4-2004

Nonparametric Estimation of Expected Shortfall

Song Xi Chen
Iowa State University, songchen@iastate.edu

Nonparametric Estimation of Expected Shortfall

Abstract
The paper evaluates the properties of nonparametric estimators of the expected shortfall, an increasingly popular risk measure in financial risk management. It is found that the existing kernel estimator based on a single bandwidth does not offer variance reduction, which is surprising considering that kernel smoothing reduces the variance of estimators for the value at risk and the distribution function. We reformulate the kernel estimator such that two different bandwidths are employed in the kernel smoothing for the value at risk and the shortfall itself. We demonstrate by both theoretical analysis and simulation studies that the new kernel estimator achieves a variance reduction. The paper also covers the practical issues of bandwidth selection and standard error estimation.

Keywords
kernel estimator, risk measures, smoothing bandwidth, value at risk, weak dependence

Disciplines
Statistics and Probability

Comments
This preprint was published as Song Xi Chen, "Nonparametric Estimation of Expected Shortfall", *Journal of Financial Econometrics* (2008): 87-107, doi: 10.1093/jjfinec/nbm019

Nonparametric Estimation of Expected Shortfall[1]

SONG XI CHEN

Department of Statistics, Iowa State University

Email:songchen@iastate.edu
Phone: (515)- 294 2729

[1]The author thanks Mr Cheng Yong Tang for computational support and acknowledge support of an National University of Singapore Academic Research Grant (R-155-000-026-112).

Running Head: Nonparametric Estimation of Expected Shortfall.

Corresponding author: Song Xi Chen
Department of Statistics, Iowa State University, Ames, IA 50011-1210
Phone: 1-515-294 2729; Email:songchen@iastate.edu

Web page for download the paper:
http://www.public.iastate.edu/%7esongchen/shortfall4a.pdf

ABSTRACT. The paper evaluates the properties of nonparametric estimators of the expected shortfall, an increasingly popular risk measure in financial risk management. It is found that the existing kernel estimator based on a single bandwidth does not offer variance reduction, which is surprising considering that kernel smoothing reduces the variance of estimators for the value at risk and the distribution function. We reformulate the kernel estimator such that two different bandwidths are employed in the kernel smoothing for the value at risk and the shortfall itself. We demonstrate by both theoretical analysis and simulation studies that the new kernel estimator achieves a variance reduction. The paper also covers the practical issues of bandwidth selection and standard error estimation.

Key Words: Kernel estimator; Risk Measures; Smoothing bandwidth; Value at Risk; Weak dependence.

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 of 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 weakly 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 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 amount of the excessive losses apart from 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 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

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

Relative to VaR, there are less statistical analyzes of ES mainly due to its late entry to the field of financial risk management, despite that it has been used by actuaries for some time. The method commonly used by actuaries was fully parametric based on a parametric loss distribution. 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) is a semiparametric approach which uses the asymptotic distribution of excedence over a high threshold to model the excessive loses and then carries out a parametric inference within the framework of the Generalized Pareto distributions. Scaillet (2003a) proposes a nonparametric kernel estimator and applies it for sensitivity analysis in the context of portfolio allocation. An advantage of the kernel method is its being fully nonparametric and hence is model robust and avoids potential model misspecification. Another advantage is that it allows a wide range of data dependence, which makes it adaptable in the context of financial losses. The kernel method has been applied to various aspects of time series analysis and financial econometrics as summarized in the recent book by Fan and Yao (2003).

In this paper we consider two nonparametric estimators of μ_p based on two estimators of ν_p respectively. The first estimator is just a simple conditional average of excessive losses over the sample quantile estimator of ν_p. The other is a kernel weighted conditional average of excessive losses larger than a kernel estimator of ν_p, which is designed to bring more data into the inference by smoothing. The goal is to achieve variance and mean square error reduction in the estimation of μ_p. Our analysis reveals that the kernel estimator considered in Scaillet (2003) does not necessarily offer a variance reduction. This is surprising considering that kernel smoothing leads to smaller variance in quantile and conditional quantile estimation for both independent (Falk, 1984; Sheather and Marron, 1990) and dependent (Cai and Roussas, 1997; Cai, 2002) observations.

The lack of variance reduction of the existing kernel estimator is due to using the same bandwidth in both the kernel VaR estimation and the kernel weighted averaging of extreme losses. To achieve a variance reduction, we reformulate the kernel estimator so that the bandwidth used to construct the weighted conditional average of excessive losses is different from that for the VaR estimation. Although like quantile and distribution function estimation the variance reduction is of the second order, the reduction can be significant in financial terms as, say a 10% , reduction can translate to a large reduction of provision in the absolute dollar term. It should be highlighted that the estimation of μ_p and ν_p is highly volatile as they are parameters defined on the tail of the loss distribution where data information is scarce. Therefore, a variance reduction in the second order is still very meaningful. The proposed kernel estimator can be applied to estimation of conditional expected shortfall in the context of Scaillet (2003b).

The paper is structured as follows. We introduce the two nonparametric estimators of the ES in Section 2. Their statistical properties are discussed in Section 3. Section 4 addresses the issue of bandwidth selection. Simulation results are reported in Section 6 followed by empirical studies in Section 7. All the technical details are given in the appendix.

2. NONPARAMETRIC ESTIMATORS

In this section we formulate two nonparametric estimators of μ_p. Let $Y_{(r)}$ is the r-th order statistic of the original negative log returns and $\hat{\nu}_p = Y_{([n(1-p)]+1)}$ be the sample quantile estimator of ν_p, which is called the historical VaR estimator in empirical finance.

The first nonparametric estimator of μ_p considered in this paper is a simple conditional average of excessive losses larger than $\hat{\nu}_p$:

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

where $I(\cdot)$ is the indicator function. To introduce the second estimator, we note that

$$\mu_p = E(Y_t | Y_t \geq \nu_p) = p^{-1} \int_{\nu_p}^{\infty} z f(z) dz. \tag{4}$$

where f is the stationary density of Y_t.

Let K be a kernel function, which is a symmetric probability density function. The standard kernel density estimator of f is $\hat{f}_h(z) = n^{-1} \sum_{t=1}^{n} K_h(z - Y_t)$ where $K_h(z) = h^{-1}K(z/h)$ and $h > 0$ is the smoothing bandwidth which controls the smoothness of the fitted curve. Let $G(t) = \int_t^{\infty} K(u)du$ and $G_h(t) = G(t/b)$. The kernel estimator of the survival function $S(x) = 1 - F(x)$ is

$$S_h(z) = n^{-1} \sum_{t=1}^{n} G_h(z - Y_t) \tag{5}$$

A kernel estimator of ν_p, denoted as $\hat{\nu}_{p,h}$ is the solution of $S_h(z) = p$, which is the estimator proposed in Gourieroux, Laurent and Scaillet (2000). Chen and Tang (2003) consider the statistical properties of $\hat{\nu}_p$ and $\hat{\nu}_{p,h}$ in the context of dependent financial returns. [2]

The kernel estimator considered in Scaillet (2003) is

$$\tilde{\mu}_{p,h} = (np)^{-1} \sum_{t=1}^{n} Y_t G_h(\hat{\nu}_{p,h} - Y_t) \tag{6}$$

[2] On related research works, Cai (2002) considers kernel estimation of conditional VaR, Fan, Gu and Zhou (2003) treat semi-parametric estimation of VaR via exponential smoothing.

which replaces the indicator function in (3) by the smoother G_h and utilizes the fact that $n^{-1}\sum_{t=1}^{n} G_h(\hat{\nu}_{p,h} - Y_t) = p$. However, our analysis given in the next section shows that $\tilde{\mu}_{p,h}$ does not have a smaller variance than $\hat{\mu}_p$. Hence the benefits of the kernel smoothing is not realized, which is contrary to the VaR estimation where the kernel estimator $\hat{\nu}_{p,h}$ reduces the variance of $\hat{\nu}_p$. The lack of variance reduction by $\tilde{\mu}_{p,h}$ can be explained by the fact that the ES is a parameter defined upon another parameter ν_p and requires different amount of smoothing from that used in the kernel estimation of ν_p. In fact our analysis shows that the benefits of the kernel estimation of ν_p cancels out with those in the kernel smoothing of the excessive losses if the same bandwidth is employed.

