Journal of Financial Econometrics, 2005, Vol. 3, No. 2, 227–255

Nonparametric Inference of Value-at-Risk for Dependent Financial Returns

SONG XI CHEN
Iowa State University

CHENG YONG TANG
Iowa State University

ABSTRACT
The article 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: α-mixing, kernel estimation, sample quantile, spectral density estimation, standard error estimation

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 $\{Y_t\}_{t=1}^n$ 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

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

The authors thank Professor Eric Renault and the associate editor for comments and suggestion that improved the presentation of the article and acknowledge support from a National University of Singapore research grant (R-155-000-026-112) and from an Iowa State University grant. Address correspondence to Song Xi Chen, Department of Statistics, Iowa State University, Ames, IA 50011-1210, or e-mail: songchen@iastate.edu.

doi:10.1093/jjfinec/nbi012

which specifies the smallest amount of loss such that the probability of the loss in market value being larger than ν_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 autoregressive conditional heteroskedastic (ARCH) or generalized ARCH (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 as well as data dependence. The advantages of the parametric approaches lay in their easy interpretation. However, they are model dependent and are subject to errors of model misspecification. Recently a VaR estimation method based on extreme value distributions (EVDs) 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 (i.i.d.) returns and dependent returns that can be expressed as a process with i.i.d. 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. Gourieroux, 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 weaker assumptions on the dynamics of the return process. 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 of data. Also, the sample size required by the nonparametric approach is comparable to that required by the EVD approach, as both approaches concentrate on the tail part of the data. When weighing between errors due to small sample size (sampling errors) and errors of model misspecification (model errors), we should go for the former, as the sampling errors can be measured mathematically, whereas the same is difficult to do for the model errors. In other words, the model risk of using a nonparametric approach is lower than that of a parametric approach.

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, yet still works when the data are independent. Developing such an approach is an objective of this article. The dependence structures that are applicable by the techniques proposed here are very wide, including autoregressive moving average (ARMA), ARCH/GARCH, stochastic volatility, and diffusion models, as long as they satisfy the α-mixing.

Another issue that the current article 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 nontrivial task. We propose an approach based on a kernel estimation of a spectral density function that 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 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 would be the case for the other approaches too. A parametric approach may be easily implemented computationally. Its main difficulty is its exposure to the model risk, which is hard to evaluate mathematically. As demonstrated by both theory and simulation studies, we show in this article 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 pth 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 would be altered by the single big loss.

The article is structured as follows. Section 1 introduces various financial return models to which the results of the article are applicable. Nonparametric VaR estimators are outlined in Section 2 and their statistical properties are investigated in Section 3. The issue of standard error estimation is considered in Section 4. Section 5 gives details of practical implementation. Simulation results are presented in Section 6, whereas empirical analyses of two financial returns series are carried out in Section 7. Section 8 gives a general discussion. All the technical details are provided in the appendix.

1 DEPENDENCE STRUCTURE AND MODELS

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

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

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

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

1.1 Linear Processes

For a linear causal process (which includes ARMA models),

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

with i.i.d. innovation $\{\xi_s\}_{s=0}^{\infty}$, Gorodeskii (1977) showed that the process is α-mixing under certain conditions and established the rate for the α-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 α-mixing.

1.2 Markov Processes

Consider a Markov process

$$Y_t = m(\overline{Y}_{t-1,p}) + \sigma(\overline{Y}_{t-1,p})\epsilon_t, \tag{2}$$

where $\overline{Y}_{t-1,p} = (Y_{t-1}, \ldots, Y_{t-p})$ are p-lagged values of Y_t and $\{\epsilon_t\}_{t=1}^T$ are i.i.d. random variables. Here $m(\cdot)$ and $\sigma^2(\cdot)$ are, respectively, the conditional mean and volatility functions of Y_t given $\overline{Y}_{t-1,p}$. The model includes ARCH(p) models. Masry and Tjøstheim (1995) prove that the series is geometric ergodic and α-mixing under some mild conditions.

1.3 GARCH Models

Let $\overline{Y}_{t-1}$ denote the sigma-field generated by $\{Y_i\}_{i\leq t-1}$. The GARCH (p,q) model introduced by Bollerslev (1986) for the return can be specified as follows:

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

where c, α_i, and β_j are all positive parameters. Carrasco and Chen (2002) provides general conditions that ensure the above GARCH (p, q) model as a special case of the generalized hidden Markov models is geometric β-mixing [see Doukhan (1994) for its definition and relationship to α-mixing] and hence geometric α-mixing. They also give more explicit conditions for various GARCH (1,1) models.

1.4 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-homogeneous diffusion model for the log-return is specified by the following stochastic differential equation:

$$d\tilde{Y}_t = \mu(\tilde{Y}_t)dt + \sigma(\tilde{Y}_t)dW_t, \tag{3}$$

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 $t_j = i\Delta$ for some $\Delta > 0$. For a family of diffusion model

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

Genon-Catalot, Jeantheau, and Laredo (2000) have given restrictions on the parameters (ν, α, β) such that $\{\tilde{Y}_{j\Delta}\}_{j=1}^{T}$ is geometric α-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 α-mixing.

1.5 Stochastic Volatility Models

Stochastic volatility models are extensions of the one-factor diffusion model of Equation (3) that allows the volatility of the log-price process $(\tilde{Y}_t)_{t\geq 0}$ to be driven by another diffusion model as follows:

$$d\tilde{Y}_t = \sigma_t dW_t,\ V_t = \sigma_t^2, \text{ and } dV_t = b(V_t)dt + a(V_t)dB_t, \tag{4}$$

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 hidden Markov chain, that the series $\{\tilde{Y}_{j\Delta}\}$ is geometric α-mixing under certain conditions, which implies that the log-return series is geometric α-mixing too.

