

3-2004

# Nonparametric Inference of Value at Risk for Dependent Financial Returns

Song Xi Chen

*Iowa State University*, [songchen@iastate.edu](mailto:songchen@iastate.edu)

Cheng Yong Tang

*National University of Singapore*

Follow this and additional works at: [http://lib.dr.iastate.edu/stat\\_las\\_preprints](http://lib.dr.iastate.edu/stat_las_preprints)

 Part of the [Statistics and Probability Commons](#)

---

## Recommended Citation

Chen, Song Xi and Tang, Cheng Yong, "Nonparametric Inference of Value at Risk for Dependent Financial Returns" (2004). *Statistics Preprints*. 35.

[http://lib.dr.iastate.edu/stat\\_las\\_preprints/35](http://lib.dr.iastate.edu/stat_las_preprints/35)

This Article is brought to you for free and open access by the Statistics at Iowa State University Digital Repository. It has been accepted for inclusion in Statistics Preprints by an authorized administrator of Iowa State University Digital Repository. For more information, please contact [digirep@iastate.edu](mailto:digirep@iastate.edu).

---

# Nonparametric Inference of Value at Risk for Dependent Financial Returns

## **Abstract**

The paper considers nonparametric estimation of Value at Risk (VaR) and associated standard error estimation for dependent financial returns. Theoretical properties of the kernel VaR estimator are investigated in the context of dependence. The presence of dependence affects the variance of the VaR estimates and has to be taken into consideration in order to obtain adequate assessment of their variation. An estimation procedure of the standard errors is proposed based on kernel estimation of the spectral density of a derived series. The performance of the VaR estimators and the proposed standard error estimation procedure are evaluated by theoretical investigation, simulation of commonly used models for financial returns, and empirical studies on real financial return series.

## **Keywords**

$\alpha$ -mixing, kernel estimation, sample quantile, spectral density estimation, standard error estimation

## **Disciplines**

Statistics and Probability

## **Comments**

This preprint was published as Song Xi Chen and Cheng Yong Tang, "Nonparametric Inference of Value-at-Risk for Dependent Financial Returns", *Journal of Financial Econometrics* (2005): 227-255, doi: [10.1093/jjfinec/nbi012](https://doi.org/10.1093/jjfinec/nbi012)

**NONPARAMETRIC INFERENCE OF VALUE AT RISK  
FOR DEPENDENT FINANCIAL RETURNS**

SONG XI CHEN

Department of Statistics, Iowa State University

and

Cheng Yong Tang

Department of Statistics and Applied Probability,  
National University of Singapore

Corresponding author: Song Xi Chen

Department of Statistics, Iowa State University, Ames, IA 50011-1210

Phone: 1-515-294 2729; Email:stacsx@nus.edu.sg

## ABSTRACT

The paper considers nonparametric estimation of Value at Risk (VaR) and associated standard error estimation for dependent financial returns. Theoretical properties of the kernel VaR estimator are investigated in the context of dependence. The presence of dependence affects the variance of the VaR estimates and has to be taken into consideration in order to obtain adequate assessment of their variation. An estimation procedure of the standard errors is proposed based on kernel estimation of the spectral density of a derived series. The performance of the VaR estimators and the proposed standard error estimation procedure are evaluated by theoretical investigation, simulation of commonly used models for financial returns, and empirical studies on real financial return series.

KEY WORDS:  $\alpha$ -mixing; kernel estimation; sample quantile; spectral density estimation; standard error estimation.

## 1. INTRODUCTION

Value at Risk (VaR) is a popular measure of market risk associated with an asset or a portfolio of assets. It has been chosen by the Basel Committee on Banking Supervision as a benchmark risk measure and has been used by financial institutions for asset management and minimization of risk. Let  $\{X_t\}_{t=1}^n$  be the market value of an asset over  $n$  periods of a time unit, and let  $Y_t = \log(X_t/X_{t-1})$  be the log-returns. Suppose  $\{Y_t\}_{j=1}^n$  is a strictly stationary dependent process with marginal distribution function  $F$ . Given a positive value  $p$  close to zero, the  $1 - p$  level VaR is

$$(1.1) \quad \nu_p = \inf\{u : F(u) \geq p\},$$

which specifies the smallest amount of loss such that the probability of the loss in market value being larger than  $\nu_p$  is less than  $p$ . Comprehensive discussions on VaR are available in Duffie and Pan (1997) and Jorion (2001), and references therein.

Early estimators of VaR are based on parametric models for the return distribution  $F$ , for instance Gaussian or  $t$ -distributions. A more sophisticated parametric approach based on ARCH or GARCH models has been developed under the trademarks of RiskMetrics, KMV and Creditmetrics, which are able to resemble to certain degrees the fat-tail phenomenon of financial returns. The advantages of the parametric approaches lay in their easy interpretation. However, they are model dependent and are subject to errors of model mis-specification. Recently, a VaR estimation method based on the extreme value distributions (EVD) is gaining popularity. The EVD approach fits the extreme tail part of data by a Generalized Pareto distribution. The approach is based on the Balkema-de Haan-Pickands theorem (Balkema

and de Haan, 1974) for independent and identically distributed high exceedances. The situations where the EVD approach is suitable include independent and identically distributed returns and dependent returns that can be expressed as a process with independent and identically distributed innovations, see Embrechts, Resnick and Samorodnitsky (1999) and McNeil and Frey (2000) for comprehensive reviews.

Model-free nonparametric estimation of VaR has been proposed by Dowd (2001) based on the sample quantile, which is commonly called the historical VaR. Gouriéroux, Laurent and Scaillet (2000) introduce nonparametric kernel VaR estimators. These nonparametric estimators have the advantages of (i) being free of distributional assumptions on  $Y_t$  while being able to capture fat-tail and asymmetry distribution of returns automatically; and (ii) imposing much weaker assumptions on the dynamics of the return process and allowing data “speak for themselves”. A potential limitation of nonparametric methods may be the requirement of a reasonable sample size to ensure good performance. However, our simulation results reported in Section 7 indicate that the nonparametric approach produces reasonable VaR estimates for sample sizes of 125, which corresponds roughly to six months data. Also, the sample size required by the nonparametric approach should be comparable to that required by the EVD approach, as both approaches concentrate on the tail part of the data.

That financial return series are subject to data dependence is a known reality in empirical finance, which was the motivation behind proposing ARCH/GARCH models, along with the observation that the returns tend to have heavy tails. Recently, Bellini and Figá-Talamanca (2002) have shown, by carrying out a nonparametric runs test, that financial time series exhibit quite strong tail dependence even for large threshold levels. This calls for a more general approach for VaR estimation, that is able to cater for dependence while still works when the data are independent. Developing such an approach is an objective of this paper. The dependence structures which are applicable by the techniques proposed here are very wide, including ARMA, ARCH/GARCH, stochastic volatility and diffusion models. And most importantly, the detail of the dependence structure can be unknown as long as it satisfies  $\alpha$ -mixing.

Another issue that the current paper wishes to address is the provision of standard errors for VaR estimates. It appears that users of VaR have not paid much attention to the standard errors associated with their estimates. As a consequence, their VaR estimates are subject to uncertain risk themselves. Providing the standard errors is not only practically important, as it provides a measure of risk for the VaR, but also an interesting statistical problem as the dependence makes the variance estimation a non-trivial task. We propose an approach based on a kernel estimation of a spectral density function which can capture all the covariances induced by the dependence.

There are some concerns about the nonparametric VaR approach. One concern is that extreme quantiles are very difficult to estimate nonparametrically and would require large numbers of observations. Indeed, extreme quantiles generally require larger sample sizes to estimate as the amount of data information is thin in the tail part of the distribution. However, this is the case for the other approaches too. A parametric approach may be easily implemented computationally. Its main difficulty is its exposure to uncertain model errors which are hard to evaluate mathematically. As demonstrated by both theory and simulation studies, we show in this paper that extreme quantiles can be estimated effectively by the nonparametric kernel method. Another concern with the nonparametric approach is that the VaR estimates are volatile whenever a large loss enters the sample. We do not think this is a valid concern. Take the sample VaR estimator at level  $1 - p$  as an example. As it is the  $p$ -th sample quantile, it is unchanged unless there are more than  $[pn]$  big new losses entering the return series; here  $[a]$  is the integer part of the real number  $a$ . If  $n > 100$ , a single big loss does not alter the 99% sample VaR estimate; and the robustness increases when  $n$  becomes larger. In contrast, both the parametric and the EVD based VaR estimates are altered by a single big loss.

The paper is structured as follows. We introduce in Section 2 various financial return models to which the results of the paper are applicable. Nonparametric VaR estimators are outlined in Section 3, and their statistical properties are investigated in Section 4. The issue of standard error estimation is considered in Section 5. Section 6 give details of practical implementation. Simulation results which describe the empirical performance of the proposed methods are presented in Section 7, whereas an empirical analysis of two financial returns series is carried out in Section 8. Section 9 gives a general discussion. All the technical details are provided in the Appendix.

## 2. DEPENDENCE STRUCTURE AND MODELS

Let us first introduce the concept of mixing for dependent processes. For the log return series  $\{Y_t\}_{j=1}^n$ , let  $\mathcal{F}_k^l$  be the  $\sigma$ -algebra of events generated by  $\{Y_t, k \leq t \leq l\}$  for  $l \geq k$ . The  $\alpha$ -mixing coefficient introduced by Rosenblatt (1956) is

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

The series is said to be  $\alpha$ -mixing if  $\lim_{k \rightarrow \infty} \alpha(k) = 0$ . The dependence described by  $\alpha$ -mixing is the weakest, as it is implied by other types of mixing; see Doukhan (1994) for comprehensive discussions on mixing and related topics. The return series is said to be geometric  $\alpha$ -mixing, if there exist some constants  $c > 0$  and  $\rho \in (0, 1)$  such that  $\alpha(k) \leq c\rho^k$

for  $k \geq 1$ .

The following commonly used financial econometric models generate series  $\{Y_t\}_{j=1}^n$  which are geometric  $\alpha$ -mixing and hence to which the results of this paper may be applicable.

**Linear Processes.** For a linear causal process (which includes ARMA models)

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

with independent and identically distributed innovation  $\{\xi_s\}_{s=0}^{\infty}$ , Gorodeskii (1977) showed that the process is  $\alpha$ -mixing under certain conditions and established the rate for the  $\alpha$ -mixing coefficient. Pham and Tran (1985) show that if each coefficient  $g_t$  of the process is  $O(\gamma^t)$ ,  $0 < \gamma < 1$ , then the process is geometric  $\alpha$ -mixing.

**Markov Processes.** Consider a Markov process

$$(2.1) \quad Y_j = m(\bar{Y}_{j,p}) + \sigma(\bar{Y}_{j,p})\epsilon_j \quad \text{for } i = 1, \dots, T$$

where  $\bar{Y}_{j,p} = (Y_{j-1}, \dots, Y_{j-p})$  are  $p$ -lagged values of  $Y_t$  and  $\{\epsilon_j\}_{j=1}^T$  are independent and identically distributed random variables. Here  $m(\cdot)$  and  $\sigma^2(\cdot)$  are respectively the conditional mean and volatility functions of  $Y_t$  given  $\bar{Y}_{j,p}$ . The model includes ARCH(p) models. Masry and Tjøstheim (1995) prove that the series is geometric ergodic and  $\alpha$ -mixing under some mild conditions.

