On estimating the conditional expected shortfall

Franco Peracchi[1,2,*,†] and Andrei V. Tanase[1]

[1]*Tor Vergata University*, *Rome*, *Italy*
[2]*EIEF*, *Rome*, *Italy*

SUMMARY

Unlike the value at risk, the expected shortfall is a coherent measure of risk. In this paper, we discuss estimation of the expected shortfall of a random variable Y_t with special reference to the case when auxiliary information is available in the form of a set of predictors X_t. We consider three classes of estimators of the conditional expected shortfall of Y_t given X_t: a class of fully non-parametric estimators and two classes of analog estimators based, respectively, on the empirical conditional quantile function and the empirical conditional distribution function. We study their sampling properties by means of a set of Monte Carlo experiments and analyze their performance in an empirical application to financial data. Copyright © 2008 John Wiley & Sons, Ltd.

1. INTRODUCTION

A measure of risk is coherent if it simultaneously satisfies the following properties: sub-additivity, monotonicity, positive homogeneity, and translation invariance (see Artzner *et al.* [1] and Delbaen [2]). As described in Acerbi and Tasche [3], sub-additivity may be violated by the value at risk (VaR). Although popular in financial applications, because it gives a lower bound on the loss made in the worst α percent of the cases during a prespecified period, the VaR is therefore not a coherent measure of risk. In addition, the practical usefulness of the VaR is limited by the fact that it tells us nothing about the potential size of the loss in the worst-case scenario.

Even before the introduction of the VaR, the expected value of the left tail of the returns on a risky asset has been proposed as an alternative measure of risk. This quantity, variously known as the expected shortfall or the tail conditional expectation or the tail conditional mean, measures the loss that one may expect to make in the worst α percent of the cases. Formally, let

---

*Correspondence to: Franco Peracchi, Faculty of Economics, University of Rome 'Tor Vergata', via Columbia 2, 00133 Rome, Italy.
†E-mail: franco.peracchi@uniroma2.it

the random variable (rv) Y_t represent the returns on a given asset during a prespecified period. Then the VaR at level α is equal to the αth quantile of Y_t, $Q(\alpha)$, and the α-level expected shortfall is $\tau(\alpha) = E(Y_t \mid Y_t \leqslant Q(\alpha))$, where $E(Y_t \mid A_t)$ denotes the conditional expectation of Y_t given the event A_t. Unlike the VaR, the expected shortfall takes into account all possible losses that exceed the severity level corresponding to the VaR. Besides satisfying all the properties of a coherent measure of risk (see Acerbi and Tasche [3]), the expected shortfall is continuous with respect to α regardless of the underlying distribution of Y_t and therefore is not too sensitive to small changes in α.

In this paper, we consider the important question of how to estimate the expected shortfall when auxiliary information about asset returns Y_t is provided by a set of predictors. We therefore depart from Bassett *et al.* [4] and Koenker [5, pp. 289–292], who are concerned with using the expected shortfall for asset allocation problems. We focus on the simple case when asset returns may be represented by a continuous rv Y_t with a continuous and strictly increasing distribution function (df) conditional on a set of predictors X_t, which represents the information available at time $t-1$. If this conditional df (cdf) is known, at least up to a finite-dimensional parameter, then one may in some cases compute the expected shortfall analytically. For example, this is the case when the conditional distribution of Y_t given X_t is Gaussian. Notwithstanding the apparent simplicity of the problem, in practice we face two issues. First, even when the cdf of Y_t has a known shape, the integral that defines the expected shortfall may be hard to compute. Second, we might not even know the shape of the cdf of Y_t.

