The Journal of Risk (15–32) Volume 12/Number 4, Summer 2010

Kernel quantile-based estimation of expected shortfall

Keming Yu
Department of Mathematical Sciences, Brunel University, Uxbridge, Middlesex UB8 3PH, UK; email: Keming.Yu@brunel.ac.uk

Abdallah K. Ally
Department of Mathematical Sciences, Brunel University, Uxbridge, Middlesex UB8 3PH, UK; email: mapgaaa@brunel.ac.uk

Shanchao Yang
School of Mathematics, Guangxi Normal University, People's Republic of China; email: scyang@mailbox.gxnu.edu.cn

David J. Hand
Institute for Mathematical Sciences, Department of Mathematics, Imperial College, South Kensington Campus, London SW7 2AZ, UK; email: d.j.hand@imperial.ac.uk

Since its proposal as an alternative risk measure to value-at-risk (VaR), expected shortfall (ES) has attracted a great deal of attention in financial risk management, primarily owing to its coherent properties. Recently, there has been an upsurge of research on the estimation of ES from a non-parametric perspective. The focus of this paper is on a few kernel-based ES estimators, including jackknife-based bias-correction estimators that have theoretically been documented to reduce bias. Bias reduction is particularly effective in reducing the tail estimation bias as well as the consequential bias that arises in kernel smoothing and finite-sample fitting and, thus, serves as a natural approach to the estimation of extreme quantiles of asset price distributions. By taking advantage of ES as an integral of the quantile function, a new type of ES estimator is proposed. To compare the performance of the estimators, a series of comparative simulation studies are presented and the methods are applied to real data. An estimator that has an analytical form turned out to perform the best.

1 INTRODUCTION

Risk measures play an important role in market risk management. The most common risk measure is value-at-risk (VaR), which is defined as the maximum potential loss of a given portfolio over a prescribed holding period at a given confidence level. Its use was recommended by the Basel Committee in 1996 (Basel Committee on

Banking Supervision (1996)) and also in the latest proposed Basel II standards (Basel Committee on Banking Supervision (2003)). Following these guidelines, financial regulators all over the world have adopted VaR to design capital adequacy standards for banks and financial institutions. In addition, financial firms have adopted VaR for internal risk management and the allocation of resources (see Danielsson *et al* (2005)).

Artzner *et al* (1997) introduced the concept of coherent risk measures (see also Artzner *et al* (1999) and Artzner (1999)). They argued that a risk measure should satisfy four desirable properties: monotonicity, subadditivity, positive homogeneity and translation invariance.

Artzner *et al* (2002) further extended the concept of coherence to multi-period risk measurement, explaining why and how to deal with the definition, acceptability, computation and management of risk in a genuinely multi-temporal way. They pointed out that VaR is not a coherent risk measure because it does not satisfy the subadditivity condition. This implies that the risk of a portfolio, when measured by VaR, can be larger than the sum of the standalone risks of its components.

In order to construct a risk measure that is both coherent and easy to compute and estimate, the expected shortfall (ES) was proposed and discussed by Acerbi *et al* (2001). Expected shortfall is defined as the expectation of losses above VaR. Acerbi and Tasche (2002a) then provided an integral representation of ES, while Acerbi and Tasche (2002b) showed that ES arises in a natural way from the estimation of the "average of the $100\alpha\%$ worst losses" in a sample of returns to a portfolio.

The estimation of VaR has received a great deal of attention in the literature, and in recent years there has been a growing interest in the estimation of ES from a non-parametric angle; see Scaillet (2004, 2005) and Chen (2008).

Basically, these authors propose a two-step non-parametric method that involves, first, non-parametric estimation of VaR and, then, non-parametric estimation of the expectation of a truncated variable at VaR. Fermanian and Scaillet (2005) also explored some interesting applications of these methods in credit risk environments.

The above estimators, while interesting, may have substantial bias arising from the boundary effects of kernel estimation for small probability levels. In this paper, we note that ES can be expressed as an integral of the quantile function. By virtue of this expression, studying the analytical properties of ES becomes analytically tractable. Based on the integral, we develop a one-step non-parametric method of ES estimation and consider some kernel-based estimators with bias correction.

The layout of the paper is as follows. Section 2 introduces the quantile function-based ES estimator. Section 3 lists a few (existing or new) kernel-based estimators of ES, including the proposed one-step kernel estimator and the existing two-step kernel estimator of Scaillet (2004) and Chen (2008). In Section 4 we compare six kernel-based ES estimators through Monte Carlo experiments. Finally, in Section 5, we present an empirical illustration of the estimators on two financial indexes.

2 EXPECTED SHORTFALL

Let X be the random variable describing the future value of the profit or loss of a portfolio at some fixed time horizon T from today, and let $\alpha \in (0, 1)$ be a probability level. Usually, α stands for the loss probability. The quantile with level α is defined to be:

$$q(\alpha) = \sup\{x \mid P(X \leq x) \leq \alpha\} \tag{2.1}$$

and the VaR is defined by:

$$VaR(\alpha) = -q(\alpha) \tag{2.2}$$

If the truncated mean exists, ie, if $E[X^-] < \infty$ where $X^- = \max(-X, 0)$, then the tail conditional expectation (TCE) and the expected shortfall (ES) are defined, respectively, as:

$$TCE(\alpha) = -E\{X \mid X \leq q(\alpha)\} \tag{2.3}$$

and:

$$ES(\alpha) = -\frac{1}{\alpha}\{E[XI(X \leq q(\alpha))] - q(\alpha)[P(X \leq q(\alpha)) - \alpha]\} \tag{2.4}$$

Delbaen (1998) reported that the TCE does not, in general, satisfy subadditivity. Acerbi *et al* (2001) proved that $ES(\alpha)$ is subadditive but that $TCE(\alpha)$ generally is not. However, if X is a continuous random variable, then:

$$ES(\alpha) = TCE(\alpha) = -\frac{1}{\alpha}E[XI(X \leq q(\alpha))] \tag{2.5}$$

Based on the structure of truncated expectation, Scaillet (2004) presented a nonparametric kernel estimator of the ES associated with a portfolio, and derived the asymptotic properties of the kernel estimator and its first-order derivative with respect to portfolio allocation, in the context of a stationary process satisfying strong mixing conditions. The asymptotic performance and optimal bandwidth of the kernel estimator were considered by Chen (2008). These estimation methods involve two bandwidth selections: one for estimating $q(\alpha)$ and the other for estimating $ES(\alpha)$.