To achieve a variance reduction, we construct a new estimator by first using a bandwidth b which is different from h to obtain the kernel VaR estimator $\hat{\nu}_{p,b}$. Then, we replace f, ν_p and p by $\hat{f}_h(z)$, $\hat{\nu}_{p,b}$ and $S_h(\hat{\nu}_{p,b})$ respectively in (4) to obtain

$$\hat{\mu}_{p,h,b} = \frac{\sum_{t=1}^{n} \int_{\hat{\nu}_{p,b}}^{\infty} zK_h(z - Y_t)dz}{\sum_{t=1}^{n} \int_{\hat{\nu}_{p,b}}^{\infty} K_h(z - Y_t)dz}.$$

We note here that replacing p by $S_h(\hat{\nu}_{p,b})$ in the denominator is to provide a better weighting in the case of using two different bandwidths. By defining $G_{1h}(t) = \int_{t/h}^{\infty} uK(u)du$,

$$\hat{\mu}_{p,h,b} = \frac{\sum_{t=1}^{n} \{Y_t G_h(\hat{\nu}_{p,b} - Y_t) + hG_{1h}(\hat{\nu}_{p,b} - Y_t)\}}{\sum_{t=1}^{n} G_h(\hat{\nu}_{p,b} - Y_t)}. \quad (7)$$

3. MAIN RESULTS

In this section we evaluate the theoretical properties of the ES estimators $\hat{\mu}_p$, $\tilde{\mu}_{p,h}$ and $\hat{\mu}_{p,h,b}$. 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.

We assume the following conditions:

(i) There exists a $\rho \in (0, 1)$ such that $\alpha(k) \le C\rho^k$ for all $k \ge 1$.

(ii) $E(Y_t^{2+\delta}) \leq C$ for some $\delta > 0$; f has continuous second derivatives in $\mathcal{B}(\nu_p)$, a neighborhood of ν_p; F_k, the joint distribution function of (Y_1, Y_{k+1}), has all its second partial derivatives bounded in $\mathcal{B}(\nu_p)$.

(iii) Let r be either b or h. We assume $r \to 0$, $nr^{3-\beta} \to \infty$ for any $\beta > 0$ and $nr^4 \log^2(n) \to 0$ as $n \to \infty$.

(iv) K is a univariate symmetric probability density function satisfying the moment conditions $\int_{-1}^{1} uK(u)du = 0$ and $\int_{-1}^{1} u^2 K(u)du = \sigma_k^2$.

Condition (i) means that the time series is geometric α-mixing, which is known to be 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 case for ARCH model. Conditions (ii) and (iv) are standard regularity conditions. Condition (iii) specifies a range of allowable bandwidths h and b which includes $O(n^{-1/3})$, the optimal order that minimizes the mean square error of $\hat{\mu}_{p,h,b}$.

The following theorem establishes a Bahadur type expansion for $\hat{\mu}_p$.

Theorem 1. Under conditions (i) and (ii), for an arbitrarily small positive κ,

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

From the above expansion, it may be derived by employing the delta method that

$$E(\hat{\mu}_p) = \mu_p + O(n^{-3/4}) \quad \text{and} \quad Var(\hat{\mu}_p) = n^{-1}p^{-1}\sigma_0^2(p;n) + o(n^{-1}) \tag{8}$$

where $\sigma_0^2(p;n) = \{Var\{(Y_1 - \nu_p)I(Y_1 \geq \nu_p)\} + 2\sum_{k=1}^{n-1}\gamma(k)\}$ and $\gamma(k) = Cov\{(Y_1 - \nu_p)I(Y_1 \geq \nu_p), (Y_{k+1} - \nu_p)I(Y_{k+1} \geq \nu_p)\}$ for positive integers k. Assumption (i) and the Davydov inequality imply that $\sigma_0^2(p,n)$ is finite.

Theorem 2. Under conditions (i)-(iv),

$$Bias(\hat{\mu}_{p,h,b}) = -\tfrac{1}{2}p^{-1}\sigma_k^2 h^2\{(\nu_p + \mu_p)f'(\nu_p) - f(\nu_p)\} + \tfrac{1}{2}p^{-1}\sigma_k^2 b^2 f'(\nu_p)(\nu_p + \mu_p) + o(h^2 + b^2) \quad \text{and} \tag{9}$$

$$Var(\hat{\mu}_{p,h,b}) = p^{-1}n^{-1}\sigma_0^2(p,n) - 2n^{-1}p^{-2}(\nu_p - \mu_p)^2 f(\nu_p) \times \tag{10}$$

$$\times[b\{c_k(1) - c_k(b/h)\} + h\{c_k(1) - c_k(h/b)\}] + o(n^{-1}h). \tag{11}$$

where $c_k(t) = \int_{-\infty}^{\infty} uK(u)du \int_{-\infty}^{tu} K(v)dv$.

It is noted that both $\hat{\mu}_p$ and $\hat{\mu}_{p,h,b}$ share the same leading order variance term. We note also that for any $t > 0$

$$t\{c_k(1) - c_k(t)\} + \{c_k(1) - c_k(t^{-1})\} \geq 0$$

and the equality is achieved only at $t = 1$. Thus, for any positive b and h as long as $h \neq b$, there is a second order variance reduction in the kernel estimator relative to $\hat{\mu}_p$. However, if $h = b$, $b\{c_k(1) - c_k(b/h)\} + \{c_k(1) - c_k(h/b)\} = 0$, which means no variance reduction for the estimator $\tilde{\mu}_{p,h}$ as confirmed formally in the following corollary whose proof is contained in that of Theorem 2 and hence will not be specially specified.

Corollary 1. Under conditions (i)-(iv),

$$\begin{aligned} Bias(\tilde{\mu}_{p,h}) &= \tfrac{1}{2}p^{-1}\sigma_k^2 h^2 f(\nu_p) + o(h^2) \quad \text{and} & (12)\\ Var(\tilde{\mu}_{p,h}) &= p^{-1}n^{-1}\sigma_0^2(p,n) + o(n^{-1}h). & (13) \end{aligned}$$

To gather empirical evidence for the corollary, we carry out simulation under an AR and an ARCH model whose details are given in (19) and (20) in Section 5 as part of a large scale simulation study. Figures 1 and 2 display the variance and mean square errors of $\tilde{\mu}_{p,h}$ 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 ES and VaR estimators $\hat{\mu}_p$ and $\hat{\nu}_p$. Although the sample size considered in these figures is 250, the same pattern of results are observed for the sample sizes of 500 as well. It is striking to see that $\tilde{\mu}_{p,h}$ has larger variance and, to a large extent, larger mean square error too than $\hat{\mu}_p$ for both models. This means that smoothing with a single bandwidth for ES estimation is actually counter-productive. In contrast, the kernel VaR estimator $\hat{\nu}_{p,h}$ delivers both variance and mean square error reduction as predicted by the established theory.

Combine (9) and (10), the mean square error of $\hat{\mu}_{p,h,b}$ is

$$\begin{aligned} MSE(\hat{\mu}_{p,h,b}) &= p^{-1}n^{-1}\sigma_0^2(p,n)\} + \tfrac{1}{4}p^{-2}\sigma_k^4\{(h^2 - b^2)(\nu_p + \mu_p)f'(\nu_p) - h^2 f(\nu_p)\}\} \\ &- 2n^{-1}p^{-2}(\nu_p - \mu_p)^2 f(\nu_p)[b\{c_k(1) - c_k(b/h)\} + h\{c_k(1) - c_k(h/b)\}] \\ &+ o(n^{-1}h + h^4 + b^2). \qquad (14) \end{aligned}$$

The optimal b and h that minimize the second order terms of $MSE(\hat{\mu}_{p,h,b})$ is

$$h^\star = t_0^{-1} b^\star \qquad (15)$$

$$\begin{aligned} b^\star &= 2^{2/3} n^{-1/3} (\nu_p - \mu_p)^{2/3} \sigma_k^{-4/3} \{(\nu_p + \mu_p) f^{'}(\nu_p)\}^{-2/3} \{c_k(1) - c_k(t_0)\}^{1/3} \times \\ & \left\{ \frac{(\nu_p + \mu_p) f^{'}(\nu_p)}{f(\nu_p) - (\nu_p + \mu_p) f^{'}(\nu_p)} \left(\frac{c_k(1) - c_k(t_0)}{c_k(1) - c_k(t_0^{-1})} \right)^3 + \left(\frac{c_k(1) - c_k(t_0)}{c_k(1) - c_k(t_0^{-1})} \right) \right\}^{-1/3} \end{aligned} \quad (16)$$

where, by letting $\beta = \{f(\nu_p) - (\nu_p + \mu_p) f^{'}(\nu_p)\}\{(\nu_p + \mu_p) f^{'}(\nu_p)\}^{-1}$, t_0 is the solution of

$$t = \beta \frac{c_k(1) - c_k(t^{-1})}{c_k(1) - c_k(t)} \quad (17)$$

which is solvable upon given K and β. We note that in the context of the loss distribution, $f^{'}(\nu_p)$ should be negative, both ν_p and μ_p are positive and $f(\nu_p)$ is usually small relative to $|(\nu_p + \mu_p) f^{'}(\nu_p)|$. Hence, β should be within $(-1, 0)$. We note also that $\{c_k(1) - c_k(t_0)\}\{c_k(1) - c_k(t_0^{-1})\} < 0$. These collectively mean that t_0 is positive. It can be checked along the same line that both $b^\star$ and $h^\star$ are positive as well.