**Continuous-Time Diffusion Models.** Continuous-time models are effective tools for modeling continuous evolution of asset value processes over time. Here  $(\tilde{Y}_t)_{t \geq 0}$  is the log-price process in which the index  $t$  takes value continuously within  $[0, \infty)$ . A time-homogenous diffusion model for the log-return is specified by the following stochastic differential equation:

$$(2.2) \quad d\tilde{Y}_t = \mu(\tilde{Y}_t)dt + \sigma(\tilde{Y}_t)dW_t$$

where  $\mu(\cdot)$  and  $\sigma(\cdot)$  are the drift and diffusion functions respectively, and  $W_t$  is a Brownian motion independent of  $\tilde{Y}_t$ . Although  $(\tilde{Y}_t)_{t \geq 0}$  is continuous in time, what we observe is a discrete sample path  $\{\tilde{Y}_{j\Delta}\}_{j=1}^n$  at equally spaced time points  $t_j = j\Delta$  for some  $\Delta > 0$ . For a family of diffusion model

$$d\tilde{Y}_t = \alpha(\beta - \tilde{Y}_t)dt + c\tilde{Y}_t^v dW_t$$

Genon-Catalot, Jeantheau and Laredo (2000) have given restrictions on the parameters  $(v, \alpha, \beta)$  such that  $\{\tilde{Y}_{j\Delta}\}_{j=1}^T$  is geometric  $\alpha$ -mixing. This implies that the log-return series  $\{Y_j\}_{j=1}^n$  where  $Y_j = \tilde{Y}_{j\Delta} - \tilde{Y}_{(j-1)\Delta}$  is geometric  $\alpha$ -mixing.

**Stochastic Volatility Models.** Stochastic volatility models are extensions of the one factor diffusion model (2.2) that allows the volatility of the log-price process  $(\tilde{Y}_t)_{t \geq 0}$  to be driven by another diffusion model as follows:

$$(2.3) \quad d\tilde{Y}_t = \sigma_t dW_t, \quad V_t = \sigma_t^2 \quad \text{and} \quad dV_t = b(V_t)dt + a(V_t)dB_t$$

where  $(W_t, B_t)_{t \geq 0}$  is a two-dimensional Brownian motion,  $(V_t)_{t \geq 0}$  is a positive diffusion and  $V_0 = \eta$  is a positive random variable independent of  $(W_t, B_t)_{t \geq 0}$ . The stochastic volatility  $V_t$  is not directly observable. Genon-Catalot, Jeantheau and Laredo (2000) show for a discretely observed sample path  $\{\tilde{Y}_{j\Delta}\}_{j=1}^T$ , by treating the model as a hidden Markov chain, that the series  $\{\tilde{Y}_{j\Delta}\}$  is geometric  $\alpha$ -mixing under certain conditions, which implies that the log-return series is geometric  $\alpha$ -mixing.

### 3. NONPARAMETRIC ESTIMATION OF VaR

Let  $F_n(x) = n^{-1} \sum_{j=1}^n I(Y_j \leq x)$  be the empirical distribution function of the return series  $\{Y_t\}$ , where  $I(\cdot)$  is the indicator function. The historical VaR estimator proposed by Dowd (2001) is  $\hat{\nu}_{ps} = Y_{([np]+1)}$  where  $Y_{(r)}$  is the  $r$ -th order statistic. It is just the sample quantile estimator commonly used in statistics by replacing  $F$  with  $F_n$  in (1.1). And for this reason, it is called the sample VaR estimator in this paper. It is a consistent estimator of  $\nu_p$  for  $\alpha$ -mixing data (Yoshihara, 1995). However, as the VaR is an extreme quantile situated in the tail region of the distribution where the amount of data information is thin, it will be beneficial to carry out kernel smoothing on the empirical distribution  $F_n$ . The smoothing essentially leads to an estimator which is a weighted average of the order statistics around  $Y_{([np]+1)}$  rather than relying on a single order statistic. Studies done for independent and identically distributed data, for instance Falk (1981) and Sheather and Marron (1990), showed that the variance of the sample quantile estimator is reduced by kernel smoothing. For pairwise positively or negatively quadrant dependent data, Cai and Roussas (1997) studied various asymptotic properties of the kernel quantile estimator. In this paper, we focus on  $\alpha$ -mixing series and study the effects of smoothing on the bias and variance of the kernel estimator.

Let  $G(x) = \int_{-\infty}^x K(u)du$  be the distribution function of a kernel function  $K$  which is a symmetric probability density function. A kernel estimator of  $F(x)$  replaces the indicator function  $I$  in the formulation of  $F_n$  by the smoother  $G$ , i.e.

$$(3.1) \quad \hat{F}_{n,h}(x) = n^{-1} \sum_{j=1}^n G\left(\frac{x - Y_j}{h}\right)$$

where  $h$  is a smoothing bandwidth that controls the amount of smoothness in the estimation of  $F$ . A kernel estimator of  $\nu_p$ , denoted as  $\hat{\nu}_{p,h}$ , is obtained by inverting  $\hat{F}_{n,h}(x) = p$ , such that  $\hat{\nu}_{p,h}$  satisfies

$$(3.2) \quad n^{-1} \sum_{j=1}^n G\left(\frac{\hat{\nu}_{p,h} - Y_t}{h}\right) = p.$$

This kernel VaR estimator, first introduced by Gouriéroux, Laurent and Scaillet (2000) in the context of VaR estimation, can be viewed as a smoothed version of  $\hat{\nu}_p$ .

In studying the properties of the kernel VaR estimator, we assume the following conditions:

**A.1** The process  $\{Y_t\}_{t=1}^n$  is strictly stationary and  $\alpha$ -mixing, and there exists a  $\rho \in (0, 1)$  such that  $\alpha(k) \leq C\rho^k$  for all  $k \geq 1$ ; each  $Y_t$  is continuously distributed with  $f$  and  $F$  as its density and distribution functions respectively.

**A.2**  $f(\nu_p) > 0$  and  $f$  has continuous second derivative in a neighborhood  $\mathcal{B}(\nu_p)$  of  $\nu_p$ ; the second partial derivatives of  $F_k$ , which is the joint distribution function of  $(Y_1, Y_{k+1})$   $k \geq 1$ , are all bounded in  $\mathcal{B}(\nu_p)$  uniformly with respect to  $k$ .

**A.3**  $K$  is a univariate probability density function, has continuous bounded second derivative and satisfies the following moment conditions:

$$\int_{-\infty}^{\infty} uK(u)du = 0 \quad \text{and} \quad \int_{-\infty}^{\infty} u^2K(u)du = \sigma_K^2.$$

**A.4** The smoothing bandwidth  $h$  satisfies  $h \rightarrow 0$ ,  $nh^{3-\beta} \rightarrow \infty$  for any  $\beta > 0$  and  $nh^4 \log^2(n) \rightarrow 0$  as  $n \rightarrow \infty$ .

The stationarity and geometric  $\alpha$ -mixing as assumed in A.1 are satisfied by those models discussed in the previous section under certain conditions. A.2 contains standard conditions for quantile estimation whereas conditions in A.3 and A.4 are commonly imposed conditions in kernel smoothing. In particular, conditions in A.4 specify a range for the bandwidth  $h$ .

#### 4. PROPERTIES OF THE NONPARAMETRIC VaR ESTIMATORS

Let us first outline some existing results on the sample VaR estimator  $\nu_p$ . Yoshihara (1995) established the following Bahadur representation under  $\alpha$ -mixing:

$$(4.1) \quad \hat{\nu}_p - \nu_p = \frac{F_n(\nu_p) - p}{f(\nu_p)} + O\{n^{-3/4} \log(n)\} \quad a.s.$$

and showed that

$$(4.2) \quad \text{Var}(\hat{\nu}_p) = n^{-1} f^{-2}(\nu_p) \sigma^2(p; n) \{1 + o(1)\}$$

where  $\sigma^2(p; n) = \{p(1-p) + 2 \sum_{k=1}^{n-1} (1-k/n) \gamma(k)\}$  and  $\gamma(k) = \text{Cov}\{I(Y_1 < \nu_p), I(Y_{k+1} < \nu_p)\}$  for positive integers  $k$ . The Bahadur representation implies strong convergence of  $\hat{\nu}_p$  to  $\nu_p$  and also indicates that

$$(4.3) \quad E(\hat{\nu}_p) = \nu_p + O(n^{-3/4}).$$

A key quantity that describes the variance properties of the kernel estimator  $\hat{\nu}_{p,h}$  is  $\sigma_h^2(p; n) = \{p(1-p) + 2 \sum_{k=1}^{n-1} (1-k/n) \gamma_h(k)\}$ , where  $\gamma_h(k) = \text{Cov}\{G\left(\frac{\nu_p - Y_1}{h}\right), G\left(\frac{\nu_p - Y_{k+1}}{h}\right)\}$ . The following lemma indicates its relationship with  $\sigma^2(p; n)$ .

LEMMA 1. *Under conditions A.1-A.4,  $|\sigma_h^2(p; n) - \sigma^2(p; n)| = o(h)$ .*

Now let us study the properties of the kernel estimator  $\hat{\nu}_{p,h}$ . The strong convergence similar to that of  $\hat{\nu}_p$  is considered in the following Theorem 1, whereas bias and variance properties are treated in Theorem 2.

THEOREM 1: *Under conditions A.1-A.4, as  $n \rightarrow \infty$ ,  $\{n^{1/2} \log^{-1}(n)\}(\hat{\nu}_{p,h} - \nu_p) \rightarrow 0$  a.s..*

THEOREM 2: *Under conditions A.1-A.4, as  $n \rightarrow \infty$ ,*

$$(4.4) \quad E(\hat{\nu}_{p,h}) = \nu_p - \frac{1}{2}h^2\sigma_k^2 f'(\nu_p)f^{-1}(\nu_p) + o(h^2).$$

$$(4.5) \quad Var(\hat{\nu}_{p,h}) = n^{-1}f^{-2}(\nu_p)\sigma_h^2(p; n) - 2n^{-1}hf^{-1}(\nu_p)b_k + o(h/n)$$

where  $b_k = \int uK(u)G(u)du$ .

Theorem 2 and (4.2) indicate that both the kernel and the sample VaR estimators share the same asymptotic variance as Lemma 1 shows  $\sigma_h^2(p; n) = \sigma^2(p; n) + o(h)$ . The dependence of the return series is expressed in both  $\sigma(p; n)$  and  $\sigma_h^2(p; n)$  via the covariances  $\gamma(k)$  and  $\gamma_h(k)$ . For independent data,  $\sum_{k=1}^{n-1} \gamma_h(k) = 0$  which means that the asymptotic variance is simply  $n^{-1}p(1-p)f^{-2}(\nu_p)$ . If there is dependence in the data which is ignored, the asymptotic variance of the nonparametric VaR estimates would be wrongly regarded as  $n^{-1}p(1-p)f^{-2}(\nu_p)$ , which can result in a wrong assessment of the variability.

Both Theorem 2 and (4.2) indicate that the rates of convergence of the nonparametric VaR estimators are  $n^{-1/2}$ , which is the same as the parametric VaR estimators. The appearance of  $f^2(\nu_p)$  in the denominators of the leading variance terms indicates that the variability gets larger as  $p$  gets smaller, provided  $f$  is monotone in the left tail which is usually the case. It also nicely spells out the difficulty we are facing in VaR estimation.