In this paper we shall use the fact that ES can be expressed as an integral of a quantile function via a simple integral transformation:

$$ES(\alpha) = -\frac{1}{\alpha}\int_0^{\alpha} q(p)\,\mathrm{d}p \tag{2.6}$$

See, for example, Pflug and Römisch (2007).

This expression makes studying the analytical properties of ES mathematically tractable. For instance, it is clear from (2.6) that $ES(\alpha)$ is continuous in α, while this is not obvious from (2.4). Moreover, $ES(\alpha)$ can be estimated via an explicit computation as long as the quantile function $q(p)$ is specified; thus the estimation is both simple and fast.

3 NON-PARAMETRIC ESTIMATION OF EXPECTED SHORTFALL

3.1 Kernel estimation

The basic kernel density estimator is given by $\hat{f}_h(x) = (1/n)\sum_{i=1}^{n} K_h(x - x_i)$, where $x_1, x_2, \ldots, x_n \sim f$ is a sample from the distribution with density f. Here $K_h(u) = h^{-1}K(h^{-1}u)$, h is the smoothing parameter, and $K(u)$ is taken to be a symmetric probability density function (see Silverman (1986)).

Let F be the distribution function of f. The basic kernel estimator of F is given by $\hat{F}_h(x) = (1/n)\sum_{i=1}^{n} A_h(x - x_i)$, where $A(x) = \int_{-\infty}^{x} K(u)\,\mathrm{d}u$ is the integrated kernel and $A_h(u) = A(u/h)$.

Let $q(p)$ be a p% quantile of F (where $0 < p < 1$). Then there are three basic kernel quantile function estimators of q_p, given by:

1) distribution function-based estimation from the inverse solution of $F_h(x) = p$;
2) density function-based estimation from the integral solution of $\int_{-\infty}^{q(x)} f_h(x)\,\mathrm{d}x = p$;
3) a kernel-weighted sum of order statistics, $\hat{q}(p) = \sum_{i=1}^{n}[\int_{i-1/n}^{i/n} K_h(t - p)\,\mathrm{d}t]x_{(i)}$, where $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ are the sample order statistics.

The above three kernel quantile estimators have equivalent asymptotic performance, in terms of the following asymptotic mean square error (AMSE) given by Cheng and Sun (2006):

$$\text{AMSE} = \frac{p(1-p)}{n}u^2(p) + \frac{h^4}{4}(u'(p)\sigma_K)^2 - \frac{h}{n}u^2(p)\psi(K) \tag{3.1}$$

where $u = q'$, $u' = q''$, $\sigma_K^2 = \int_{-\infty}^{\infty} t^2 K(t)\,\mathrm{d}t$ and $\psi(K) = 2\int yK(y)A(y)\,\mathrm{d}y$. The finite-sample performance of the three kernel quantile estimators is studied in Section 4.

Let $\hat{q}(p)$ be the kernel estimator of $q(p)$; then the two-step kernel estimation of ES from the basic definition (2.5) is given by $\widetilde{ES}_h(\alpha) = (1/n\alpha)\sum_{i=1}^{n} x_i A_h(\hat{q}(\alpha) - x_i)$ (Scaillet (2004); Chen (2008)), where $\hat{q}(\alpha)$ is the solution of $F_h(x) = \alpha$.

Alternatively, a one-step kernel estimator of ES from (2.6) is given by $\widehat{ES}_h(\alpha) = -(1/\alpha)\int_0^{\alpha} \hat{q}(p)\,\mathrm{d}p$.

Clearly, $\widetilde{ES}_h(\alpha)$ is a sum-type estimator whereas $\widehat{ES}_h(\alpha)$ is an integral-type estimator.

From asymptotic results on kernel estimation (see Parzen (1979)), the bias of all these estimators is $O(h^2)$ as $h \to 0$. In terms of theoretical properties, it is difficult to compare the performance of a one-step kernel estimator with that of a two-step one, but in the appendix we do provide some analytical properties of these estimators, such as their asymptotic bias and mean-square errors. Under certain conditions, these asymptotic properties do recommend $\widehat{ES}_h$ over $\widetilde{ES}_h(\alpha)$. In order to evaluate and compare the finite-sample performance of these estimators, in Section 4 we employ Monte Carlo simulations to calculate the bias of all the proposed estimators for a number of different models.

3.2 Bias reduction of kernel estimators

According to the kernel-based jackknife rule (see Jones and Signorini (1997), among others), if $\hat{\eta}_h$ is the kernel estimator of η_h with bias $O(h^2)$, then as $h \to 0$, $\tilde{\eta} = 2\hat{\eta}_h - \hat{\eta}_{\sqrt{2}h}$ theoretically improves bias from $O(h^2)$ to $O(h^4)$. There is usually data scarcity in the tail of a distribution, especially in the far tail. Consequently, for small p, proper kernel estimation of the quantile $q(p)$ is difficult. The simulation study in Section 4 shows that the bias-reduction technique is particularly effective for the estimation of $ES(\alpha)$ when α is small.

3.3 One- and two-step kernel estimators

For the simulation and empirical studies we will compare the following six ES estimators.

We consider the two-step sum-type kernel estimator and its bias-reduced form, as follows.

$Est1(2S)$: $\widetilde{ES}_h(\alpha) = (1/(n\alpha)) \sum_{i=1}^{n} x_i A_h(\hat{q}(\alpha) - x_i)$, where $\hat{q}(\alpha)$ is a kernel estimator of $q(\alpha)$.

$Est2(2S)_{br(1)}$: the bias-reduced version of $Est1(2S)$, ie, $\widetilde{ES}_r(\alpha) = 2\widetilde{ES}_h(\alpha) - \widetilde{ES}_{\sqrt{2}h}(\alpha)$.

Two-step kernel estimation requires selecting the smoothing parameter h twice. The selection of h is of crucial importance and is well known to be a difficult task, especially when smoothing the tails of underlying distributions with possible data scarcity. Hence, the fewer times h is selected the simpler the estimator.

We also consider the following three one-step kernel estimators.

$Est3(1S)_{dens}$: $\widehat{ES}_f(\alpha) = -(1/\alpha) \int_0^{\alpha} \hat{q}(p)\,\mathrm{d}p$, with $\hat{q}(p)$ being estimated by a kernel density function.

$Est4(1S)_{dist}$: $\widehat{ES}_F(\alpha) = -(1/\alpha) \int_0^{\alpha} \hat{q}(p)\,\mathrm{d}p$, with $\hat{q}(p)$ being estimated by a kernel distribution function.

