Nonparametric Estimation of Expected Shortfall[1]

SONG XI CHEN

Iowa State University

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 *Mathematical 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 estimation of the VaR where smoothing has been shown to produce reduction in both the variance and the mean square error of estimation. Therefore, the simpler ES estimator based on the sample average of excessive losses is attractive for the shortfall estimation.

Key Words: Expected shortfall; Kernel estimator; Risk Measures; Value at Risk; Weakly dependent.

[1]The author thanks Cheng Yong Tang for computational support. Address of correspondence: Department of Statistics, Iowa State University, Ames, IA 50011-1210, USA; email: songchen@iastate.edu

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

$$\nu_p = \inf\{u : F(u) \geq 1-p\} \tag{1}$$

which is the $(1-p)$-th quantile of the lose distribution F. The VaR specifies a level of excessive losses such that the probability of a loss larger than ν_p is less than p. See Duffie and Pan (1997) and Jorion (2001) for the financial background, statistical inference and applications of VaR. A major shortcoming of the VaR in addition to not being a coherent risk measure is that it provides no information on the extent of excessive losses other than specifying a level that defines the excessive losses. In contrast, ES is a risk measure which is not only coherent but also informative on the extend of losses larger than ν_p.

The ES associated with a confidence level $1-p$, denoted as μ_p, is the conditional expectation of a loss given that the loss is larger than ν_p, that is

$$\mu_p = E(Y_t|Y_t > \nu_p). \tag{2}$$

Estimation of the ES can be carried out by assuming a parametric loss distribution, which is the method commonly used in actuary studies. Frey and McNeil (2002) propose a binomial mixture model approach to estimate ES and VaR for a large balanced portfolio. The extreme value theory approach (Embrechts, Kluppelberg and Mikosch, 1997) can be viewed as a semiparametric approach which uses the asymptotic distribution of exceedance over a high threshold to model the excessive loses and then carries out a parametric inference within the framework of the Generalized Pareto distributions. Recently, Scaillet (2004) proposes 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 plays 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 Rootzén, 1983). An empirical study by Bellini and Figá-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 $\hat{\nu}_p$ 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 $\hat{\nu}_p$ where $\hat{\nu}_p = Y_{([n(1-p)]+1)}$ is the sample VaR (quantile) estimator of ν_p and $Y_{(r)}$ is the r-th order statistic of $\{Y_t\}_{t=1}^n$. Hence, it is

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

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 $G(t) = \int_t^\infty K(u)du$ 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

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

A kernel estimator of ν_p, denoted as $\hat{\nu}_{p,h}$, is the solution of $S_h(z) = p$, as proposed in Gourieroux, Laurent and Scaillet (2000).[2] By replacing the indicator function and $\hat{\nu}_p$ with the smoother G_h and $\hat{\nu}_{p,h}$ respectively in (3) Scaillet (2004) proposed the following kernel estimator

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

[2] 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 conditional quantile estimation. A kernel estimator of conditional ES is proposed in Scaillet (2003). See Fan and Yao (2003) for other applications of the kernel method for nonlinear time series analysis.

Based on the improvement of the kernel VaR estimator $\hat{\nu}_{p,h}$ over $\hat{\nu}_p$, it is expected that the kernel ES estimator $\hat{\mu}_{p,h}$ would improve the estimation accuracy of the unsmoothed estimator $\hat{\mu}_p$. 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 σ-algebra of events generated by $\{Y_t, k \le t \le l\}$ for $l > k$. The α-mixing coefficient introduced by Rosenblatt (1956) is

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

The series is said to be α-mixing if $\lim_{k\to\infty} \alpha(k) = 0$. The dependence described by the α-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 $\alpha(k) \le C\rho^k$ for all $k \ge 1$ 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 ν_p; the joint distribution functions of (Y_1, Y_{k+1}) F_k for $k \ge 1$ have all its second partial derivatives bounded in $\mathcal{B}(\nu_p)$; $E(|Y_t|^{2+\delta}) \le C$ for some $\delta > 0$ and a positive constant C.

(iii) K is a symmetric probability density satisfying the moment conditions $\int_{-1}^1 uK(u)du = 0$ and $\int_{-1}^1 u^2K(u)du = \sigma_K^2 > 0$, and K has bounded and Lipschitz continuous derivative.

(iv) h satisfies $h \to 0$, $nh^{3-\beta} \to \infty$ for any $\beta > 0$ and $nh^4 \log^2(n) \to 0$ as $n \to \infty$.

Condition (i) means that the time series is geometric α-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 α-mixing for ARCH model; and Genon–Catalot, Jeantheau and Larédo (2000) for diffusion models. Conditions (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 $n \to \infty$.

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 κ,

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

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 $n \to \infty$

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

The theorem indicates that the asymptotic variance of $\hat{\mu}_p$ is $\sigma_0^2(p;n)/(np)$, which is the variance of $p^{-1}\{n^{-1}\sum_{i=1}^{n}(Y_t - \nu_p)I(Y_t \geq \nu_p) - p(\mu_p - \nu_p)\}$, 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 $n \to \infty$

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

and furthermore,

$$\begin{aligned} Bias(\hat{\mu}_{p,h}) &= -\tfrac{1}{2}p^{-1}\sigma_K^2 h^2 f(\nu_p) + o(h^2) \quad \text{and} \qquad (9)\\ Var(\hat{\mu}_{p,h}) &= p^{-1}n^{-1}\sigma_0^2(p;n) + o(n^{-1}h). \qquad (10) \end{aligned}$$

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 μ_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 ν_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 $n^{-1}h$ 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 μ_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

$$\text{an AR(1) model:} \quad Y_t = 0.5Y_{t-1} + \epsilon_t, \epsilon_t \stackrel{iid}{\sim} N(0,1); \tag{11}$$

$$\text{an ARCH(1) model:} \quad Y_t = 0.5Y_{t-1} + \epsilon_t, \quad \epsilon_t^2 = 4 + 0.4\epsilon_{t-1}^2 + \eta_t, \quad \eta_t \stackrel{iid}{\sim} N(0,1). \tag{12}$$

We are interested in estimating $\mu_{0.01}$, the 99% ES. In constructing the kernel estimator, the Gaussian kernel $K(u) = \frac{1}{\sqrt{2\pi}} \exp(-u^2/2)$ is employed. The sample size considered in the simulation are 250 and 500. The number of simulation is 1000.

