Journal of Financial Econometrics, 2008, 87–107

Nonparametric Estimation of Expected Shortfall

Song Xi Chen
Iowa State University

ABSTRACT
The expected shortfall is an increasingly popular risk measure in financial risk management and it 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.

KEYWORDS: expected shortfall, kernel estimator, risk measures, value at risk, weakly dependent

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.[1] The sub-additivity

The author thanks the Editor-in-Chief Professor Eric Renault, an Associate Editor and two referees for valuable comments and suggestions which have improved the presentation of the paper. The author also thanks Cheng Yong Tang for valuable computational support and acknowledges support of a National Science Foundation Grant (DMS-0604563). Address for correspondence: Department of Statistics, Iowa State University, Ames, IA 50011-1210, email: songchen@iastate.edu

[1] The sub-additivity of a risk measure means that the risk for the sum of two independent risky events is not greater than the sum of the risks of the two events.

doi: 10.1093/jjfinec/nbm019
Advance Access publication November 28, 2007

implies convexity of a risk measure which is used to define a risk measure being coherent; see Artzner, Delbaen, Eber and Heath (1999) for details. In contrast, ES is coherent (Föllmer and Schied, 2002) 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 tth period. Suppose $\{Y_t\}_{t=1}^n$ is a stationary process with the stationary 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 loss 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 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 that is not only coherent but also informative on the extent of losses greater 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 exceedances over a high threshold to model the excessive losses and then carries out a parametric inference within the framework of the generalized Pareto distributions. Recently, Scaillet (2004) has proposed a nonparametric kernel estimator and applied it to sensitivity analysis in the context of portfolio allocation.

An advantage of the nonparametric method is that it is 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 the tail part of the loss distribution. However, data are generally sparse in the tail and hence finding a proper parametric loss model that 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 dependence on the variance of ES estimation can be clearly spelled out in the variance formula. This is different from the extreme-value approach as the latter effectively treats high exceedances as independent 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), 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. The other 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. It was hoped that the kernel smoothing would produce a more accurate estimator, like the case of VaR estimation studied by Chen and Tang (2005).

A main finding of the current paper is that the variance and the mean square error of the kernel estimator proposed by Scaillet (2004) is not necessarily smaller than those of the sample weighted average estimator. This is because the second order variance term of the kernel ES estimator vanishes instead of taking a negative value as in the case of VaR estimator. This indicates no meaningful variance reduction due to the kernel smoothing. As kernel smoothing introduces a bias, the lack of variance reduction makes the smoothing not worthwhile as the overall mean square error increases. Another finding is that the weighted average estimator has the same asymptotic variance as the kernel estimator. Therefore, for estimation of the ES, the sample weighted average of excessive losses is attractive as it is easy to compute as far as point estimation is concerned. 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 that these different effects of kernel smoothing happen is that the unconditional ES is effectively a mean parameter, which can be estimated accurately by simple averaging.

The paper is structured as follows. We introduce the two nonparametric ES estimators in Section 1. Their statistical properties are discussed in Section 2. Variance estimation for the purpose for supplying standard errors for the ES estimates is 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.

1 NONPARAMETRIC ESTIMATORS

The first nonparametric estimator of the ES considered in this paper is

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

which is a weighted average of excessive losses larger than $\hat{\nu}_p$ where $I(\cdot)$ is the indicator function, $\hat{\nu}_p = Y_{([n(1-p)]+1)}$ is the sample VaR (quantile) estimator of ν_p and $Y_{(r)}$ is the rth order statistic of $\{Y_t\}_{t=1}^{n}$.

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 by 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 Equation (3), Scaillet (2004) proposed the following kernel estimator

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

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.