Based on Theorem 2 and Lemma 1, the mean square error (MSE) of  $\hat{\nu}_{p,h}$  is

$$\begin{aligned} \text{MSE}(\hat{\nu}_{p,h}) &= n^{-1}f^{-2}(\nu_p)\sigma^2(p; n) - 2f(\nu_p)b_khn^{-1} + \frac{1}{4}h^4\sigma_k^4\{f'(\nu_p)f^{-1}(\nu_p)\}^2 \\ &\quad + o(h/n + h^4). \end{aligned}$$

Then, the optimal bandwidth that minimizes  $\text{MSE}(\hat{\nu}_{p,h})$  is

$$(4.6) \quad h_{opt} = \left\{ \frac{2f^3(\nu_p)b_k}{\sigma_k^4 f'^2(\nu_p)} \right\}^{1/3} n^{-1/3}$$

which can be estimated by plugging-in the estimates of  $f(\nu_p)$  and  $f'(\nu_p)$ , an issue will be discussed in the next section. Substituting  $h_{opt}$  into (4.7), the optimal MSE is then

$$(4.7) \quad \text{MSE}(\hat{\nu}_{p,h}) = n^{-1}\sigma^2(p; n) - 32^{-2/3}b_k^{4/3}\sigma_k^{-4/3}f^2(\nu_p)\{f'(\nu_p)\}^{-4/3}n^{-4/3} + o(n^{-4/3})$$

which indicates a reduction to MSE of the second order.

The following theorem establishes the asymptotic normality of  $\hat{\nu}_{p,h}$ .

**THEOREM 3:** *Under conditions A.1-A.4, as  $n \rightarrow \infty$ ,*

$$\sqrt{n}(\hat{\nu}_{p,h} - \nu_p) \xrightarrow{d} N\left(0, \sigma^2(p)f^{-2}(\nu_p)\right)$$

where  $\sigma^2(p) = \lim_{n \rightarrow \infty} \sigma^2(p; n)$  whose existence is guaranteed by condition A.1.

Theorem 3 can be used to construct asymptotic confidence intervals for  $\nu_p$  as well as to carry out tests on hypotheses regarding  $\nu_p$ .

## 5. STANDARD ERRORS OF VaR ESTIMATES

Regardless of which VaR estimator we use, a standard error has to be attached in order to gain information on its variability. It seems that practitioners have not paid their due attention to the issue of standard errors. It is not uncommon to see VaR estimates presented without attaching standard errors. As a result, their estimates are subject to uncertain risk themselves. The issue is very important here as VaR estimates are subject to high variability, and it is crucial to have knowledge of this variability. Providing standard errors is not only practically important, as it provides a measure of risk for the VaR, but also an interesting statistical problem as the dependence makes the variance estimation a non-trivial task. We propose an approach based on a kernel estimation of the spectral density function which can capture all the covariances induced by the dependence.

The key is to estimate  $\sigma_h^2(p; n) = p(1-p) + 2 \sum_{k=1}^{n-1} \gamma_h(k)$ . Although each  $\gamma_h(k)$  may be estimated consistently, adding these  $(n-1)$  estimates together does not yield even a consistent estimator of  $\sigma_h^2(p; n)$ . The route we are going to take is to estimate the spectral density of the derived series  $\{Z_t\} = \left\{G\left(\frac{\nu_p - Y_t}{h}\right)\right\}$ , where  $h$  is the bandwidth used in the kernel VaR estimation and is regarded as a given quantity in this section.

Let  $i$  be the imaginary number in complex analysis, and

$$\phi(\lambda) = (2\pi)^{-1} \sum_{k=-\infty}^{\infty} \gamma_h(k) \exp(-ik\lambda) \quad \text{for } \lambda \in [-\pi, \pi]$$

be the spectral density of  $\{Z_t\}$ . From the Davydov inequality,  $|\gamma_h(k)| \leq C\alpha(k)$  for some constant  $C > 0$ . Thus, from A.1  $\sum_{k=0}^{\infty} |\gamma_h(k)| < \infty$ , which in turn implies that  $\phi(0)$  is finite and hence the derived series is weakly dependent. According to Brockwell and Davis (1991, Corollary 4.3.2),  $\lim_{n \rightarrow \infty} \{\sigma_h^2(p; n) - 2\pi\phi(0)\} = 0$ . Hence the estimation of  $\sigma_h^2(p; n)$  can be achieved by estimating  $\phi(0)$ .

Define

$$(5.1) \quad I_n(\omega_j) = n^{-1} \left| \sum_{j=1}^n Z_j e^{-ij\omega_j} \right|^2, \quad j = 0, \pm 1, \dots, \pm[n/2]$$

where  $\omega_j = 2\pi j/n \in [-\pi, \pi]$  are the Fourier frequencies. Let  $T = \{\pm 1, \pm 2, \dots, \pm ([n/2] - 1)\}$  which excludes 0 as  $I_n(0)$  has different asymptotic behaviors from other  $I_n(\omega_j)$ . According to Theorem 5.2.6 of Brillinger (1981) for any  $j \in T$

$$(5.2) \quad I_n(\omega_j) = (2\pi)\phi(\omega_j)E_j + R_j$$

where  $\{E_j\}_{j \in T}$  are independent standard exponential random variables, and  $\{R_j\}_{j \in T}$  are asymptotically negligible terms.

We note that  $\{Z_j\}$  are not observable due to the involvement of the unknown  $\nu_p$ . Let  $\hat{Z}_j = G\left(\frac{\hat{\nu}_p - Y_t}{h}\right)$ , and  $\hat{I}_n(\omega_j)$  be the periodograms defined on  $\{\hat{Z}_j\}$  by replacing  $Z_j$  by  $\hat{Z}_j$  in (5.1). As  $\hat{Z}_j = Z_j + o_p(1)$  uniformly for all the  $j$ 's, it may be shown that  $\hat{I}_n(\omega_j) = I_n(\omega_j) + o_p(1)$  uniformly for all  $j \in T$ . Hence from (5.3),

$$(5.3) \quad \hat{I}_n(\omega_j) = (2\pi)\phi(\omega_j)E_j + R'_j$$

where  $R'_j$  are asymptotically negligible.

A commonly used approach for spectral density estimation, for instance that used in Fan and Gijbels (1996) and Lee (1999) is to taking logarithm on both sides of (5.3) and ignore  $\{R'_j\}$ ,

$$(5.4) \quad \log\{\hat{I}_n(\omega_j)/(2\pi)\} = \log\{\phi(\omega_j)\} + \log(E_j) \quad \text{for } j \in T.$$

Since  $E_j$  are standard exponentially distributed,  $E\{\log(E_j)\} = -0.57721$  (the Euler constant) and  $Var\{\log(E_j)\} = \pi^2/6$ . Let

$$\eta_j = \log(E_j) + 0.57721, \quad W_j = \log\{\hat{I}_n(\omega_j)/(2\pi)\} + 0.57721 \quad \text{and} \quad m(\omega) = \log\{\phi(\omega)\}.$$

Then (5.4) can be approximated by the following fixed design nonparametric regression:

$$(5.5) \quad W_j = m(\omega_j) + \eta_j \quad j \in T$$

where  $\{\eta_j\}_{j \in T}$  are independent and identically distributed with zero mean and variance  $\sigma_\eta^2 = \pi^2/6$ . The idea is to estimate  $m(0) = \log\{\phi(0)\}$  by kernel smoothing.

The Nadaraya–Waston (NW) estimator of  $m(\omega)$  based on another kernel  $K_1$  and a smoothing bandwidth  $b$  is

$$(5.6) \quad \hat{m}_b(\omega) = \frac{\sum_{j \in T} K_1\left(\frac{\omega - \omega_j}{b}\right) W_j}{\sum_{j \in T} K_1\left(\frac{\omega - \omega_j}{b}\right)}$$

where  $b \rightarrow 0$  and  $nb \rightarrow \infty$  as  $n \rightarrow \infty$ . Then, a kernel estimator of  $\phi(0)$  is

$$(5.7) \quad \hat{\phi}(0) = \exp\{\hat{m}_b(0)\}.$$

Fan and Gijbels (1996) considered a local linear kernel estimation of spectral density. As the design points are fixed at the Fourier frequencies  $\omega_j$ s and zero is not a boundary point (negative frequencies are used), the NW estimator is as good as the local linear estimator in this particular situation.

Standard results in nonparametric regression show that  $\hat{m}_b(0) \xrightarrow{p} m(0)$ , and thus  $\hat{\phi}(0) \xrightarrow{p} \phi(0)$  as  $n \rightarrow \infty$ . Furthermore,

$$\begin{aligned} E\{\hat{m}_b(0)\} &= m(0) - \frac{1}{2}b^2\sigma_{k_1}^2 m''(0) + o(b^2) \\ &= m(0) - \frac{1}{2}b^2\sigma_{k_1}^2 \{\phi''(0)\phi^{-1}(0) - \phi'(0)\phi^{-2}(0)\} + o(b^2), \\ \text{Var}\{\hat{m}_b(0)\} &= \frac{R(K_1)\pi^2}{6nb} + o\{(nb)^{-1}\}, \end{aligned}$$

where  $\sigma_{k_1} = \int u^2 K_1(u) du$  and  $R(K_1) = \int K_1^2(u) du$ . It may be shown by employing the delta-method that

$$\begin{aligned} E\{\hat{\phi}(0)\} &= \phi(0) + \frac{1}{2}b^2\sigma_{k_1}^2 \phi''(0) + o(b^2) + o(b^2) \\ \text{Var}\{\hat{m}_b(0)\} &= \frac{\phi^2(0)R(K_1)\pi^2}{6nb} + o\{(nb)^{-1}\} \end{aligned}$$

where  $\phi''(0) = (2\pi)^{-1} \sum_{k=-\infty}^{\infty} k^2 \gamma(k)$  is finite as implied by A.1 and the Davydov inequality. The optimal bandwidth  $b$  that minimizes the mean square error of  $\hat{\phi}(0)$  is then

$$(5.8) \quad b_{opt} = \left\{ \frac{\phi^2(0)R(K)\pi^2}{6\sigma_{k_1}^4 \{\phi''(0)\}^2} \right\}^{1/5} n^{-1/5}.$$

After estimating  $\phi(0)$  and ignoring the second order difference in the variance between  $\hat{\nu}_{p,h}$  and  $\hat{\nu}_p$ ,

$$\sqrt{\frac{2\pi\hat{\phi}(0)}{n\hat{f}^2(\hat{\nu}_{p,h})}}$$

can be regarded as the common standard error for both VaR estimates where  $\hat{f}(\cdot)$  is an estimator of  $f(\cdot)$ . The estimation of  $f$  will be discussed in the next section.

The above procedure for obtaining the standard error is for the case of  $\phi(0) \neq 0$ . If  $\phi(0) = 0$ , then the variance of  $\nu_{p,h}$  will be a smaller order of  $n^{-1}$ . To estimate the variance in this case, we need to develop a new variance expansion and then the similar plug-in method as just outlined can be adapted.

## 6. PRACTICAL IMPLEMENTATION

In this section we discuss issues related to implementing the kernel VaR and its standard error estimation.