Figures 1 and 2 display the variance and mean square errors of $\hat{\mu}_p$ and the kernel VaR estimator $\hat{\nu}_{p,h}$ over a set of bandwidth values. For comparison, the figures also include the variance and mean square errors of the unsmoothed VaR estimator $\hat{\nu}_p$ and the kernel estimator $\hat{\nu}_{p,h}$ respectively. Although the sample size considered in these figures is 250, the same pattern of results is observed for the sample size 500 as well. It is observed that $\hat{\mu}_{p,h}$ has a larger variance and, to a large extent, a larger MSE too than $\hat{\mu}_p$ for both models. In contrast, the kernel VaR estimator $\hat{\nu}_{p,h}$ delivers both variance and mean square error reduction as revealed in Chen and Tang (2005). This confirms that there is no need to smooth the data for ES estimation.

5. EMPIRICAL STUDY

We apply the proposed kernel estimator to estimate the ES of two financial time series. The two financial series are the CAC 40 and the Dow Jones series from October 1st 2001 to September 30th 2003, which consist of 500 observations (2 years data). The log-return series are displayed in Figure 3 together with their sample auto-correlation functions (ACF). To confirm the existence of dependence, we carry out the Box-Pierce test with the test statistic $Q = n\sum_{k=1}^{29} \hat{\gamma}^2(k)$ where $\hat{\gamma}(k)$ is the sample auto-correlation for lag k. The statistic Q takes value 51.146 for the CAC 40 and 43.001 for Dow Jones, which produces p-values of 0.0068 for CAC 40 and 0.0455 for Dow Jones respectively. Therefore, the dependence is significant for both series at 5% significant level.

We carry out analysis over three periods on each series, which are the first year (2001-2002), the second year (2002-2003), and the entire two year (2001-2003), respectively. Table 1 presents the ES estimates $\hat{\mu}_{0.01}$ and their standard errors. The standard errors are obtained via a kernel estimation of the spectral density of $\{(Y_t - \nu_p)G_h(Y_t - \nu_p)\}$, which resembles the approach for obtaining standard errors for kernel VaR estimation considered in Chen and Tang (2005). The table also provides the kernel estimates for the 99% VaR. It is observed that for both indices the year 2001-2002 had the largest estimates (risk) of the ES and the VaR, and hence the highest risk, which reflected the high volatility after the burst of Internet bubble and the September 11. The level of risk settles down in the year 2002-2003. It is interesting to see that the CAC was more risky than Dow Jones as the estimates of the ES and the VaR were all larger than her counterparts in Dow Jones. The variability of the ES estimate for the Dows was much higher than that of the CAC in the year 2001-2002. It seems that this high variability migrated to CAC in the second year. We observed as expected the variability for the ES estimates based on the entire two year observations were smaller than those of each individual year.

We then extend the analysis for 20 equally spaced levels of p ranging from 0.01 to 0.03. The kernel estimates of $\hat{\mu}_p$ and their 95% confidence bands are displayed in Figure 5. The confidence bands are constructed by adding and subtracting 1.96 times the standard errors. These plots show that as expected the ES estimate declines as p increases. For both indices, the year 2001-2002 experienced the largest risk than the year 2002-2003. It reveals again that the CAC is more risky that Dow Jones as the ES estimates are always larger than those of Dows for each of the three time periods and at each fixed p level.

APPENDIX

Throughout this section we use C and C_i to denote generic positive constants. The proof of Theorems 1 and 2 requires the following lemmas.

LEMMA 1. Under Condition (i), $P\left(|\hat{\nu}_p - \nu_p| \geq \epsilon_n\right) \to 0$ exponentially fast as $n \to \infty$.

PROOF: We only give the proof for $\tilde{\nu}_p = \hat{\nu}_p$ as that for $\hat{\nu}_{p,b}$ can be treated similarly,

Let $C_1 = \inf_{x \in [\nu_p - \epsilon_n, \nu_p + \epsilon_n]} f(x)$. It is easily shown that

$$\begin{aligned} & P\left(|\hat{\nu}_p - \nu_p| \geq \epsilon_n\right) \qquad\qquad (A.1) \\ \leq\ & P\{|F_n(\nu_p + \epsilon_n) - F(\nu_p + \epsilon_n)| > C_1\epsilon_n\} + P\{|F_n(\nu_p - \epsilon_n) - F(\nu_p - \epsilon_n)| > C_1\epsilon_n\}. \end{aligned}$$

Let $X_i = I(Y_t < \nu_p + \epsilon_n) - F(\nu_p + \epsilon_n)$. Clearly $E(X_i) = 0$ and $|X_i| \leq 2$. Choose $q = b_0 n\epsilon_n$,

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

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

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

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

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

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

which converges to zero exponentially fast too. This completes the proof of Lemma 1. □.