In this paper we introduce two classes of analog estimators based on two alternative representations of the conditional expected shortfall, either as an integral of the conditional quantile function (cqf) of Y_t given X_t or as an integral of the cdf of Y_t given X_t. Both representations open the way to estimation of the expected shortfall by replacing the population cqf and the population cdf by suitable estimates. Estimators based on the first representation are easily interpretable, as the passage from the unconditional to the conditional case is very intuitive, and are particularly simple to obtain when the cqf is assumed to be linear in parameters. Unfortunately, estimated linear conditional quantiles may cross each other, especially when evaluated in the tails of the distribution of X_t. As for the second representation, the approach in Peracchi [6] makes it possible to naturally impose monotonicity of the estimated cdf and therefore avoid the quantile crossing problem that might arise with the first representation. Although quite straightforward, these two classes of estimators appear to be novel.

The remainder of the paper is organized as follows. Section 2 presents the formal definitions of expected shortfall and conditional expected shortfall and illustrates their computation in a few examples. Section 3 shows how the expected shortfall and the conditional expected shortfall may be estimated. Section 4 presents some Monte Carlo evidence. Section 5 presents an empirical illustration. Finally, Section 6 summarizes and concludes.

2. THE EXPECTED SHORTFALL

Let Y_t be a real-valued rv with a continuous and strictly increasing df $F(y) = \Pr\{Y_t \leqslant y\}$. Its quantile function (qf) is defined as $Q(p) = \inf\{y : F(y) \geqslant p\}$, with $p \in (0, 1)$. Since F is continuous and strictly increasing, Q is also continuous and strictly increasing. Further, $F(Q(p)) = p$ and $Q(F(y)) = y$, and so $Q(p) = F^{-1}(p)$ and $F(y) = Q^{-1}(y)$.

If Y_t has a finite mean, then the mean of Y_t conditional on $Y_t \leqslant c$, where c is any real number, is defined as

$$E(Y_t \mid Y_t \leqslant c) = \frac{1}{F(c)} \int_{-\infty}^{c} y \, dF(y) \tag{1}$$

The α-level expected shortfall of Y_t, with $0 < \alpha < 1$, is therefore

$$\tau(\alpha) = E(Y_t \mid Y_t \leqslant Q(\alpha)) = \frac{1}{\alpha} \int_{-\infty}^{Q(\alpha)} y \, dF(y) \tag{2}$$

where $Q(\alpha)$ is the αth quantile of Y_t. In financial applications, Y_t is the return on a given asset and $\tau(\alpha)$ gives the expected value of a loss (negative return) that exceeds $Q(\alpha)$, the VaR at level α.

Acerbi and Tasche [3] establish the main properties of the expected shortfall. If $\tau_X(\alpha)$ denotes the α-level expected shortfall of a real-valued rv X with continuous and strictly increasing df, they show that (i) $\tau_X(\alpha) \leqslant 0$, if $X \leqslant 0$ (monotonicity), (ii) $\tau_{X+Y}(\alpha) \geqslant \tau_X(\alpha) + \tau_Y(\alpha)$ (sub-additivity), (iii) $\tau_{bX} = b \tau_X(\alpha)$ for $b \geqslant 0$ (positive homogeneity), and (iv) $\tau_{X+c}(\alpha) = \tau_X(\alpha) + c$ for $c \in \mathbb{R}$ (translation invariance). Thus, the α-level expected shortfall is a coherent risk measure.

2.1. Alternative representations

Since F is continuous and strictly increasing, a change of variable from $F(y)$ to p gives

$$E(Y_t \mid Y_t \leqslant c) = \frac{1}{F(c)} \int_{F(-\infty)}^{F(c)} F^{-1}(p) \, dp = \frac{1}{F(c)} \int_0^{F(c)} Q(p) \, dp$$

Thus we have the equivalent representation

$$\tau(\alpha) = \frac{1}{\alpha} \int_0^{\alpha} Q(p) \, dp \tag{3}$$

This representation is particularly convenient when the quantiles of an rv Y_t have a closed-form expression. For example, suppose that Y_t may be represented as $Y_t = \mu + \sigma U_t$ for some $\mu \in \Re$ and $\sigma > 0$, where U_t is an rv with continuous and strictly increasing df G. Because in this case $F(y) = G((y - \mu)/\sigma)$, it follows immediately that $Q(p) = \mu + \sigma \zeta(p)$, where $\zeta(p) = G^{-1}(p)$ is the pth quantile of U_t. Therefore,

$$\tau(\alpha) = \frac{1}{\alpha} \int_0^{\alpha} [\mu + \sigma \zeta(p)] \, dp = \mu + \sigma \tau^*(\alpha)$$

where $\tau^*(\alpha) = \alpha^{-1} \int_0^{\alpha} \zeta(p) \, dp$ is the α-level expected shortfall of U_t.