2 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 $\hat{\nu}_p = Y_{([np]+1)},$ where $Y_{(r)}$ is the rth order statistic. It is just the sample quantile estimator commonly used in statistics by replacing F with F_n in Equation (1). And for this reason, it is called the sample VaR estimator in this article. It is a consistent estimator of ν_p for α-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 in thin, it will be beneficial to carry out kernel smoothing on the empirical distribution F_n. The smoothing essentially leads to an estimator that is a weighted average of the order statistics around $Y_{([np]+1)}$ rather than relying on a single-order statistic. Studies done for i.i.d. data [e.g., Falk (1984) 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 article we focus on α-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, that is,

$$\hat{F}_{n,h}(x) = n^{-1}\sum_{j=1}^{n} G\left(\frac{x - Y_j}{h}\right), \tag{5}$$

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

$$n^{-1}\sum_{j=1}^{n} G\left(\frac{\hat{\nu}_{p,h} - Y_j}{h}\right) = p. \tag{6}$$

This kernel VaR estimator, first introduced by Gourieroux, 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:

Assumption 1. The process $\{Y_t\}_{t=1}^{n}$ is strictly stationary and α-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.

Assumption 2. $f(\nu_p) > 0$ and f has continuous second derivative in a neighborhood $\mathcal{B}(\nu_p)$ of ν_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.

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

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

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

The stationarity and geometric α-mixing assumed in Assumption 1 are satisfied by those models discussed in the previous section under certain conditions.

Assumption 2 contains standard conditions for quantile estimation, whereas conditions in Assumption 3 and 4 are commonly imposed conditions in kernel smoothing. In particular, conditions in Assumption 4 specify a range for the bandwidth h.

3 PROPERTIES OF THE NONPARAMETRIC VaR ESTIMATORS

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

$$\hat{\nu}_p - \nu_p = \frac{F_n(\nu_p) - p}{f(\nu_p)} + O\left\{n^{-3/4}\log(n)\right\} \ a.s., \tag{1}$$

and showed that

$$\mathrm{var}(\hat{\nu}_p) = n^{-1} f^{-2}(\nu_p)\sigma^2(p;\, n)\ \{1 + o(1)\}, \tag{2}$$

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

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

A key quantify that describes the variance properties of the kernal estimator $\hat{\nu}_{p,h}$ is $\sigma_h^2(p;\, n) = \left\{p(1-p) + 2\sum_{k=1}^{n-1}(1-k/n)\gamma_h(k)\right\}$, where $\gamma_h(k) = \mathrm{cov}\left\{G\left(\frac{\nu_p - Y_1}{h}\right), G\left(\frac{\nu_p - Y_{k+1}}{h}\right)\right\}$. The following lemma indicates that $\sigma^2(p;\, n)$ differs from the unsmoothed $\sigma^2(n)$ by an amount of $o(h)$.

Lemma 1. Under the conditions of Assumptions 1–4, $|\sigma_h^2(p;\, n) - \sigma^2(p;\, n)| = o(h)$. The strong convergence of the kernel estimator $\hat{\nu}_{p,h}$ to ν_p at a rate is provided in the following theorem.

Theorem 1. Under the conditions of Assumptions 1–4, $\hat{\nu}_{p,h} = \nu_p + o\{n^{1/2}\log^{-1}(n)\}$ *a.s.* The following theorem provides details on the bias and variance of the kernel estimator.

Theorem 2. Under the conditions of Assumptions 1–4, as $n \to \infty$,

$$E(\hat{\nu}_{p,h}) = \nu_p - \frac{1}{2}h^2\sigma_K^2\ f'(\nu_p)f^{-1}(\nu_p) + o(h^2). \tag{4}$$

$$\mathrm{var}(\hat{\nu}_{p,h}) = n^{-1}f^{-2}(\nu_p)\sigma_h^2(p;\, n) - 2n^{-1}hf^{-1}(\nu_p)b_K + o(h/n), \tag{5}$$

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

Remark 1. The difficulty in the VaR estimation is clearly spelt out by the appearance of $f^2(\nu_p)$ in the denominator of the leading variance terms of both nonparametric estimators as given in Equation (2) and Equation (5). It clearly demonstrates that the variability of VaR estimates increases when p gets smaller as $f(\nu_p)$ tend to become thinner. The latter reflects the reality of fewer return observations around ν_p.

Remark 2. From Lemma 1 and Equation (5), we have

$$\text{var}(\hat{\nu}_{p,h}) = n^{-1}f^{-2}(\nu_p)\sigma^2(p;\, n) - 2n^{-1}hf^{-1}(\nu_p)b_K + o(h/n), \tag{6}$$

which together with Equation (2) indicates that both the kernel and the sample VaR estimators share the same leading asymptotic variance term. However, the kernel estimator reduces the variance in the second order of h/n as $b_k > 0$. This second-order reduction is still significant considering that (i) the data are thin in the tail and (ii) even a second-order reduction can still be significant in financial terms as, say, a 10% reduction can translate to a large reduction of provision in the absolute dollar term.

Remark 3. The impact of data dependence on the variance of VaR estimators is clearly felt through either $\sigma^2(p;\, n)$ or $\sigma_h^2(p;\, n)$ via the covariances $\gamma(k)$ and $\gamma_h(k)$, respectively. 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 and it is ignored, the asymptotic variance of the nonparametric VaR estimates would be wrongly regarded as $n^{-1}p(1-p)f^{-2}(\nu_p)$, resulting in a wrong assessment of the variability.

Let $\text{MSE}(\hat{\nu}_{p,h}) =: E(\hat{\nu}_{p,h} - \nu_p)^2$ be the mean square error (MSE) of $\hat{\nu}_{p,h}$. From Theorem 2 and Equation (6),

$$\text{MSE}(\hat{\nu}_{p,h}) = n^{-1}f^{-2}(\nu_p)\sigma^2(p;\, n) - 2f(\nu_p)b_K hn^{-1} + \frac{1}{4}h^4\sigma_K^4\left\{f'(\nu_p)f^{-1}(\nu_p)\right\}^2 + o(h/n + h^4).$$

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

$$h_{opt} = \left\{\frac{2f^3(\nu_p)b_K}{\sigma_K^4 f'^2(\nu_p)}\right\}^{1/3} n^{-1/3}, \tag{7}$$

which can be estimated by plugging in the estimates of $f(\nu_p)$ and $f'(\nu_p)$,an issue that will be discussed in Section 5. Substituting h_{opt} into Equation (8), the optimal MSE is then

$$\text{MSE}(\hat{\nu}_{p,h}) = n^{-1}\sigma^2(p;\, n) - 3\left(2^{-2/3}\right)b_K^{4/3}\sigma_K^{-4/3}f^2(\nu_p)\{f'(\nu_p)\}^{-4/3}n^{-4/3} + o\left(n^{-4/3}\right), \tag{8}$$

which indicates a reduction to the MSE of the second order.