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

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

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

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

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

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

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

$$|I_{tk1}| \le \sqrt{E(\hat\nu_p-\nu_p)^2 P\left(|\hat\nu_p-\nu_p|\ge n^{-a}\eta\right)}.$$

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

$$I_{tk1} \to 0 \quad \text{exponentially fast.} \tag{A.5}$$

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

$$I_{t12} \le \int_{\nu_p}^{\nu_p + n^{-a}\eta} dv(z - \nu_p)f(z)dz = O(n^{-2a}).$$

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,

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

for an arbitrarily small positive κ, which in turn implies

$$E\Big[n^{-1}\sum(Y_t - \nu_p)\{I(Y_t \ge \hat{\mu}_p) - I(Y_t \ge \nu_p)\}\Big] = o(n^{-1+\kappa}). \tag{A.7}$$

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

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

Note that

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

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

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

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

$$E\{(Y_t-\nu_p)^2 I(\hat\nu_p \le Y_t < \nu_p) I(\hat\nu_p \ge \nu_p - n^{-a}\eta)\} = O(n^{-3a}) \quad \text{and}$$
$$E\{(Y_t-\nu_p)^2 I(\hat\nu_p > Y_t \ge \nu_p) I(\hat\nu_p < \nu_p + n^{-a}\eta)\} = O(n^{-3a}).$$

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,

$$Var\Big[n^{-1}\sum_{i=1}^{n}(Y_t-\nu_p)\{I(Y_t \ge \hat\mu_p) - I(Y_t \ge \nu_p)\}\Big] = o(n^{-3/2+\kappa}). \tag{A.8}$$

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 $\hat\eta = (nh)^{-1}\sum_{i=1}^n Y_t K_h(\nu_p - Y_t)$. Under the conditions (i)-(iv),

$$\text{(a)} \quad Cov\Big[\hat\beta, \{p - S_h(\nu_p)\}\{\hat f(\nu_p) - f(\nu_p)\}\Big] = o(n^{-1}h),$$
$$\text{(b)} \quad Cov\Big[\hat\beta, (\hat\eta - \eta)\{p - S_h(\nu_p)\}\Big] = o(n^{-1}h),$$
$$\text{(c)} \quad Cov\Big[\{p - S_h(\nu_p)\}, (\hat\eta - \eta)\{p - S_h(\nu_p)\}\Big] = o(n^{-1}h).$$

PROOF: We only present the proof of (a) as the proofs for the others are similar. Define $\beta = E(\hat\beta)$. Let $\hat\beta - \beta = n^{-1}\sum\psi_1(Y_t)$, $\hat f(\nu_p) - f(\nu_p) = n^{-1}\sum\psi_2(Y_t) + O(h^2)$ and $p - \hat F_h(\nu_p) = n^{-1}\sum\psi_3(Y_t) + O(h^2)$ for some functions ψ_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),

$$\begin{aligned} A \;&=:\; |E\Big[(\hat\beta - \beta)\{p - S_h(\nu_p)\}\{\hat f(\nu_p) - f(\nu_p)\}\Big] \\ &\le\; n^{-2}\sum_{i\ge1, j\ge1, i+j\le n} |E\{\psi_1(Y_1)\psi_2(Y_i)\psi_3(Y_{i+j})\}|[6] + O(n^{-1}h^4 + n^{-2}h^2) \end{aligned} \tag{A.9}$$

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 δ. From the Davydov inequality,

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

Since $|\psi_3(Y_{i+j})| \leq 2$ and $E|\psi_2(Y_i)|^{2+\delta} \leq Ch^{-1-\delta}$,

$$||\psi_2(Y_i)\psi_3(Y_{i+j})||_q \leq C||\psi_2(Z_{i+j})||_q \leq Ch^{-\frac{1+\delta}{2+\delta}}.$$

This and the fact that $||\psi(Y_1)||_p = E^{1/p}|\psi_1(Y_1)|^p \leq C$ lead to

$$|E\{\psi_1(Y_1)\psi_2(Y_i)\psi_3(Y_{i+j})\}| \leq 12Ch^{-\frac{1+\delta}{2+\delta}}\alpha^{\frac{\delta}{2+\delta}}(i).$$

Similarly, $|E\{\psi_1(Y_1)\psi_2(Y_i)\psi_3(Y_{i+j})\}| \leq 12Ch^{-\frac{1+\delta}{2+\delta}}\alpha^{\frac{\delta}{2+\delta}}(j)$. Therefore,

$$|E\{\psi_1(Y_1)\psi_2(Y_i)\psi_3(Y_{i+j})\}| \leq 12Ch^{-\frac{1+\delta}{2+\delta}}\min\{\alpha^{\frac{\delta}{2+\delta}}(i), \alpha^{\frac{\delta}{2+\delta}}(j)\}. \qquad \text{(A.10)}$$

From (A.9) and (A.10), and the fact that $\alpha(k)$ is monotonic no increasing,

$$\begin{aligned} A &\leq Cn^{-2}h^{-\frac{1+\delta}{2+\delta}}\sum_{j=1}^{n-1}(2j-1)\alpha^{\frac{\delta}{2+\delta}}(j) + O(n^{-1}h^4 + n^{-2}h^2) \\ &= O(n^{-2}h^{-\frac{1+\delta}{2+\delta}}) + o(n^{-1}h) = o(n^{-1}h) \end{aligned}$$

since $\sum j\alpha^{\frac{\delta}{2+\delta}}(j) < \infty$ as implied by Condition (i). □

LEMMA 4. Under the conditions (i)-(v) and for $l_1, l_2 = 0$ or 1,

$$\begin{aligned} |\sum_{k=1}^{n-1}(1-k/n)\Big[&Cov\{Y_1^{l_1}G_h(\nu_p - Y_1), Y_{k+1}^{l_2}G_h(\nu_p - Y_{k+1})\} \\ &-Cov\{Y_1^{l_1}I(Y_1 > \nu_p), Y_{k+1}^{l_2}I(Y_{k+1} > \nu_p)\}\Big] = o(h). \end{aligned}$$