The commonly employed stochastic models in financial data modeling and risk assessment can generate data to which the proposed ES estimation may be applied. These models include the linear process

$$Y_t = \sum_{s=0}^{\infty} g_{t-s} \xi_s$$

with independent and identically distributed innovation $\{\xi_s\}_{s=0}^{\infty}$; the Markov process

$$Y_t = m(\overline{Y}_{t-1,p}) + \sigma(\overline{Y}_{t-1,p})\epsilon_t$$

where $\overline{Y}_{t-1,p} = (Y_{t-1}, \ldots, Y_{t-p})$ are p-lagged values of Y_t and $\{\epsilon_t\}_{t=1}^{T}$ are independent and identically distributed random variables, and $m(\cdot)$ and $\sigma^2(\cdot)$ are respectively the conditional mean and volatility functions of Y_t given $\overline{Y}_{t-1,p}$; the GARCH (p, q) model

$$E(Y_t|\overline{Y}_{t-1}) = 0 \quad \text{and} \quad \text{Var}(Y_t|\overline{Y}_{t-1}) = h_t =: c + \sum_{i=1}^{q} \alpha_i Y_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j}$$

where c, α_i, and β_j are all positive parameter; as well as the continuous-time diffusion models and the stochastic volatility models.

[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 (2005). See Fan and Yao (2003) for other applications of the kernel method for nonlinear time series analysis.

2 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 $\{Y_t, k \leq t \leq l\}$ 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 a comprehensive discussion. 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 stationary distribution F of the stationary process $\{Y_t\}$ is absolutely continuous with probability density f which has continuous second derivatives in $\mathcal{B}(\nu_p)$, a neighborhood of ν_p; for $k \geq 1$, F_k, the joint distribution functions of (Y_1, Y_{k+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 $\int_{-1}^1 uK(u)du = 0$ 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; some of them are listed at the end of last section. For instance Carrasco and Chen (2002) 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 to conditions imposed by other authors.

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

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

Assumption (i) and the Davydov inequality (see Bosq, 1998) 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,

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

This is a Bahadur-type expansion (Bahadur, 1966) which leads to the following theorem regarding the asymptotic normality of $\hat{\mu}_p$.

Theorem 1. *Under conditions (i) and (ii), as $n \to \infty$*

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

This theorem indicates that the asymptotic variance of $\hat{\mu}_p$ is $\sigma_0^2(p;n)/(np^2)$, 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 expansion (6). The dependence in the original time series is reflected in the asymptotic variance through the covariance in $\sigma_0^2(p;n)$. This means that we need to accommodate the dependence in further statistical inference for the shortfall estimation; see Section 3 for estimation of the variance. We note also that the effective sample size for the ES estimation is np^2. 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) and (iv), as $n \to \infty$*

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

and furthermore,

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

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

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 $\hat{\mu}_{p,h}$ converge 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 story different 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 a bias that leads to an overall increase in the mean square error. Therefore, for the purpose of estimating the ES, the kernel smoothing is counterproductive. 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 estimated well by the sample mean.

It should be noted that our above conclusion 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 as shown in the next section will play a significant role in estimating $\sigma_0^2(p;n)$. For estimation of conditional ES (Scaillet, 2005), smoothing is needed due to the involvement of conditioning variables.

3 STANDARD ERRORS

In this section we introduce a method of obtaining standard errors for the non-parametric ES estimates considered earlier. Although it has not been advised for point estimation of ES, smoothing is needed for variance estimation so as to supply standard errors for the ES estimates. A similar approach is used in Chen and Tang (2005) for obtaining standard errors for VaR estimates.

Let ϕ be the spectral density of $\{(Y_t - \nu_p)I(Y_t \geq \nu_p)\}$. From Brockwell and Davis (1991),

$$\lim_{n\to\infty} n\mathrm{Var}\left\{n^{-1}\sum_{i=1}^{n}(Y_t - \nu_p)I(Y_t \geq \nu_p)\right\} = 2\pi\phi(0)$$

which means that the leading order $\mathrm{Var}(\hat{\mu}_p)$ is $2\pi\phi(0)(np^2)^{-1}$. Hence, the key is estimating $\phi(0)$.

Let $Z_t = (Y_t - \hat{\nu}_p)I(Y_t \geq \hat{\nu}_p)$ for $t = 1, \ldots, n$. We propose estimating $\phi(0)$ by smoothing a set of sample periodograms close to the zero frequency of $\{Z_t\}_{t=1}^{n}$. One may use $G_h(Y_t - \nu_p)$ to replace $I(Y_t \geq \hat{\nu}_p)$ in order to reduce the variability in the estimation of $\phi(0)$. Let

$$I_n(\omega_j) = n^{-1}\left|\sum_{t=1}^{n} Z_j e^{-it\omega_j}\right|^2, \quad j = 0, \pm 1, \ldots, \pm[n/2] \tag{11}$$

be the sample periodograms at frequency $\omega_j = 2\pi j/n \in [-\pi, \pi]$ for $j \in T = \pm 1, \ldots, \pm[n/2]$.