The following theorem establishes the asymptotic normality of $\hat{\nu}_{p,h}$, which can be used to construct asymptotic confidence intervals for ν_p as well as to carry out tests on hypotheses regarding ν_p.

Theorem 3. Under the conditions of Assumptions 1–4, as $n \to \infty$,

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

where $\sigma^2(p) = \lim_{n\to\infty}\sigma^2(p; n)$, whose existence is guaranteed by Assumption 1.

4 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 nontrivial 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 $\{Z_t\}_{t=1}^n = \left\{G\left(\frac{\nu_p - Y_t}{h}\right)\right\}_{t=1}^n$, 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

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

be the spectral density of $\{Z_t\}_{t=1}^n$. From the Davyadov inequality, $|\gamma_h(k)| \le C\alpha(k)$ for some constant $C > 0$. Thus, from Assumption 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), $\lim_{n\to\infty}\left\{\sigma_h^2(p; n) - 2\pi\phi(0)\right\} = 0$. Hence the estimation of $\sigma_h^2(p; n)$ can be achieved by estimating $\phi(0)$.

Define

$$I_n(\omega_j) = n^{-1}\left|\sum_{l=1}^{n} Z_l e^{-il\omega_j}\right|^2, \ j = 0, \pm 1, \ldots, \pm[n/2], \qquad (9)$$

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

$$I_n(\omega_j) = (2\pi)\phi(\omega_j)E_j + R_j, \tag{10}$$

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 $\{Z_t\}_{t=1}^n$ are not observable due to the involvement of the unknown ν_p. Let $\hat{Z}_t = G\left(\frac{\hat{\nu}_p - Y_t}{h}\right)$, and $\hat{I}_n(\omega_j)$ be the periodograms defined on $\left\{\hat{Z}_t\right\}_{t=1}^n$ by replacing Z_t by $\hat{Z}_t$ in Equation (9). As $\hat{Z}_t = Z_t + o_p(1)$uniformly for all t'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 Equation (10),

$$\hat{I}_n(\omega_j) = (2\pi)\phi(\omega_j)E_j + R_j', \tag{11}$$

where R_j' are asymptotically negligible. However, a single periodogram does not lead to a consistent estimation of the spectral density. Smoothing the periodograms over neighboring frequencies are needed. There is a substantial amount of literature on kernel estimation of spectral density in statistics [Brockwell and Davis (1991)] and in econometrics [Andrews and Monahan (1992)].

A commonly used approach, for instance, that used in Fan and Gijbels (1996) and Lee (1997), is to take the logarithm on both sides of Equation (11) and ignore $\{R_j'\}$

$$\log\left\{\hat{I}_n(\omega_j)/(2\pi)\right\} = \log\{\phi(\omega_j)\} + \log(E_j) \text{ for } j \in T. \tag{12}$$

Note that $E\{\log(E_j)\} = -0.57721$ (the Euler constant) and $\text{var}\{\log(E_j)\} = \pi^2/6$. Let

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

Then Equation (12) can be approximated by the following fixed-design nonparametric regression:

$$W_j = m(\omega_j) + \eta_j \; j \in T \tag{13}$$

where $\{\eta_j\}_{j\in T}$ are i.i.d. 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

$$\hat{m}_b(\omega) = \frac{\sum_{j\in T} K_1\left(\frac{\omega-\omega_j}{b}\right)W_j}{\sum_{j\in T} K_1\left(\frac{\omega-\omega_j}{b}\right)}, \tag{14}$$

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

$$\hat{\phi}(0) = \exp\{\hat{m}_b(0)\}. \tag{15}$$

Fan and Gijbels (1996) considered a local linear kernel estimation of spectral density. As the design points are fixed at the Fourier frequencies ω_j 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, for instance, those given in Andrews and Monahan (1992), show that $\hat{m}_b(0) \xrightarrow{p} m(0)$, and thus $\hat{\phi}(0) \xrightarrow{p} \phi(0)$ as $n \to \infty$.

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 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.

5 PRACTICAL IMPLEMENTATION

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

5.1 Kernel VaR Estimation

As mentioned in Section 4, 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 Equation (7) 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 σ_K^2 are known after choosing K. The approach we use in 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 ν_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 that is situated in the tail of the distribution. In particular, let

$$\omega_{\gamma,\mu,\sigma}(x) = \frac{1}{\sigma}\left(1 + \gamma\frac{x-\mu}{\sigma}\right)^{-\left(1+\frac{1}{\gamma}\right)} \tag{16}$$

be the density of a GP distribution with a scale parameter σ, a shape parameter γ, and a truncation level μ [see Reiss and Thomas (2001) for comprehensive discussions on the theory and applications of GP distributions]. For a 99% VaR, we fit the lower 5% of the data to a GP model, which means taking $\mu = \hat{\nu}_{0.05}$. For other levels of VaR, μ 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, $\omega_{\hat{\gamma},\hat{\sigma},\hat{\nu}_{0.05}}(\hat{\nu}_p)$ and $\omega'_{\hat{\gamma},\hat{\sigma},\hat{\nu}_{0.05}}(\hat{\nu}_p)$, which then lead to a practically useful h.

5.2 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

$$R(b) = \frac{1}{n}\sum_{j\in T} q_{nj}\{m(\omega_j) - \hat{m}_b(\omega_j)\}^2, \tag{17}$$

be 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

$$r(b) = \frac{1}{n}\sum_{j\in T} q_{nj}\{W_j - \hat{m}_b(\omega_j)\}^2 + \frac{\pi^2}{6}\left(1 - \frac{4\pi K(0)}{nb}\right)\sum_{j\in T} q_{nj}. \tag{18}$$

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 $\hat{f}(\nu_p) = \omega_{\hat{\gamma},\hat{\sigma},\hat{\nu}_{0.05}}(\hat{\nu}_{ph})$, which is a by-product of the h-bandwidth selection discussed earlier.