PROOF: The case of $l_1 = l_2 = 0$ has been proved in Chen and Tang (2005) and the proofs for the other cases are almost the same, and hence are not given here. □

PROOF OF THEOREM 1: Let $\phi_1(t) = n^{-1}\sum_{i=1}^n Y_t I(Y_t \geq t)$ and $\phi_2(t) = n^{-1}\sum_{i=1}^n I(Y_t \geq t)$. Then, $\hat{\mu}_p = \phi_1(\hat{\nu}_p)/\phi_2(\hat{\nu}_p)$. Note that $E\{\phi_1(\nu_p)\} = p\mu_p$, $E\{\phi_2(\nu_p)\} = p$ and $\phi_2(\hat{\nu}_p) = ([np]+1)/n$. From Lemma 2, for an arbitrarily small positive κ,

$$\phi_1(\hat{\nu}_p) = \phi_1(\nu_p) + \nu_p\{\phi_2(\hat{\nu}_p) - \phi_2(\nu_p)\} + o_p(n^{-3/4+\kappa}). \qquad \text{(A.11)}$$

These lead to

$$\begin{aligned} \hat{\mu}_p &= \mu_p + p^{-1}\{\phi_1(\nu_p) - p\mu_p\} + p^{-1}\nu_p\{p - \phi_2(\nu_p)\} + o_p(n^{-3/4+\kappa}) \qquad \text{(A.12)} \\ &= \mu_p + p^{-1}\{n^{-1}\sum_{i=1}^n (Y_t - \nu_p)I(Y_t \geq \nu_p) - p(\mu_p - \nu_p)\} + o_p(n^{-3/4+\kappa}). \end{aligned}$$

We are to employing the blocking technique and Bradley's Lemma in establishing the asymptotic normality. Write $\sigma_0^{-1}(p;n)\hat{\mu}_p - \mu_p = n^{-1}\sum_{i=1}^n T_{i,n} + o_p(n^{-3/4+\kappa})$ where $T_{i,n} = \sigma_0^{-1}(p;n)p^{-1}\{(Y_i - \nu_p)I(Y_i \geq \nu_p) - p(\mu_p - \nu_p)\}$.

Let k and k' be respectively positive integers such that $k' \to \infty$, $k'/k \to 0$ and $k/n \to 0$ as $n \to \infty$. Let r be a positive integer so that $r(k+k') \leq n < r(k+k'+1)$. Define the large blocks

$$V_{j,n} = T_{(j-1)(k+k')+1,n} + \cdots + T_{(j-1)(k+k')+k,n} \quad \text{for} \quad j = 1, 2, \cdots, r;$$

the smaller blocks

$$V'_{j,n} = T_{(j-1)(k+k')+k+1,n} + \cdots + T_{(j-1)(k+k')+k+k',n} \quad \text{for} \quad j = 1, 2, \cdots, r$$

and the residual block $\delta_n = T_{r(k+k')+1,n} + \cdots + T_{n,n}$. Then

$$S_n =: n^{-1/2}\sum_{i=1}^n T_{i,n} = n^{-1/2}\sum_{j=1}^r V_{jn} + n^{-1/2}\sum_{j=1}^r V'_{jn} + n^{-1/2}\delta_n =: S_{n,1} + S_{n,2} + S_{n,3}.$$

We note that $E(S_{n,2}) = E(S_{n,3}) = 0$ and as $n \to \infty$,

$$\begin{aligned} Var(S_{n,2}) &= \frac{r\sigma_0^2(p;k')}{np\sigma_0^2(p;n)}\{1+o(1)\} \to 0 \quad \text{and} \\ Var(S_{n,3}) &= \frac{(n - r(k+k'))\sigma_0^2(p;n-r(k+k'))}{np\sigma_0^2(p;n)}\{1+o(1)\} \to 0. \end{aligned}$$

Therefore, for $l = 2$ and 3

$$S_{n,l} \xrightarrow{p} 0 \quad \text{as} \quad n \to \infty. \tag{A.13}$$

We are left to prove the asymptotic normality of $S_{n,1}$. From Bradley's lemma, there exist independent and identically distributed random variables $W_{j,n}$ such that each $W_{j,n}$ is identically distributed as $V_{j,n}$ and

$$\begin{aligned} P\left(|V_{j,n} - W_{j,n}| \leq \epsilon\sqrt{n}/r\right) &\leq 18\epsilon^{-2/5}r^{2/5}n^{-1/5}\{E(V_{j,n}^2)\}^{1/5}\alpha(k') \\ &\leq C_1\epsilon^{-2/5}n^{-1/5}r^{2/5}k^{1/5}\alpha(k') \leq C_2\epsilon^{-2/5}r^{1/5}\alpha(k'). \end{aligned} \tag{A.14}$$

Let $\Delta_n = S_{n,1} - n^{-1/2}\sum_{j=1}^r W_{j,n}$. Then

$$P(|\Delta_n| > \epsilon) \leq \sum_{j=1}^r P\left(|V_{j,n} - W_{j,n}| \leq \epsilon\sqrt{n}/r\right) \leq C_3\epsilon^{-2/5}r^{6/5}\rho^{k'} \tag{A.15}$$

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 $n \to \infty$. Hence

$$\Delta_n \xrightarrow{p} 0 \quad \text{as} \quad n \to \infty. \tag{A.16}$$

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

By applying the inequality estbalished in Yokoyama (1980) and the construction of $W_{j,n}$, we have $E(W_{j,n})^4 = E(V_{j,n}^4) \le C_1 k^2$ and $Var(W_{j,n}) = E(V_{j,n}^2) \le C_2 k$. Thus,

$$\frac{\sum E|W_{jn}|^4}{\{rVar(W_{1n})\}^2} \le \frac{C_3 rk}{r^2 k^2} \to 0$$

as $n \to \infty$, which is the Liapounov condition for the central limit theorem of triangular arrays. Therefore,

$$n^{-1/2} \sum_{j=1}^{r} W_{j,n} \xrightarrow{d} N(0,1) \quad \text{as} \quad n \to \infty. \tag{A.17}$$

Thus, the proof of the theorem is completed by combining (A.13), (A.16), (A.17) and the Slutsky theorem.