When the equation $F(y) = p$ does not have a closed-form solution, the existence and uniqueness theorem for first-order ordinary differential equations (see Hirsch and Smale [7]) guarantees that the solution exists and is unique provided that F is continuous and strictly increasing. In these cases, computation of $\tau(\alpha)$ must typically be carried out by numerical methods.

Another equivalent representation of $\tau(\alpha)$ is in terms of the df of Y_t. Under regularity conditions, integrating (1) by parts, we get

$$E(Y_t \mid Y_t \leqslant c) = c - \int_{-\infty}^{c} \frac{F(y)}{F(c)} \, dy$$

Therefore,

$$\tau(\alpha) = Q(\alpha) - \frac{1}{\alpha} \int_{-\infty}^{Q(\alpha)} F(y)\,\mathrm{d}y \qquad (4)$$

This shows that the expected shortfall $\tau(\alpha)$ is larger, in absolute value, than the VaR $Q(\alpha)$.

2.2. Relationship to other concepts

The expected shortfall is closely related to other concepts, such as the mean excess function and the Lorenz curve.

The mean excess function (or mean residual life function) is the mean excess over a threshold c, that is

$$e(c) = E(Y_t - c \mid Y_t \leqslant c) = E(Y_t \mid Y_t \leqslant c) - c$$

This quantity is an important tool in financial risk management and in various other fields, such as medicine (see Embrechts *et al.* [8, pp. 294–303]). Evaluating the mean excess function at $c = Q(\alpha)$ gives

$$e(Q(\alpha)) = \tau(\alpha) - Q(\alpha)$$

which is just the difference between the expected shortfall and the VaR.

The Lorenz curve is commonly used in economics to describe the distribution of income and is associated with measures of inequality such as the Gini coefficient. In this case, Y_t is typically taken to be a non-negative rv with finite, non-zero mean μ. The Lorenz curve is defined as

$$L(\alpha) = \frac{1}{\mu} \int_0^\alpha Q(p)\,\mathrm{d}p, \quad 0 < \alpha < 1$$

and so, from (3),

$$L(\alpha) = \frac{\alpha}{\mu}\tau(\alpha)$$

The generalized Lorenz curve [9] is the Lorenz curve scaled up by the mean and is equal to

$$GL(\alpha) = \int_0^\alpha Q(p)\,\mathrm{d}p = \alpha\tau(\alpha), \quad 0 < \alpha < 1$$

If the non-negative rv Y_t represents individual income, then $GL(\alpha)$ simply cumulates individual incomes up to the αth quantile.

2.3. Examples

We illustrate computation of the expected shortfall for two leading examples, the normal distribution and a finite mixture of normals.

If $Y_t \sim \mathcal{N}(\mu, \sigma^2)$, then we can write $Y_t = \mu + \sigma U_t$, where $U_t \sim \mathcal{N}(0, 1)$. By standard results

$$E(Y_t \mid Y_t \leqslant c) = \mu + \sigma E(U_t \mid U_t \leqslant c^*) = \mu - \sigma\frac{\phi(c^*)}{\Phi(c^*)}$$

where $c^* = (c - \mu)/\sigma$, and ϕ and Φ, respectively, denote the density and the df of a standard normal. If we set c equal to $Q(\alpha) = \mu + \sigma \Phi^{-1}(\alpha)$, then $c^* = (Q(\alpha) - \mu)/\sigma = \Phi^{-1}(\alpha)$ and we obtain

$$\tau(\alpha) = \mu - \frac{\sigma}{\alpha} \phi(\Phi^{-1}(\alpha)) \tag{5}$$

Since $Q(p) = \mu + \sigma \Phi^{-1}(p)$, we equivalently have

$$\tau(\alpha) = \mu + \frac{\sigma}{\alpha} \int_0^\alpha \Phi^{-1}(p) \, \mathrm{d}p$$