Let $W_j = \log\{I_n(\omega_j)/(2\pi)\} + 0.57721$ and $m(\omega) = \log\{\phi(\omega)\}$. Following the lines of Fan and Gijbels (1996) and Chen and Tang (2005), a Nadaraya–Waston kernel estimator of $m(0)$ based on a symmetric kernel K_1 and a smoothing bandwidth λ is

$$\hat{m}_\lambda(\omega) = \frac{\sum_{j\in T} K_1\left(\frac{\omega-\omega_j}{\lambda}\right) W_j}{\sum_{j\in T} K_1\left(\frac{\omega-\omega_j}{\lambda}\right)} \tag{12}$$

where $\lambda \to 0$ and $n\lambda \to \infty$ as $n \to \infty$. Then, an estimator of $\phi(0)$ is $\hat{\phi}(0) = \exp\{\hat{m}_\lambda(0)\}$.

An important issue here is the selection of λ. An objective function we may use in guiding the bandwidth selection is to minimize

$$R(\lambda) = \frac{1}{n} \sum_{j \in T} q_{nj} \{m(\omega_j) - \hat{m}_\lambda(\omega_j)\}^2 \tag{13}$$

by defining weights $q_{nj} = I(|j| \leq [k_n])$ where k_n is an n-dependent integer. We choose $k_n = [0.05n]$, which means that only the 10% sample periodograms close to the zero frequency are considered. It may be shown that an unbiased estimate of $R(\lambda)$ is

$$r(\lambda) = \frac{1}{n} \sum_{j \in T} q_{nj} \{W_j - \hat{m}_\lambda(\omega_j)\}^2 - \frac{\pi^2}{6} \left(1 - \frac{4\pi K(0)}{n\lambda}\right) \sum_{j \in T} q_{nj}. \tag{14}$$

Ignoring the term not involving λ, the object function to be minimized for λ selection is

$$\frac{1}{n} \sum_{j \in T} q_j \{W_j - \hat{m}_\lambda(\omega_j)\}^2 + \frac{2\pi^3 K(0)}{3n\lambda} \sum_{j \in T} q_j.$$

The proposed standard error estimation method with the proposed bandwidth selection will be applied in analyses of some financial data sets in Section 5.

4 SIMULATION STUDY

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

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

$$\text{an AR(1) model:} \quad Y_t = 0.5Y_{t-1} + \epsilon_t, \epsilon_t \overset{iid}{\sim} N(0, 1); \tag{15}$$

$$\text{an ARCH(1) model:} \quad Y_t = 0.5Y_{t-1} + \epsilon_t, \quad \epsilon_t^2 = 4 + 0.4\epsilon_{t-1}^2 + \eta_t, \quad \eta_t \overset{iid}{\sim} N(0, 1). \tag{16}$$

We are interested in estimating $\mu_{0.01}$, the 99% ES. In constructing the kernel estimator, the Gaussian kernel $K(u) = (1/\sqrt{2\pi}) \exp(-u^2/2)$ 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 bias, 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 bias, 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. One feature that is worth noting from Figures 1 and 2 is that a large bandwidth increased both the variance and MSE of the kernel ES estimator. At the same time, the impact of a large bandwidth on the bias was quite