## 6.1 Kernel VaR Estimation

As mentioned in Section 5, we suggest using a kernel supported on  $R$  rather than a compact kernel in order to facilitate standard error estimation. The Gaussian kernel is a natural choice. What is left to decide is the selection of  $h$ . The theoretically optimal  $h$  given in (4.6) is

$$h_{opt} = \left\{ \frac{2f^3(\nu_p)b_k}{\sigma_k^4 f'^2(\nu_p)} \right\}^{1/3} n^{-1/3}.$$

Here  $b_k$  and  $\sigma_k^2$  are known after choosing  $K$ . The approach we use is the plug-in method, that is to obtain  $h$  by plugging-in estimates of  $f(\nu_p)$  and  $f'(\nu_p)$  into the above formula. First of all, we replace  $\nu_p$  by the sample VaR  $\hat{\nu}_p$ . The method of reference to a parametric distribution, which is a simple and commonly used bandwidth selection method in kernel smoothing, is used to obtain estimates of  $f$  and  $f'$ . A natural candidate for the reference distribution is the Generalized Pareto(GP) distribution as we are concerned with an extreme quantile which is situated in the tail of the distribution. In particular let

$$(6.1) \quad w_{\gamma,\mu,\sigma}(x) = \frac{1}{\sigma} \left( 1 + \gamma \frac{x - \mu}{\sigma} \right)^{-(1+\frac{1}{\gamma})},$$

be the density of a GP distribution with a scale parameter  $\sigma$ , a shape parameter  $\gamma$ , and a truncation level  $\mu$ ; see Reiss and Thomas (2001) for comprehensive discussions on the theory and applications of GP distributions. For a 99% VaR, we fit the lower five percent of the data to a GP model, which means taking  $\mu = \hat{\nu}_{0.05}$ . For other levels of VaR,  $\mu$  should be adjusted accordingly. Let  $\hat{\sigma}$  and  $\hat{\gamma}$  be the method of moment estimates of the parameters. Then, the estimates of  $f(\nu_p)$  and  $f'(\nu_p)$  are respectively  $w_{\hat{\gamma},\hat{\sigma},\hat{\nu}_{0.05}}(\hat{\nu}_p)$  and  $w'_{\hat{\gamma},\hat{\sigma},\hat{\nu}_{0.05}}(\hat{\nu}_p)$ , which then lead to a practically useful  $h$ .

## 6.2 Standard Error Estimation

The kernel  $K_1$  can be any kernel, compactly supported or otherwise. The main issue is the selection of  $b$ . For bandwidth selection, the objective function we want to minimize with respect to  $b$  is

$$(6.2) \quad R(b) = \frac{1}{n} \sum_{j \in T} q_{nj} \{m(\omega_j) - \hat{m}_b(\omega_j)\}^2$$

by defining weights  $q_{nj} = I(|j| \leq [k_n])$  where  $k_n$  is an integer depending on  $n$ . We choose  $k_n = [0.05n]$  which means that only the 10% periodograms close to the zero frequency are considered. This is natural as we are interested in estimation of  $\phi(0)$  only. Again, we eliminate  $I_n(0)$  by choosing  $j \in T$ , as  $I_n(0)$  has entirely different asymptotics. A derivation

presented in the appendix shows that an unbiased estimate of  $R(b)$  is

$$(6.3) \quad r(b) = \frac{1}{n} \sum_{j \in T} q_{nj} \{W_j - \hat{m}_b(\omega_j)\}^2 + \frac{\pi^2}{6} \left(1 - \frac{4\pi K(0)}{nb}\right) \sum_{j \in T} q_{nj}.$$

Ignoring the term not involving  $b$ , the object function needing to be minimized is then

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

On the estimation of  $f(\nu_p)$ , for simplicity we choose  $\hat{f}(\nu_p) = w_{\hat{\gamma}, \hat{\sigma}, \hat{\nu}_{0.05}}(\hat{\nu}_{ph})$  which is a by-product of the  $h$ -bandwidth selection discussed earlier.

## 7. SIMULATION RESULTS

In this section we report results from a simulation study designed to evaluate the performance of the nonparametric VaR estimators and their standard error estimation for six commonly used financial time series models, which offers a wide range of dependent structures. The models considered are

- (i) AR(1) model:  $Y_t = 0.5Y_{t-1} + \epsilon_t$ ,  $\epsilon_t \stackrel{iid}{\sim} N(0, 1)$ ;
- (ii) AR(2) model:  $Y_t = 0.9Y_{t-1} - 0.2Y_{t-2} + \epsilon_t$ ,  $\epsilon_t \stackrel{iid}{\sim} N(0, 1)$ ;
- (iii) MA(2) model:  $Y_t = \epsilon_t + 0.65\epsilon_{t-1} + 0.24\epsilon_{t-2}$ ,  $\epsilon_t \stackrel{iid}{\sim} N(0, 1)$ ;
- (iv) ARCH(1) model:  $Y_t = 0.5Y_{t-1} + \epsilon_t$ ,  $\epsilon_t^2 = 4 + 0.5\epsilon_{t-1}^2 + \eta_t$ ,  $\eta_t \stackrel{iid}{\sim} N(0, 1)$ ;
- (v) Stochastic Volatility (SV) model:

$$\begin{aligned} Y_t &= V_t \epsilon_t, \quad \log(V_t) = 0.6 \log(V_{t-1}) + \eta_t, \quad \epsilon_t \stackrel{iid}{\sim} N(0, 1), \\ \eta_t &\stackrel{iid}{\sim} N(0, 0.5), \quad Cov(\epsilon_t, \eta_t) = 0.5 \quad \text{and} \quad Cov(\epsilon_t, \eta_{t-j}) = 0 \quad \text{for } j > 0; \end{aligned}$$

- (vi) Diffusion Model:  $d\tilde{Y}_t = 0.4(2 - \tilde{Y}_t)dt + \sqrt{50}dB_t$ ,  $B_t$  is the Brownian Motion and  $Y_t = \tilde{Y}_t - \tilde{Y}_{t+\Delta}$  with  $\Delta = 1/250$  (daily returns).

The generation of the AR and MA series is straight forward. To generate the ARCH series, we generate an iid series  $\{\delta_t\}$  such that  $P[\delta_t = \pm 1] = 0.5$ , and another series  $Z_t = 4 + 0.5Z_{t-1} + \eta_t$  where  $\eta_t \stackrel{iid}{\sim} N(0, 1)$ . The innovations for the ARCH series are then  $\epsilon_t = \delta_t \sqrt{Z_t}$ . For the SV series, we generate two iid standard normal series  $\{\epsilon_t\}$  and  $\{\zeta_t\}$ , and let  $\eta_t = 0.5\epsilon_t + 0.5\zeta_t$ , which are respectively the innovations of the SV model. The rest follows the formulae of the model. The diffusion  $\tilde{Y}_t$  is generated, from  $\tilde{Y}_{t-\Delta}$ , from the transitional

density  $N(\tilde{Y}_{t-\Delta}e^{-0.4\Delta} + 2(1 - e^{-0.4\Delta}), 50(1 - e^{0.8\Delta})/0.8)$  whereas  $\tilde{Y}_0$  is generated from the stationary distribution  $N(2, 50/0.8)$ . It should be noted that the ARCH and SV models are only asymptotically stationary. Therefore, we pre-run the series for 1000 times in each simulation before the real series being started. The exact VaR values for the ARCH and SV models are obtained, based on 10,000 simulation of the real models, whereas those for the AR, MA and the diffusion models can be obtained from the known stationary distributions.

We choose the Gaussian kernel  $K(u) = \frac{1}{\sqrt{2\pi}}exp(-u^2/2)$  as the kernel for estimating VaR and the biweight kernel  $K_1(u) = \frac{15}{16}(1 - u^2)^2I(|u| \leq 1)$  for estimating  $\phi(0)$ . The bandwidths  $h$  and  $b$  are chosen according to the procedures outlined in the previous section. The sample size ranges from 125 to 2000, which corresponds to data ranging from six months to 8 years.

The results on the bias, the standard deviation (SD) and the root mean square error (RMSE) of the nonparametric estimates of 99% VaR are reported in Tables 1-6 for the six models based on 5000 simulations. The tables also include the estimated standard errors of the kernel VaR estimates. We find that both nonparametric VaR estimators produce quite satisfactory results for all those models considered. It is very assuring to see that the proposed standard error estimation procedure offers quite accurate prediction to the real standard deviation of the kernel and sample VaR estimates, even when the sample size is small. We observe that the bias, the standard deviations and the root mean square errors all decrease as  $n$  increases which indicates the proposed VaR estimation methods are consistent. The kernel estimates have less RMSE than their sample VaR counterparts, confirming our theory given in (4.7). However, the reduction in RMSE is not very large for large samples which reflects our early prediction that the reduction is of second order only. We note that the kernel estimates have smaller bias than the sample VaR estimates while having almost the same variance (the kernel variance is only slightly smaller). This suggests that the proposed data-driven  $h$ -bandwidth may be slightly smaller. For a larger  $h$ , the variance of the kernel would be smaller and in return the bias would be larger. This is a common phenomenon in kernel smoothing.

## 8. EMPIRICAL STUDIES

In this section we apply the nonparametric VaR inference procedures to analyze the daily log-return series of the Nasdaq index and Microsoft from January 1st 1999 to December 31st 2002, which consist of four years data ( $n = 1000$ ). These two return series are displayed in Figure 1 (a) and (c) respectively. The sample auto-correlation functions are plotted in Figure 1 (b) and (d), which indicate dependence in these two series. To formally confirm it, we calculate the Box-Pierce Q-test statistics  $Q = n \sum_{k=1}^{26} \hat{\gamma}(k)^2$  where  $\hat{\gamma}(k)$  is the sample

auto-correlation. If the series are independent,  $Q$  should be distributed as  $\chi_{26}^2$ . The  $p$ -value of these two series are 0.004 and 0.041 respectively, indicating that both series are dependent and the dependence in Nasdaq is stronger than that in Microsoft. This is not surprising, as an index tends to be more persistent than a single stock.

To gain insights into the dynamics of these two return series, we plot in Figures 2 and 3 the kernel estimates of the return densities for the two series based on, respectively, each of the four one year segments of the series, each of the two year segments and the entire four years data. The bandwidths used to draw these kernel density estimates are given by the default values of S-plus for the Gaussian kernel. There are some yearly variations in the kernel density estimates as shown in Panels (a) of Figures 2 and 3, especially between, before and after 2000. This reflects well the burst of the Internet bubble. These may indicate the returns not being stationary. However, it may be also due to variations. Further investigation is needed on the issue. The density estimates based on the two-year data are much more stable, and they are not that different from the density estimates based on the four year data especially in the left tail.

We carried out the 99% VaR estimation using the kernel VaR estimation for the first and last two years, and the entire four year of these two series, respectively. Standard errors for the VaR estimates are obtained by applying the proposed spectral density estimation method. These results are summarized in Figure 4, which displays bars centered at the kernel VaR estimates whose length is 3.92 times the estimated standard error. So they can be regarded as a kind of 95% confidence intervals for the real VaR. For Nasdaq, the  $h$  and  $b$  bandwidths were respectively 0.003 and 0.8216 for the 1999-2000 sub-series, 0.003 and 0.9433 for the 2001-2002 sub-series, and 0.002 and 0.4251 for the entire series. For Microsoft, the bandwidths were respectively 0.003 and 0.6856 for the 1999-2000 sub-series, 0.005 and 0.7233 for the 2001-2002 sub-series, and 0.002 and 0.3995 for the entire series. We do not present the sample VaR estimates as they are very close to the kernel estimates.

For the Nasdaq returns, we see little change among the three kernel VaR estimates and these were all around 6% after removing the negative sign. There were some variations in the VaR estimates for Microsoft. In particular, the estimate for the two years 1999-2000 was at 7.48%, much higher than the other two estimates. For both series, the variability of the VaR estimation was higher for the first two year period than for the last two year period. There was substantial reduction in variation when the length of the series increased from two to four years.