Further, from (4), we also have

$$\tau(\alpha) = \mu + \sigma \Phi^{-1}(\alpha) - \frac{1}{\alpha} \int_{-\infty}^{\mu + \sigma \Phi^{-1}(\alpha)} \Phi\left(\frac{y - \mu}{\sigma}\right) \mathrm{d}y$$

Our second example is a finite mixture of normals. This is an important example because any continuous distribution may be approximated arbitrarily well by a mixture of J normal distributions (see McLachlan and Peel [10]). Thus, a finite mixture of normals provides a flexible and tractable way of allowing for asymmetry, skewness and heavy tails. For simplicity, we consider the case when $J = 2$, that is, the rv Y_t has a distribution that is a mixture of a $\mathcal{N}(\mu_1, \sigma_1^2)$ and a $\mathcal{N}(\mu_2, \sigma_2^2)$ distribution, with mixing probabilities $\pi_1 = \pi$ and $\pi_2 = 1 - \pi$, respectively. In this case, the df of $\bar{Y}_t$ is equal to

$$F(y) = \pi \Phi\left(\frac{y - \mu_1}{\sigma_1}\right) + (1 - \pi) \Phi\left(\frac{y - \mu_2}{\sigma_2}\right)$$

Although we do not have a closed-form expression for the quantiles of a normal mixture, they can easily be evaluated numerically.

Let $c_j^* = (c - \mu_j)/\sigma_j$ and $F_j(c) = \Phi(c_j^*)$, for $j = 1, 2$, and define $F(c) = \pi F_1(c) + (1 - \pi) F_2(c)$ and $\theta(c) = \pi F_1(c)/F(c)$. After some algebra, Equation (1) becomes

$$E(Y_t \,|\, Y_t \leqslant c) = \theta(c) \mu_1(c) + [1 - \theta(c)] \mu_2(c)$$

where $\mu_j(c) = \mu_j - \sigma_j \phi(c_j^*)/\Phi(c_j^*)$, $j = 1, 2$. It follows that

$$\tau(\alpha) = \theta(\alpha) \tau_1(\alpha) + [1 - \theta(\alpha)] \tau_2(\alpha)$$

where $\tau_j(\alpha) = \mu_j - \sigma_j \phi(c_j^*)/\Phi(c_j^*)$, $c_j^* = [Q(\alpha) - \mu_j]/\sigma_j$, with $j = 1, 2$, and $\theta(\alpha) = \pi \Phi(c_1^*)/\alpha$. Thus, the α-level expected shortfall of Y_t is a convex combination of the expected shortfalls of the two normal components of the mixture. This result is easily generalized to the case when Y_t is a mixture of $J \geqslant 2$ normals with mixing probabilities $\pi_1, \ldots, \pi_J$ that are positive and add up to one.

2.4. Conditional expected shortfall

The previous results are easily generalized to the case when additional information about Y_t is provided by a set of predictors X_t, which represents the information available at time $t - 1$ and may contain a finite number of lags of Y_t.

