a = -1/2 + \gamma$  where  $\gamma > 0$  is arbitrarily small,

$$E(W_t) = o(n^{-1+\kappa}) \quad (\text{A.6})$$

for an arbitrarily small positive  $\kappa$ , which in turn implies

$$E\left[n^{-1} \sum (Y_t - \nu_p)\{I(Y_t \geq \hat{\nu}_p) - I(Y_t \geq \nu_p)\}\right] = o(n^{-1+\kappa}). \quad (\text{A.7})$$

We now consider  $Var(W_i)$ . For  $a \in (0, 1/2)$ ,

$$\begin{aligned} E(W_t^2) &= E\left[(Y_t - \nu_p)^2\{I(Y_t \geq \hat{\nu}_p) - 2I(Y_t \geq \hat{\nu}_p)I(Y_t \geq \nu_p) + I(Y_t \geq \nu_p)\}\right] \\ &= E\left[(Y_t - \nu_p)^2\{I(\nu_p > Y_t \geq \hat{\nu}_p) + I(\hat{\nu}_p > Y_t \geq \nu_p)\}\right] \\ &= E\left[(Y_t - \nu_p)^2 I(\hat{\nu}_p \leq Y_t < \nu_p)\{I(\hat{\nu}_p \geq \nu_p - n^{-a}\eta) + I(\hat{\nu}_p < \nu_p - n^{-a}\eta)\}\right] \\ &+ E\left[(Y_t - \nu_p)^2 I(\hat{\nu}_p > Y_t \geq \nu_p)\{I(\hat{\nu}_p \geq \nu_p + n^{-a}\eta) + I(\hat{\nu}_p < \nu_p + n^{-a}\eta)\}\right]. \end{aligned}$$

Note that

$$\begin{aligned} E\{I(\hat{\nu}_p \leq Y_t < \nu_p)I(\hat{\nu}_p \leq \nu_p - n^{-a}\eta)\} &\leq P(|\hat{\nu}_p - \nu_p| \geq n^{-a}\eta) \quad \text{and} \\ E\{I(\hat{\nu}_p > Y_t \geq \nu_p)I(\hat{\nu}_p > \nu_p + n^{-a}\eta)\} &\leq P(|\hat{\nu}_p - \nu_p| \geq n^{-a}\eta) \end{aligned}$$

which converge to zero exponentially fast as implied by Lemma 1. Applying the Cauchy-Schwartz inequality, we have

$$\begin{aligned} E\{(Y_t - \nu_p)^2 I(\hat{\nu}_p \leq Y_t < \nu_p)I(\hat{\nu}_p \leq \nu_p - n^{-a}\eta)\} &\quad \text{and} \\ E\{(Y_t - \nu_p)^2 I(\hat{\nu}_p > Y_t \geq \nu_p)I(\hat{\nu}_p > \nu_p + n^{-a}\eta)\} &\end{aligned}$$

converge to zero exponentially fast as well. Then, applying the same method that establish (A.6), we have

$$\begin{aligned} E\{(Y_t - \nu_p)^2 I(\mathcal{D}_p \leq Y_t < \nu_p) I(\mathcal{D}_p \geq \nu_p - n^{-a}\eta)\} &= O(n^{-3a}) \quad \text{and} \\ E\{(Y_t - \nu_p)^2 I(\mathcal{D}_p > Y_t \geq \nu_p) I(\mathcal{D}_p < \nu_p + n^{-a}\eta)\} &= O(n^{-3a}). \end{aligned}$$

In summary we have  $E(W_t^2) = o(n^{-3/2+\kappa})$ . This and (A.6) mean  $Var(W_t) = o(n^{-3/2+\kappa})$ . By slightly modifying the above derivation for  $Var(W_t)$ , it may be shown that for any  $t_1, t_2$   $Cov(W_{t_1}, W_{t_2}) = o(n^{-3/2+\kappa})$ . Therefore,

$$Var\left[n^{-1} \sum_{i=1}^n (Y_t - \nu_p) \{I(Y_t \geq \hat{\nu}_p) - I(Y_t \geq \nu_p)\}\right] = o(n^{-3/2+\kappa}). \quad (\text{A.8})$$

This together with (A.7) readily establishes the lemma.