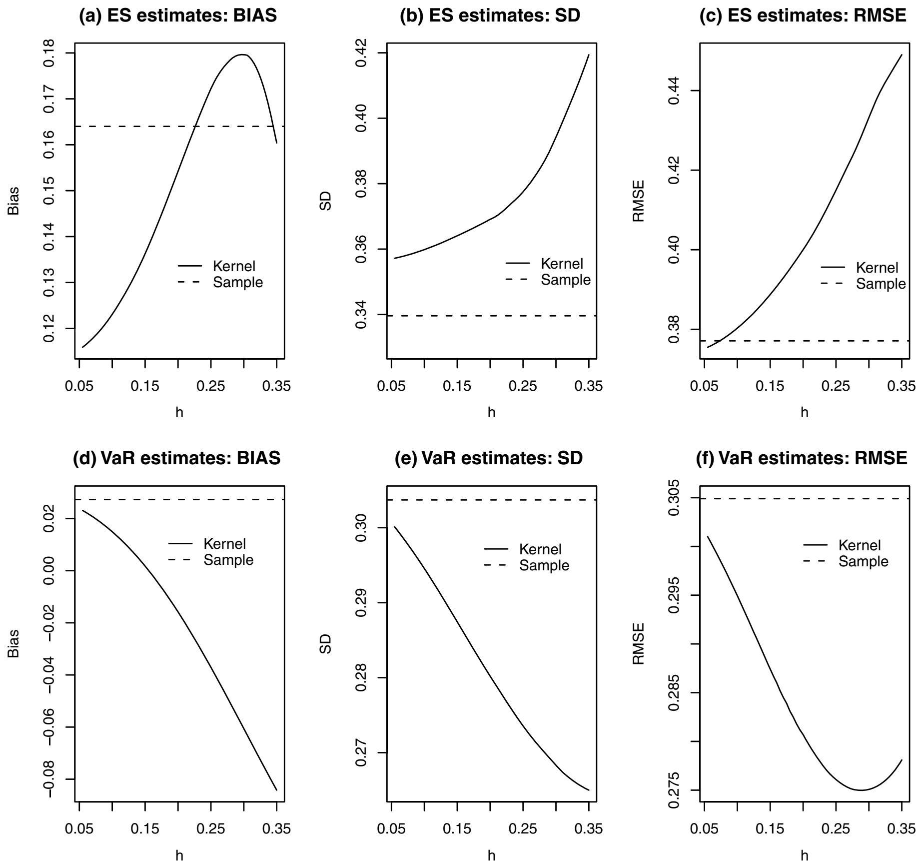

Figure 1 Simulated average standard deviation (SD) and root mean square error (RMSE) of the kernel 99% ES estimator $\hat{\mu}_{0.01,h}$ in (a) and (b) and 99% kernel VaR estimator $\hat{\nu}_{0.01,h}$ in (c) and (d), and their unsmoothed (with legend sample) counterparts $\hat{\mu}_{0.01}$ in (a) and (b) and $\hat{\nu}_{0.01}$ in (c) and (d) for the AR model with $n = 250$. And $\mu_{0.01} = 3.078$ and $\nu_{0.01} = 2.686$.

limited as shown by the drop of the bias for large h. The main revelation of the simulation is that $\hat{\mu}_{p,h}$ has a larger variance and, to a large extent, a larger MSE 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