$Est5(1S)_{order}$: $\widehat{ES}_{Kq}(\alpha) = -(1/\alpha) \int_0^{\alpha} \hat{q}(p)\,\mathrm{d}p$, with $\hat{q}(p)$ being estimated by kernel-weighted order statistics.

Note that there is an explicit expression for $Est5(1S)_{order}$, namely $\widehat{ES}_{Kq,h}(\alpha) \equiv \widehat{ES}_{Kq}(\alpha) = -(1/\alpha) \sum_{i=1}^{n} x_{(i)} \int_{(i-1)/n}^{i/n} \int_0^{\alpha} K_h(t-p)\,\mathrm{d}p\,\mathrm{d}t$.

In addition, we consider its bias-reduced version, as follows:

$$Est6(1S)_{br(5)}\colon \widehat{ES}_{Kqr}(\alpha) = 2\widehat{ES}_{Kq,h}(\alpha) - \widehat{ES}_{Kq,\sqrt{2}h}(\alpha)$$

As discussed in Remark 1 of Appendix A, the two-step kernel estimator almost always underestimates ES, but the one-step kernel estimator may not.

Henceforth, where convenient, we shall refer to the above estimators as estimators 1 to 6.

4 MONTE CARLO STUDY

Throughout this section, we fix the kernel function to be the standard normal density and employ the bandwidth selection rules proposed by Sheather and Jones (1991) and Bowman *et al* (1998) for the estimation of kernel density and distribution, respectively.

A Monte Carlo study was carried out to evaluate the performance of the six estimators of $ES(\alpha)$ using three different models.

4.1 Model 1: normal distribution

Suppose that the return of a financial asset X is normally distributed such that:

$$X \sim N(\mu - \sigma^2/2, \sigma^2) \tag{4.1}$$

Write $a =: \mu - \sigma^2/2$; then we have:

$$E[XI(X \leq q(\alpha))] = a\Phi\left(\frac{q(\alpha) - a}{\sigma}\right) - \frac{\sigma}{\sqrt{2\pi}}\,\mathrm{e}^{-([q(\alpha)-a]^2/2\sigma^2)}$$

Hence, the true value of ES is given by:

$$ES(\alpha) = \frac{\sigma^2}{\alpha}\phi\left(\frac{q(\alpha) - a}{\sigma}\right) - a \tag{4.2}$$

where and $\phi(\cdot)$ and $\Phi(\cdot)$ denote, respectively, the standard normal density and distribution functions.

4.2 Model 2: mixture normal distribution

Assume that X is the return of an asset with a mixture normal distribution:

$$f(x) = \tau f_1(x) + (1 - \tau) f_2(x) \tag{4.3}$$

where $f_1(x)$ and $f_2(x)$ are the density functions of $N(a_1, \sigma_1^2)$ and $N(a_2, \sigma_2^2)$, respectively. This is a non-elliptical distribution. Under (4.3), the true value of ES is:

$$\begin{aligned} ES(\alpha) = {} & \frac{\tau\sigma_1}{\alpha}\phi\left(\frac{q(\alpha) - a_1}{\sigma_1}\right) + \frac{(1-\tau)\sigma_2}{\alpha}\phi\left(\frac{q(\alpha) - a_2}{\sigma_2}\right) \\ & - \tau a_1 - (1-\tau)a_2 \end{aligned}$$

4.3 Model 3: mixture t distribution

Assume that X is the return of an asset with a mixture Student's t distribution:

$$f(x) = \tau f_1(x) + (1 - \tau) f_2(x) \tag{4.4}$$

TABLE 1 Mean bias comparison of the kernel quantile estimators based on cumulative distribution, density and order statistics for a sample size of 100.

	Distribution		Density		Order statistics	
	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$
Model 1	−0.0173	−0.0127	−0.0171	−0.0127	0.0056	0.0092
	(0.2883)	(0.1818)	(0.2881)	(0.1810)	(0.3433)	(0.2077)
Model 2	0.0210	0.0148	0.0210	0.0148	0.0256	0.0186
	(0.0762)	(0.0492)	(0.0762)	(0.1809)	(0.0904)	(0.0554)
Model 3	0.7270	0.4226	0.8316	0.4226	0.7526	0.5242
	(7.6468)	(0.0492)	(9.3140)	(0.1809)	(12.8185)	(2.6595)

Shown in brackets are the standard errors multiplied by 100.

where $f_1(x)$ and $f_2(x)$ are the density functions of Student's t distributions with degrees of freedom three and five, respectively. Under (4.4), the true value of ES is given by:

$$ES(\alpha) = \frac{1}{\alpha}\left(\tau \frac{4}{\sqrt{2}\pi} + (1-\tau)\frac{3}{2\sqrt{2}\pi}\right) q(\alpha)$$

Before proceeding to the calculation of ES, we empirically assess the finite-sample performance of the three kernel quantile estimators described in Section 3.1, which are known to have the same asymptotic mean square error. Table 1 shows the average bias induced by the three estimators for a sample size of 100. From the table it can be observed that there are no significant differences in results obtained from the three estimators. The table further confirms the fact that the accuracy of the kernel quantile estimators diminishes as we attempt to estimate extreme quantiles.

For all three models we now consider the ES estimation for each estimator at loss probability levels 0.01 and 0.05 and sample sizes 100 and 300, where the mean and variance in (4.1) are chosen to be 0.05 and 0.01, respectively. In response to a referee's suggestion that the kernel-based estimators be compared with other commonly used methods of estimating expected shortfall, we include in our comparison the extreme value theory (EVT) approach to estimating ES by fitting a generalized Pareto distribution to the lower tail of the distribution, as explained in McNeil and Frey (2000).

Table 2 (see the next page) reports the bias of the estimators as well as 95% confidence bands of biases, based on 1,000 replications, for a sample size of 100. The results for sample size 300 are presented in Table A.1 (see page 30) in the appendix.

Numerical results show that, in most cases, kernel-based methods tend to underestimate the theoretical ES, which is consistent with the theoretical result stated in Remark 1 of Appendix A. However, the performances of these estimators differ. First, despite similar asymptotic behavior, the four non-bias-reduced estimators (1, 3, 4 and 5) show different bias performances for finite samples of size $n = 100$ and size $n = 300$; see Remark 1 in Appendix A. In particular, at both 1% and 5% levels, the

TABLE 2 Bias of kernel estimation of ES with sample size $n = 100$, where the bias has been multiplied by 1,000.