Substituting $h^\star$ and $b^\star$ into (14), it can be readily shown that the kernel estimator $\hat{\mu}_{p,h,b}$ reduces the mean squares error of $\hat{\mu}_p$ by an amount of order $n^{-4/3}$.

4. BANDWIDTH SELECTION

The expressions in (15), (16) and (17) can be used to obtain practically useful bandwidths by plugging-in estimates of ν_p, $f(\nu_p)$ and $f^{'}(\nu_p)$. We can for the sake of bandwidth selection to employ the extreme value theory approach by approximating the conditional density of Y_t given that Y_t is larger than a high threshold η by a density of a Generalized Pareto (GP) distribution as we are concerned with extreme losses. This idea is similar to the reference to a standard distribution approach for bandwidth selection in nonparametric curve estimation.

Let

$$w_{\gamma,\eta,\sigma}(x) = \frac{1}{\sigma}(1 + \gamma \frac{x - \eta}{\sigma})^{-(1+\frac{1}{\gamma})} I(\eta \le x < \eta_1), \quad (18)$$

be the density of a GP distribution with a scale parameter σ, a shape parameter γ, and $\eta_1 = \infty$ if $\gamma > 0$ and $\eta_1 = \mu + \sigma/|\gamma|$ if $\gamma < 0$. For a 99% ES, we can fit the upper five percent of the data to a GP distribution, which means taking $\eta = \nu_{0.05}$. For other levels of ES, η should be adjusted accordingly. Let $\hat{\eta} = \hat{\nu}_{0.05}$, in the case of 99% ES, and $\hat{\sigma}$ and $\hat{\gamma}$ be the method of moment estimators under the GP distribution (Reiss and Thomas, 2001). Then, the estimates of $f(\nu_p)$ and $f^{'}(\nu_p)$ are respectively $0.05 w_{\hat{\gamma},\hat{\sigma},\hat{\nu}_{0.05}}(\hat{\nu}_p)$ and $0.05 w^{'}_{\hat{\gamma},\hat{\sigma},\hat{\nu}_{0.05}}(\hat{\nu}_p)$. Substituting the estimates of ν_p, $f(\nu_p)$ and $f^{'}(\hat{\nu}_p)$ gives an estimate $\hat{\beta}$ of β, which then leads to solving t_0 in equation (17). Finally we plug in $\hat{\mu}_p$, $\hat{\nu}_p$ and the estimates of $f(\hat{\nu}_p)$ and $f^{'}(\hat{\nu}_p)$ into (15) and (16) to obtain the plug-in bandwidths.

5. SIMULATION STUDIES

In this section we report results from a simulation study that evaluates the performance of the nonparametric ES estimators. The main objective is to confirm the results in Theorem 2, which indicates that the kernel smoothing with two different bandwidths provides more accurate inference than the unsmoothed estimator $\hat{\mu}_p$.

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

$$\text{an AR(1) model:} \quad Y_t = 0.5Y_{t-1} + \epsilon_t, \epsilon_t \overset{iid}{\sim} N(0,1); \tag{19}$$

$$\text{an ARCH(1) model:} \quad Y_t = 0.5Y_{t-1} + \epsilon_t, \quad \epsilon_t^2 = 4 + 0.4\epsilon_{t-1}^2 + \eta_t, \quad \eta_t \overset{iid}{\sim} N(0,1). \tag{20}$$

We are interested in estimating $\mu_{0.01}$, the 99% ES. In constructing the kernel estimators, the Gaussian kernel $K(u) = \frac{1}{\sqrt{2\pi}}\exp(-u^2/2)$ is employed. The bandwidth h and b are chosen within a large region specified in Figure 3 to gain information on the effect of bandwidths on the performance of the kernel estimators. The sample size considered in the simulation are 250 and 500. The number of simulation is 1000. These are also the settings for the results displayed in Figures 1 and 2 in Section 3.

Figure 3 contains contour plots of the root mean square errors (RMSE) of the kernel ES estimator $\hat{\mu}_{p,h,b}$ for the above two models. The titles of the plots contain the RMSE of the unsmoothed estimator $\hat{\mu}_p$. The results in Figure 3 can be summarized as follows. First of all, the kernel estimator has a smaller RMSE than $\hat{\mu}_p$ over a wide range of bandwidth combinations. The reduction in RMSE ranges from 10% to 15% depending on the sample size and the model. The direction of the valley in these contours is almost the same for all the models considered, and indicates a negative linear relationship between h and b. This is consistent with the theoretical finding that the kernel estimator prefers two different bandwidths. We did not observe an increase of RMSE alone $h = b$ as happened to $\tilde{\mu}_{p,h}$ in Figures 1 and 2. This is probably due to the second term in the numerator of (7) which adds favorable effect on the performance despite it is of a higher order in theory.

6. EMPIRICAL STUDY

We apply the proposed kernel estimator $\hat{\mu}_{p,h,b}$ 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 exactly 500 observations (2 years data). The log-return series are displayed in Figure 4 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 three separate analyzes on each of the series, which are based on the first year data (2001-2002), the second year data (2002-2003), and the entire two year data (2001-2003), respectively. The kernel estimates of the return density for each period of the two series are displayed in Figure 5 using the default kernel and bandwidth in Splus.

At $p = 0.01$, the bandwidth selection approach described in Section 4 is used to choose h and b, whereas the standard errors of the ES estimates are obtained by implementing the method discussed in Section 5. The prescribed bandwidths and the associated intermediate results are listed in Table 1. Table 2 presents the ES estimates $\hat{\mu}_{0.01}$ and $\hat{\mu}_{0.01,h,b}$ 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 closely to the standard error estimation for kernel VaR estimation treated in Chen and Tang (2003). 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 had higher risk than Dow Jones as the estimates of ES and Var were all larger than her counterparts in Dow Jones. The variability of the ES estimate for the Dows was 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 $\hat{\mu}_{p,h,b}$ and their 95% confidence bands are displayed in Figure 6. 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: Technical Details

The proofs of Theorems 1 and 2 require the following lemmas, whose proofs are presented first. Throughout this section we use C, C_1 ... to denote positive constants which may take different values.

Lemma 1. Let $\tilde{\nu}_p$ be either $\hat{\nu}_p$ or $\hat{\nu}_{p,b}$ and $\epsilon_n \to 0$ and $n\epsilon_n^2 \to \infty$ as $n \to \infty$, and assume $\{Y_t\}$ is geometrically α-mixing, that is $\alpha(k) \leq c\rho^k$ for $c > 0$ and $\rho \in [0, 1)$, then we have $P\left(|\tilde{\nu}_p - \nu_p| \geq \epsilon_n\right) \to 0$ exponentially fast as $n \to 0$.