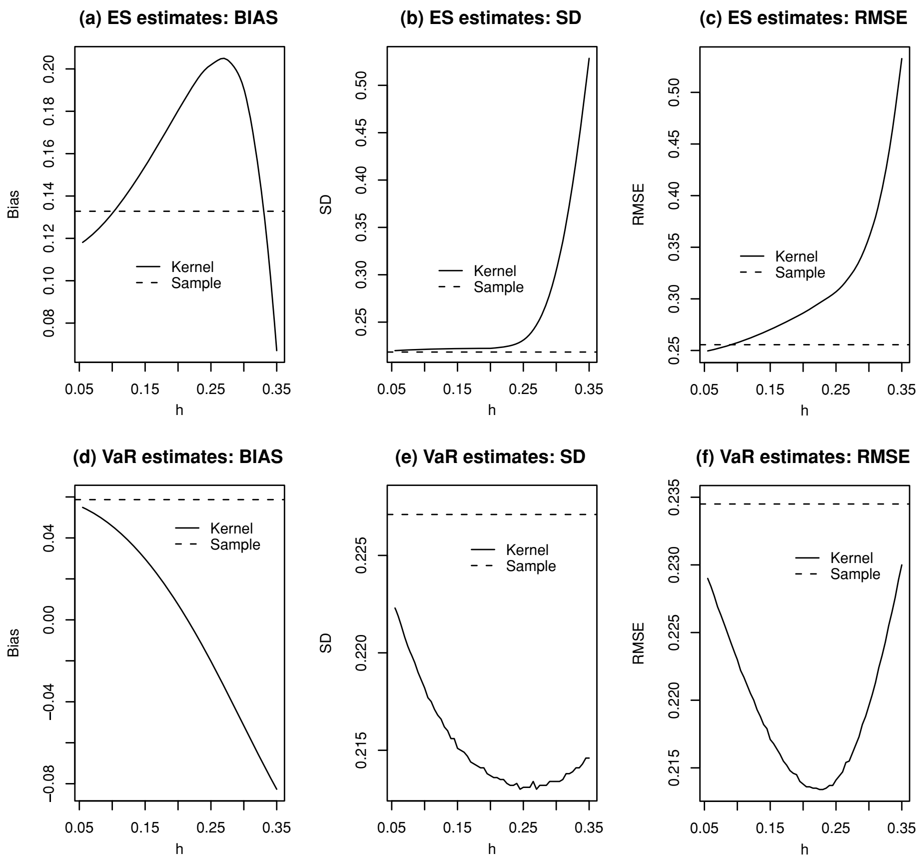

Figure 2 Simulated average standard deviation (SD) and root mean square error (RMSE) of the kernel 99% ES estimator $\hat{\mu}_{0.01,h}$ in Panels (a) and (b) and 99% kernel VaR estimator $\hat{\nu}_{0.01,h}$ in Panels (c) and (d), and their unsmoothed (with legend sample) counterparts $\hat{\mu}_{0.01}$ in (a) and (b) and $\hat{\nu}_{0.01}$ in (c) and (d) for the ARCH model with $n = 250$. And $\mu_{0.01} = 5.8042$ and $\nu_{0.01} = 5.6647$.

sample autocorrelation functions (ACFs). 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 autocorrelation 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% level of significance.

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 years (2001–2003), respectively. Table 1 presents the ES estimates $\hat{\mu}_{0.01}$ and their standard errors. The standard errors were obtained by using the approach outlined in Section 3. 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

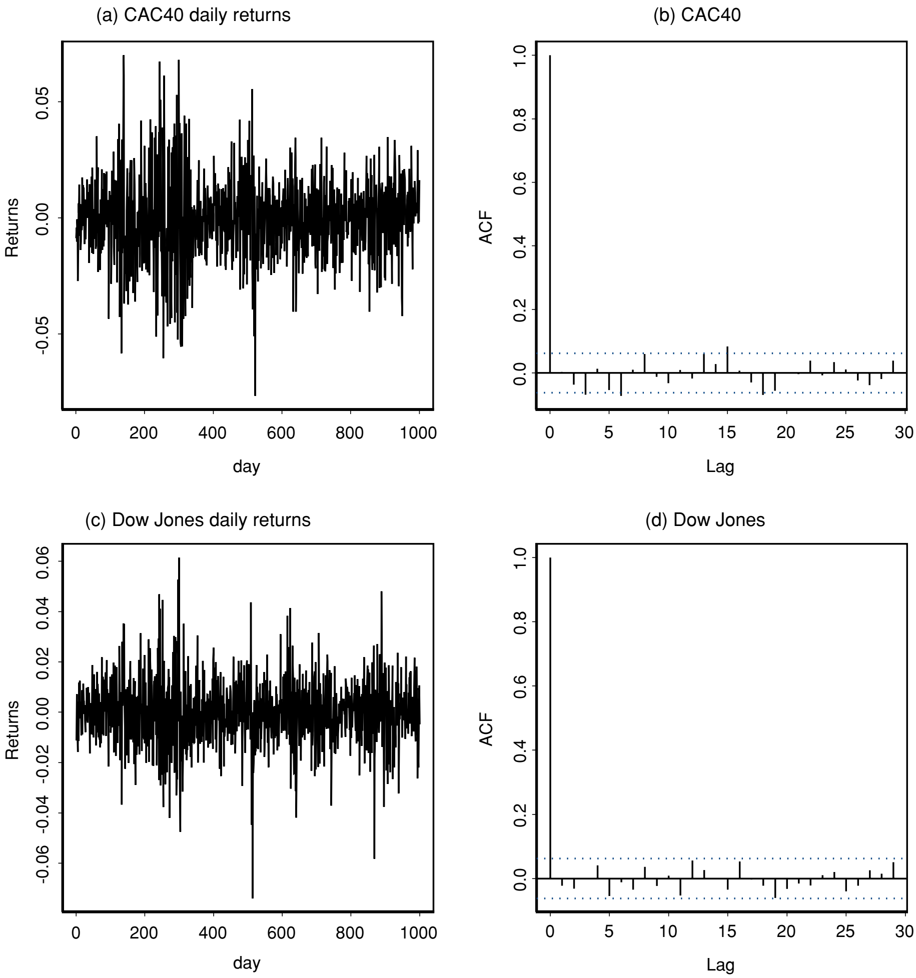

Figure 3 The two financial return series in (a) and (c) and their sample autocorrelation functions (ACF) in (b) and (d).