	Est1(2S)	**Est2(2S)$_{br(1)}$**	**Est3(1S)$_{dens}$**	**Est4(1S)$_{dist}$**	**Est5(1S)$_{order}$**	**Est6(1S)$_{br(5)}$**	**EVT**
$\alpha = 0.01$							
Model 1	−18.0	−15.5	−31.9	−19.8	−9.7	−2.7	−15.7
	(−64.6, 2.7)	(−57.7, 2.4)	(−63.3, −4.2)	(−39.3, −3.0)	(−51.4, 12.8)	(−37.4, 10.6)	(−17.4, −13.5)
Model 2	−8.7	−2.4	−25.0	−10.8	22.3	−6.3	−5.3
	(−29.2, 10.2)	(−21.7, 14.2)	(−41.4, −8.6)	(−17.3, −3.7)	(5.4, 39.1)	(−16.8, 2.0)	(−9.4, 0.8)
Model 3	−2.7	−1.9	−5.0	12.1	−2.1	−1.3	−8.2
	(−12.4, 5.9)	(−8.3, 3.3)	(−10.1, −1.1)	(2.3, 14.4)	(−7.1, −4.2)	(−4.3, 7.8)	(−20.9, 0.30)
$\alpha = 0.05$							
Model 1	−4.9	−1.5	−13.2	−9.3	−1.4	−1.4	−13.4
	(−16.1, 3.7)	(−6.4, 5.4)	(−20.7, −5.3)	(−14.8, −3.5)	(−12.0, 7.2)	(−11.3, 6.9)	(−16.0, −9.2)
Model 2	−34.6	−5.0	−60.1	−16.6	−11.1	1.8	−4.8
	(−131.2, 3.4)	(−75.7, 22.1)	(−123.6, −8.1)	(−30.8, −2.4)	(−52.3, 4.5)	(−83.4, 5.1)	(−8.8, −1.2)
Model 3	−3.1	−1.7	−3.3	−9.4	12.0	−1.1	−7.4
	(−18.5, 1.1)	(−10.4, 3.5)	(−21.3, 1.4)	(−10.7, 0.6)	(−8.9, 22.1)	(−7.6, 8.8)	(−14.1, −3.0)

Shown in brackets are the 95% bootstrapped confidence intervals.

two-step estimator 1 performed better than the one-step non-bias-reduced estimators 3 and 4 but achieved more or less the same results as the other one-step non-bias-reduced estimator, namely estimator 5. Second, by comparing the performance of estimator 2 with that of estimator 1 and the performance of estimator 6 with that of estimator 5, we observe that, in agreement with the asymptotic results, bias reduction does improve estimation accuracy for a finite sample. Additionally, it is worth noting that the biases of estimators 1 to 4 may be partially caused by the kernel estimation of $q(p)$ in the first place, whereas estimators 5 and 6 do not rely on any initial estimation of $q(p)$. Finally, it can be observed that the one-step bias-reduced estimator 6, derived from estimator 5, performs best with the smallest bias and narrower confidence bands.

Table 2 demonstrates empirically that the EVT method, like the kernel-based estimators, underestimates the true ES.

In order to gain a better understanding of the bias as a function of α, we consider 40 equally spaced levels of α ranging from 0.01 to 0.05. Figure 1 (see the next page) displays smoothed bias estimates and their 95% bootstrapped confidence bands for estimators 5 and 6, based on model 1 with $n = 100$. Graphs for estimators 1 to 4 are shown in Figure A.1 of Appendix A (see page 29). The plots of all estimators show a slight increase in bias as α decreases; however, such a gentle increase does not make the estimation accuracy weak. The wider confidence bands for the lower quantiles indicate greater uncertainty in estimating ES, primarily owing to the scarcity of data.

Comparing the biases of the estimators for $n = 100$ and $n = 300$ reveals that bias does not always decrease with large sample size. This observation is further illustrated by Figure 2 (see page 25), which shows the calculated bias for all six estimators under the normal mixture model for $n = 100$ and $n = 300$. The results for the remaining models are roughly the same and are therefore omitted. We also considered the bias-reduced versions of estimators 3 and 4 but came to the conclusion that they perform no better than estimator 6.

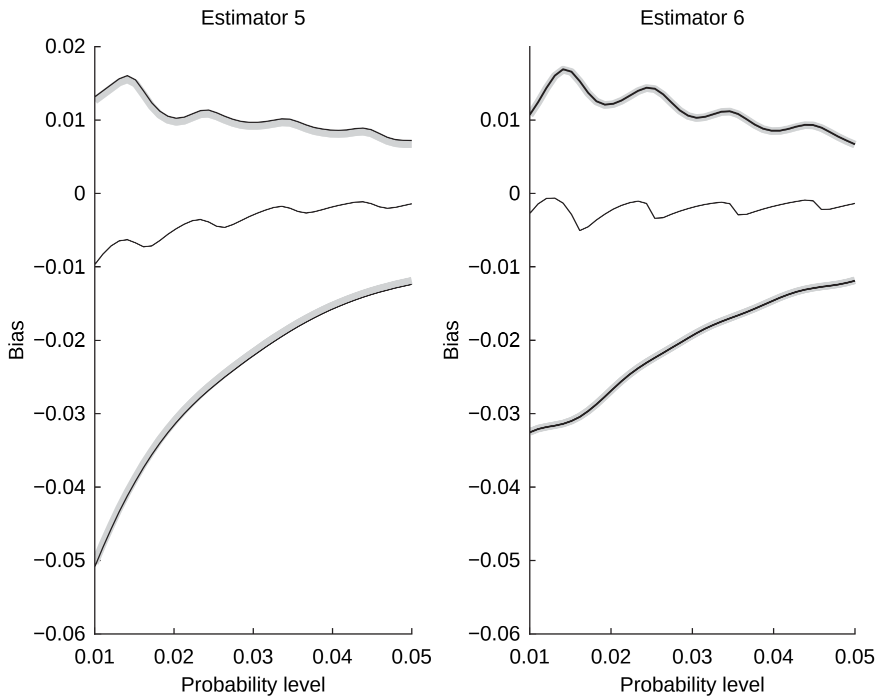
The bias performance of the bias-reduction estimators is significantly better than that of their originators. In particular, estimator 6 (ie, $\widehat{ES}_{Kqr}(\alpha) = 2\widehat{ES}_{Kq,h}(\alpha) - \widehat{ES}_{Kq,\sqrt{2}h}(\alpha)$) generally does best.

5 EMPIRICAL STUDY

In this section, the proposed kernel estimators are applied to estimate the ES of two financial series. These two financial series consist of the daily returns of the Dow Jones and S&P 500 indexes for the period from January 1, 2002 to December 31, 2004, comprising 750 observations. Figure 3 (see page 26) displays the log-returns series for the two indexes.