**LEMMA 3.** Let  $\hat{\beta} = (np)^{-1} \sum Y_t G_h(\nu_p - Y_t)$  and  $\hat{\eta} = (n)^{-1} \sum_{i=1}^n Y_t K_h(\nu_p - Y_t)$ . Under the conditions (i)-(iv),

$$\begin{aligned} (\text{a}) \quad Cov\left[\hat{\beta}, \{p - S_h(\nu_p)\} \{\hat{f}(\nu_p) - f(\nu_p)\}\right] &= o(n^{-1}), \\ (\text{b}) \quad Cov\left[\hat{\beta}, (\hat{\eta} - \eta) \{p - S_h(\nu_p)\}\right] &= o(n^{-1}), \\ (\text{c}) \quad Cov\left[\{p - S_h(\nu_p)\}, (\hat{\eta} - \eta) \{p - S_h(\nu_p)\}\right] &= o(n^{-1}). \end{aligned}$$

**PROOF:** We only present the proof of (a) as the proofs for the others are similar. Define  $\beta = E(\hat{\beta})$ . Let  $\hat{\beta} - \beta = n^{-1} \sum \psi_1(Y_t)$ ,  $\hat{f}(\nu_p) - f(\nu_p) = n^{-1} \sum \psi_2(Y_t) + O(n^{-2})$  and  $p - \hat{F}_h(\nu_p) = n^{-1} \sum \psi_3(Y_t) + O(n^{-2})$  for some functions  $\psi_j$ ,  $j = 1, 2$  and  $3$ , such that  $E\{\psi_j(Y_t)\} = 0$ . For instance,  $\psi_2(Y_t) = K_h(\nu_p - Y_t) - E\{K_h(\nu_p - Y_t)\}$  and  $\psi_3(Y_t) = G_h(\nu_p - Y_t) - E\{G_h(\nu_p - Y_t)\}$ .

Using the approach in Billingsley (1968, p 173),

$$\begin{aligned} A &=: |E\left[\left(\hat{\beta} - \beta\right) \{p - S_h(\nu_p)\} \{\hat{f}(\nu_p) - f(\nu_p)\}\right]| \\ &\leq n^{-2} \sum_{i \geq 1, j \geq 1, i+j \leq n} |E\{\psi_1(Y_1) \psi_2(Y_i) \psi_3(Y_{i+j})\}| [6] + O(n^{-4} + n^{-2}) \quad (\text{A.9}) \end{aligned}$$

where [6] indicates all the six different permutations among the three indices. Let  $p = 2 + \delta$ ,  $q = 2 + \delta$  and  $s^{-1} = 1 - p^{-1} - q^{-1}$  for some positive  $\delta$ . From the Davydov inequality,

$$|E\{\psi_1(Y_1) \psi_2(Y_i) \psi_3(Y_{i+j})\}| \leq 12 \|\psi_1(Z_1)\|_p \|\psi_2(Y_i) \psi_3(Z_{i+j})\|_q \alpha^{1/s}(\mathfrak{p}).$$



We are to employing the blo

By choosing  $r = n^a$  for  $a \in (0, 1)$  and  $k' = n^c$  such that  $c \in (0, 1 - a)$ , we can show that the left hand side of (A.15) converges to 0 as  $n \rightarrow \infty$ . Hence

$$n \xrightarrow{p} \mathbf{0} \quad \text{as } n \rightarrow \infty. \quad (\text{A.16})$$

Therefore,  $S_{n,1} = n^{-1/2} \sum_{j=1}^r W_{j,n} + o_p(\mathbf{1})$ .

By applying the inequality established in Yokoyama (1980) and the construction of  $W_{j,n}$ , we ha

Let  $\eta = E\{p^{-1}Y_t K_h(Y_t - \nu_p)\} = p^{-1} \int (\nu_p - \mathbf{y}) K(\mathbf{y}) f(\nu_p - \mathbf{y}) d\mathbf{y} = p^{-1} \nu_p f(\nu_p) + O(p^{-2})$ . Using a standard derivations for  $\alpha$ -mixing sequences for instance those given in Bosq (1998), we have  $Coy\{(np)^{-1} \sum Y_t K_h(\nu_p - Y_t), \hat{\nu}_{p,h} - \nu_p\} = O(n^{-1})$ . Hence, from (A.18),

$$\begin{aligned} E\{(np)^{-1} \sum Y_t K(\nu_p - Y_t)(\hat{\nu}_{p,h} - \nu_p)\} &= \eta E(\hat{\nu}_{p,h} - \nu_p) + O(n^{-1}) \\ &= -\frac{1}{2} p^{-1} \nu_p f'(\nu_p) \sigma_K^2 + o(p^{-2}) + O(n^{-1}) \end{aligned} \quad (\text{A.21})$$

Combine (A.19), (A.20) and (A.21),

$$E(\hat{\nu}_{p,h}) = \nu_p - \frac{1}{2} p^{-1} \sigma_K^2 f'(\nu_p) + o(p^{-2}) + O(n^{-1})$$

which establishes the bias given in (9).