after the setback in the ".com" business and the September 11. The level of risk settled 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 their counterparts in Dow Jones. The variability of the ES estimate for Dow was much higher than that of CAC in the year 2001–2002; and the situation reversed in the second year when the ES estimates of CAC became more variable. We observed as expected that the variability for the ES estimates based on the entire 2-year observations were smaller than those of each individual year (Table 1).

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

Table 1 Estimates for $\nu_{0.01}$, $\mu_{0.01}$, and standard errors (SE).

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

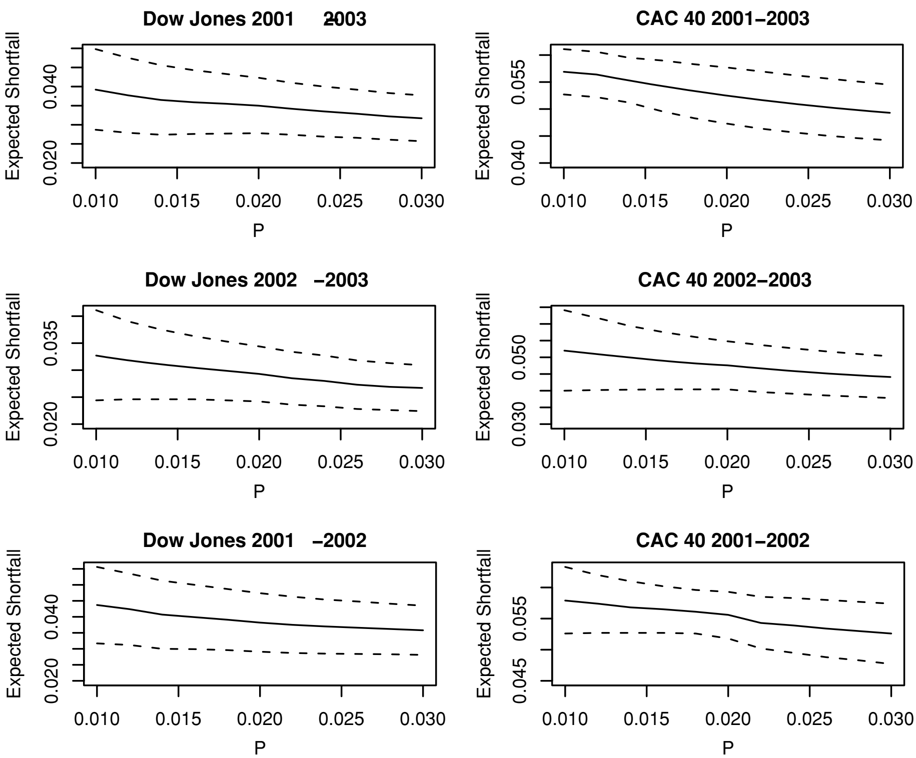

Figure 4 Expected shortfall estimates and their confidence bands for the two financial return series.

in Figure 4. The confidence bands were constructed by adding and subtracting 1.96 times the standard errors. These plots show that, as expected, the ES estimate declined as p increased. For both indices, the year 2001–2002 experienced larger risk than the year 2002–2003. It reveals again that the CAC was more volatile that Dow Jones as the ES estimates were always larger than those of Dow for each of the three time periods and at each fixed p level.