All six ES estimators investigated in this study are based on non-parametric methods and thus make no distribution assumptions. It is by now well established that one of the characteristics of real financial time series is the presence of heteroscedasticity. In order to account for the effects of changing volatility in forecasting

FIGURE 1 The estimated bias together with 95% confidence bands for $Est5(1S)_{order}$ (left panel) and $Est6(1S)_{br(5)}$ (right panel), based on model 1 with $n = 100$.



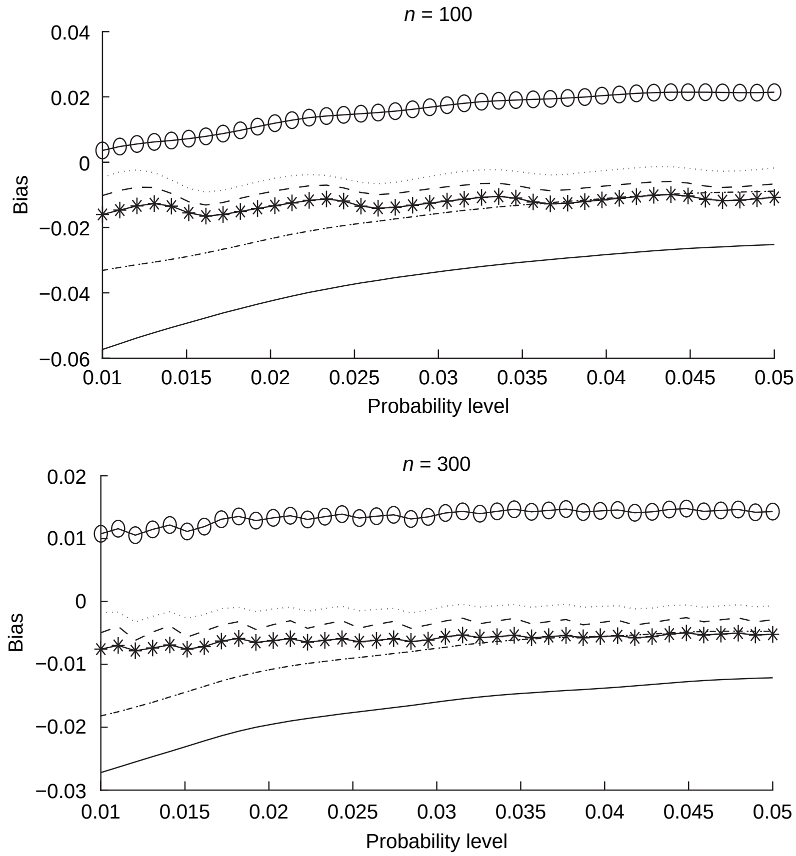
the next period's ES, the six proposed estimators could be applied after the removal of heteroscedasticity through filtering methods such as the one proposed by McNeil and Frey (2000). However, we note that the application of the filtering approach of McNeil and Frey (2000) would render the six estimators semi-parametric, because this method requires the selection of a specific model, such as the generalized autoregressive conditional heteroscedasticity (GARCH) model of Bollerslev (1986), for the conditional mean and volatility dynamics.

Assuming that the conditional mean is zero and that the distribution of the return process is constant, as is commonly the case in many financial applications, we make use of Parkinson's volatility estimator (Parkinson (1980)) in conjunction with the six proposed estimators to forecast the ES one period ahead and thus maintain the non-parametric nature of the estimators.

Let H_t and L_t denote, respectively, the highest and lowest prices on day t; then the Parkinson volatility estimator is defined as:

$$\hat{\sigma}_{P,t} = \frac{\ln(H_t) - \ln(L_t)}{\sqrt{4\ln(2)}} \tag{5.1}$$

FIGURE 2 The estimated bias, based on model 2, for sample sizes $n = 100$ (top panel) and $n = 300$ (bottom panel).

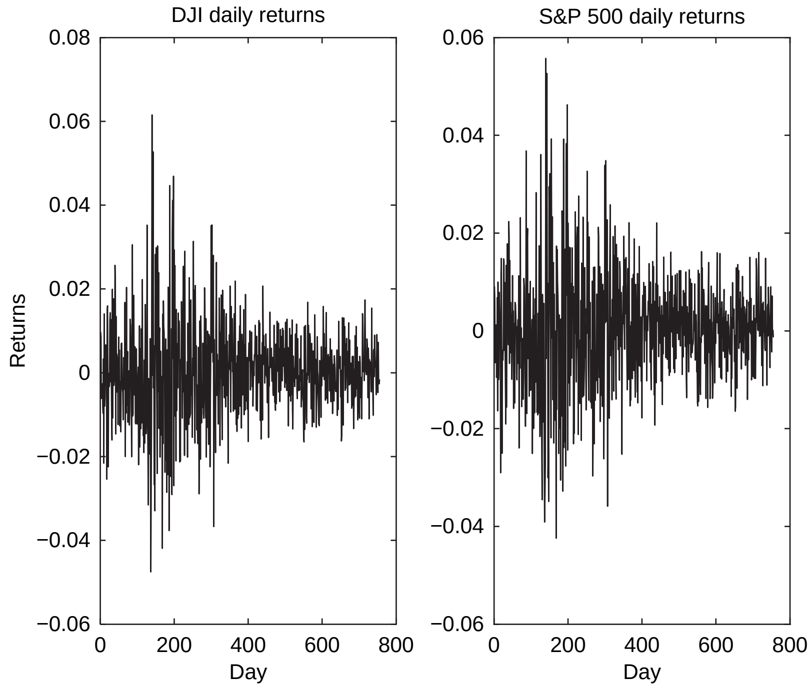


The dash–dotted, dashed, circled, solid, starred and dotted lines represent estimators 1 to 6, respectively.

Given the historical evolution of log-transformed prices of a financial asset, $\{r_t = \ln(P_t) - \ln(P_{t-1})\}_{t=1}^{T}$ where P_t denotes the closing price on day t, together with the highest and lowest prices $\{H\}_{t=1}^{T}$ and $\{L\}_{t=1}^{T}$, our approach, like that of McNeil and Frey (2000), can be split into two stages as follows.