PROOF OF THEOREM 2: We first derive (9) and (10). From derivations given in Chen and Tang (2005), $\hat{\nu}_{p,h}$ admits an expansion: $\hat{\nu}_{p,h} - \nu_p = \frac{S_h(\nu_p) - p}{f(\nu_p)} + o_p(n^{-1/2})$. From the bias of $\hat{\nu}_{p,h}$ given in Chen and Tang (2005),

$$E(\hat{\nu}_{p,h}) - \nu_p = -\tfrac{1}{2}\sigma_K^2 f'(\nu_p) f^{-1}(\nu_p) h^2 + o(h^2). \tag{A.18}$$

The kernel ES estimator

$$\hat{\mu}_{p,h,b} = (np)^{-1} \sum_{t=1}^{n} \{Y_t G_h(\nu_p - Y_t) - Y_t K_h(\nu_p - Y_t)(\hat{\nu}_{p,h} - \nu_p)\} + O_p(n^{-1}) + o_p(h^2). \tag{A.19}$$

Note that

$$\begin{aligned} E\Big[(np)^{-1} \sum \{Y_t G_h(\nu_p - Y_t)\}\Big] &= p^{-1} \int z G_h(\nu_p - z) f(z) dz \\ &= p^{-1} \int_{-\infty}^{\infty} K(u) du \{\int_{\nu_p}^{\infty} z f(z) dz + \int_{\nu_p - hu}^{\nu_p} z f(z) dz\} \\ &= \mu_p - \tfrac{1}{2} p^{-1} h^2 \sigma_K^2 \{\nu_p f'(\nu_p) + f(\nu_p)\} + o(h^3). \end{aligned} \tag{A.20}$$

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

$$
\begin{aligned}
E\{(np)^{-1}\sum Y_t K(\nu_p - Y_t)(\hat{\nu}_{p,h} - \nu_p)\} &= \eta E(\hat{\nu}_{p,h} - \nu_p) + O(n^{-1}) \\
&= -\tfrac{1}{2}p^{-1}\nu_p f'(\nu_p)h^2\sigma_K^2 + o(h^3) + O(n^{-1}). \qquad (A.21)
\end{aligned}
$$

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

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

which establishes the bias given in (9).

We now derive the variance of $\hat{\mu}_{p,h}$. 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,

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

It is easy to see that

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

Let $c_K = \int_{-\infty}^{\infty} uK(u)du \int_{-\infty}^{u} K(v)dv$. It may be shown that

$$
\begin{aligned}
Var\{Y_t G_h(\nu_p - Y_t)\} &= \int z^2 G_h^2(\nu_p - z)f(z)dz - p^2\mu_p^2 + O(h^2) \\
&= \int_{-\infty}^{\infty} K(u)du\Big[\int_{-\infty}^{u} K(v)dv\{\int_{\nu_p}^{\infty} z^2 f(z)dz + \int_{\nu_p - hv}^{\nu_p} z^2 f(z)dz\} \\
&\quad + \int_{u}^{\infty} K(v)dv\{\int_{\nu_p}^{\infty} z^2 f(z)dz + \int_{\nu_p - hu}^{\nu_p} z^2 f(z)dz\} - p^2\mu_p^2 + O(h^2) \\
&= Var\{Y_t I(Y_t \geq \nu_p)\} - 2h\nu_p^2 f(\nu_p)c_K + O(h^2). \qquad (A.23)
\end{aligned}
$$

Equation (A.23) and Lemma 3 mean

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

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 $\eta = p^{-1}\nu_p f(\nu_p) + O(h^2)$

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

From the inequality given in Yokoyama (1980) for α-mixing sequences,

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

Applying the Cauchy-Schwartz inequality and Lemma 3,

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

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

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

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

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

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

Since $Cov\{Y_t G_h(\nu_p - Y_t), G_h(\nu_p - Y_t)\} = p(1-p)\mu_p - 2\nu_p f(\nu_p)hc_K + o(h)$,

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

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

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

which establish (10).

The asymptotic normality of $\hat{\nu}_{p,h}$ 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 Figá-Talamanca, G. (2002): "Detecting and modeling tail dependence." Manuscript and a paper presented in Quantative Methods in Finance 2003 Conference, Cairns, Australia, December 2003.

Billingsley, P. (1968): *Convergence of probability measures*, New York: Wiley.

Bosq, D. (1998): *Nonparametric Statistics for Stochastic Processes*, Lecture Notes in Statistics, **110**. Heidelberg: Springer-Verlag.

Cai, Z.-W. (2002): "Regression quantiles for time series data", *Econometric Theory*, **18** 169-192.

Chen, S. X. and Tang, C.Y. (2005): "Nonparametric inference of Value at Risk for dependent financial returns", *Journal of Financial Econometrics*, **3**, 227-255.

Doukhan, P. (1994): *Mixing*. Lecture Notes in Statistics, **85**. Heielberg: Springer-Verlag.

Duffie, D. and Pan, J. (1997): "An overview of value at risk", *Journal of Derivative*, 7-49.