For comparison, we also present in the figure standard errors of the kernel VaR estimates assuming independence, which were all smaller than those under dependence. The difference between the two standard errors was the largest at the first two year period of the Nasdaq series. We also present the parametric VaR estimates based on the indepen-

dent Gaussian model with the standard errors obtained by the bootstrap. Two bootstrap resampling schemes were employed: the full nonparametric bootstrap which resamples directly from the original data by sampling with replacement and the parametric bootstrap by generating resamples from  $N(\bar{Y}, S^2)$  where  $\bar{Y}$  and  $S^2$  are respectively the sample mean and variance of the returns. The bootstrap standard errors of the parametric Gaussian VaR estimates were all much smaller than those of the kernel VaR, and indicates possible severe under-estimation of the variability. We also observe quite large discrepancies between the parametric and nonparametric VaR estimates for the Microsoft series in the period of 1999-2000.

## 9. DISCUSSIONS

Despite its popular use by financial institutions for risk management, VaR is known to be not in general a coherent measure of risk as it is not sub-additive (Artzner, Delbaen, Eber and Heath, 1999). Examples of VaR being not sub-additive have been given in Artzner, Delbaen, Eber and Heath (1999) when the return distribution  $F$  is discrete. The properties of the VaR improves when the returns are continuously distributed. It is known that VaR is sub-additive within the family of Gaussian returns. Embrechts, McNeil and Straumann (2002) show that it is also the case for the family of elliptical distributions. It may be shown by generalizing the proof of the above authors that the VaR is sub-additive within any family of distributions generated by the location-scale transformation of a distribution  $F_0$ , that is  $\mathcal{F}(x) = \{F|F(x) = F_0\left(\frac{x-a}{b}\right) \text{ for any } a, b \in R\}$  which includes the Gaussian and elliptical families as special cases, as well as the Normal Inverse Gaussian subfamily within the family of generalized hyperbolic distributions which is closely associated with continuous-time asset pricing models based on Lévy processes (Eberlein, 1999). The expected shortfall (ES) is a closely related risk measure to VaR, which is coherent. Nonparametric estimation of the ES has been considered in Scaillet (2002). Because of the close link between the VaR and the ES, an investigation of the estimation of standard errors of VaR will be helpful to the inference of ES too.

This paper has provided an evaluation on the statistical properties of the kernel and sample VaR estimators and has proposed a nonparametric procedure for the standard error estimation for a wide range of dependence structures. Considering that a bandwidth  $h$  has to be chosen for the kernel method, one may just use the simpler  $\hat{\nu}_p$ . However, the extra effort of smoothing pays off at the end, as it produces estimates with less RMSE which can translate to a large amount in financial terms. This is especially the case when the sample size is small. Another advantages of smoothing is in the standard error estimation. Our study shows that, to achieve a fixed level of accuracy, the standard error estimation based

on the unsmoothed series  $\{I(Y_i \leq \hat{\nu}_p)\}$  requires much larger sample size than that required for the smoothed series  $\{\hat{Z}_j\}$ . Smoothing significantly enhances the estimation of standard errors.

### Appendix: Technical Details

*Proof of Theorem 1:* The theorem is proved if

$$(A.1) \quad \sum_{n=1}^{\infty} P\left(|\hat{\nu}_{p,h} - \nu_p| \geq n^{-1/2} \log(n)\eta\right) < \infty$$

for any  $\eta > 0$ . Let  $\epsilon = n^{-1/2} \log(n)\eta$ . Then,

$$\begin{aligned} A &=: P(|\hat{\nu}_{p,h} - \nu_p| \geq \epsilon) \leq P(\hat{\nu}_{p,h} > \nu_p + \epsilon) + P(\hat{\nu}_{p,h} < \nu_p - \epsilon) \\ &= P\{p - F(\nu_p + \epsilon) > \hat{F}_{n,h}(\nu_p + \epsilon) - F(\nu_p + \epsilon)\} \\ &\quad + P\{p - F(\nu_p - \epsilon) < \hat{F}_{n,h}(\nu_p - \epsilon) - F(\nu_p - \epsilon)\} \end{aligned}$$

By Taylor expansion of  $F(\nu_p \pm \epsilon)$  at  $\nu_p$ ,

$$\begin{aligned} A &= P\{\hat{F}_{n,h}(\nu_p + \epsilon) - F(\nu_p + \epsilon) < -f(\nu_p + \theta_1\epsilon)\epsilon\} \\ &\quad + P\{\hat{F}_{n,h}(\nu_p - \epsilon) - F(\nu_p - \epsilon) > f(\nu_p - \theta_2\epsilon)\epsilon\} \\ &\leq P\{|\hat{F}_{n,h}(\nu_p + \epsilon) - F(\nu_p + \epsilon)| > c_1\epsilon\} + P\{|\hat{F}_{n,h}(\nu_p - \epsilon) - F(\nu_p - \epsilon)| > c_1\epsilon\} \end{aligned}$$

where  $\theta_1, \theta_2 \in (0, 1)$  and  $c_1 = \inf_{x \in [\nu_p - \epsilon, \nu_p + \epsilon]} f(x) > 0$  as implied by Condition (iv). The above equations indicate that (A.1) is established if

$$(A.2) \quad \sum_{n=1}^{\infty} P\{|\hat{F}_{n,h}(\nu_p + \lambda\epsilon) - F(\nu_p + \lambda\epsilon)| \geq n^{-1/2} \log(n)\eta\} < \infty$$

for  $\lambda = 1$  and  $-1$ .

We prove the case for  $\lambda = 1$  only as the other case is exactly the same. Notice that

$$\begin{aligned} E\{\hat{F}_{n,h}(\nu_p + \epsilon)\} - F(\nu_p + \epsilon) &= \int_{-1}^1 \{F(\nu_p + \epsilon - hu) - F(\nu_p + \epsilon)\} K(u) du \\ &= \int_{-1}^1 f'(\nu_p + \epsilon - \theta_3 hu) h^2 u^2 K(u) du \end{aligned}$$

where  $\theta_3 \in (0, 1)$ . As  $f'$  is bounded in a neighborhood of  $\nu_p$  and  $nh^4 \log^2(n) \rightarrow 0$  as assumed in condition (iv), we have

$$(A.3) \quad |E\{\hat{F}_{n,h}(\nu_p + \epsilon)\} - F(\nu_p + \epsilon)| = O(h^2) = o\{n^{-1/2} \log(n)\}.$$

Let  $Z_j = G_h(\nu_p + \epsilon - Y_j) - E\{G_h(\nu_p + \epsilon - Y_j)\}$ . Clearly,  $E(Z_j) = 0$  and  $|Z_j| \leq 2$ . Let  $q = b_0 n^{1/2} \log(n)$ ,  $p = n/(2q)$  and  $u^2(q) = \max_{0 \leq j \leq 2q-1} E\left(\sum_{l=[jp]+1}^{[(j+1)p]} Z_l\right)^2$ . From an inequality

given in Bosq (1998, Theorem 1.3) for  $\alpha$ -mixing sequences,

$$(A.4) \quad P \left( \left| \sum_{j=1}^n Z_j \right| > c_2 n \epsilon \right) \leq 4 \exp \left( -\frac{c_2^2 \epsilon^2 q}{8 \sigma^2(q)} \right) + 22 \left( 1 + \frac{8}{c_2 \epsilon} \right)^{1/2} q \alpha \{ [n/(2q)] \}$$

where  $\sigma^2(q) = 2p^{-2}u^2(q) + \epsilon = Cn^{-1/2} \log(n)$ . It is obvious that

$$(A.5) \quad 4 \exp \left( -\frac{c_1^2 \epsilon^2 q}{8 \sigma^2(q)} \right) \leq 4 \exp \{ -C(b_0) \log^2(n) \}$$

where  $C(b_0) > 0$  is a constant which is positively related to  $b_0$ . From Condition (i),

$$(A.6) \quad \begin{aligned} 22 \left( 1 + \frac{8}{c_1 \epsilon} \right)^{1/2} q \alpha \{ [n/(2q)] \} &\leq Cn^{3/4} \log^{-1/6}(n) \alpha \{ [n^{1/2} \log^{-1}(n)/2] \} \\ &\leq Cn^{3/4} \log^{-1/6}(n) \rho^{n^{1/2} \log^{-1}(n)/2} \end{aligned}$$

which converges to zero sufficiently fast. From (A.4), (A.5) and (A.6) we have

$$\sum_{n=1}^{\infty} P \left( \left| \sum_{j=1}^n Z_j \right| > c_1 n \epsilon \right) < \infty.$$

This and (A.3) imply (A.2) for  $\lambda = 1$ . Thus we complete the proof of Theorem 1.  $\square$

*Proof of Lemma 1:* The proof follows that of Lemma 2.2 of Cai and Roussas (1998) but replaces their (2.15), which is a key result under the assumption that  $\{Y_t\}$  are positively or negatively dependent random variables, by

$$(A.7) \quad \sup_{(x,y) \in R^2} |F_k(x,y) - F(x)F(y)| \leq \alpha(k)$$

which is trivially true from the definition of  $\alpha(k)$ . In particular,  $|\gamma_h(k) - \gamma(k)| \leq Ch^2$  as shown in Cai and Roussas (1998). It is fairly clear from (A.7) that  $|\gamma(k)| \leq C_1 \alpha(k)$  and  $|\gamma_h(k)| \leq C_1 \alpha(k)$ . These means

$$|\gamma_h(k) - \gamma(k)| \leq Ch^{2\tau} \alpha^{1-\tau}(k).$$

As Condition (i) means  $\sum_{j=1}^{\infty} \alpha^{1-\tau}(j) < \infty$ , we have

$$\sum_{k=1}^{n-1} (1 - k/n) |\gamma_h(k) - \gamma(k)| \leq C_1 h^{2\tau} \sum_{k=1}^{\infty} \alpha^{1-\tau}(k) = O(h^{2\tau}).$$

This completes the proof.  $\square$

We are now ready to derive the variance of  $\hat{F}_{n,h}(\nu_p)$  in the following Lemma.

LEMMA 2. Under the condition (i) and (ii),

$$\text{Var}\{\hat{F}_{n,h}(\nu_p)\} = n^{-1}\{p(1-p) + 2\sum_{k=1}^{n-1}\gamma_h(k)\} - 2hf(\nu_p) \int uK(u)G(u)du + o(h/n).$$

*Proof:* Note that

$$\text{Var}\{\hat{F}_{n,h}(\nu_p)\} = n^{-1}\text{Var}\left\{G\left(\frac{\theta - Y_1}{h}\right)\right\} + n^{-1}\sum_{k=1}^{n-1}(1 - k/n)\gamma_h(k)$$

and

$$\text{Var}\left\{G\left(\frac{\theta - Y_1}{h}\right)\right\} = p(1-p) - 2hf(\nu_p) \int uK(u)G(u)du + o(h).$$

These immediately imply the lemma.