1) Estimate $\{\hat{\sigma}_{P,t}\}_{t=1}^{T}$ and standardize the returns so that $\{z_t\}_{t=1}^{T} = \{r_t\}_{t=1}^{T} / \{\hat{\sigma}_{P,t}\}_{t=1}^{T}$; then use the proposed estimators to estimate the ES of the standardized returns, $ES_z(\alpha)$.
2) To obtain an estimate of the next day's ES, multiply the standardized expected shortfall by $\hat{\sigma}_{P,T}$; that is, take $ES(\alpha) = \hat{\sigma}_{P,T} ES_z(\alpha)$. Implicitly, this model-free approach uses the current volatility as an estimator of the next period's

FIGURE 3 Log-returns time series for the Dow Jones index (DJI) and the S&P 500 index.

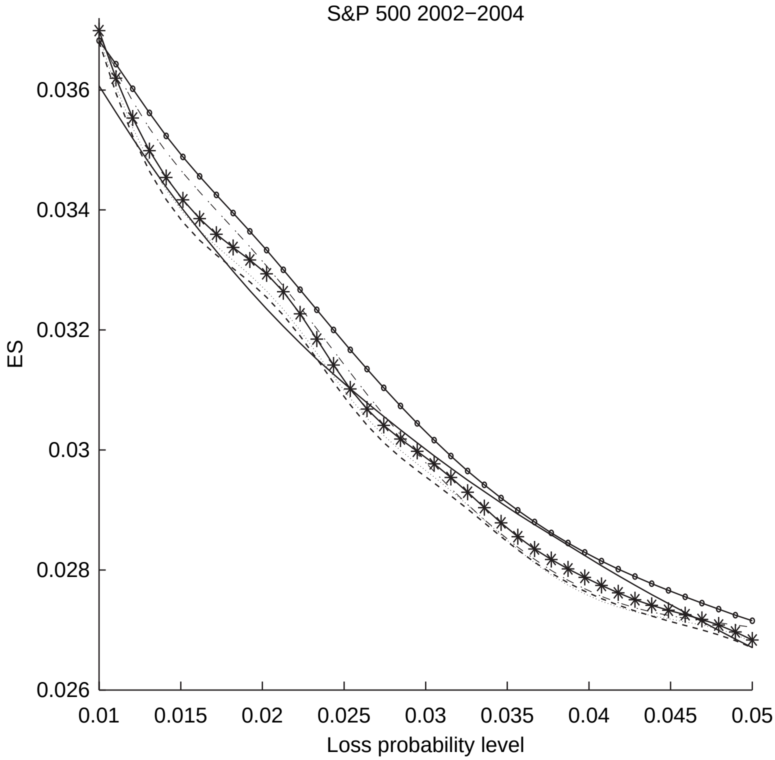


volatility and is different from the method proposed by McNeil and Frey (2000), which by construction uses a model to forecast volatility; furthermore, McNeil and Frey (2000) employed the generalized Pareto distribution to estimate the ES of the standardized residuals.

Before estimating the one-step-ahead ES from the six models, we present in Figure 4 (see the next page) ES estimates for 40 equally spaced loss probability levels ranging from 0.01 to 0.05, using the original unstandardized returns of the S&P 500 index. The estimates for the Dow Jones index exhibit the same pattern and are therefore not included. Careful investigation of the graph reveals that, in comparison to the other four estimators, estimators 2 and 6 produce very close ES estimates.

Table 3 (see the next page) shows one-step-ahead ES estimates for the Dow Jones and S&P 500 indices obtained from the six estimators using the procedure outlined above. In Table 3 we also present ES estimates obtained from the Gaussian and Student's t distributions, using the same procedure of estimating volatility. From the table it can be observed that the kernel-based and parametric ES estimators produce quite different results.

FIGURE 4 ES estimates produced by the six estimators.



The dash–dotted, dashed, solid–circled, solid, solid–starred and dotted lines represent estimators 1 to 6, respectively.

TABLE 3 One-period-ahead ES estimates for loss probability levels $\alpha = 0.01$ and $\alpha = 0.05$ obtained from kernel-based and parametric estimators, where the original value has been multiplied by 100.

	DJI		**S&P 500**	
	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$
*Est*1(2S)	0.7262	0.4177	0.9842	0.4477
*Est*2(2S)$_{\mathrm{br}(1)}$	0.5427	0.4087	0.7146	0.4258
*Est*3(1S)$_{\mathrm{dens}}$	0.6396	0.5091	0.6784	0.5780
*Est*4(1S)$_{\mathrm{dist}}$	0.5638	0.4710	0.5937	0.5345
*Est*5(1S)$_{\mathrm{order}}$	0.5581	0.4805	0.8151	0.5952
*Est*6(1S)$_{\mathrm{br}(5)}$	0.5317	0.3854	0.7091	0.5654
Normal	0.4912	0.4270	0.6737	0.5794
Student's $t_{(5)}$	0.5305	0.4487	0.7078	0.5456

6 CONCLUSIONS

Kernel smoothing is a useful non-parametric method for estimating ES. In this paper we have proposed one-step kernel estimators of ES, and through Monte Carlo simulation we have demonstrated the appealing numerical performance of these estimators in comparison to existing two-step kernel estimators. Our estimators are kernel quantile-based ones. Several of the estimators proposed in this paper, such as estimators 1, 3, 4 and 5, are shown to be fast, efficient and valid for the estimation of ES. Estimator 5 can even be computed explicitly. The failure of an estimator of ES is often due not to substandard methodology but rather to the inaccurate and difficult estimation in the tail of a distribution. It was noted that application of a bias-reduction technique leads to estimators (namely, estimators 2 and 6) that show improved accuracy relative to their non-bias-reduced versions (estimators 1 and 5). In particular, estimator 6, a computationally efficient new bias-reduction estimator with an explicit expression, is demonstrated to be unrivaled in terms of reducing ES estimation bias.

APPENDIX A ANALYTIC PROPERTIES OF $\widetilde{ES}_h(\alpha)$ AND $\widehat{ES}_h(\alpha)$

We use the following basic assumptions from Chen and Tang (2005).

A1) The process $\{X_i : 1 \le i \le n\}$ is strictly stationary and α-mixing, and there exists a $\rho \in (0, 1)$ such that the mixing coefficient satisfies $\alpha(k) \le C\rho^k$ for all $k \ge 1$. The random variable X_1 is continuously distributed, with f and F as its density and distribution functions, respectively.

A2) We have $f(q(p)) > 0$ whenever $p \in (0, 1)$ and f has continuous second derivative in a neighborhood $B(q(p))$ of $q(p)$. The second partial derivatives of F_k, which form the joint distribution function of (X_1, X_{k+1}) for $k \ge 1$, are all bounded in $B(q(p))$ uniformly with respect to k.