Embrechts, P., Klueppelberg, C., and Mikosch, T. (1997): *Modeling Extremal Events for Insurance and Finance*, Berlin: Springer-Verlag.

Fan, J. and Gu, J. (2003): Semiparametric Estimation of Value-at-Risk, *Econometrics Journal*, **6**, 261-290.

Fan, J. and Yao, Q. (2003): *Nonlinear Time Series: Nonparametric and Parametric Methods*, New York: Springer-Verlag.

Föllmer, H. and Schied, A. (2002): "Convex measures of risk and trading constraints", *Finance and Stochastics*, **6**, 429–447.

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.

Gouriéroux, 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 Rootzén, 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): *Statistical Analysis of Extreme Values: with applications to insurance, finance, hydrology, and other fields*, *2nd ed.*, Boston: Birkhäuser Verlag.

Rosenblatt, M. (1956): "A central limit theorem and a the α-mixing condition", *Proc. Nat. Acad. Sc. U.S.A.*, **42**, 43-47.

Scaillet, O. (2003): "Nonparametric estimation of conditional expected shortfall", HEC Geneve 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 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.

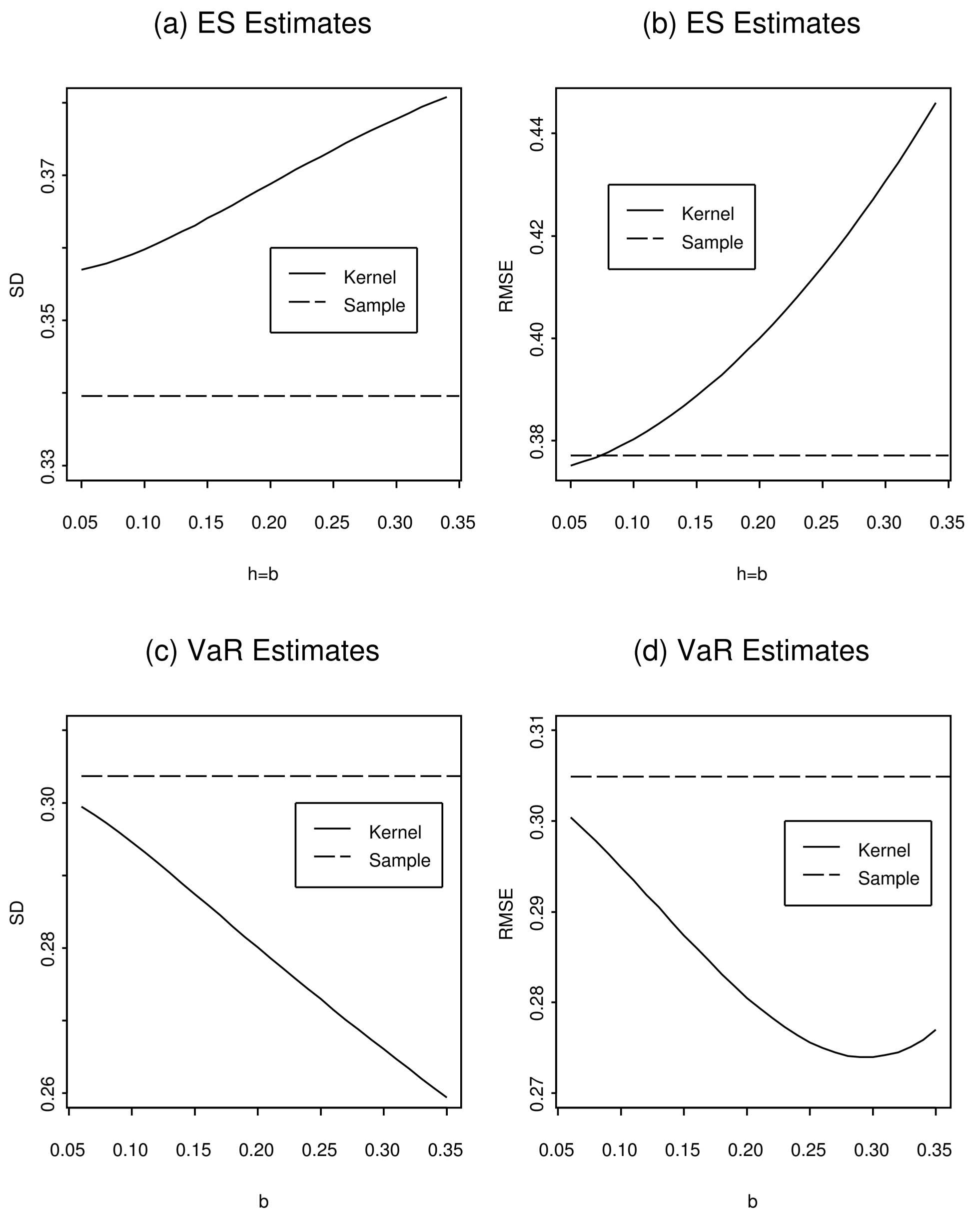

Figure 1. Simulated average standard deviation (SD) and root mean square error (RMSE) of the kernel 99% ES estimator $\hat{\mu}_{0.01,h}$ in (a) and (b) and 99% kernel VaR estimator $\hat{\nu}_{p,b}$ in (c) and (d), and their unsmoothed (legended as sample) counterparts $\hat{\mu}_{0.01}$ in (a) and (b) and $\hat{\nu}_{0.01}$ in (c) and (d) for the AR model with $n = 250$.