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

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

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

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

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)-(iii) and for any $\kappa > 0$,

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

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

$$\begin{aligned} I_{t1} &= E\{(Y_t - \nu_p)I(\nu_p \leq Y_t < \hat{\nu}_p)I(\hat{\nu}_p > \nu_p)\} \quad \text{and} \\ I_2 &= E\{(Y_t - \nu_p)I(\hat{\nu}_p \leq Y_t < \nu_p)I(\hat{\nu}_p < \nu_p)\}. \end{aligned}$$

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

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

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

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

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

$$|I_{t12}| \leq E\{(Y_t - \nu_p)I(\nu_p \leq Y_t < \nu_p + n^{-a}\eta)\}.$$

This means

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

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

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

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

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

Note that

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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 $l_1 = l_2 = 0$ can be proved by has been proved in Chen and Tang (2003) 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 $\phi_2(\hat{\nu}_p) = p + O_p(n^{-1})$. From Lemma 2, for an arbitrarily small positive κ,

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

These lead to

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

Proof of Theorem 2

We first derive (9). From derivations given in Chen and Tang (2003), $\hat{\nu}_{p,b}$ admits an expansion: $\hat{\nu}_{p,b} - \nu_p = \frac{S_b(\nu_p)-p}{\hat{f}_b(\nu_p)} + o_p(n^{-1/2})$. Let $m_\nu = E(\hat{\nu}_{p,b})$. From the bias of $\hat{\nu}_{p,b}$ given in Chen and Tang (2003),

$$m_\nu - \nu_p = \tfrac{1}{2}\sigma_k^2 f'(\nu_p) f^{-1}(\nu_p) b^2 + o(b^2). \tag{A.13}$$

By Taylor expansion

$$S_h(\hat{\nu}_{p,b}) = S_h(\nu_p) - \hat{f}_h(\nu_p)(\hat{\nu}_{p,b} - \nu_p) + O_p(n^{-1}). \tag{A.14}$$

Let $m_S = E\{S_h(\hat{\nu}_{p,b})\}$. From (A.13)

$$m_S = S(m_\nu) + \tfrac{1}{2} f'(m_\nu)\sigma_k^2 h^2 + o(h^2) = p + \tfrac{1}{2}\sigma_k^2 f'(\nu_p)(h^2 - b^2) + o(h^2 + b^2). \tag{A.15}$$

Moreover, from (A.14),

$$\begin{aligned} S_h(\hat{\nu}_{p,b}) - m_S &= S_h(\nu_p) - m_S - \hat{f}_h(\nu_p)(\hat{\nu}_{p,b} - \nu_p) + O_p(n^{-1}) \\ &= n^{-1}\sum_{t=1}^n G_h(\nu_p - Y_t) - p - \hat{f}_h(\nu_p) f_{n,b}^{-1}(\nu_p)\{n^{-1}\sum_{t=1}^n G_b(\nu_p - Y_t) - p\} \\ &\quad + O_p(n^{-1} + h^2 + b^2) \\ &= n^{-1}\sum_{t=1}^n \{G_h(\nu_p - Y_t) - G_b(\nu_p - Y_t)\} + O_p(n^{-1} + h^2 + b^2) \end{aligned} \tag{A.16}$$

These and (7) mean the kernel ES estimator

$$\begin{aligned} \hat{\mu}_{p,h,b} &= (nm_S)^{-1}\sum_{t=1}^n \{Y_t G_h(\hat{\nu}_{p,b} - Y_t) + hG_{1h}(\hat{\nu}_{p,b} - Y_t)\}\{1 - \frac{S_h(\hat{\nu}_{p,b}) - m_S}{m_S}\} + O_p(n^{-1}) \\ &= A_1 - A_2 + A_3 + A_4 + O_p(n^{-1}) \end{aligned}$$

where

$$\begin{aligned} A_1 &= (nm_S)^{-1}\sum_{t=1}^n \{Y_t G_h(\nu_p - Y_t) - Y_t K_h(\nu_p - Y_t)(\hat{\nu}_{p,b} - \nu_p)\}, \\ A_2 &= m_S^{-2} n^{-1}\sum_{t=1}^n Y_t G_h(\nu_p - Y_t) n^{-1}\sum_{t=1}^n \{G_h(\nu_p - Y_t) - G_b(\nu_p - Y_t)\}, \\ A_3 &= (nm_S)^{-1} h\sum_{t=1}^n G_{1h}(\nu_p - Y_t)\Big[1 - m_S^{-1} n^{-1}\sum_{t=1}^n \{G_h(\nu_p - Y_t) - G_b(\nu_p - Y_t)\}\Big], \\ A_4 &= (nm_S)^{-1} h\sum_{t=1}^n G'_{1h}(\nu_p - Y_t)(\hat{\nu}_{p,b} - \nu_p). \end{aligned} \tag{A.17}$$

Note that $A_1 = (nm_S)^{-1}\sum_{t=1}^n Y_t G_h(\nu_p - Y_t) - (nm_S)^{-1}\sum_{t=1}^n Y_t K_h(\nu_p - Y_t)(\hat{\nu}_{p,b} - \nu_p)$ and

$$\begin{aligned} E\Big[(np)^{-1}\sum\{Y_t G_h(\nu_p - Y_t)\}\Big] &= p^{-1}\int zG_h(\nu_p - z)f(z)dz \\ &= p^{-1}\int_{-\infty}^{\infty} K(u)du\{\int_{\nu_p}^{\infty} zf(z)dz + \int_{\nu_p - hu}^{\nu_p} zf(z)dz\} \\ &= \mu_p - \tfrac{1}{2}p^{-1}h^2\sigma_k^2\{v_p f'(\nu_p) + f(\nu_p)\} + O(h^3). \qquad (A.18) \end{aligned}$$

Let $\eta = E\{m_S^{-1} Y_t K_h(Y_t - \nu_p)\} = p^{-1}\int(\nu_p - hu)K(u)f(\nu_p - hu)du = p^{-1}\nu_p f(\nu_p) + O(h^2)$. As $Cov\{(np)^{-1}\sum Y_t K_h(\nu_p - Y_t), \hat{\nu}_{p,b} - \nu_p)\} = O(n^{-1})$,

$$\begin{aligned} E\{(np)^{-1}\sum Y_t K(\nu_p - Y_t)(\hat{\nu}_{p,b} - \nu_p)\} &= \eta E(\hat{\nu}_{p,b} - \nu_p) + O(n^{-1}) \\ &= -\tfrac{1}{2}p^{-1}\nu_p f'(\nu_p)b^2\sigma_k^2 + O(hb^2 + n^{-1}). \, (A.19) \end{aligned}$$

Combine (A.15), (A.18) and (A.19),

$$E(A_1) \;=\; \mu_p + \tfrac{1}{2}p^{-1}\sigma_k^2\Big[\nu_p f'(\nu_p)b^2 - \{\nu_p f'(\nu_p) + f(\nu_p)\}h^2\Big] + o(h^2 + b^2).$$

Using the same technique, we have

$$E(A_2) \;=\; \tfrac{1}{2}p^{-1}\mu_p\sigma_k^2 f'(\nu_p)(h^2 - b^2) + o(h^2 + b^2) \quad \text{and}$$

$$E(A_3) = p^{-1}h^2\sigma_k^2 f(\nu_p) + o(h^2 + b^2) + O(n^{-1}h^2),$$

It may be shown that

$$E\{h(np)^{-1}\sum G'_{1h}(\nu_p - Y_t)\} \;=\; -h^2p^{-1}\int_{-\infty}^{\infty} uK(u)f(\nu_p - hu)du = O(h^3) \quad (A.20)$$

which leads to $E(A_4) = O(h^5 + n^{-1})$. In summary,

$$\begin{aligned} E(\hat{\mu}_{p,h,b}) &= \mu_p - \tfrac{1}{2}p^{-1}\sigma_k^2h^2\{\nu_p f'(\nu_p) + f(\nu_p)\} + \tfrac{1}{2}p^{-1}\sigma_k^2 b^2\nu_p f'(\nu_p) \\ &- \tfrac{1}{2}p^{-1}\mu_p\sigma_k^2 f'(\nu_p)(h^2 - b^2) + o(h^2 + b^2) + O(n^{-1}) \end{aligned}$$

which derives (9).