APPENDIX: PROOFS

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), $P(|\hat{\nu}_p - \nu_p| \geq \epsilon_n) \to 0$ exponentially fast as $n \to \infty$.*

Proof. We only give the proof for $\tilde{\nu}_p = \hat{\nu}_p$ as that for $\hat{\nu}_{p,h}$ 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

$$P(|\hat{\nu}_p - \nu_p| \geq \epsilon_n) \leq P\{|F_n(\nu_p + \epsilon_n) - F(\nu_p + \epsilon_n)| > C_1\epsilon_n\} + P\{|F_n(\nu_p - \epsilon_n) - F(\nu_p - \epsilon_n)| > C_1\epsilon_n\}. \tag{A.1}$$

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

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

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

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

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

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

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

Lemma 2. *Under the Conditions (i) and (ii) and for any $\kappa > 0$,*

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

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 $E(W_t) =: -I_{t_1} + I_{t_2}$ where

$$I_{t_1} = E\{(Y_t - \nu_p)I(\nu_p \leq Y_t < \hat{\nu}_p)I(\hat{\nu}_p > \nu_p)\} \quad \text{and}$$
$$I_{t_2} = E\{(Y_t - \nu_p)I(\hat{\nu}_p \leq Y_t < \nu_p)I(\hat{\nu}_p < \nu_p)\}.$$

Furthermore, let $I_{t_1} = I_{t_{11}} + I_{t_{12}}$ and $I_{t_2} = I_{t_{21}} + I_{t_{22}}$ where, for $a \in (0, 1/2)$ and $\eta > 0$,

$$\begin{aligned} I_{t_{11}} &= E\{(Y_t - \nu_p)I(\nu_p \le Y_t < \hat{\nu}_p)I(\hat{\nu}_p \ge \nu_p + n^{-a}\eta)\}, \\ I_{t_{12}} &= E\{(Y_t - \nu_p)I(\nu_p \le Y_t < \hat{\nu}_p)I(\nu_p < \hat{\nu}_p < \nu_p + n^{-a}\eta)\}, \\ I_{t_{21}} &= E\{(Y_t - \nu_p)I(\nu_p > Y_t \ge \hat{\nu}_p)I(\hat{\nu}_p \le \nu_p - n^{-a}\eta)\} \quad \text{and} \\ I_{t_{22}} &= E\{(Y_t - \nu_p)I(\nu_p > Y_t \ge \hat{\nu}_p)I(\nu_p > \hat{\nu}_p > \nu_p - n^{-a}\eta)\}. \end{aligned}$$

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

$$|I_{t_{k1}}| \le \sqrt{E(\hat{\nu}_p - \nu_p)^2 P(|\hat{\nu}_p - \nu_p| \ge n^{-a}\eta)}.$$

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

$$I_{t_{k1}} \to 0 \quad \text{exponentially fast.} \tag{A.5}$$

To evaluate $I_{t_{12}}$, we note that $|I_{t_{12}}| \le E\{(Y_t - \nu_p)I(\nu_p \le Y_t < \nu_p + n^{-a}\eta)\}$. This means

$$I_{t_{12}} \le \int_{\nu_p}^{\nu_p + n^{-a}\eta} dv(z - \nu_p)f(z)dz = O(n^{-2a}).$$

Using exactly the same approach we can show that $I_{t_{22}} = O(n^{-2a})$ 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

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

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

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

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 $\text{Var}(W_t) = o(n^{-3/2+\kappa})$. By slightly modifying the above derivation for $\text{Var}(W_t)$, it may be shown that for any t_1, t_2 $\text{Cov}(W_{t_1}, W_{t_2}) = o(n^{-3/2+\kappa})$. Therefore,

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

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

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

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),

$$A =: |E[(\hat{\beta} - \beta)\{p - S_h(\nu_p)\}\{\hat{f}(\nu_p) - f(\nu_p)\}]$$
$$\le n^{-2}\sum_{i\ge 1, j\ge 1, i+j\le n} |E\{\psi_1(Y_1)\psi_2(Y_i)\psi_3(Y_{i+j})\}|[6] + O(n^{-1}h^4 + n^{-2}h^2) \quad \text{(A.9)}$$

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

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

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

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

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

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

From (A.9) and (A.10), and the fact that $\alpha(k)$ is monotonic non-increasing,