6 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: $Y_t = 0.5Y_{t-1} + \epsilon_{t,}\ \epsilon_t \overset{iid}{\sim} N(0,\ 1)$;

(ii) AR(2) model: $Y_t = 0.9Y_{t-1} - 0.2Y_{t-2} + \epsilon_t,\ \epsilon_t \overset{iid}{\sim} N(0,\ 1)$;

(iii) MA(2)model: $Y_t = \epsilon_t + 0.65\epsilon_{t-1} + 0.24\epsilon_{t-2}\ \epsilon_t \overset{iid}{\sim} N(0,\ 1)$;

(iv) ARCH(1) model: $Y_t = 0.5Y_{t-1} + \epsilon_t\epsilon_t^2 = 4 + 0.5\epsilon_{t-1}^2 + \eta_t,\ \eta_t \overset{iid}{\sim} N(0,\ 1)$;

(v) Stochastic volatility (SV) model:

$Y_t = V_t\epsilon_t,\ \log(V_t) = 0.6\ \log(V_{t-1}) + \eta_t,\ \epsilon_t \overset{iid}{\sim} N(0, 1)$,

$\eta_t \overset{iid}{\sim} N(0{,}0.5)$, $\text{cov}(\epsilon_t, \eta_t) = 0.5$, and $\text{cov}\left(\epsilon_t, \eta_{t-j}\right) = 0$ for $j > 0$;

(vi) Diffusion model: $d\tilde{Y}_t = 0.4\left(2 - \tilde{Y}_t\right)dt + \sqrt{50}\ dB_t$, Bt 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 straightforward. To generate the ARCH series, we generate an i.i.d. 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 $\eta_t \overset{iid}{\sim} N(0{,}1)$. The innovations for the ARCH series are then $\epsilon_t = \delta_t\sqrt{Z_t}$. For the SV series, we generate two i.i.d. standard normal series $\{\epsilon_t\}$ and $\{\zeta_t\}$, and let $\eta_t = 0.5\epsilon_t + 0.5\varsigma_{t,}$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 $N\left(\tilde{Y}_{t-\Delta}e^{-0.4\Delta} + 2\left(1 - e^{-0.4\Delta}\right), 50\left(1 - e^{0.8\Delta}\right)/0.8\right)$, 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 prerun 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 $K(u) = \frac{1}{\sqrt{(2\pi)}}exp\left(-u^2/2\right)$ as the kernel for estimating VaR and the biweight kernel $K_1(u) = \frac{15}{16}\left(1 - u^2\right)^2 I(|u| \le 1)$ 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 6 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 an accurate prediction of the real standard deviation of the kernel and sample VaR estimates, even when the sample size is small. We observe that the bias, the

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

	Kernel			Sample			Est. SD	
N	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

	Kernel			Sample			Est. SD	
N	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

	Kernel			Sample			Est. SD	
N	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

standard deviations, and the RMSEs 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 Equation (8). 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

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

	Kernel			Sample			Est. SD	
N	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 kernal and sample quantile estimates for 99% VaR for the SVM model with the true 99% VaR at 2.383659

	Kernel			Sample			Est. SD	
N	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

	Kernel			Sample			Est. SD	
N	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

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.

7 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 1, 1999, to December 31, 2002, which consist of four years of data (n = 1000). These two return series are displayed in Figure 1 a and c, respectively. The sample autocorrelation functions for the square of the returns are plotted in Figure 1b and d, which indicate substantial amount of dependence in these two series. To formally confirm it, we calculate the Box-Pierce Q-test statistic $Q = n\sum_{k=1}^{26}\hat{\gamma}(\mathrm{k})^2$, where $\hat{\gamma}(\mathrm{k})$ is the sample autocorrelation. If the series are independent, Q should be distributed as χ^2_{26}. It is found that Q = 367.42 for the Nasdaq and Q = 87.43 for Microsoft. Both are overwhelmingly larger than $\chi^2_{26}(0.01) = 45.64$, the 1% percentile of χ^2_{26}. The p-value of these two series are almost zero, indicating strong dependence in both of the returns series.

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 of 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 also be 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 years of 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 years 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 subseries, 0.003 and 0.9433 for the 2001–2002 subseries, 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 subseries, 0.005 and 0.7233 for the 2001–2002 subseries, 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