We now turn to the derivation of (10). We would like to establish first that

$$Cov(A_i, A_j) = o(n^{-1}h) \quad \text{for } i = 3 \text{ and } 4, j = 1, 2, 3 \text{ and } 4. \qquad (A.21)$$

As h appears in both A_3 and A_4, it is easily seen that

$$Cov(A_i, A_j) = O(n^{-1}h^2) \quad \text{for } i, j = 3 \text{ and } 4. \tag{A.22}$$

Let $\hat{\xi} = h(nm_S)^{-1}\sum G_{1h}(\nu_p - Y_t)$ and $\xi = E(\hat{\xi})$. As $\xi = O(h^2)$ and note that $\eta = E\{(np)^{-1}\sum_{t=1}^{n} Y_t G_h(\nu_p - Y_t)\}$. Hence

$$\begin{aligned} Cov(A_1, A_3) &= Cov\{(nm_S)^{-1}\sum_{t=1}^{n} Y_t G_h(\nu_p - Y_t), h(nm_S)^{-1}\sum G_{1h}(\nu_p - Y_t)\} \\ &- \eta Cov\{\hat{\nu}_{p,b}, h(nm_S)^{-1}\sum G_{1h}(\nu_p - Y_t)\} + o(n^{-1}h). \end{aligned}$$

Note that for $l = 0$ or $l \geq 1$, $Cov\{Y_1 G_h(\nu_p - Y_1), G_{1h}(\nu_p - Y_{l+1})\} = O(h)$. It may be shown by using the α-mixing condition and the Davydov inequality that

$$\begin{aligned} & Cov\{(nm_S)^{-1}\sum_{t=1}^{n} Y_t G_h(\nu_p - Y_t), h(np)^{-1}\sum G_{1h}(\nu_p - Y_t)\} \\ = {} & h(np^2)^{-1}\Big[Cov\{Y_1 G_h(\nu_p - Y_1), G_{1h}(\nu_p - Y_1)\} \\ & +2\sum_{l=1}^{n-1}(1 - l/n)Cov\{Y_1 G_h(\nu_p - Y_1), G_{1h}(\nu_p - Y_{l+1})\}\Big]\{1 + O(h^2)\} = O(n^{-1}h^2). \end{aligned} \tag{A.23}$$

Using the same arguments, we have $Cov\{\hat{\nu}_{p,b}, h(nm_S)^{-1}\sum G_{1h}(\nu_p - Y_t)\} = O(n^{-1}hb)$. Hence, $Cov(A_1, A_3) = o(n^{-1}h)$. Similarly, $Cov(A_2, A_3) = o(n^{-1}h)$ and $Cov(A_i, A_4) = o(n^{-1}h)$ for $i = 1$ and 2. These and (A.22) imply (A.21), which in turn means that

$$Var(\hat{\mu}_{p,b,h}) = Var(A_1) + Var(A_2) - 2Cov(A_1, A_2) + o(n^{-1}h). \tag{A.24}$$

From the definition of A_1,

$$\begin{aligned} Var(A_1) &= Var\{(nm_S)^{-1}\sum Y_t G_h(\nu_p - Y_t)\} + Var\{\hat{\eta}(\hat{\nu}_{p,b} - \nu_p)\} \\ &- 2Cov\{(nm_S)^{-1}\sum Y_t G_h(\nu_p - Y_t), \hat{\eta}(\hat{\nu}_{p,b} - \nu_p)\}. \end{aligned} \tag{A.25}$$

It is easy to see that

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

A detail analysis reveals that

$$\begin{aligned} Var\{Y_t G_h(\nu_p - Y_t)\} &= \int z^2 G_h^2(\nu_p - z) f(z) dz - p^2 \mu_p^2 + O(h^2) \\ &= \int_{-\infty}^{\infty} K(u) du \Big[\int_{-\infty}^{u} K(v) dv \{ \int_{\nu_p}^{\infty} z^2 f(z) dz + \int_{\nu_p - hv}^{\nu_p} z^2 f(z) dz \} \\ &\quad + \int_{u}^{\infty} K(v) dv \{ \int_{\nu_p}^{\infty} z^2 f(z) dz + \int_{\nu_p - hu}^{\nu_p} z^2 f(z) dz \} - p^2 \mu_p^2 + O(h^2) \\ &= Var\{Y_t I(Y_t \geq \nu_p)\} - 2h\nu_p^2 f(\nu_p) c_k(1) + O(h^2). \qquad (A.26) \end{aligned}$$

This and Lemma 4 mean

$$Var\{(nm_S)^{-1} \sum Y_t G_h(\nu_p - Y_t)\} = p^{-2} Var\{\phi_1(\nu_p)\} - 2n^{-1} h \nu_p^2 f(\nu_p) c_k(1) + o(n^{-1}h). \quad (A.27)$$

The second term on the right hand side of (A.25) is

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

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

$$\eta^2 Var(\hat{\nu}_{p,b}) = p^{-2} \nu_p^2 Var\{n^{-1} \sum_{t=1}^{n} I(Y_t > \nu_p)\} - 2p^{-2} n^{-1} b \nu_p^2 f(\nu_p) c_k(1) + o(n^{-1}b). \quad (A.28)$$

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

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

Applying the Cauchy-Schwartz inequality and Lemma 3,

$$Var\{(\hat{\eta} - \eta)(\hat{\nu}_{p,b} - \nu_p)\} = O(n^{-2}h^{-3/2}) = o(n^{-1}h) \text{and} \quad (A.29)$$

$$Cov\{\eta(\hat{\nu}_{p,b} - \nu_p), p^{-1}(\hat{\eta} - \eta)(\hat{\nu}_{p,b} - \nu_p)\} = o(n^{-1}h). \quad (A.30)$$

Combine (A.28), (A.29) and (A.30),

$$Var(\hat{\eta}(\hat{\mu}_{p,b} - \nu_p)\} = p^{-2} \nu_p^2 Var\{\phi_2(\nu_p)\} - 2p^{-2} n^{-1} b \nu_p^2 f(\nu_p) c_k(1) + o(n^{-1}b). \quad (A.31)$$

The covariance term on the right hand side of (A.25) is

$$Cov\{(nm_S)^{-1} \sum_{t=1}^{n} Y_t G_h(\nu_p - Y_t), \hat{\eta}(\hat{\mu}_{p,b} - \nu_p)\}$$

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

From Lemma 4 and the fact that

$$
Cov\{Y_t G_h(\nu_p - Y_t), G_b(\nu_p - Y_t)\} = p\mu_p + \nu_p f(\nu_p)\{bc_k(b/h) + hc_k(h/b)\} + o(h+b),
$$

$$
\begin{aligned}
&Cov\{(nm_S)^{-1}\sum_{t=1}^{n} Y_t G_h(\nu_p - Y_t), (nm_S)^{-1}\sum_{i=1}^{n} Y_t K_h(\nu_p - Y_t)(\hat{\nu}_{p,b} - \nu_p)\} \\
&= p^{-2}\nu_p Cov\{\phi_1(\nu_p), \phi_2(\nu_p)\} + n^{-1}p^{-2}\nu_p^2 f(\nu_p)\{bc_k(b/h) + hc_k(h/b)\} + o(h+b) \quad (A.32)
\end{aligned}
$$

Substitute (A.28), (A.31) and (A.32) to (A.25), we have

$$
\begin{aligned}
Var(A_1) &= p^{-1}n^{-1}\sigma_0^2(p,n) - 2\nu_p^2 f(\nu_p)p^{-2}n^{-1}\Big[b\{c_k(1) - c_k(b/h)\} + h\{c_k(1) - c_k(h/b)\}\Big] \\
&\quad + o\{n^{-1}(h+b)\}. \quad (A.33)
\end{aligned}
$$

Employing the same techniques exhibits above, we can show that

$$
\begin{aligned}
Var(A_2) &= p^{-2}\mu_p^2\Big[Var\{S_{n,h}(\nu_p)\} + Var\{S_{n,b}(\nu_p) - 2Cov\{S_{n,h}(\nu_p), S_{n,b}(\nu_p)\}\Big] \\
&= p^{-2}\mu_p^2 n^{-1} f(\nu_p)\Big[\{b\{c_k(1) - c_k(b/h)\} + h\{c_k(1) - c_k(h/b)\}\Big] \text{and} \quad (A.34) \\
Cov(A_1, A_2) &= 2p^{-2}\mu_p\nu_p f(\nu_p)n^{-1}\Big[\{b\{c_k(1) - c_k(b/h)\} + h\{c_k(1) - c_k(h/b)\}\Big] \quad (A.35)
\end{aligned}
$$

It is apparent that (A.33), (A.34) and (A.35) implies (10). □

REFERENCES

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

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.

Cai, Z.-W. and Roussas, G. G. (1997), Smoothed estimate of quantile under association, *Statistics and Probability Letters*, **36**, 275-287.

Cai, Z.-W. and Roussas, G. G. (1998), Efficient estimation of a distribution function under quadrant dependence, *Scandinavian Journal of Statistics*, **25**, 211-224.

Chen, S. X. and Tang, C.Y. (2003). Nonparametric inference of Value at Risk for dependent financial returns. Research report, Department of Statistics, Iowa State University.

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.

Falk, M. (1984), Relative deficiency of kernel type estimators of quantiles, *Annals of Statistics*, **12**, 261-268.

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

Fan, J., Gu, J. and G. Zhou (2003), Semiparametric Estimation of Value-at-Risk, *Econometrics Journal*, to appear.

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* to appear.

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.

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.