Let $F(y \,|\, x) = \Pr\{Y_t \leqslant y \,|\, X_t = x\}$ and $Q(p \,|\, x) = \inf\{y : F(y \,|\, x) \geqslant p\}$ be, respectively, the cdf and the cqf of Y_t given $X_t = x$. Assuming that $F(\cdot \,|\, x)$ is continuous and strictly increasing for all x,

we have that $F(Q(p|x)|x) = p$ and so $Q(p|x) = F^{-1}(p|x)$ for all x. The conditional mean of Y_t given $Y_t \leqslant c$ and $X_t = x$ is defined as

$$E(Y_t | Y_t \leqslant c, X_t = x) = \frac{1}{F(c|x)} \int_{-\infty}^{c} y \, dF(y|x)$$

Since $F(\cdot|x)$ is continuous and strictly increasing for all x, we have the equivalent representations

$$E(Y_t | Y_t \leqslant c, X_t = x) = \frac{1}{F(c|x)} \int_{0}^{F(c|x)} Q(p|x) \, dp$$

$$= c - \int_{-\infty}^{c} \frac{F(y|x)}{F(c|x)} \, dy$$

The α-level conditional expected shortfall of Y_t given $X_t = x$ is obtained by setting $c = Q(\alpha|x)$. Thus,

$$\tau(\alpha|x) = \frac{1}{\alpha} \int_{-\infty}^{Q(\alpha|x)} y \, dF(y|x)$$

$$= \frac{1}{\alpha} \int_{0}^{\alpha} Q(p|x) \, dp$$

$$= Q(\alpha|x) - \frac{1}{\alpha} \int_{-\infty}^{Q(\alpha|x)} F(y|x) \, dy \qquad (6)$$

For example, if the conditional distribution of Y_t given $X_t = x$ is normal, with mean $\mu(x)$ and variance $\sigma(x)^2$, then, from (5),

$$\tau(\alpha|x) = \mu(x) - \frac{\sigma(x)}{\alpha} \phi(\Phi^{-1}(\alpha))$$

3. ESTIMATION

We now consider alternative approaches to estimation of the α-level expected shortfall $\tau(\alpha)$ and the α-level conditional expected shortfall $\tau(\alpha|x)$. The basic idea is to exploit the equivalence between (2), (3) and (4) in the unconditional case, and the equivalence (6) in the conditional case.

3.1. Estimation of the expected shortfall

Given a random sample $Y_1, \ldots, Y_T$ from a distribution with df F and qf Q, the expected shortfall $\tau(\alpha)$ may simply be estimated by replacing F and Q in (2) and (4) by their empirical counterparts, namely the empirical df (edf) $\hat{F}$, defined on the real line by

$$\hat{F}(y) = \begin{cases} 0 & \text{if } y < Y_{(1)} \\ t/T & \text{if } Y_{(t)} \leqslant y < Y_{(t+1)}, \ t = 1, \ldots, T-1 \\ 1 & \text{if } y \geqslant Y_{(T)} \end{cases}$$

where $Y_{(1)} \leqslant \cdots \leqslant Y_{(T)}$ are the sample-order statistics, and the empirical qf (eqf) $\hat{Q}$, defined on the unit interval $(0, 1)$ by

$$\hat{Q}(p) = Y_{(t)} \quad \text{if } \frac{t-1}{T} < p \leqslant \frac{t}{T}, \quad t = 1, \ldots, T$$

Because $\hat{Q}(p) = \inf\{y : \hat{F}(y) \geqslant p\}$, $p \in (0, 1)$, it follows that

$$\hat{\tau}(\alpha) = \frac{1}{\alpha} \int_0^\alpha \hat{Q}(p)\,\mathrm{d}p = \hat{Q}(\alpha) - \frac{1}{\alpha} \int_{-\infty}^{\hat{Q}(\alpha)} \hat{F}(y)\,\mathrm{d}y = \frac{1}{\alpha T} \sum_{t=1}^{[\alpha T]} Y_{(t)} + \left(1 - \frac{[\alpha T]}{\alpha T}\right) Y_{([\alpha T]+1)} \qquad (7)$$

where $[\alpha T]$ denotes the integer part of αT. Thus, $\hat{\tau}(\alpha)$ is a linear combination of extreme-order statistics. Unlike standard L-estimators, however, both the number and the nature of the terms in the linear combination change with the sample size. If αT is an integer, then $\tau(\alpha) = (\alpha T)^{-1} \sum_{t=1}^{[\alpha T]} Y_{(t)}$. Note that this estimator coincides with the maximum likelihood estimator under the assumption that the distribution of Y_t conditional on $Y_t \leqslant Q(\alpha)$ is exponential on the negative half-line.