LEMMA 3: Under the condition (i) and (ii), for  $l, k = 1$  or  $2$  and  $l + k \geq 3$ ,

$$(A.8) \quad \text{Cov}\left[\{p - \hat{F}_{n,h}(\nu_p)\}^l, \{p - \hat{F}_{n,h}(\nu_p)\}^k\right] = o(h/n).$$

*Proof:* Let  $\mu_h(x) = E\{\hat{F}_{n,h}(x)\}$ . It can be shown that  $\mu_h(\nu_p) - p = O(h^2)$ . From Theorem 2 of Yokoyama (1980) under the geometric strong mixing condition

$$(A.9) \quad E|\mu_h(\nu_p) - \hat{F}_{n,h}(\nu_p)|^r = Cn^{-r/2}$$

for any positive integer  $r$  and a positive constant  $C$ . We only prove for the case of  $l = 2$  and  $k = 2$  since that for  $l = 1$  and  $k = 2$  is slightly simpler. Note that

$$\text{Cov}\left[\{p - \hat{F}_{n,h}(\nu_p)\}^2, \{p - \hat{F}_{n,h}(\nu_p)\}^2\right] = E\{p - \hat{F}_{n,h}(\nu_p)\}^4 - E^2\{p - \hat{F}_{n,h}(\nu_p)\}^2$$

From Lemma 2,  $E\{p - \hat{F}_{n,h}(\nu_p)\}^2 = E^2\{p - \mu_h(\nu_p)\}^2 + \text{Var}\{\hat{F}_{n,h}(\nu_p)\} = O(n^{-1} + h^4)$ . This together with (A.9) implies

$$\text{Cov}\left[\{p - \hat{F}_{n,h}(\nu_p)\}^2, \{p - \hat{F}_{n,h}(\nu_p)\}^2\right] = O(n^{-2} + h^2n^{-3/2} + h^4n^{-1} + h^8) = o(h/n).$$

Thus, complete the proof.  $\square$

*Proof of Theorem 2:* Since (4.4) can be easily established via a standard derivation of the bias, we only derive the variance part of the theorem. Let

$$\hat{f}_{n,h}(x) = (nh)^{-1}\sum_{t=1}^n K\left(\frac{\hat{\nu}_{p,h} - Y_t}{h}\right) \quad \text{and} \quad \hat{f}'_{n,h}(x) = (nh^2)^{-1}\sum_{t=1}^n K'\left(\frac{\hat{\nu}_{p,h} - Y_t}{h}\right)$$

be the kernel estimators of the density  $f(x)$  and the density derivative  $f'(x)$  respectively.

Expand  $\hat{F}_{n,h}(\hat{\theta}_p)$  at  $\nu_p$ ,

$$(A.10) p = n^{-1} \hat{F}_{n,h}(\nu_p) + \hat{f}_{n,h}(\nu_p)(\hat{\nu}_{p,h} - \nu_p) + \frac{1}{2} \hat{f}'_{n,h}\{\nu_p + \theta(\hat{\nu}_{p,h} - \nu_p)\}(\hat{\nu}_{p,h} - \nu_p)^2$$

where  $\theta \in (0, 1)$ . From Lemma 2.1 of Bosq (1998)

$$\hat{f}(\nu_p) = f(\nu_p) + O\{(nh)^{-1/2} \log(n)\} \quad \text{a.s.}$$

Slightly modifying Theorem 2.2 of Bosq (1998) from density estimation to density derivative estimator, it can be proved that

$$\sup_{\theta \in [0,1]} |\hat{f}'\{\nu_p + \theta(\hat{\nu}_{p,h} - \nu_p)\} - f'\{\nu_p + \theta(\hat{\nu}_{p,h} - \nu_p)\}| = O_p\{(nh^2)^{-1/2} \log(n)\}.$$

Since  $f$  has bounded second derivative near  $\nu_p$  as implied by Condition (iv), we have

$$\sup_{\theta \in [0,1]} |f'\{\nu_p + \theta(\hat{\nu}_{p,h} - \nu_p)\} - f'(\nu_p)| = O_p\{n^{-1/2} \log(n)\}.$$

Also, Theorem 1 implies  $p - \hat{F}_{n,h}(\nu_p) = o\{n^{-1/2} \log(n)\}$  a.s. These results imply that, by inverting the expansion (A.10),

$$(A.11) \quad \begin{aligned} \hat{\nu}_{p,h} - \nu_p &= \frac{p - \hat{F}_{n,h}(\nu_p)}{\hat{f}_{n,h}(\nu_p)} - \frac{1}{2} \frac{\hat{f}'_{n,h}\{\nu_p + \theta(\hat{\nu}_{p,h} - \nu_p)\}}{\hat{f}_{n,h}^3(\nu_p)} (p - \hat{F}_{n,h}(\nu_p))^2 + O_p\{n^{-3/2} \log^3(n)\} \\ &= \frac{p - \hat{F}_{n,h}(\nu_p)}{\hat{f}_{n,h}(\nu_p)} - \frac{1}{2} \frac{f'(\nu_p)(p - \hat{F}_{n,h}(\nu_p))^2}{f^3(\nu_p)} + O_p\{n^{-3/2} h^{-1} \log^3(n)\}. \end{aligned}$$

From Lemma 3,

$$(A.12) \quad \begin{aligned} Var(\hat{\nu}_{p,h}) &= Var\left\{\frac{p - \hat{F}_{n,h}(\nu_p)}{\hat{f}_{n,h}(\theta_q)}\right\} - Cov\left\{\frac{p - \hat{F}_{n,h}(\nu_p)}{\hat{f}_{n,h}(\theta_q)}, \frac{f'(\nu_p)(p - \hat{F}_{n,h}(\nu_p))^2}{f^3(\nu_p)}\right\} \\ &\quad + Var\left\{\frac{f'(\nu_p)(p - \hat{F}_{n,h}(\nu_p))^2}{\hat{f}_{n,h}^3(\nu_p)}\right\} + o(h/n) \\ &= Var\left\{\frac{p - \hat{F}_{n,h}(\nu_p)}{\hat{f}_{n,h}(\theta_q)}\right\} + o(h/n) \end{aligned}$$

Employing the delta method,

$$Var\left\{\frac{p - \hat{F}_{n,h}(\nu_p)}{\hat{f}_{n,h}(\theta_q)}\right\} = f^{-2}(\nu_p) Var\{\hat{F}_{n,h}(\nu_p)\} + o(h/n).$$

This together with Lemmas 1 and 2 leads to (4.5).  $\square$

*Proof of Theorem 3:* From (A.11) and note that  $\hat{f}_{n,h}(\nu_p) = f(\nu_p) + O_p\{(nh)^{-1/2} + h^2\}$ ,

$$\hat{\nu}_{p,h} - \nu_p = \frac{p - \hat{F}_{n,h}(\nu_p)}{f(\nu_p)} + O_p(n^{-1}h^{-1/2} + n^{-1/2}h^2).$$

Hence we only need to prove the asymptotic normality of  $\hat{F}_{n,h}(\nu_p) - p = n^{-1} \sum_{i=1}^n T_{i,n} + E\{G_h(\nu_p - Y_1) - p$  where  $T_{i,n} = G_h(\nu_p - Y_i) - E\{G_h(\nu_p - Y_i)\}$ .

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

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

the smaller blocks

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

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

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

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

$$\text{Var}(S_{n,2}) = \frac{r\sigma^2(p)}{nf^2(\nu_p)} \{1+o(1)\} \rightarrow 0 \quad \text{and} \quad \text{Var}(S_{n,3}) = \frac{(n - r(k+k'))\sigma^2(p)}{nf^2(\nu_p)} \{1+o(1)\} \rightarrow 0.$$

Therefore, for  $l = 2$  and  $3$

$$(A.13) \quad S_{n,l} \xrightarrow{p} 0 \quad \text{as } n \rightarrow \infty.$$

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

$$(A.14) \quad P(|V_{j,n} - W_{j,n}| \leq \epsilon\sqrt{n}/r) \leq 18\epsilon^{-1}rn^{-1/2}\|V_{j,n}\|_{\infty}^{1/2}\alpha(k') \leq C_1\epsilon^{-1}n^{-1/2}k^{1/2}\alpha(k').$$

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

$$(A.15) \quad \begin{aligned} P(|\Delta_n| > \epsilon) &\leq \sum_{j=1}^r P(|V_{j,n} - W_{j,n}| \leq \epsilon\sqrt{n}) \\ &\leq C_1r^{3/2}n^{-1/2}(rk)^{1/2}\alpha(k') \leq C_2r^{3/2}\rho^{k'} \end{aligned}$$

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

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

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

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

$$\frac{\sum E|W_{j,n}|^4}{\{r \text{Var}(W_{1n})\}^2} \leq \frac{C_3 r k}{r^2 k^2} \rightarrow 0$$

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

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

It may be shown by checking on the variance of  $V_{j,n}$  that  $\kappa^2 = \sigma^2(p)$ . Thus, the proof of the theorem is completed by combining (A.13), (A.16) and (A.17).

*Derivation of (6.3):* Recall that

$$(A.18) \quad W_j = m(\omega_j) + \eta_j$$

where  $m(\omega) = \log(\phi(\omega))$ ,  $W_j = \log\{I_n(\omega_j)/(2\pi)\} + 0.57721$  and  $\eta_j$  are independent zero mean random variables with variance  $\pi^2/6$ ; and

$$\hat{m}_b(\omega_i) = \frac{\sum_{j \in T} K\left(\frac{\omega_i - \omega_j}{b}\right) W_j}{\sum_{j \in T} K\left(\frac{\omega_i - \omega_j}{b}\right)} =: \sum_{j \in T} w_{b,j} W_j$$

It is obvious that  $E(W_i^2) = m^2(\omega_i) + \pi^2/6$ , and

$$\begin{aligned} E\{W_i \hat{m}_b(\omega_i)\} &= E\left[\{m(\omega_i) + \eta_i\} \sum_{j \in T} w_{b,j} W_j\right] = E\{m(\omega_i) + \eta_i\} \sum_{j \in T} w_{b,j} \{m(\omega_j) + \eta_j\} \\ &= m(\omega_i) \sum_{j \in T} w_{b,j} m(\omega_j) + E(w_{b,i} \pi^2/6) = m(\omega_i) E\{\hat{m}_b(\omega_i)\} + E\left\{\frac{K(0)\pi^2}{6nb\hat{\eta}(\omega_i)}\right\} \end{aligned}$$

where  $\hat{\eta}(\omega)$  is the kernel density estimator of  $\eta(\omega) \equiv 1/(2\pi)$ . As  $E\{\hat{\eta}^{-1}(\omega_i)\} = \eta^{-1}(\omega_i) + O\{(nb)^{-1} + b^4\}$ ,

$$\begin{aligned} E\{W_i - \hat{m}_b(\omega_i)\}^2 &= m^2(\omega_i) + \pi^2/6 - 2\left[m(\omega_i)E\{\hat{m}_b(\omega_i)\} + \frac{K(0)\pi^3}{3nb}\right] + E\hat{m}_b^2(\omega_i) \\ &= E\{m(\omega_i) - \hat{m}_b(\omega_i)\}^2 + \frac{\pi^2}{6}\left(1 - \frac{4\pi K(0)}{nb}\right) + O(nb)^{-2} \end{aligned}$$

From the above derivation, we have shown that

$$r(b) = \frac{1}{n} \sum_{j \in T} q_j \{W_j - \hat{m}_b(\omega_j)\}^2 - \frac{\pi^2}{6} \left(1 - \frac{4\pi K(0)}{nb}\right) \sum_{j \in T} q_j$$

is an unbiased estimate of the weighted risk function.