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. (2003a), Nonparametric estimation and sensitivity analysis of expected shortfall, *Mathematical Finance* to appear.

Scaillet, O. (2003b), Nonparametric estimation of conditional expected shortfall, HEC Geneve mimeo.

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.

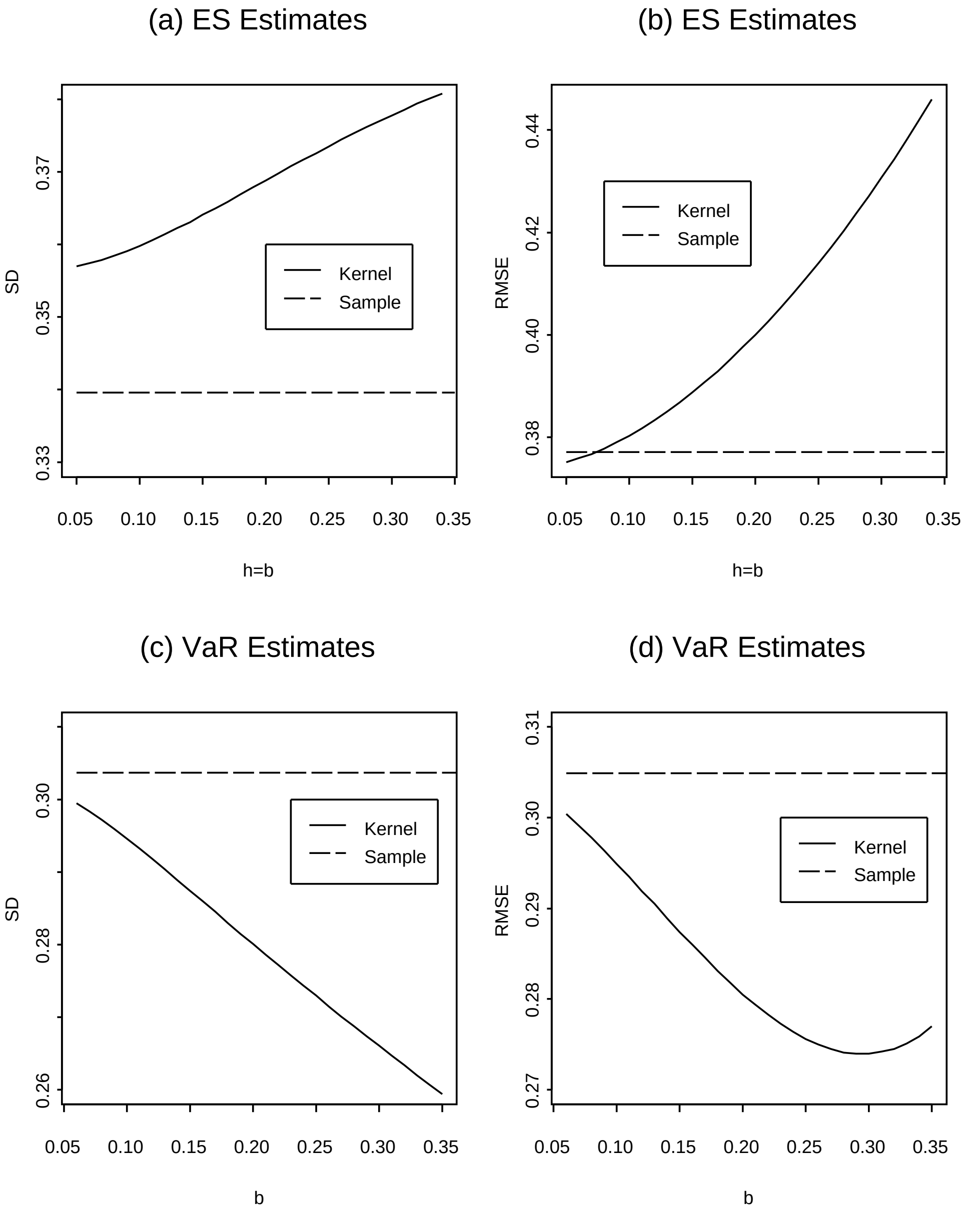


Figure 1. Simulated average standard deviation (SD) and root mean square error (RMSE) of the kernel 99% ES estimator $\hat{\mu}_{0.01,h}$ in Panels (a) and (b) and 99% kernel VaR estimator $\hat{\nu}_{p,b}$ in Panels (c) and (d), and their unsmoothed (legended as sample) counterparts $\hat{\mu}_{0.01}$ in Panels (a) and (b) and $\hat{\nu}_{0.01}$ in Panels (c) and (d) for the AR model (19) with $n = 250$.

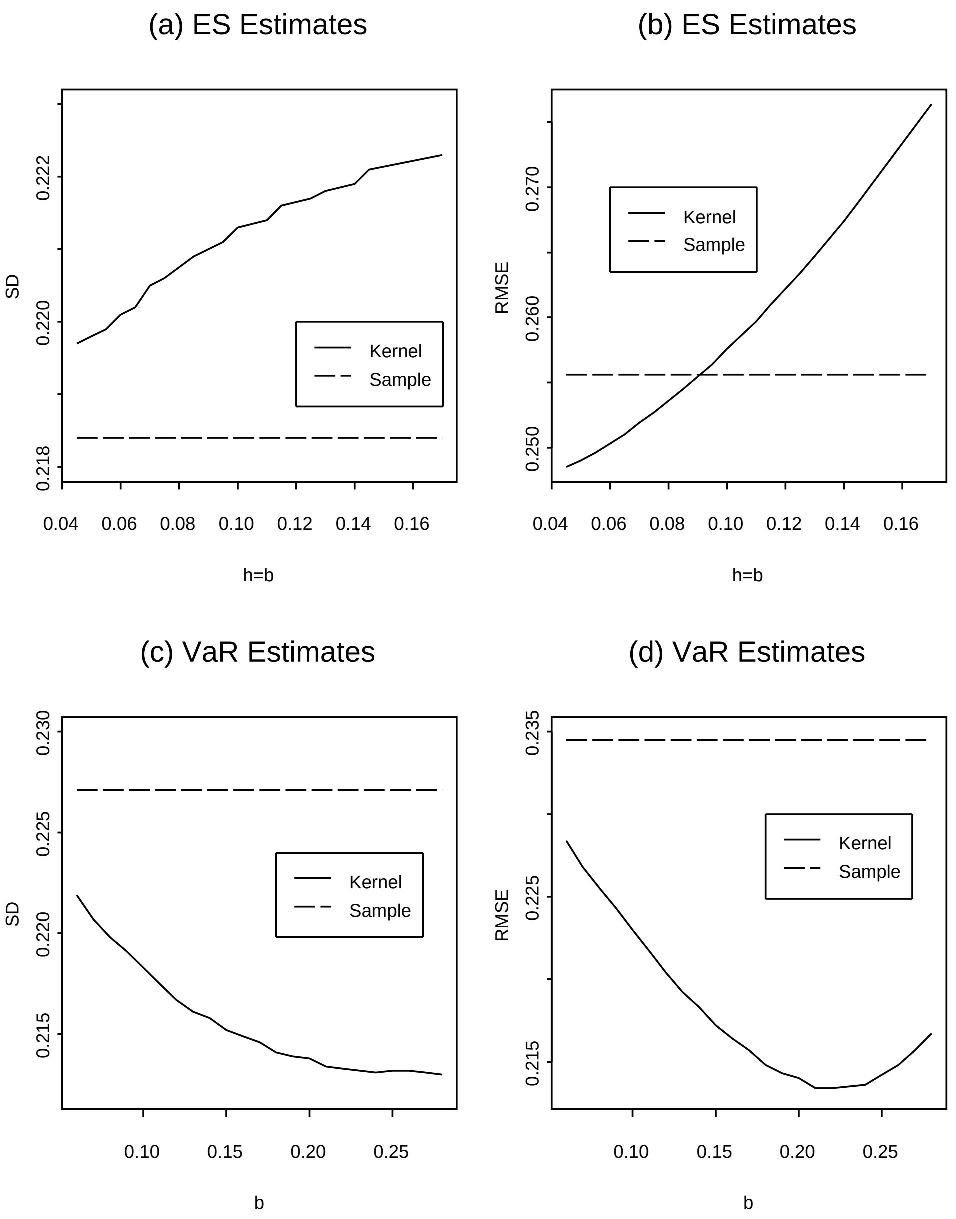


Figure 2. Simulated average standard deviation (SD) and root mean square error (RMSE) of the kernel 99% ES estimator $\hat{\mu}_{0.01,h}$ in Panels (a) and (b) and 99% kernel VaR estimator $\hat{\nu}_{p,b}$ in Panels (c) and (d), and their unsmoothed (legended as sample) counterparts $\hat{\mu}_{0.01}$ in (Panels (a) and (b)) and $\hat{\nu}_{0.01}$ in Panels (c) and (d) for the ARCH model (20) with $n = 250$.