We now derive the variance of  $\hat{\nu}_{p,h}$ . Let  $A_1 = (np)^{-1} \sum_{t=1}^n \{Y_t G_h(\nu_p - Y_t) - Y_t K_h(\nu_p - Y_t)(\hat{\nu}_{p,h} - \nu_p)\}$  be the leading order term of the expansion (A.19).

Then,

$$\begin{aligned} Var(A_1) &= Var\{(np)^{-1} \sum Y_t G_h(\nu_p - Y_t)\} + Var\{\hat{\nu}_{p,h}(\hat{\nu}_{p,h} - \nu_p)\} \\ &\quad - 2Coy\{(np)^{-1} \sum Y_t G_h(\nu_p - Y_t), \hat{\nu}_{p,h}(\hat{\nu}_{p,h} - \nu_p)\}. \end{aligned} \quad (\text{A.22})$$

It is easy to see that

$$\begin{aligned} &Var\{(np)^{-1} \sum Y_t G_h(\nu_p - Y_t)\} \\ &= n^{-1} p^{-2} \left[ Var\{Y_t G_h(\nu_p - Y_t)\} + 2 \sum_{k=1}^{n-1} (1 - k/n) Coy\{Y_1 G_h(\nu_p - Y_1), Y_{k+1} G_h(\nu_p - Y_{k+1})\} \right]. \end{aligned}$$

Let  $c_K = \int_{-\infty}^{\infty} \mathbf{y} K(\mathbf{y}) d\mathbf{y} \int_{-\infty}^{\infty} K(\mathbf{y}) d\mathbf{y}$ . It may be shown that

$$\begin{aligned} Var\{Y_t G_h(\nu_p - Y_t)\} &= \int z^2 G_h^2(\nu_p - z) f(z) dz - p^2 \nu_p^2 + O(p^{-2}) \\ &= \int_{-\infty}^{\infty} K(\mathbf{y}) d\mathbf{y} \left[ \int_{-\infty}^u K(\mathbf{y}) d\mathbf{y} \left\{ \int_{\nu_p}^{\infty} z^2 f(z) dz + \int_{\nu_p - hu}^{\nu_p} z^2 f(z) dz \right\} \right. \\ &\quad \left. + \int_u^{\infty} K(\mathbf{y}) d\mathbf{y} \left\{ \int_{\nu_p}^{\infty} z^2 f(z) dz + \int_{\nu_p - hu}^{\nu_p} z^2 f(z) dz \right\} - p^2 \nu_p^2 + O(p^{-2}) \right] \\ &= Var\{Y_t I(Y_t \geq \nu_p)\} - 2 \nu_p^2 f(\nu_p) c_K + O(p^{-2}). \end{aligned} \quad (\text{A.23})$$

Equation (A.23) and Lemma 3 mean

$$Var\{(np)^{-1} \sum Y_t G_h(\nu_p - Y_t)\} = p^{-2} Var\{\phi_1(\nu_p)\} - 2n^{-1} \nu_p^2 f(\nu_p) c_K + o(n^{-1}). \quad (\text{A.24})$$

The second term on the right hand side of (A.22) is

$$\begin{aligned} & Var\{\eta(\hat{\nu}_{p,h} - \nu_p)\} + (\hat{\eta} - \eta)(\hat{\nu}_{p,h} - \nu_p)\} \\ = & \eta^2 Var(\hat{\nu}_{p,h}) + 2\eta Cov\{\hat{\nu}_{p,h}, (\hat{\eta} - \eta)(\hat{\nu}_{p,h} - \nu_p)\} + Var\{(\hat{\eta} - \eta)(\hat{\nu}_{p,h} - \nu_p)\}. \end{aligned}$$

It may be shown by using the fact that  $\eta = p^{-1}\nu_p f(\nu_p) + O(n^{-2})$

$$\eta^2 Var(\hat{\nu}_{p,h}) = p^{-2}\nu_p^2 Var\{n^{-1}\sum_{t=1}^n I(Y_t > \nu_p)\} - 2p^{-2}n^{-1}b\nu_p^2 f(\nu_p)c_K + o(n^{-1}b). \quad (\text{A.25})$$

From the inequality given in Yokoyama (1980) for  $\alpha$ -mixing sequences,

$$E(\hat{\nu}_{p,h} - \nu_p)^4 \leq Cn^{-2} \quad \text{and} \quad E(\hat{\eta} - \eta)^4 = O(n^{-2-3}).$$