## REFERENCES

- ARTZNER, P., DELBAEN, F., EBER, J-M. AND HEATH, D. (1999). "Coherent measures of risk." *Mathematical Finance*, **9**, 203-228.
- BELLINI, F. AND FIGÁ-TALAMANCA, G. (2002). "Detecting and modeling tail dependence." Manuscript and a paper presented in Quantative Methods in Finance 2003 Conference, Cairns, Australia, December 2003.
- BALKEMA, A. A. AND DE HAAN, L. (1974). "Residual life times at great age." *Annals of Probability*, **2**, 792-804.
- BRILLINGER, D. R. (1981). *Time Series, Data Analysis and Theory*, New York: McGraw-Hill.
- BROCKWELL, P. J. AND DAVIS, R. A. (1991). *Time Series: Theory and Methods*. New York: Springer.
- BOSQ, D. (1998). *Nonparametric Statistics for Stochastic Processes*, Lecture Notes in Statistics, **110**. Heidelberg: Springer-Verlag.
- 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.
- DOUKHAN, P. (1994). *Mixing*. Lecture Notes in Statistics, **85**. Heidelberg: Springer-Verlag.
- DUFFIE, D. AND PAN, J. (1997). "An overview of value at risk." *Journal of Derivatives*, 7-49.
- DOWD, K. (2001). "Estimating VaR with order statistics." *Journal of Derivatives*, 23-30.
- EBERLEIN, E (1999). "Application of generalized hyperbolic Lévy motions to finance." Manuscript.
- EMBRECHTS, P., MCNEIL, A., STRAUMANN, D. (2002) "Correlation and dependence in risk management: properties and pitfalls." In: *Risk Management: Value at Risk and Beyond*, ed. M.A.H. Dempster, Cambridge University Press, Cambridge, pp. 176-223.
- EMBRECHTS, P., RESNICK, S. AND SAMORODNITSKY, G. (1999). "Extreme value theory as a risk management tool." *North American Actuarial Journal*, 3(2), 30-41.
- 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*. Chapman and Hall, London.

- GENON-CARALOT, V., JEANTHEAU, T. AND LAREDO, C. (2000). “Stochastic Volatility models as hidden Markov models and statistical applications.” *Bernoulli*, **6**, 1051–1079.
- GORODESKII, V. V. (1977). “On the strong mixing property for linear sequences.” *Theory of Probability And Its Applications*, **22**, 411-413.
- GOURIÉROUX, C., SCAILLET, O. AND LAURENT, J.P. (2000). “Sensitivity analysis of Values at Risk.” *Journal of Empirical Finance*, **7**, 225-245.
- LEE, T. (1997). A simple span selector for periodogram smoothing. *Biometrika*, **84**, 965-969.
- JORION, P. (2001). *Value at Risk*, 2nd Edition, McGraw-Hill, New York.
- MASRY, E. AND TJØSTHEIM, D. (1995). “Nonparametric estimation and identification of nonlinear ARCH time series.” *Econometric Theory*, **11**, 258-289.
- MCNEIL, A. J. AND FREY, R. (2000) “Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach.” *Journal of Empirical Finance*, **7**, 271-300.
- PHAM, T. D. AND TRAN, L. T. (1985). “Some mixing properties of time series models.” *Stoch. Process. Appl.*, **19**, 297-303.
- 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 the  $\alpha$ -mixing condition.” *Proc. Nat. Acad. Sc. U.S.A.*, **42**, 43-47.
- SCAILLET, O. (2002) “Nonparametric estimation and sensitivity analysis of expected shortfall.” forthcoming in *Mathematical Finance*.
- 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.

Table 1: 99% VaR Kernel and sample quantile estimates for 99% VaR for the AR(1) model with the true 99% VaR at 2.686235 .

$N$	Kernel			Sample			Est. SD	
	Bias	SD	RMSE	Bias	SD	RMSE	$\hat{SD}$	SD
125	-0.1137	0.3984	0.4143	-0.1374	0.3989	0.4288	0.3808	0.2634
250	-0.0333	0.3055	0.3073	-0.0362	0.3076	0.3098	0.3014	0.1939
500	-0.0316	0.2153	0.2176	-0.0561	0.2156	0.2227	0.2130	0.0854
1000	-0.0065	0.1551	0.1553	-0.0303	0.1553	0.1582	0.1554	0.0434
2000	0.0073	0.1092	0.1095	-0.0152	0.1093	0.1104	0.1116	0.0220

Table 2: 99% VaR Kernel and sample quantile estimates for 99% VaR for the AR(2) model with the true 99% VaR at 3.589669 .

$N$	Kernel			Sample			Est. SD	
	Bias	SD	RMSE	Bias	SD	RMSE	$\hat{SD}$	SD
125	-0.2083	0.6416	0.6745	-0.2317	0.6418	0.6825	0.6465	0.3940
250	-0.0618	0.4958	0.4997	-0.0641	0.4976	0.5017	0.4848	0.3673
500	-0.0556	0.3518	0.3561	-0.0800	0.3520	0.3609	0.3395	0.1695
1000	-0.0162	0.2536	0.2541	-0.0400	0.2536	0.2566	0.2478	0.0862
2000	0.0080	0.1815	0.1817	-0.0148	0.1816	0.1822	0.1806	0.0447

Table 3: 99% VaR Kernel and sample quantile estimates for 99% VaR for the MA(2) model with the true 99% VaR at 2.830220 .

$N$	Kernel			Sample			Est. SD	
	Bias	SD	RMSE	Bias	SD	RMSE	$\hat{SD}$	SD
125	-0.1143	0.4217	0.4369	-0.1382	0.4218	0.4439	0.4053	0.2883
250	-0.0302	0.3240	0.3255	-0.0327	0.3262	0.3279	0.3251	0.2096
500	-0.0295	0.2272	0.2291	-0.0540	0.2274	0.2337	0.2293	0.0927
1000	-0.0057	0.1618	0.1620	-0.0297	0.1619	0.1646	0.1653	0.0449
2000	0.0099	0.1132	0.1137	-0.0124	0.1132	0.1138	0.1189	0.0223

Table 4: 99% VaR Kernel and sample quantile estimates for 99% VaR for the ARCH(1) model with the true 99% VaR at 5.664672 .

$N$	Kernel			Sample			Est. SD	
	Bias	SD	RMSE	Bias	SD	RMSE	$\hat{SD}$	SD
125	-0.1413	0.3196	0.3495	-0.1639	0.3202	0.3597	0.2973	0.1602
250	-0.0559	0.2061	0.2135	-0.0576	0.2079	0.2157	0.2200	0.1254
500	-0.0282	0.1580	0.1606	-0.0514	0.1581	0.1661	0.1531	0.0822
1000	-0.0047	0.1159	0.1160	-0.0263	0.1161	0.1190	0.1100	0.0410
2000	0.0032	0.0822	0.0823	-0.0156	0.0823	0.0837	0.0815	0.0210

Table 5: 99% VaR Kernel and sample quantile estimates for 99% VaR for the SVM model with the true 99% VaR at 2.383659 .

$N$	Kernel			Sample			Est. SD	
	Bias	SD	RMSE	Bias	SD	RMSE	$\hat{SD}$	SD
125	-0.0837	0.6768	0.6820	-0.1079	0.6770	0.6856	0.6391	0.5694
250	0.0146	0.5292	0.5294	0.0121	0.5313	0.5315	0.5774	0.3315
500	-0.0416	0.3640	0.3664	-0.0800	0.3641	0.3701	0.3486	0.1322
1000	-0.0067	0.2728	0.2730	-0.0313	0.2729	0.2764	0.2510	0.0714
2000	0.0098	0.1937	0.1940	-0.0139	0.1936	0.1940	0.1790	0.0362

Table 6: 99% VaR Kernel and sample quantile estimates for 99% VaR for the VASICEK model with the real 99% VaR at 1.040374 .

$N$	Kernel			Sample			Est. SD	
	Bias	SD	RMSE	Bias	SD	RMSE	$\hat{SD}$	SD
125	-0.0190	0.1352	0.1366	-0.0426	0.1353	0.1418	0.1614	0.1019
250	-0.0043	0.1010	0.1012	-0.0070	0.1026	0.1028	0.1041	0.0545
500	0.0051	0.0694	0.0695	-0.0193	0.0695	0.0721	0.0754	0.0244
1000	0.0146	0.0498	0.0519	-0.0094	0.0498	0.0507	0.0551	0.0122
2000	0.0167	0.0352	0.0389	-0.0058	0.0351	0.0366	0.0392	0.0060

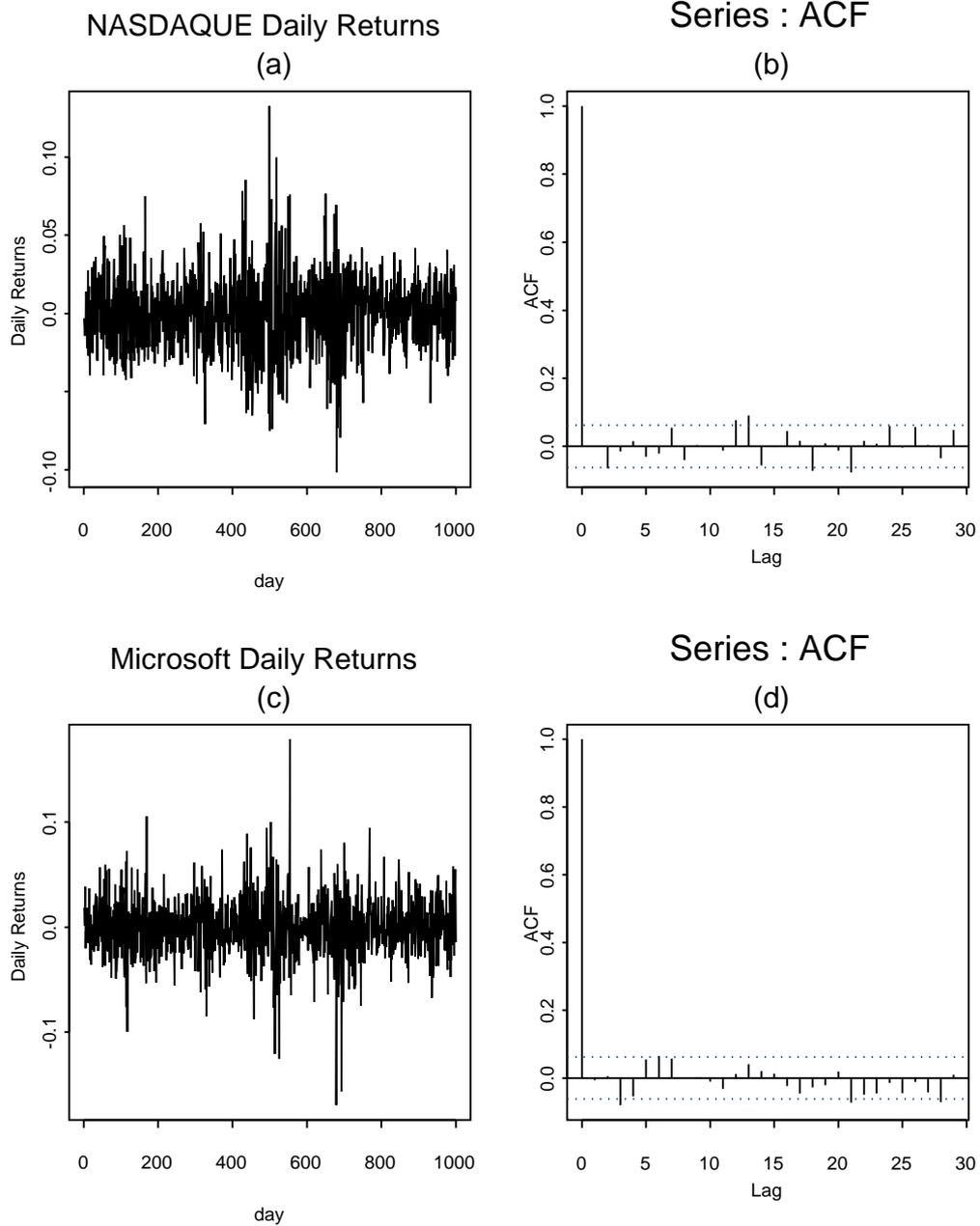
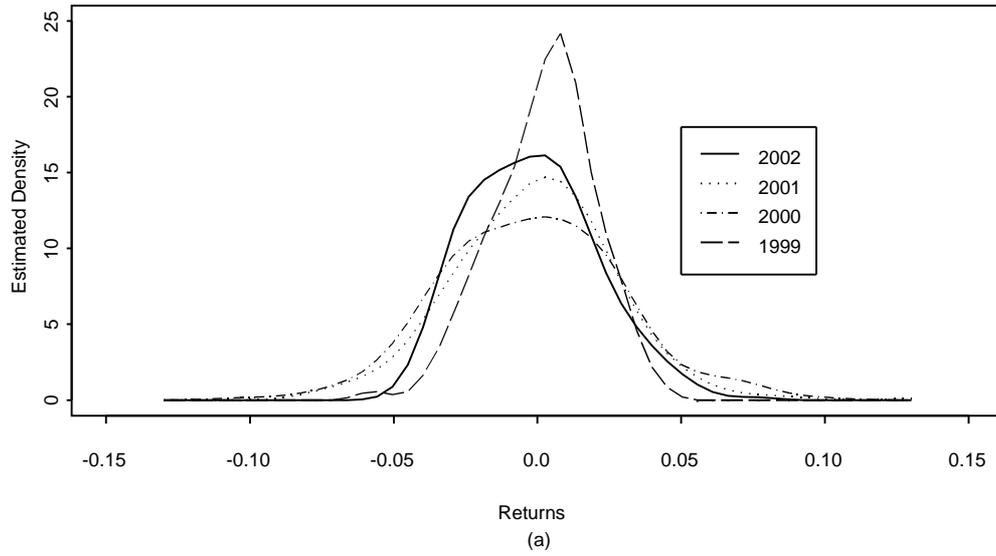