A3) The kernel function $G(t) = \int_{\infty}^{t} K(u)\,\mathrm{d}u$, where K is a univariate probability density function, has continuous bounded second derivative and satisfies the following moment conditions:

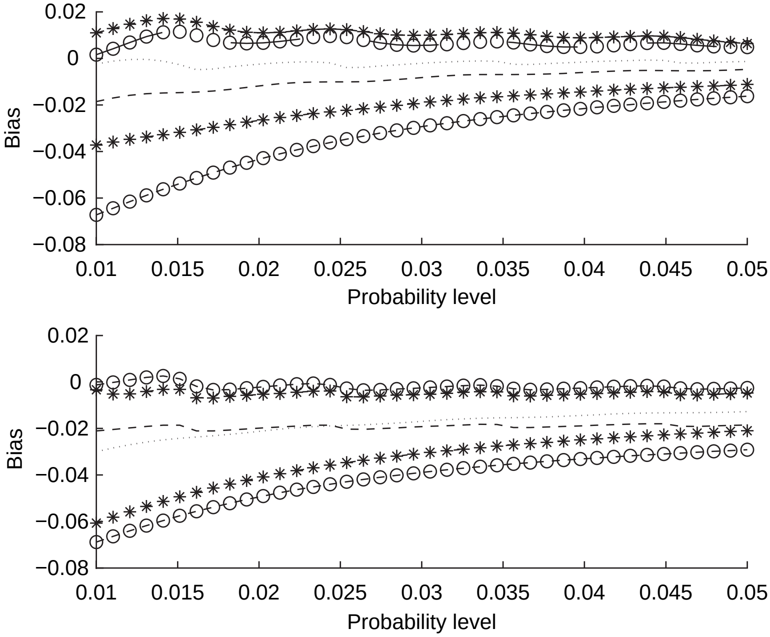
$$\int_{-\infty}^{\infty} uK(u)\,\mathrm{d}u = 0 \quad \text{and} \quad \int_{-\infty}^{\infty} u^2 K(u)\,\mathrm{d}u = \sigma_K^2 < \infty$$

A4) The smoothing bandwidth h satisfies $h \to 0$, $nh^{3-\tau} \to \infty$ for any $\tau > 0$ and $nh^4 \log^2(n) \to 0$ as $n \to \infty$.

Under the assumptions A1–A4, Chen and Tang (2005) gave the mean square error of $\hat{q}(p)$ as follows:

$$\begin{aligned} MSE(\hat{q}(p)) &= n^{-1}\sigma^2(p; n) f^{-2}(q(p)) - 2n^{-1} h b_K f^{-1}(q(p)) \\ &\quad + \frac{1}{4} h^4 \sigma_K^4 \{f'(q(p)) f^{-1}(q(p))\}^2 + o\left(\frac{h}{n} + h^4\right) \end{aligned} \tag{A.1}$$

FIGURE A.1 The estimated bias of estimators 1 to 4, based on model 1 with $n = 100$ and α taking equally spaced values between 0.01 and 0.05.



The top panel displays the estimated bias of $Est1(2S)$ (dashed line) and $Est2(2S)_{br(1)}$ (dotted line); the bottom panel shows the estimated bias of $Est3(1S)_{dens}$ (dotted line) and $Est4(1S)_{dist}$ (dashed line). In both panels, the 95% confidence bands for the dotted and dashed lines are given, respectively, by starred and circled lines.

where $b_K = \int_{-\infty}^{\infty} uK(u)G(u)\,\mathrm{d}u$, $\sigma^2(p; n) = \{p(1-p) + 2\sum_{k=1}^{n-1}(1-k/n)\gamma(k)\}$ and $\gamma(k) = \operatorname{cov}\{I(X_1 < q(p)), I(X_{k+1} < q(p))\}$ for positive integers k. Clearly, $\sigma^2(p; n) = p(1-p)$ for an independent process.

From (A.1), it is easy to get an upper bound for the mean square error of the one-step estimator $\widehat{ES}_h(\alpha)$. Indeed, we have:

$$\begin{aligned} MSE(\widehat{ES}_h(\alpha)) \leq{} & \frac{2}{n\alpha^2}\left(\int_\beta^\alpha \sigma^2(p; n) f^{-2}(q(p))\,\mathrm{d}p\right) \\ & + \frac{h^4\sigma_K^4}{4\alpha^2}\left[\int_\beta^\alpha f'(q(p)) f^{-1}(q(p))\,\mathrm{d}p\right]^2 \\ & - \frac{4hb_K}{n\alpha^2}\left(\int_\beta^\alpha f^{-1}(q(p))\,\mathrm{d}p\right) + o(h/n + h^4) \qquad \text{(A.2)} \end{aligned}$$

Owing to the mathematical complexity, it is difficult to compare the two types of estimators by their mean square errors. However, we may be able to compare their

TABLE A.1 Bias of the kernel estimation of ES with sample size $n = 300$, where the bias has been multiplied by 1,000.

	Est1(2S)	**Est2(2S)$_{\mathbf{br(1)}}$**	**Est3(1S)$_{\mathbf{dens}}$**	**Est4(1S)$_{\mathbf{dist}}$**	**Est5(1S)$_{\mathbf{order}}$**	**Est6(1S)$_{\mathbf{br(5)}}$**	**EVT**
$\alpha = 0.01$							
Model 1	−5.1	−4.9	−8.1	−2.2	6.3	−1.3	−16.3
	(−19.2, 2.6)	(−10.1, 6.7)	(−16.3, −2.1)	(−3.8, −1.2)	(−8.5, 13.6)	(−7.3, 2.1)	(−37.3, −5.3)
Model 2	−17.7	−18.0	−27.7	−7.4	21.2	−4.8	−4.27
	(−67.7, 13.2)	(−62.9, 15.3)	(−52.5, −7.1)	(−12.6, −3.4)	(−16.5, 51.0)	(−25.2, 7.1)	(−10.4, 7.2)
Model 3	−149.4	−144.5	−260.4	−83.5	124.5	−45.4	−126.4
	(−912.2, 315.0)	(−942.6, 338.1)	(−696.8, 29.3)	(−182.7, −5.3)	(26.6, 217.1)	(−360.4, 119.8)	(−136.8, −117.2)
$\alpha = 0.05$							
Model 1	−1.4	−2.4	−3.5	−1.5	7.5	−0.9	−11.3
	(−4.7, 2.2)	(−3.1, 1.8)	(−5.2, −2.4)	(−1.9, −0.7)	(5.2, 10.4)	(−1.5, 1.3)	(−30.3, −0.34)
Model 2	−4.4	−4.4	−12.1	−5.2	2.6	−2.9	−2.7
	(−16.3, 5.8)	(−17.2, 6.0)	(−18.5, −6.4)	(−8.2, −2.9)	(17.0, 35.8)	(−8.1, 2.2)	(−17.7, 6.4)
Model 3	−70.8	−65.4	−181.2	−77.2	71.1	−41.4	−22.5
	(−304.2, 151.9)	(−290.1, 153.0)	(−307.6, −0.67)	(−122.2, −40.4)	(48.1, 96.8)	(−141.8, 44.5)	(−68.6, 21.3)