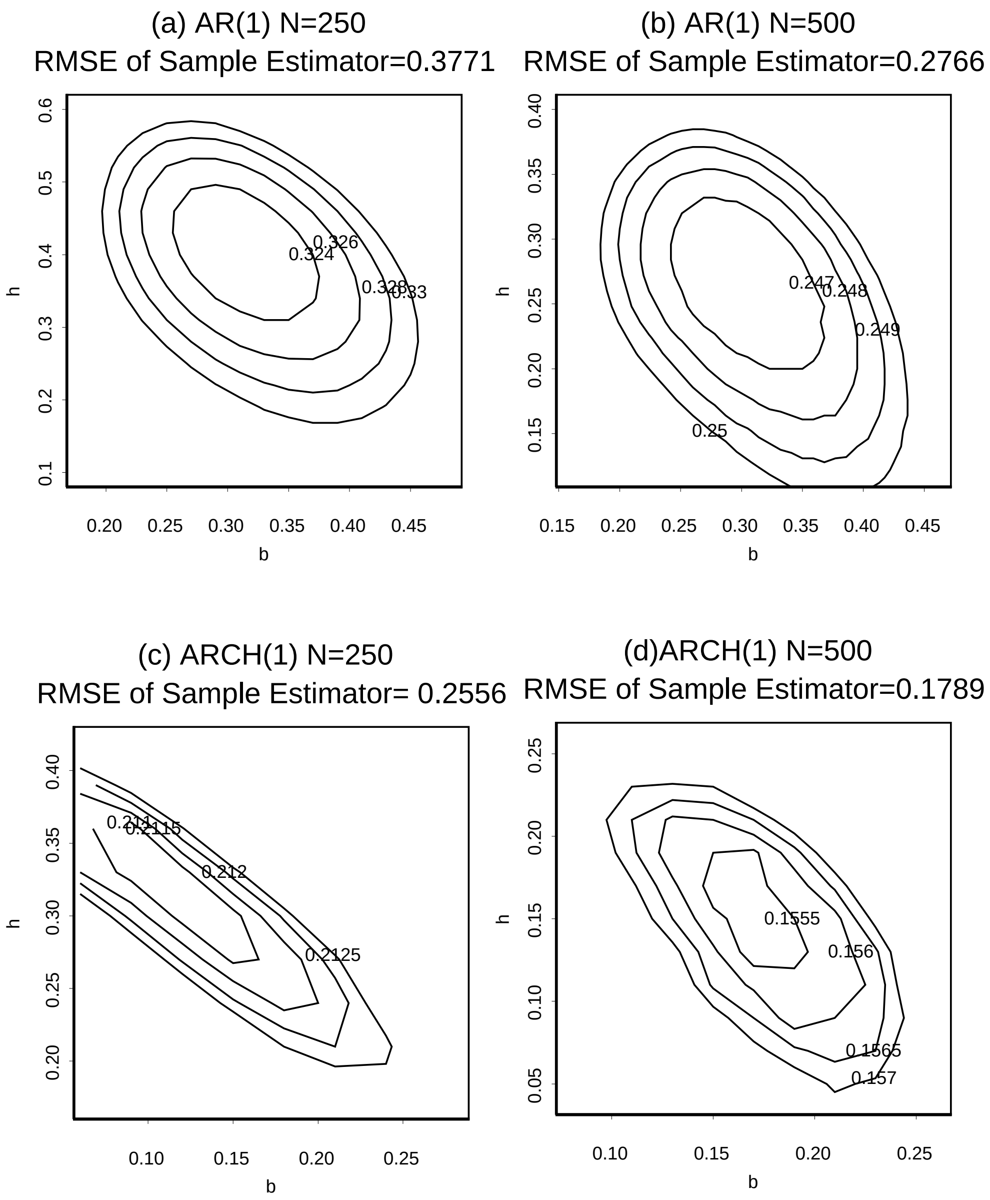


Figure 3. Contours plots of the root mean square errors (RMSE) for the AR(1) and the ARCH models.

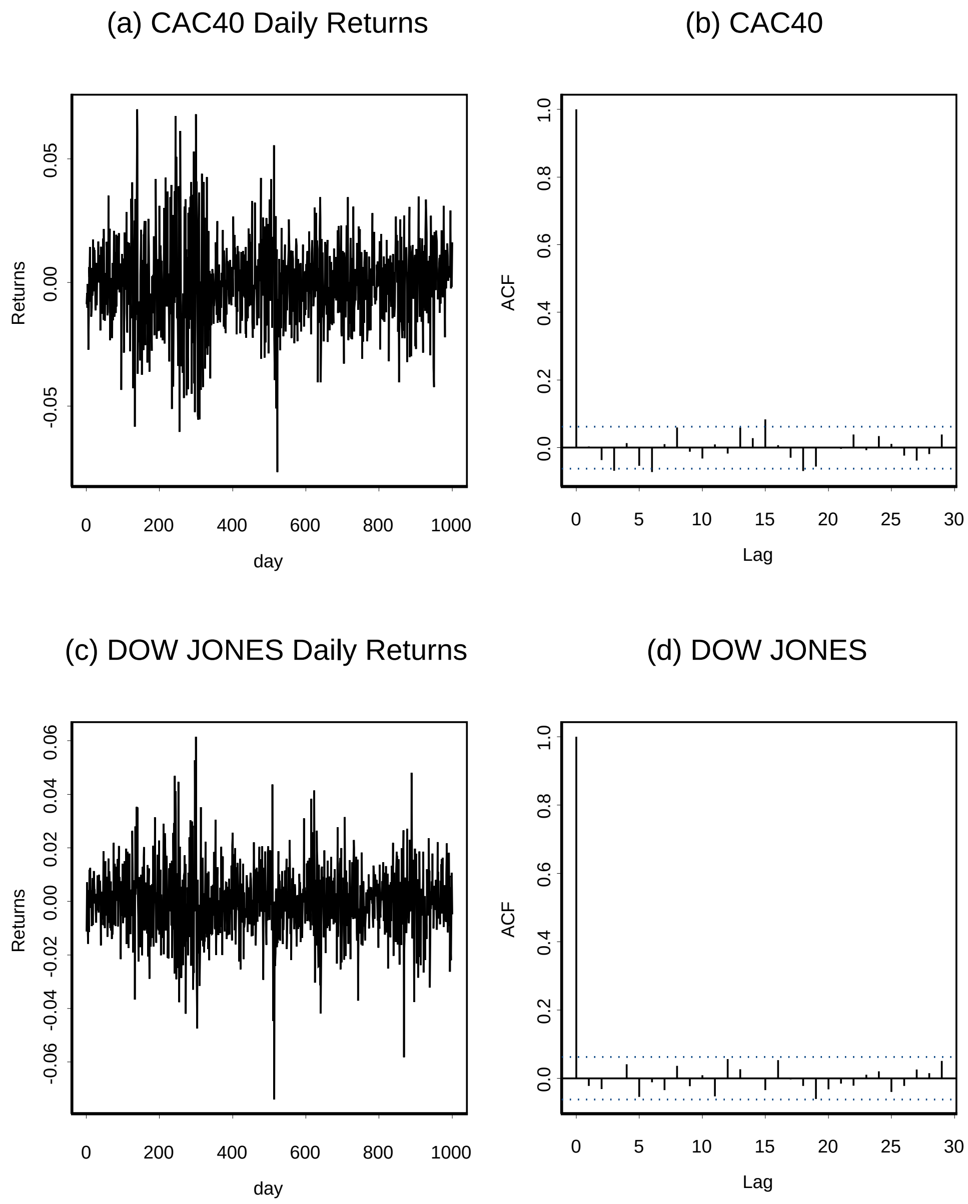
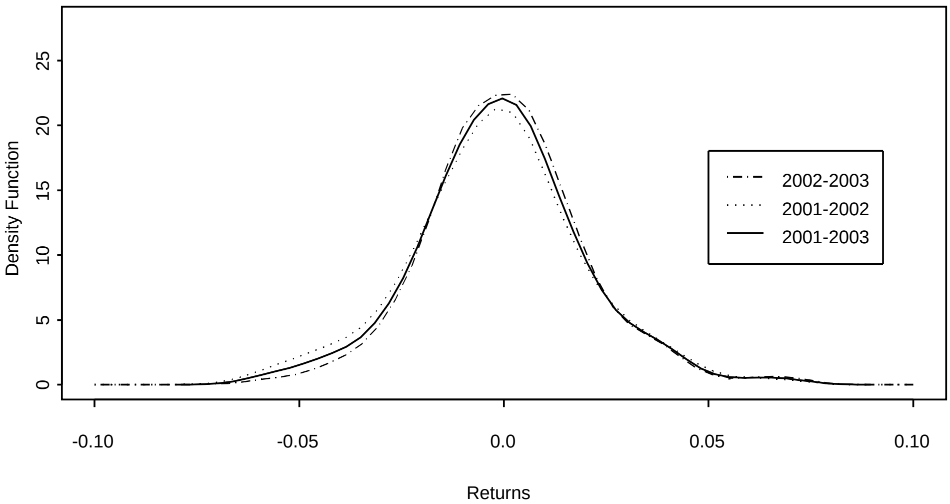