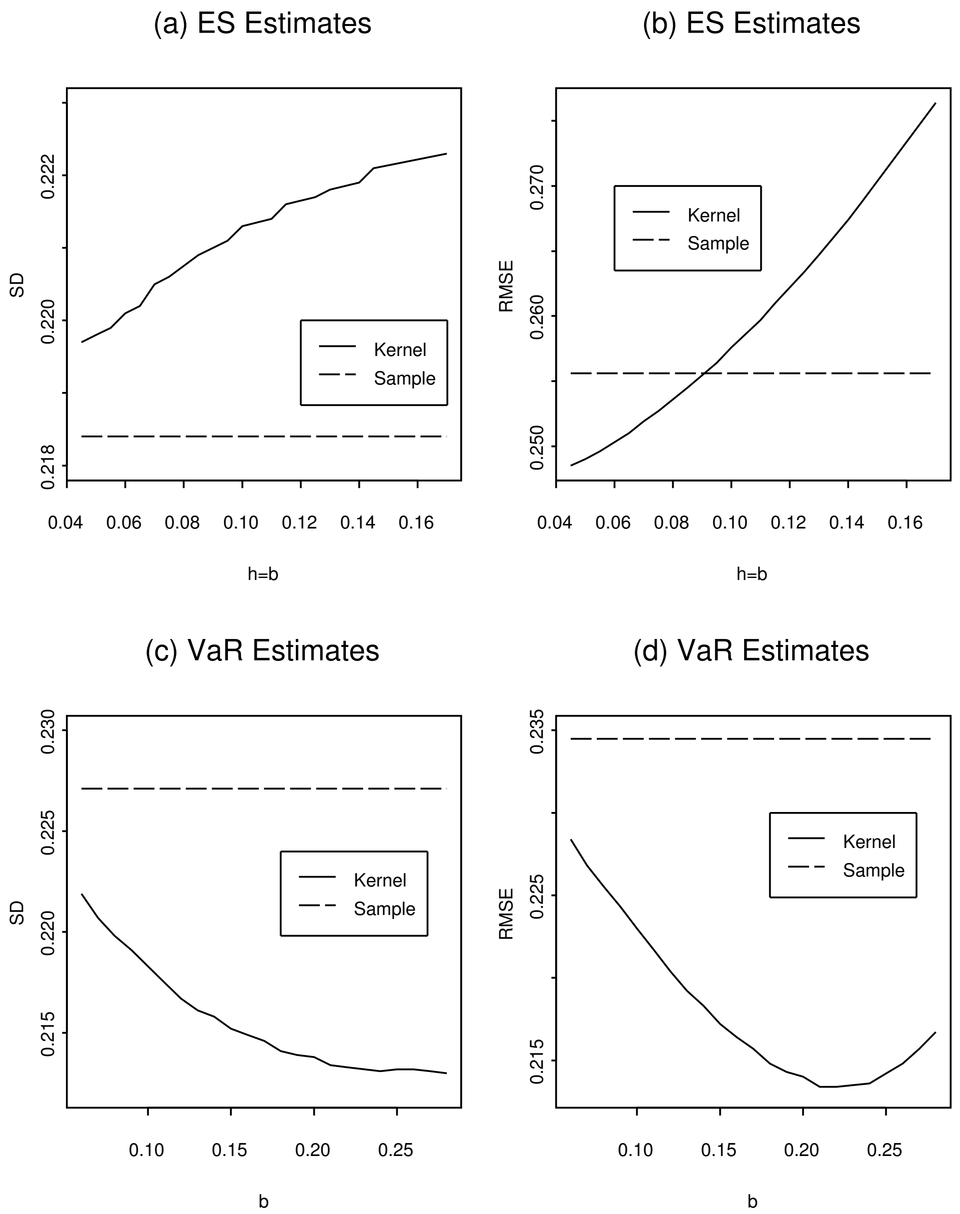

Figure 2. Simulated average standard deviation (SD) and root mean square error (RMSE) of the kernel 99% ES estimator $\hat{\mu}_{0.01,h}$ *in Panels (a) and (b) and 99% kernel VaR estimator* $\hat{\nu}_{p,b}$ *in Panels (c) and (d), and their unsmoothed (legended as sample) counterparts* $\hat{\mu}_{0.01}$ *in (a) and (b) and* $\hat{\nu}_{0.01}$ *in (c) and (d) for the ARCH model with* $n = 250$.