Shown in brackets are the 95% bootstrapped confidence intervals.

biases. In fact, from Chen and Tang (2005) and Chen (2008), we have:

$$\mathrm{Bias}(\hat{q}(p)) = -\tfrac{1}{2}h^2\sigma_K^2 f'(q(p))f^{-1}(q(p)) + o(h^2) \tag{A.3}$$

and:

$$\mathrm{Bias}(\widetilde{ES}_h(\alpha)) = -\frac{1}{2\alpha}h^2\sigma_K^2 f(q(\alpha)) + o(h^2) \tag{A.4}$$

Using the fact that $q'(p) = f^{-1}(q(p))$ and Equation (A.3), we get:

$$\mathrm{Bias}(\widehat{ES}_h(\alpha)) = \frac{1}{2\alpha}h^2\sigma_K^2[f(q(\alpha)) - \lim_{\beta\to 0} f(q(\beta))] + o(h^2) \tag{A.5}$$

REMARK 1 Asymptotically, Equation (A.4) shows that the two-step kernel estimator always underestimates ES; however, the integral-type estimator may or may not underestimate ES, depending on the sign of the factor $1 - \lim_{\beta\to 0} f(q(\beta))/f(q(\alpha))$ in Equation (A.5).

REMARK 2 The condition $|1 - \lim_{\beta\to 0} f(q(\beta))/f(q(\alpha))| \leq 1$, which implies that the one-step estimator $\widehat{ES}_h(\alpha)$ has smaller bias than the two-step estimator $\widetilde{ES}_h(\alpha)$, holds for some (but not all) distributions. In fact, we make the following observations.

1) Because $f(q(\beta))$ tends to the left end of the support of $f(x)$ when β tends to zero, and $f(x)$ is usually increasing in the left tail, we have $f(q(\beta)) > 0$ and $1 - f(q(\beta))/f(q(\alpha)) < 1$ provided that $f'(q(\beta)) > 0$.
2) It can be seen that $f(q(\beta))/f(q(\alpha)) \geq 0$ when β is very small. For example, under the normal distribution with $f(x) = 1/(\sqrt{2\pi}\sigma)\exp(-((x-a)^2/2\sigma^2))$, we have $0 < f(q(\beta))/f(q(\alpha)) < 1$.

REFERENCES

Acerbi, C., Nordio, C., and Sirtori, C. (2001). Expected shortfall as a tool for financial risk management. Working Paper, GloriaMundi.org, February.

Acerbi, C., and Tasche, D. (2002a). On the coherence of expected shortfall. *Journal of Banking and Finance* **26**, 1487–1503.

Acerbi, C., and Tasche, D. (2002b). Expected shortfall: a natural coherent alternative to value at risk. *Economic Notes* **31**(2), 379–388.

Artzner, P. (1999). Application of coherent risk measures to capital requirements in insurance. *North American Actuarial Journal* **3**(2), 11–25.

Artzner, P., Delbaen, F., Eber, J. M., and Heath, D. (1997). Thinking coherently. *Risk* **10**, 68–71.

Artzner, P., Delbaen, F., Eber, J. M., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance* **9**(3), 203–228.

Artzner, P., Delbaen, F., Eber, J. M., Heath, D., and Ku, H. (2002). Coherent multiperiod risk measurement. Working Paper, ETH Zurich.

Basel Committee on Banking Supervision (1996). Amendment to the Capital Accord to incorporate Market Risks, Bank for International Settlements. URL: http://www.bis.org/publ/bcbs24.htm.

Basel Committee on Banking Supervision (2003). The New Basel Capital Accord, Bank for International Settlements. URL: http://www.bis.org/bcbs/cp3ov.pdf.

Bollerslev, T. (1986). Generalised autoregressive conditional heteroscedasticity. *Journal of Econometrics* **31**(3), 307–327.

Bowman, A., Hall, P., and Prvan, T. (1998). Bandwidth selection for the smoothing of distribution functions. *Biometrika* **85**, 799–808.

Chen, S. X. (2008). Nonparametric estimation of expected shortfall. *Journal of Financial Econometrics* **6**(1), 87–107.

Chen, S. X., and Tang, C. Y. (2005). Nonparametric inference of value at risk for dependent financial returns. *J. Financial Econometrics* **3**(2), 227–255.

Cheng, M.-Y., and Sun, S. (2006). Bandwidth selection for kernel quantile estimation. *Journal of the Chinese Statistical Association* **44**, 271–295.

Danielsson, J., Jorgensen, B. N., Sarma, M., and Vries, C. G. (2005). Sub-additivity re-examined: the case for value-at-risk. Working Paper, Eurandom.

Delbaen, F. (1998). Coherent risk measures on general probability spaces. Working Paper, ETH Zurich. URL: http://www.math.ethz.ch/∼delbaen/ftp/preprints/RiskMeasuresGeneralSpaces.pdf.

Fermanian, J.-D., and Scaillet, O. (2005). Sensitivity analysis of VaR and expected shortfall for portfolios under netting agreements. *Journal of Banking and Finance* **29**, 927–958.

Jones, M. C., and Signorini, D. F. (1997). A comparison of higher-order bias kernel density estimators. *Journal of the American Statistical Association* **92**, 1063–1073.

McNeil, A., 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.

Parkinson, M. (1980). The extreme value method for estimating the variance of the rate of return. *Journal of Business* **53**, 61–68.

Parzen, E. (1979). Nonparametric statistical data modeling. *Journal of the American Statistical Association* **74**, 105–121.

Pflug, G. C., and Römisch, W. (2007), *Modeling, Measuring and Managing Risk*. World Scientific, Singapore.

Scaillet, O. (2004). Nonparametric estimation and sensitivity analysis of expected shortfall. *Mathematical Finance* **14**(1), 115–129.

Scaillet, O. (2005). Nonparametric estimation of conditional expected shortfall. *Revue Assurances et Gestion des Risques (Insurance and Risk Management Journal)* **74**, 639–660.

Sheather, S. J., and Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society*, Series B **53**, 683–690.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*, Chapman & Hall, New York.