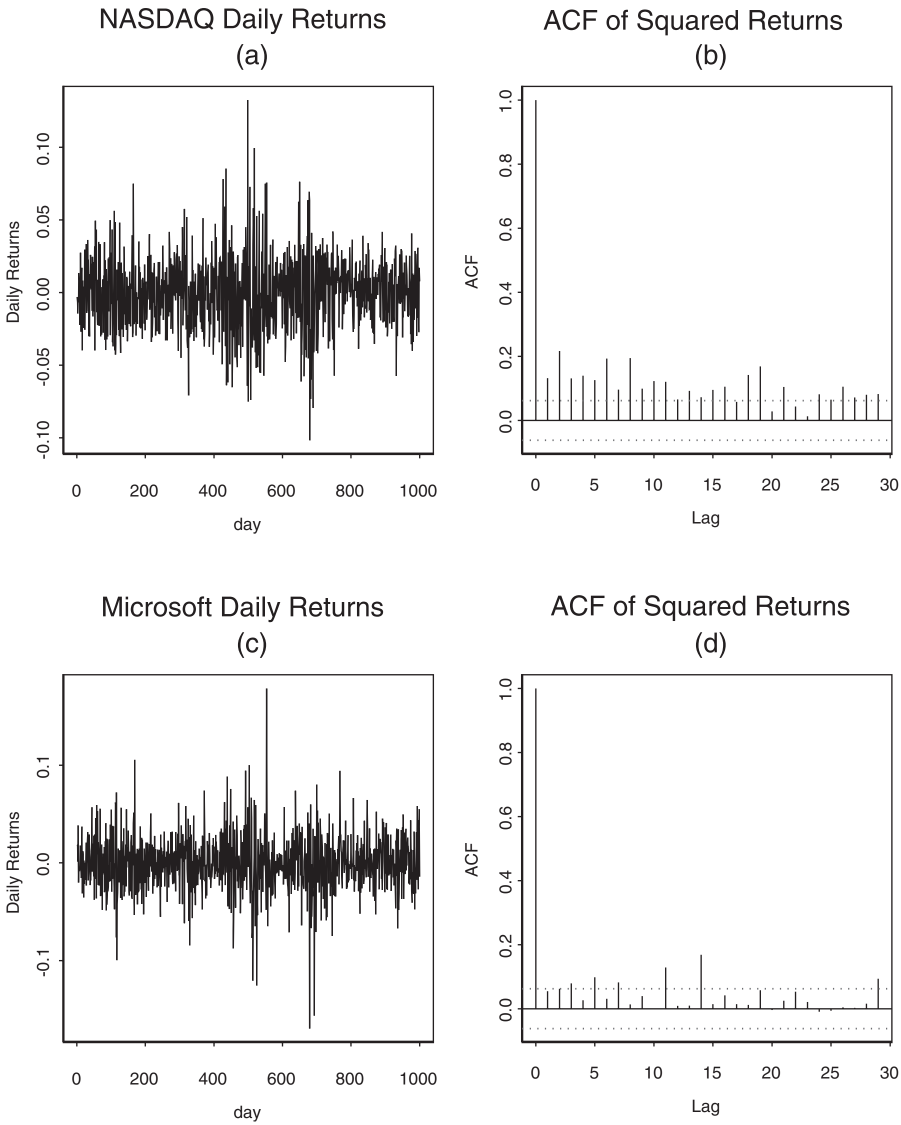

Figure 1 Daily log-return series for Nasdaq (a) and Microsoft (b) from January 1, 1999, to December 31, 2002; sample autocorrelation functions for Nasdaq (c) and Microsoft (d).

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

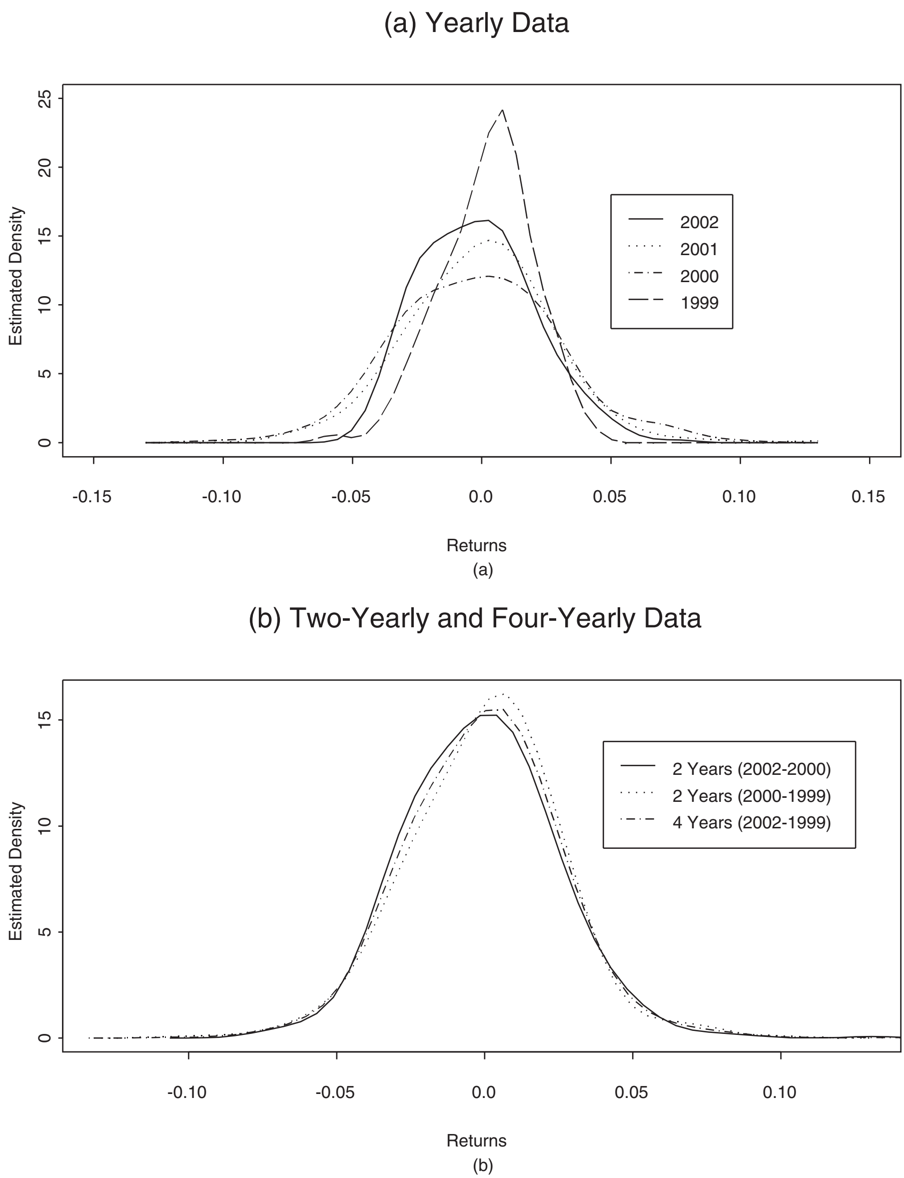

Figure 2 Kernel density estimates for Nasdaq returns.

estimates based on the independent 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 $N(\overline{Y}, S^2)$ where $\overline{Y}$ and S^2 are, respectively, the sample mean and