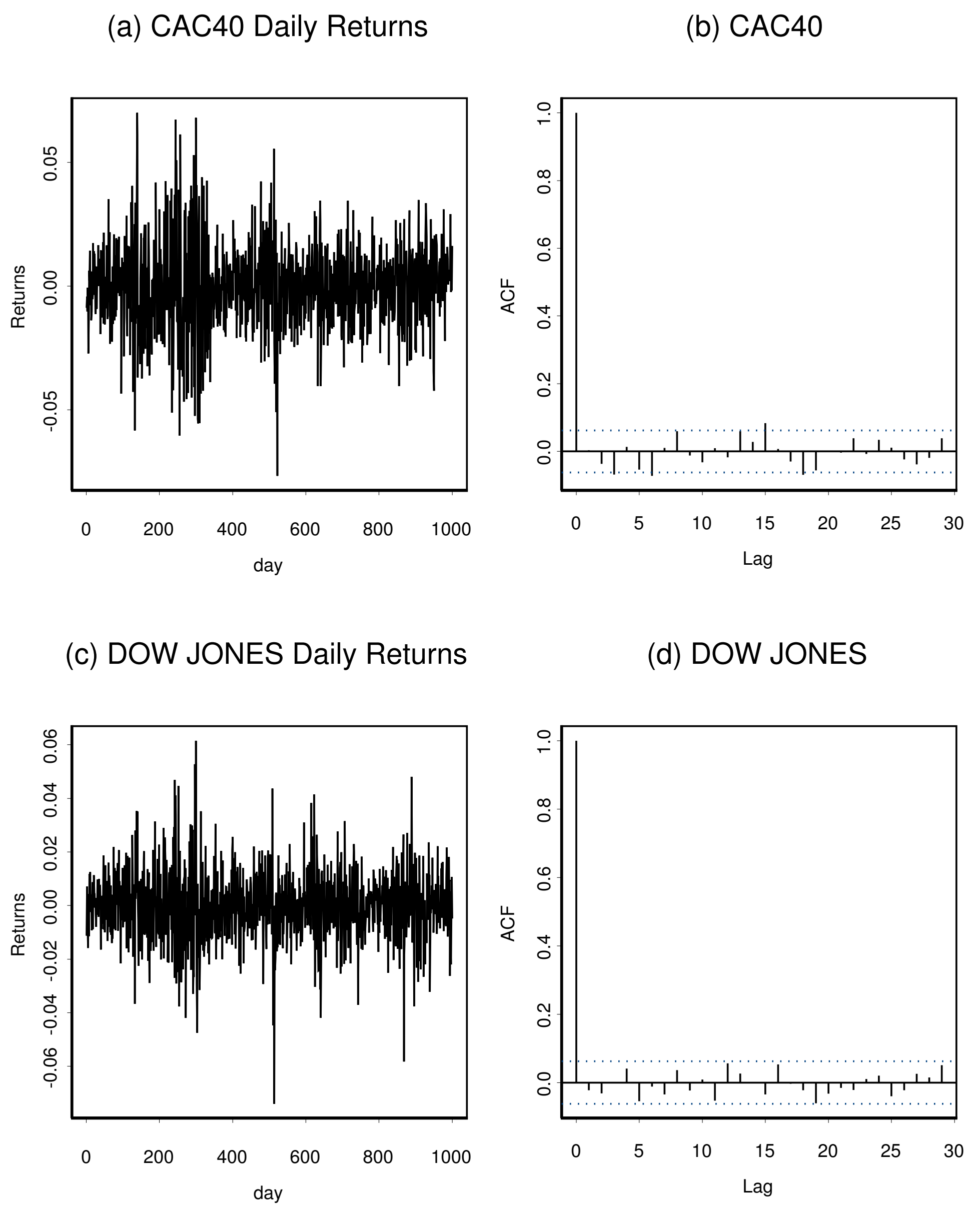

Figure 3. The two financial return series in (a) and (c) and their sample auto-correlation functions (ACF) in (b) and (d).

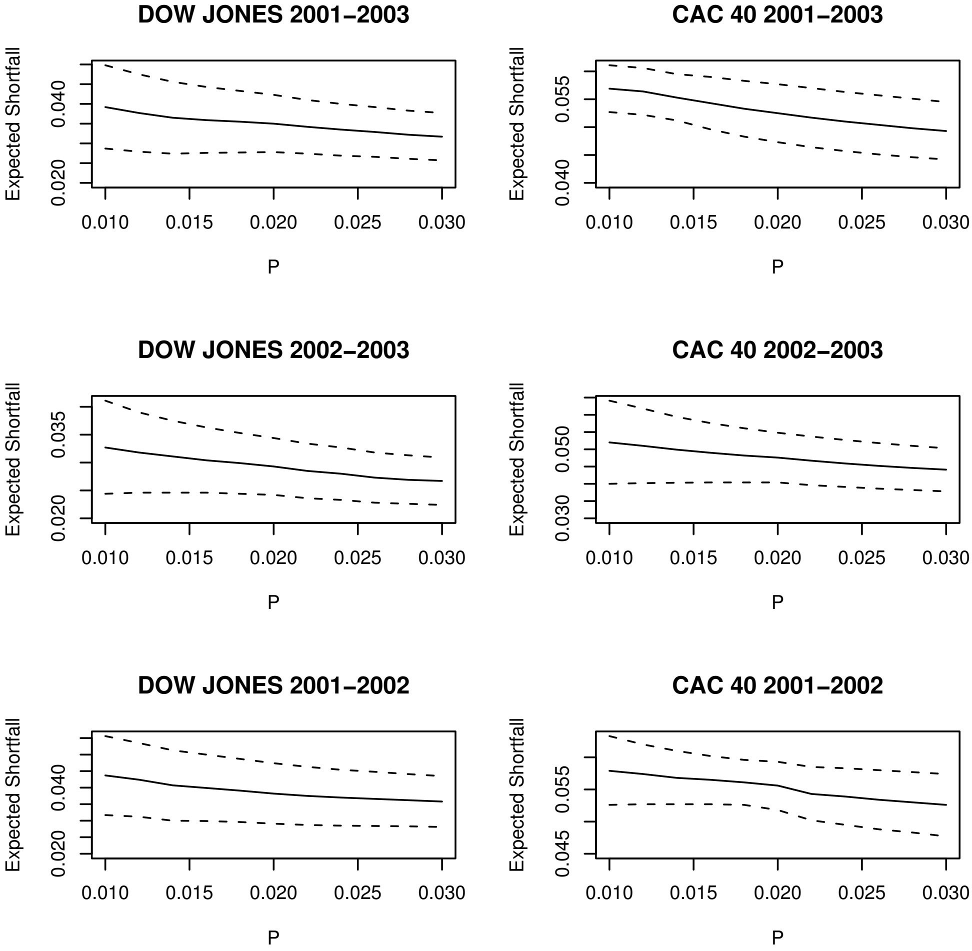

Figure 4. Expected shortfall estimates and their confidence bands for the two financial return series.

Table 1. Estimates for $\nu_{0.01}$, $\mu_{0.01}$ and Standard Errors (S.E.)

	CAC			Dow Jones		
Year	$\hat{\nu}_{0.01}$	$\hat{\mu}_p$	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