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

since $\sum j\alpha\delta/(2+\delta)(j) < \infty$ as implied by Condition (i). ■

Lemma 4. *Under the conditions (i)–(v) and for* $l_1, l_2 = 0$ *or* 1,

$$\Bigg|\sum_{k=1}^{n-1}(1-k/n)\big[\mathrm{Cov}\big\{Y_1^{l_1}G_h(\nu_p - Y_1), Y_{k+1}^{l_2}G_h(\nu_p - Y_{k+1})\big\} \\ - \mathrm{Cov}\big\{Y_1^{l_1}I(Y_1 > \nu_p), Y_{k+1}^{l_2}I(Y_{k+1} > \nu_p)\big\}\big] = o(h).$$

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 $\phi_1(t) = n^{-1}\sum_{i=1}^n Y_t I(Y_t \ge t)$ and $\phi_2(t) = n^{-1}\sum_{i=1}^n I(Y_t \ge t)$. 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,

$$\phi_1(\hat\nu_p) = \phi_1(\nu_p) + \nu_p\{\phi_2(\hat\nu_p) - \phi_2(\nu_p)\} + o_p\big(n^{-3/4+\kappa}\big). \tag{A.11}$$

These lead to

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

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

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 $r(k+k') \le n < r(k+k'+1)$. Define the large blocks

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

the smaller blocks

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

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 $n \to \infty$,

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

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 (see Bosq, 1998), 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

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

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

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

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 $\mathrm{Var}(W_{j,n}) = E(V_{j,n}^2) \le C_2 k$. Thus,

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

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: $\hat{\nu}_{p,h} - \nu_p = (S_h(\nu_p) - p)/f(\nu_p) + o_p(n^{-1/2})$. 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

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

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 $\mathrm{Cov}\{(np)^{-1}\sum Y_t K_h(\nu_p - Y_t), \hat{\nu}_{p,h} - \nu_p\} = O(n^{-1})$. Hence, from (A.18),

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

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,

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

It is easy to see that

$$\mathrm{Var}\left\{(np)^{-1}\sum Y_t G_h(\nu_p - Y_t)\right\} = n^{-1}p^{-2}\Big[\mathrm{Var}\{Y_t G_h(\nu_p - Y_t)\} + 2\sum_{k=1}^{n-1}(1-k/n) \times \mathrm{Cov}\{Y_1 G_h(\nu_p - Y_1), Y_{k+1}G_h(\nu_p - Y_{k+1})\}\Big].$$

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

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

Equation (A.23) and Lemma 3 mean

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

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

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

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

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

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

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

Applying the Cauchy–Schwartz inequality and Lemma 3,

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

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

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

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

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

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

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

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

Substituting (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

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

which establishes (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.

Received July 12, 2006; revised April 16, 2007; accepted September 24, 2007.

REFERENCES

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

Bahadur, R. R. (1966). “A note on quantiles in large samples.” *The Annals of Mathematical Statistics 37*, 577–580.

Bellini, F., and G. Figá-Talamanca. (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.

Brockwell, P. J., and R. A. Davis. (1991). *Time Series: Theory and Methods*. New York: Springer.

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

Carrasco, M., and X. Chen. (2002). "Mixing and Moment Properties of Various GARCH and Stochastic Volatility Models." *Econometrics Theory 18*, 17–39.

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

Duffie, D., and J. Pan. (1997). "An overview of value at risk." *Journal of Derivative 4*, 7–49.

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

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

Fan, J., and I. Gijbels. (1996). *Local Polynomial Modeling and Its Applications*. London: Chapman and Hall.

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

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

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

Frey, R., and A. J. McNeil. (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., O. Scaillet, and J. P. Laurent. (2000). "Sensitivity analysis of Values at Risk." *Journal of Empirical Finance 7*, 225–245.

Jorion, P. (2001). *Value at Risk*, 2nd ed. New York: McGraw-Hill.

Leadbetter, M. R., G. Lindgren, and H. Rootzén. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Berlin: Springer.

Rosenblatt, M. (1956). "A central limit theorem and a the α-mixing condition." *Proceedings of the National Academy of Science, USA 42*, 43–47.

Scaillet, O. (2004). "Nonparametric estimation and sensitivity analysis of expected shortfall." *Mathematical Finance 14*, 115–129.

Scaillet, O. (2005). "Nonparametric estimation of conditional expected shortfall." *Insurance and Risk Management Journal 74*, 639–660.

Sheather, S. J., and J. S. Marron. (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.