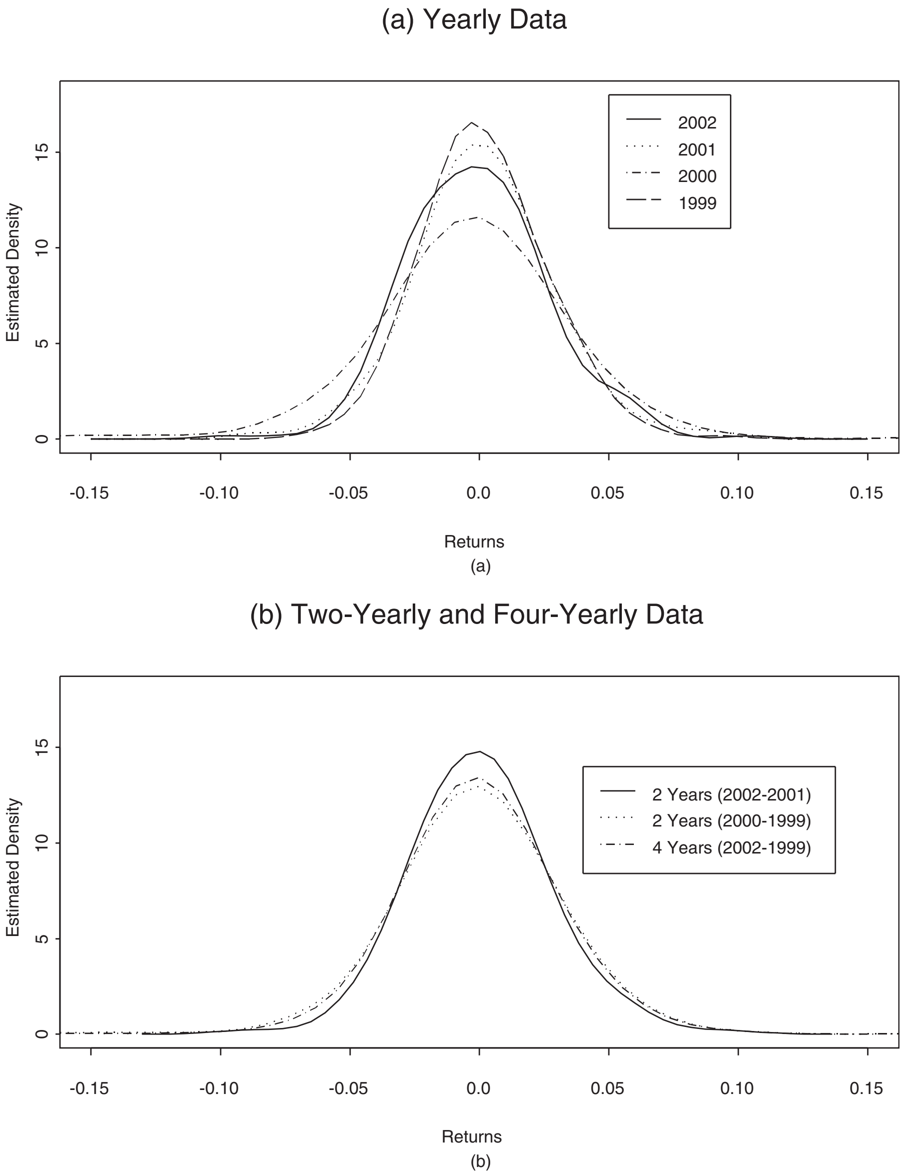

Figure 3 Kernel density estimates for Microsoft returns.

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 indicate possible severe underestimation of the variability. We also observe quite large discrepancies between the parametric and nonparametric VaR estimates for the Microsoft series in the period 1999–2000.

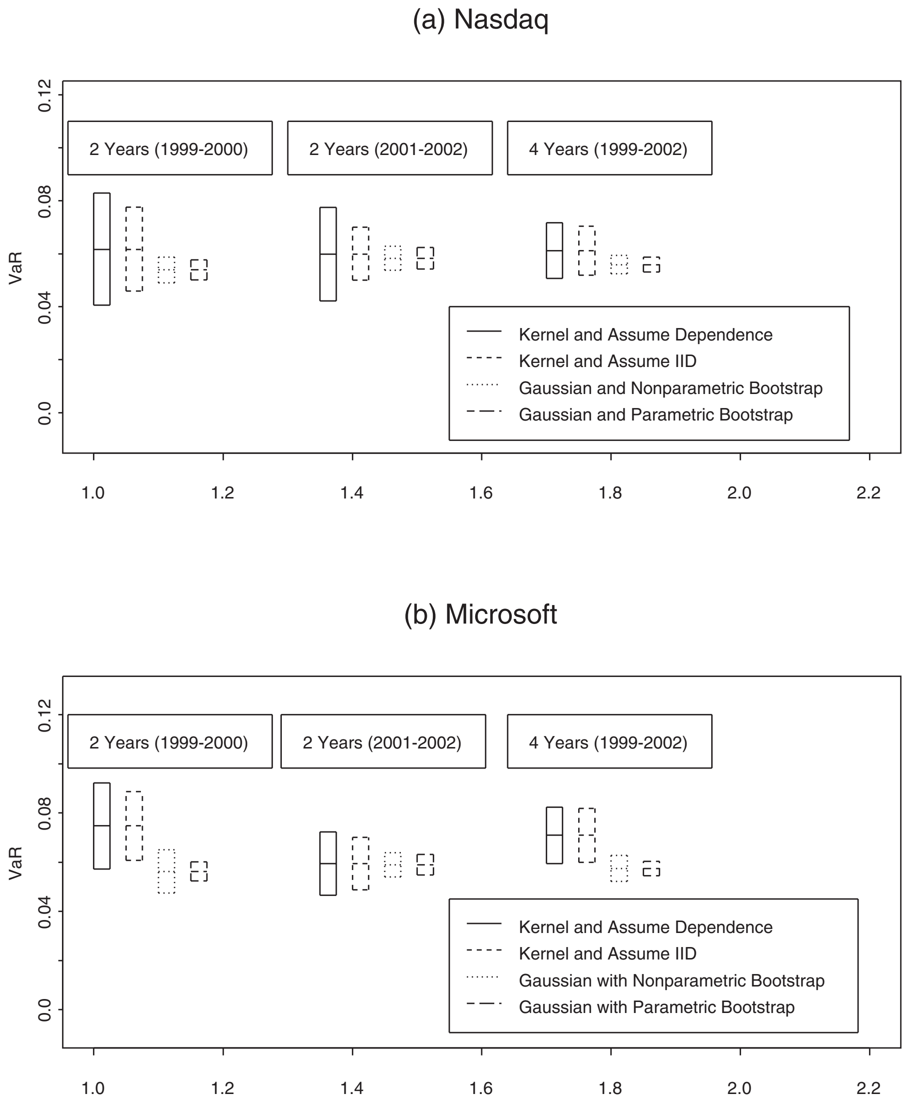

Figure 4 Kernel and parametric VaR estimates Microsoft (a) and Microsoft (b). The VaR estimates are reported without the negative sign.