Applying the Cauchy-Schwartz inequality and Lemma 3,

$$Var\{(\hat{\eta} - \eta)(\hat{\nu}_{p,h} - \nu_p)\} = O(n^{-2-3/2}) = o(n^{-1}) \quad \text{and} \quad (\text{A.26})$$

$$Cov\{\eta(\hat{\nu}_{p,h} - \nu_p), p^{-1}(\hat{\eta} - \eta)(\hat{\nu}_{p,h} - \nu_p)\} = o(n^{-1}). \quad (\text{A.27})$$

Combine (A.25), (A.26) and (A.27),

$$Var\{\hat{\eta}(\hat{\nu}_{p,h} - \nu_p)\} = p^{-2}\nu_p^2 Var\{\phi_2(\nu_p)\} - 2p^{-2}n^{-1}\nu_p^2 f(\nu_p)c_K + o(n^{-1}). \quad (\text{A.28})$$

From Lemma 3, the covariance term on the right hand side of (A.22) is

$$\begin{aligned} & Cov\{(np)^{-1}\sum_{t=1}^n Y_t G_h(\nu_p - Y_t), \hat{\eta}(\hat{\nu}_{p,h} - \nu_p)\} \\ = & Cov\{(np)^{-1}\sum_{t=1}^n Y_t G_h(\nu_p - Y_t), \eta f^{-1}(\nu_p)n^{-1}\sum_{t=1}^n G_h(\nu_p - Y_t)\} + o(n^{-1}) \\ = & (np^2)^{-1}\nu_p \left[ Cov\{Y_t G_h(\nu_p - Y_t), G_h(\nu_p - Y_t)\} \right. \\ & \left. + 2\sum_{k=1}^{n-1} (1 - k/n) Cov\{Y_1 G_h(\nu_p - Y_1), G_h(\nu_p - Y_{k+1})\} \right] + o(n^{-1}) \end{aligned}$$

Since  $Cov\{Y_t G_h(\nu_p - Y_t), G_h(\nu_p - Y_t)\} = p(1-p)\nu_p - 2\nu_p f(\nu_p)c_K + o(n^{-1})$ ,

$$\begin{aligned} & Cov\{(np)^{-1}\sum_{t=1}^n Y_t G_h(\nu_p - Y_t), (np)^{-1}\sum_{i=1}^n Y_i K_h(\nu_p - Y_i)(\hat{\nu}_{p,h} - \nu_p)\} \\ = & n^{-1}p^{-2}\nu_p Cov\{\phi_1(\nu_p), \phi_2(\nu_p)\} - 2n^{-1}p^{-2}\nu_p^2 f(\nu_p)c_K + o(n^{-1}). \end{aligned} \quad (\text{A.29})$$

Substitute (A.25), (A.28) and (A.29) to (A.22), we note that all the second order terms of  $O(n^{-1})$  cancel out each other and therefore

$$\text{Var}(\hat{\rho}_p) = p^{-1}n^{-1}\sigma_0^2(p, n) + o(n^{-1}), \quad (\text{A.30})$$

which establish (10).

The asymptotic normality of  $\hat{\rho}_{p,h}$  can be established from (A.19) by using the same blocking method as that in the proof of Theorem 1.

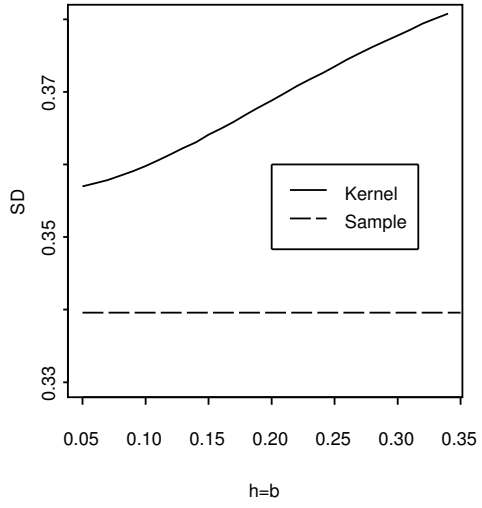
#### REFERENCES

Artzner, P., Delbaen, F., Eber, J-M. and Heath, D. (1999): "Coherent measures of risk", *Mathematical Finance*, 9, 203-228.