Figure 4. The two financial return series in (a) and (c) and their sample auto-correlation functions (ACF) in (b) and (d).

(a) CAC40



(b) DOW JONES

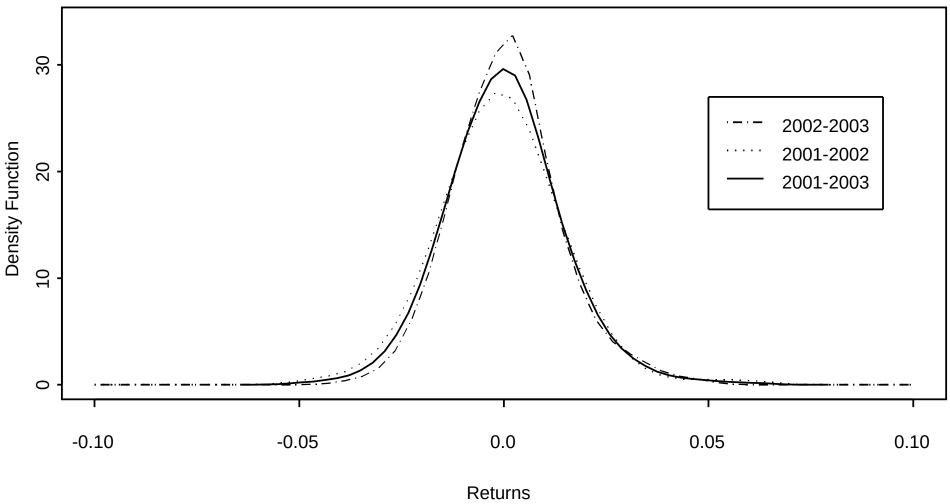


Figure 5. kernel density estimates for the two financial return series.

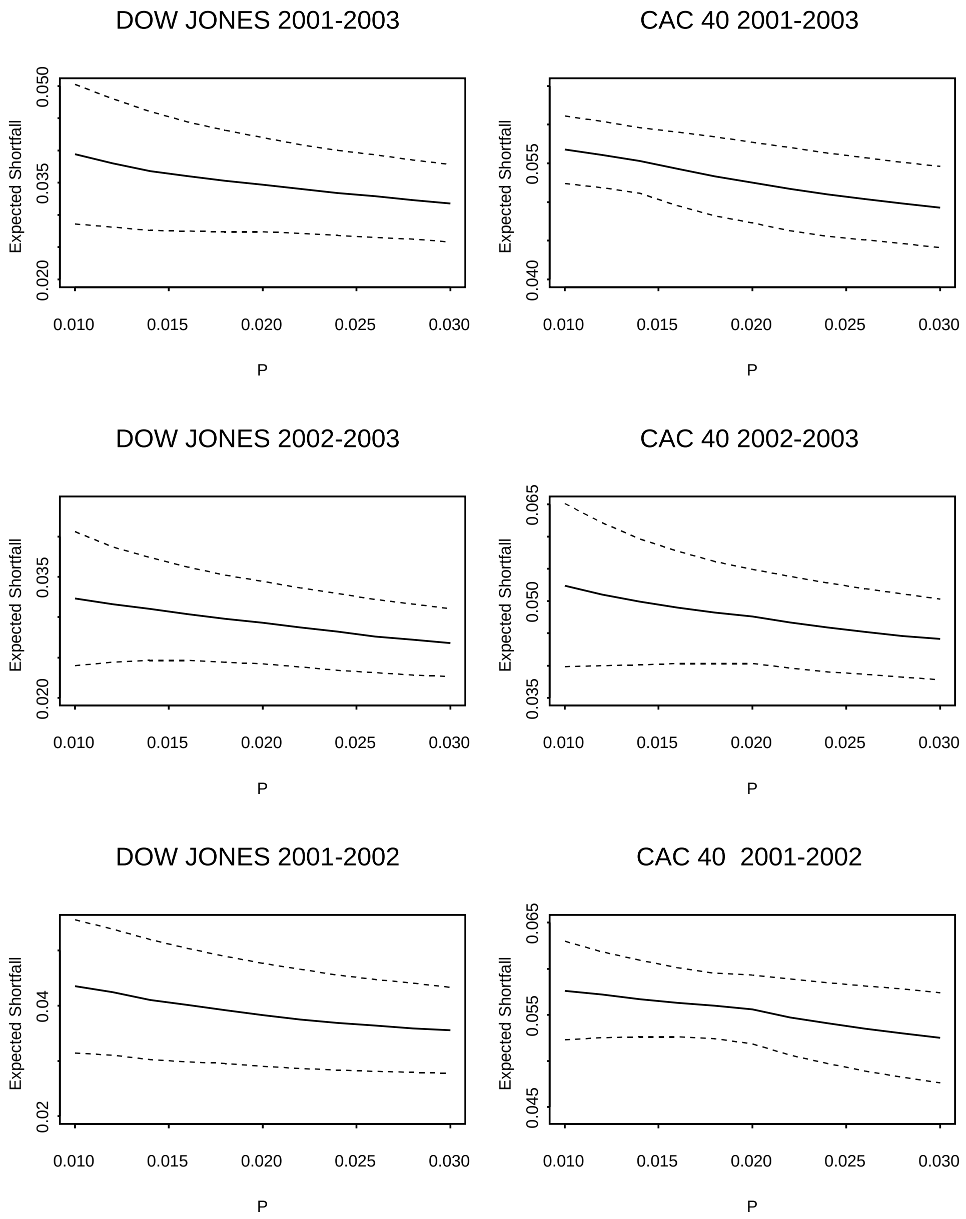


Figure 6. Expected shortfall estimates and their confidence bands for the two financial return series.

Table 1. Intermediate and Final Results in Bandwidth Selection

(a) CAC 40

Year	$\hat{\gamma}$	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\beta}$	t_0	b	h
2001-2002	-0.3189	0.0409	0.0104	-1.0750	1.1552	0.0003	0.0004
2002-2003	-0.3588	0.0284	0.0149	-1.1564	1.3334	0.0015	0.0019
2001-2003	-0.5274	0.0340	0.0165	-1.1303	1.2752	0.007	0.008

(b) DOW JONES

Year	$\hat{\gamma}$	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\beta}$	t_0	b	h
2001-2002	-0.0858	0.0241	0.0082	-1.0965	1.2014	0.0009	0.0011
2002-2003	0.0758	0.0201	0.0041	-1.0730	1.1510	0.0004	0.0005
2001-2003	-0.0151	0.0216	0.0068	-1.0965	1.2015	0.0007	0.0008

Table 2. Estimates for $\mu_{0.01}$ and Standard Errors (S.E.)

	CAC				Dow Jones			
Year	$\hat{\nu}_{p,b}$	$\hat{\mu}_p$	$\hat{\mu}_{p,h,b}$	S.E.	$\hat{\nu}_{p,b}$	$\hat{\mu}_p$	$\hat{\mu}_{p,h,b}$	S.E.
2001-2002	0.0553	0.0571	0.0576	0.0027	0.0377	0.0424	0.0435	0.0062
2002-2003	0.0443	0.0510	0.0524	0.0065	0.0287	0.0316	0.0323	0.0042
2001-2003	0.0532	0.0560	0.0568	0.0022	0.0323	0.0381	0.0394	0.0055