8 DISCUSSIONS

Despite its popular use by financial institutions for risk management, VaR is known to be, in general, a noncoherent measure of risk, as it is not subadditive [Artzner et al. (1999)]. Examples of VaR being not subadditive have been given in Artzner et al. (1999) when the return distribution F is discrete. The properties of the VaR improve when the returns are continuously distributed. It is known that VaR is subadditive 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 subadditive within any family of distributions generated by the location-scale transformation of a distribution F_0, that is, $\mathcal{F}(x) = \{F|F(x) = F_0\left(\frac{x-a}{b}\right)$ for any a, $b \in R\}$, which includes the Gaussain 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 (2004). 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 article 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 may translate to large savings in absolute dollar terms. This is especially the case when the sample size is small. Another advantage 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 $\{I(Y_t \leq \hat{\nu}_p)\}$ requires a much larger sample size than that required for the smoothed series $\left\{G\left(\frac{\hat{\nu}_{p,h}-Y_t}{h}\right)\right\}$. Smoothing significantly enhances the estimation of standard errors.

APPENDIX: PROOFS

A.1 Proof of Lemma 1

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

$$\sup_{(x,y)\in R^2} |F_k(x,y) - F(x)F(y)| \leq \alpha(k), \tag{A.1}$$

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 Equation (A.1) that $|\gamma(k)| \leq C_1\alpha(k)$ and $|\gamma_h(k)| \leq C_1\alpha(k)$. These means

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

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

$$\sum_{k-1}^{n-1}(1-k/n)|\gamma_h(k)-\gamma(k)|\leq C_1h^{2\tau}\sum_{k=1}^{\infty}\alpha^{1-\tau}(k)=O(h^{2\tau}).$$

This completes the proof. ■

A.2 Proof of Theorem 1

The theorem is proved if

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

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

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

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

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

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 Equation (A.2) is established if

$$\sum_{n=1}^{\infty}P\Big\{|\hat{F}_{n,h}(\nu_p+\lambda\epsilon)-F(\nu_p+\lambda\epsilon)|\geq n^{-1/2}\log(n)\eta\Big\}<\infty \tag{A.3}$$

for $\lambda=1$ an -1.

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

$$\begin{aligned}E\Big\{\hat{F}_{n,h}(\nu_p+\epsilon)-F(\nu_p+\epsilon)\Big\}&=\int_{-1}^{1}\{F(\nu_p+\epsilon-hu)-F(\nu_p+\epsilon)\}K(u)du\\ &=\int_{-1}^{1}f'(\nu_p+\epsilon-\theta_3hu)h^2u^2K(u)du,\end{aligned}$$

where $\theta_3\in(0,1)$. As f' is bounded in a neighborhood of ν_p and $nh^4\log^2(n)\to0$ as assumed in condition (iv), we have

$$|E\left\{\hat{F}_{n,h}\left(\nu_p+\epsilon\right)\right\}-F\left(\nu_p+\epsilon\right)|=O\left(h^2\right)=o\left\{n^{-1/2}\log(n)\right\}. \tag{A.4}$$

Let $Z_j = G_h(\nu_p + \epsilon - Y_j) - E\{G_h(\nu_p + \epsilon - Y_j)\}$. Clearly, $E(Z_j) = 0$ and $|z_j| \le 2$. Let $q = b_0 n^{1/2}\log(n)$, $p = n/(2q)$, and $u^2(q)=\max_{0\le j\le 2q-1} E\left(\sum_{l=[jp]+1}^{[(j=1)_p]} Z_l\right)^2$. From an inequality given in Bosq (1998:Theorem 1.3) for α-mixing sequences,

$$P\left(|\sum_{j=1}^{n} Z_j| > c_2 n\epsilon\right) \le 4\exp\left(-\frac{c_2^2\epsilon^2 q}{8\sigma^2(q)}\right)+22\left(1+\frac{8}{c_2\epsilon}\right)^{1/2} q\alpha\{[n/(2q)]\}, \tag{A.5}$$

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

$$4\exp\left(-\frac{c_1^2\epsilon^2 q}{8\sigma^2(q)}\right) \le 4\exp\{-C(b_0)\log^2(n)\}, \tag{A.6}$$

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

$$\begin{aligned}22\left(1+\frac{8}{c_1\epsilon}\right)^{1/2} q\alpha\{[n/(2q)]\} &\le Cn^{3/4}\log^{-1/6}(n)\alpha\left\{\left[n^{1/2}\log^{-1}(n)/2\right]\right\} \\ &\le Cn^{3/4}\log^{-1/6}(n)\rho^{n^{1/2}\log^{-1}(n)/2},\end{aligned} \tag{A.7}$$

which converges to zero sufficiently fast. From Equations (A.5), (A.6), and (A.7), we have

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

This and Equation (A.4) imply Equation (A.3) for $\lambda = 1$. Thus we complete the proof of Theorem 1. ■

A.3 Proof of Theorem 2

We derive the variance of $\hat{F}_{n,h}(\nu_p)$in the following lemma.

Lemma 2. Under conditions (i) and (ii),

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

Proof. Note that

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

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 conditions (i) and (ii), for l, k = 1 or 2 and 1+ k ≥ 3,

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

Proof. Let $\mu_h(x) = E\left\{\hat{F}_{n,h}(x)\right\}$. 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

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

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

$$\begin{aligned}\text{cov}\left[\left\{p - \hat{F}_{n,h}(\nu_p)\right\}^2, \left\{p - \hat{F}_{n,h}(\nu_p)\right\}^2\right] &= E\left\{p - \hat{F}_{n,h}(\nu_p)\right\}^4 \\ &\quad - E^2\left\{p - \hat{F}_{n,h}(\nu_p)\right\}^2.\end{aligned}$$

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

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

Thus completes the proof. ■

Proof of Theorem 2

Since Equation (4) can be easily established via a standard derivation of the bias, we only derive the variance part of the theorem. Let

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

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

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

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

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

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

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

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

Since f has bounded second derivative near ν_p as implied by condition (iv), we have

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

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 of Equation (A.10),

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

From Lemma 3,

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

Employing the delta method,

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

This together with Lemmas 1 and 2 leads to Equation (5). ■

A.4. Proof of Theorem 3

From Equation (A.11) and note that $\hat{f}_{n,h}(\nu_p) = f(\nu_p) + O_p\left\{(nh)^{1/2} + h^2\right\}$,