Figure 1. Daily log-return series for Nasdaq (a) and Microsoft (b) from January the 1st 1999 to December 31st of 2002; sample auto-correlation functions for Nasdaq (c) and Microsoft (d).

(a) Yearly Data



(b) Two-Yearly and Four-Yearly Data

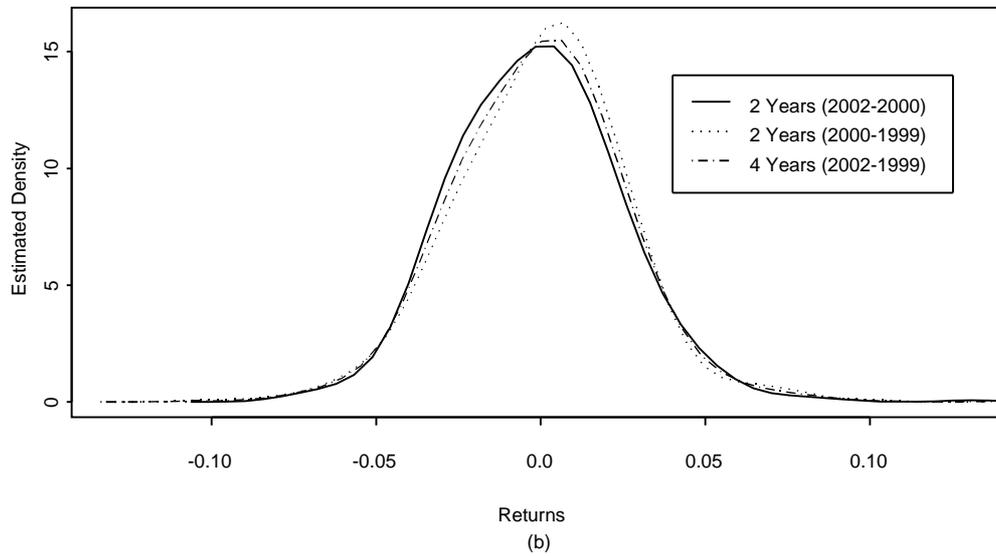
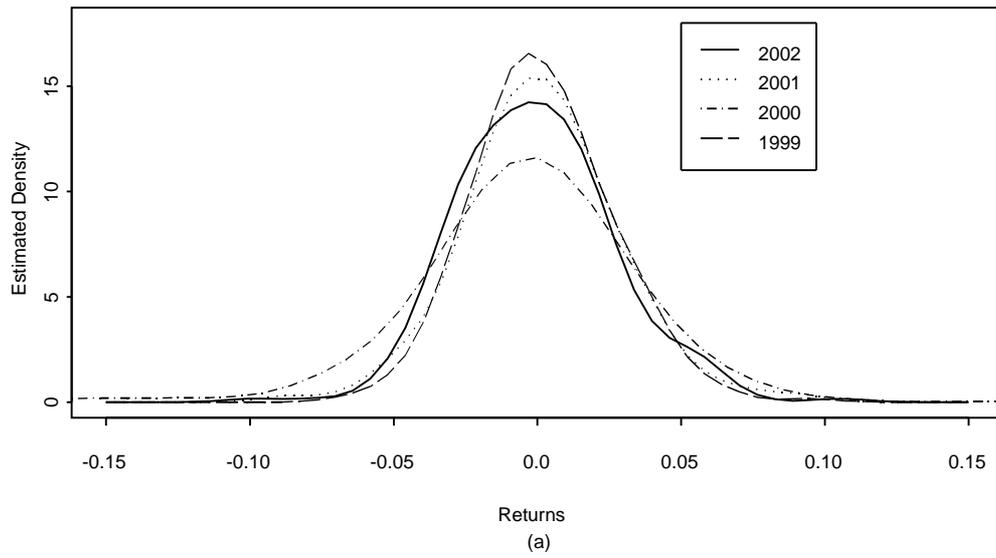


Figure 2. Kernel density estimates for Nasdaq returns.

(a) Yearly Data



(b) Two-Yearly and Four-Yearly Data

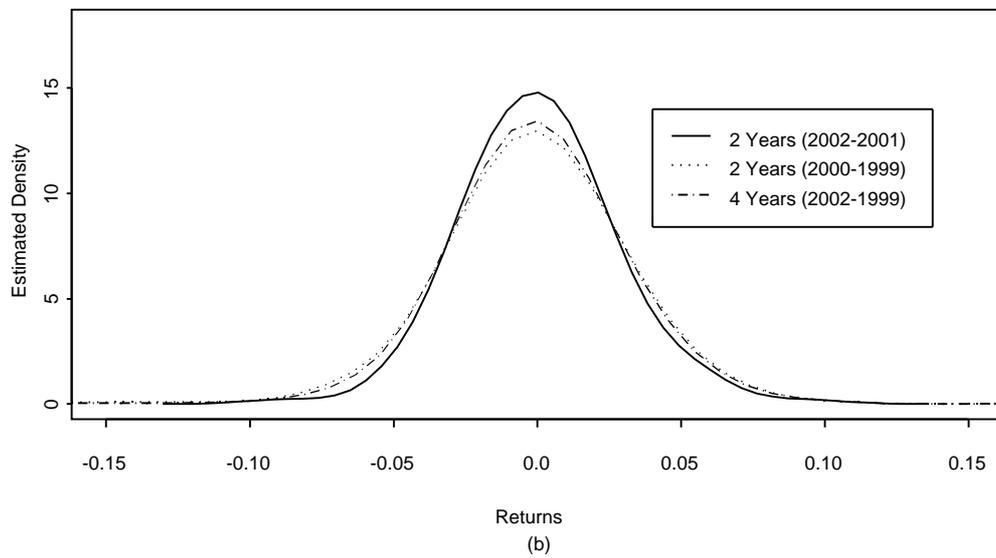
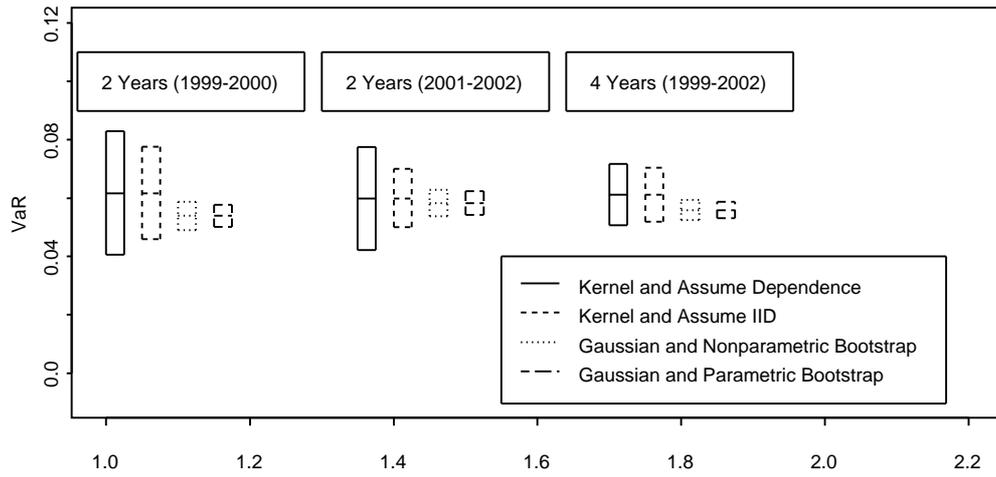


Figure 3. Kernel density estimates for Microsoft returns.

(a) Nasdaq



(b) Microsoft

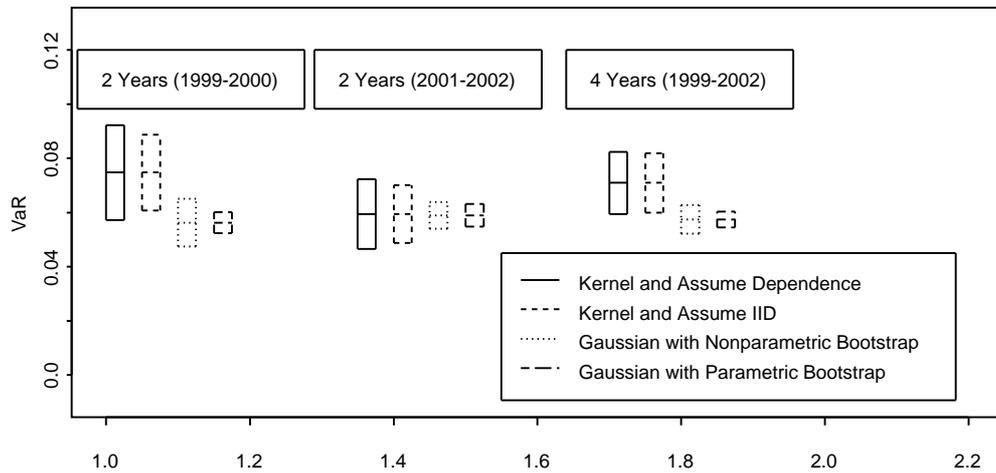


Figure 4. Kernel and Parametric VaR estimates Microsoft (a) and Microsoft (b). The VaR estimates are reported without the negative sign.