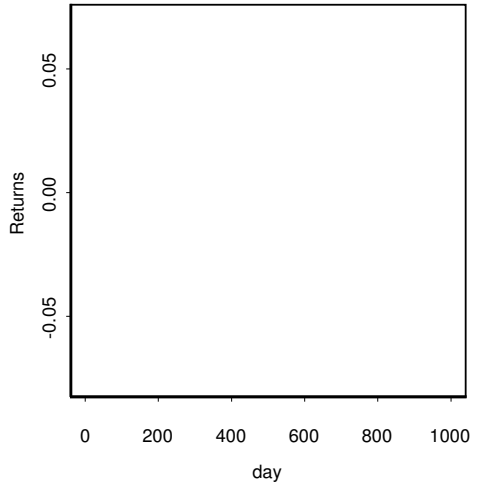
Bellini, F. and Figa-Talamanca, G. (2002): "Detecting and modeling tail dependence." Manuscript and a paper presented in Quantitative Methods in Finance 2003 Conference, Cairns, Australia, 19-21 June 2003.

- Frey, R. and McNeil, A. J. (2002): "VaR and expected shortfall in portfolios of dependent credit risks: conceptual and practical insights", *Journal of Banking and Finance*, **26**, 1317-1334.
- Genon-Caralot, V., T. Jeantheau, and C. Laredo (2000): "Stochastic Volatility Models as Hidden Markov Models and Statistical Applications," *Bernoulli*, **6**, 1051-1079.
- Gourieroux, C., Scaillet, O. and Laurent, J.P. (2000): "Sensitivity analysis of Values at Risk", *Journal of Empirical Finance*, **7**, 225-245.
- Jorion, P. (2001): *Value at Risk*, 2nd Edition, New York: McGraw-Hill.
- Leadbetter, M. R., Lindgren, G. and Rootzen, H. (1983): *Extremes and Related Properties of Random Sequences and Processes*. Springer: Berlin.
- Masry, E. and Tjøstheim, D. (1995): "Nonparametric estimation and identification of nonlinear arch time series," *Econometric Theory*, **11**, 258-289.
- Reiss, R. D. and Thomas, M. (2001): *Stochastic Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields*, 2nd ed., Boston: Birkhauser Verlag.
- Rosenblatt, M. (1956): "A central limit theorem and a the  $\alpha$ -mixing condition", *Proc. Nat. Acad. Sci. USA*, **42**, 43-47.
- Scaillet, O. (2003): "Nonparametric estimation of conditional expected shortfall", HEC Geneva mimeo.
- Scaillet, O. (2004): "Nonparametric estimation and sensitivity analysis of expected shortfall", *Mathematical Finance*, **14**, 115-129.
- Sheather, S. J. and Marron, J. S. (1990): "Kernel quantile estimators", *Journal of the American Statistical Association*, **85**, 410-416.
- Yokoyama, R. (1980): "Moment bounds for stationary mixing sequences", *Probability Theory and Related Fields*, **52**, 45-87.
- Yoshihara, K. (1995): "The Bahadur representation of sample quantiles for sequences of strongly mixing random variables", *Statistics and Probability Letters*, **24**, 299-304.

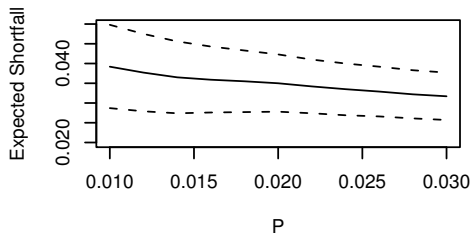
(a) ES Estimates



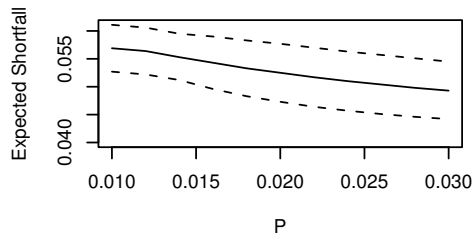




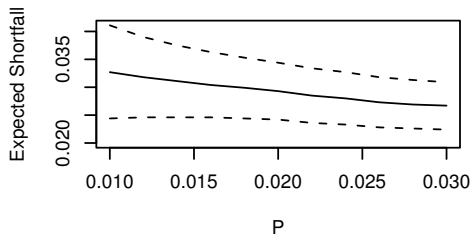
DO ON  $\lambda^2$  1<sup>2</sup>




CAC  $\lambda^2$  1<sup>2</sup>



DO ON  $\lambda^2$  2<sup>2</sup>



*T*  Estimates for  $\nu_{0.01}$   $\mu_{0.01}$  and Standard Errors S.E

Year	CAC			Dow Jones		
	$\hat{\nu}_{0.01}$	$\hat{\mu}_p$	S.E.	$\hat{\nu}_{0.01}$	$\hat{\mu}_p$	S.E.
2001-2002	0.0553	0.0571	0.0059	0.0378	0.0424	0.0135
2002-2003	0.0461	0.0510	0.0207	0.0292	0.0316	0.0231
2001-2003	0.0531	0.0567	0.0027	0.0331	0.0394	0.0065