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

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

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

the smaller blocks

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

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

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

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

$$\begin{aligned} \text{var}(S_{n,2}) &= \frac{r\sigma^2(p)}{nf^2(\nu_p)}\{1 + o(1)\} \to 0 \text{ and } \text{var}(S_{n,3}) \\ &= \frac{(n - r(k + k'))\sigma^2(p)}{nf^2(\nu_p)}\{1 + o(1)\} \to 0. \end{aligned}$$

Therefore, for $l = 2$ and 3

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

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

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

Let $\Delta_n = S_{n,1} - n^{-1/2}\sum_{j=1}^{r} W_{j,n}$. Then (A.15)

$$p(|\Delta_n| > \epsilon) \le \sum_{j=1}^{r} p\big(|V_{j,n} - W_{j,n}| \le \epsilon\sqrt{n}\big)$$
$$\le C_1 r^{3/2} n^{-1/2} (rk)^{1/2} \alpha(\mathrm{k}') \le C_2 r^{3/2} \rho^{k'}.$$

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

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

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

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

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

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

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

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 Equations (A.13), (A.16), and (A.17).

A.5 Derivation of Equation (18)

Recall that

$$W_j = m\big(\omega_j\big) + \eta_j, \tag{A.18}$$

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

$$\hat{m}_b(\omega_i) = \frac{\sum_{j \in T} K\big(\frac{\omega_i - \omega_j}{b}\big) W_j}{\sum_{j \in T} K\big(\frac{\omega_i - \omega_j}{b}\big)} =: \sum_{j \in T} \omega_{b,i} W_j$$

It is obvious that $E\big(W_{\mathrm{i}}^2\big) = m^2(\omega_{\mathrm{i}}) + \pi^2/6$, and

$$E\{W_i \hat{m}_b(\omega_i)\} = E\left[\{m(\omega_i) + \eta_i\} \sum_{j \in T} \omega_{b,j} W_j\right]$$
$$= E\{m(\omega_i) + \eta_i\} \sum_{j \in T} \omega_{b,j} \Big\{m\big(\omega_j\big) + \eta_j\Big\}$$

$$= m(\omega_i)\sum_{j\in T}\omega_{b,j}m(\omega_j) + E(\omega_{b,i}\pi^2/6)$$

$$= m(\omega_i)E\{\hat{m}_b(\omega_i)\} + E\frac{K(0)\pi^2}{6nb\hat{\eta}(\omega_{\mathrm{i}})},$$

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

$$E\{W_i - \hat{m}_b(\omega_i)\}^2 = m^2(\omega_i) + \pi^2/6 - 2\left[m(\omega_i)E\{\hat{m}_b(\omega_i)\} + \tfrac{K(0)\pi^3}{3nb}\right] + E\hat{m}_b^2(\omega_i)$$

$$= E\{m(\omega_i) - \hat{m}_b(\omega_i)\}^2 + \frac{\pi^2}{6}\left(1 - \frac{4\pi K(0)}{nb}\right) + O(nb)^{-2}.$$

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.

Received March 26, 2004; revised January 5, 2005; accepted January 12, 2005

REFERENCES

Andrews, W. K., and J. C. Monahan. (1992). "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator." *Econometrica 60*, 953–966.

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

Balkema, A. A., and L. de Haan. (1974). "Residual Life Times at Great Age." *Annals of Probability 2*, 792–804.

Bellini, F., and G. Figá-Talamanca. (2002). "Detecting and Modeling Tail Dependence." Presented at the Quantative Methods in Finance 2003 Conference, Cairns, Australia, December 2003.

Bollerslev, T. (1986). "Generalized Autoregressive Conditional Heteroscedasticity." *Journal of Econometrics 31*, 307–327.

Brillinger, D. R. (1981). *Time Series, Data Analysis and Theory*. New York: McGraw-Hill.

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

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

Cai, Z.-W., and G. G. Roussas. (1997). "Smoothed Estimate of Quantile Under Association." *Statistics and Probability Letters 36*, 275–287.

Cai, Z.-W., and G. G. Roussas. (1998). "Efficient Estimation of a Distribution Function Under Quadrant Dependence." *Scandinavian Journal of Statistics 25*, 211–224

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

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

Duffie, D., and Pan, J. (1997). " An Overview of Value at Risk." *Journal of Derivatives* vol?, 7–49.

Dowd, K. (2001). "Estimating VaR with Order Statistics." *Journal of Derivatives* vol?, 23–30.

Eberlein, E. (1999). "Application of Generalized Hyperbolic Lévy Motions to Finance." Manuscript.

Embrechts, P., A. McNeil, and D. Straumann. (2002) "Correlation and Dependence in Risk Management: Properties and Pitfalls." In: M.A.H. Dempster, ed. *Risk Management: Value at Risk and Beyond.* Cambridge: Cambridge University Press.

Embrechts, P., S. Resnick, and G. Samorodnitsky. (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 I. Gijbels. (1996). *Local Polynomial Modeling and Its Applications*. London: Chapman & Hall.

Genon-Catalot, V., T. Jeantheau, and C. Laredo. (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., O. Scaillet, and J. P. Laurent. (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 ed. New York: McGraw-Hill.

Masry, E., and D. Tjøstheim. (1995). "Nonparametric Estimation and Identification of Nonlinear ARCH Time Series." *Econometric Theory 11*, 258–289.

McNeil, A. J., and R. Frey. (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 L. T. Tran. (1985). "Some Mixing Properties of Time Series Models." *Stochastic Processesand Application 19*, 297–303.

Reiss, R. D., and M. Thomas. (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 α-Mixing Condition." *Proceedings of the National Academy of Sciences, U.S.A. 42*, 43–47.

Scaillet, O. (2004) "Nonparametric Estimation and Sensitivity Analysis of Expected Shortfall." *Mathematical Finance 14*, 115–129.

Sheather, S. J., and J. S. Marron. (1990). "Kernel Quantile Estimators." *Journal of American Statistical Association 85*, 410–416.

Yokoyama, R. (1980). "Moment Bounds for Stationary Mixing Sequences." *Probability Theory and Related Fields 52*, 45–87.

Yoshihara, K. (1995). "The Bahadur Representation of Sample Quantiles for Sequences of Strongly Mixing Random Variables." *Statistics and Probability Letters 24*, 299–304.