The study of the asymptotic distribution of estimators of the form (7) has been carried out by Csörgö *et al.* [11]. In particular, they provide necessary and sufficient conditions for $\hat{\tau}(\alpha)$ to be asymptotically normal. More precisely, they show that if and only if certain conditions on the limiting behavior of the smallest- and largest-order statistics in the sum on the right-hand side of (7) are satisfied (see their Corollary 1), then

$$\sqrt{T}(\hat{\tau}(\alpha) - \tau(\alpha)) \xrightarrow{d} \mathcal{N}(0, AV(\alpha))$$

as $T \to \infty$, where

$$AV(\alpha) = \int_0^\alpha \int_0^\alpha [\min(s, t) - st]\,\mathrm{d}Q(s)\,\mathrm{d}Q(t) \qquad (8)$$

In fact, Csörgö *et al.* [11] establish, more generally, the asymptotic properties of weighted sums of extreme-order statistics of the form

$$\tilde{\tau}(\alpha) = \sum_{i=1}^I w_i Y_{(i)} \qquad (9)$$

where $w_1, \ldots, w_I$ is a set of weights and the number of terms I in the weighted sum depends on the sample size and satisfies $I \to \infty$ and $I/T \to \alpha$.

3.2. Estimation of the conditional expected shortfall

In this section we consider the case when we also have available data on a vector X_t of predictors of Y_t, which may include a finite number of lags of Y_t. After briefly discussing non-parametric estimation, we propose two classes of analog estimators based, respectively, on estimates of the cqf $Q(p\,|\,x)$ and the cdf $F(y\,|\,x)$.

3.2.1. Non-parametric estimation. A simple class of fully non-parametric estimators of $\tau(\alpha\,|\,x)$ is local versions of (7), that is, averages of the smallest sample-order statistics over a neighborhood

of x defined by a suitably chosen kernel function $K(\cdot)$. This corresponds to the class of estimators of the form

$$\bar{\tau}(\alpha \,|\, x) = \frac{\sum_{t=1}^{T} Y_t K_t(x) \mathbb{1}\{Y_t \leqslant \hat{Q}(\alpha \,|\, x)\}}{\sum_{t=1}^{T} K_t(x) \mathbb{1}\{Y_t \leqslant \hat{Q}(\alpha \,|\, x)\}}$$

where $K_t(x) = K((X_t - x)/h)$ is the kernel weight, h is a fixed bandwidth, and $\hat{Q}(\alpha \,|\, x)$ is some estimator of the conditional quantile $Q(\alpha \,|\, x)$. Consistency of $\bar{\tau}(\alpha \,|\, x)$ requires the bandwidth h to go to zero as $T \to \infty$, but at a slower rate than T. Automatic choice of the bandwidth h is a topic for future research. Because of the curse-of-dimensionality problem, this fully non-parametric estimator is unlikely to perform well when the X_t is a vector of predictors, unless the sample size T is extremely large.

For the empirically more relevant case when X_t is a vector with several components, result (6) suggests two classes of analog estimators of $\tau(\alpha \,|\, x)$, namely

$$\hat{\tau}(\alpha \,|\, x) = \frac{1}{\alpha} \int_0^{\alpha} \hat{Q}(p \,|\, x) \, \mathrm{d}p \tag{10}$$

and

$$\tilde{\tau}(\alpha \,|\, x) = \hat{Q}(\alpha \,|\, x) - \frac{1}{\alpha} \int_{-\infty}^{\hat{Q}(\alpha \,|\, x)} \hat{F}(y \,|\, x) \, \mathrm{d}y \tag{11}$$

where $\hat{Q}(p \,|\, x)$ is some estimator of $Q(p \,|\, x)$ and $\hat{F}(y \,|\, x)$ is some estimator of $F(y \,|\, x)$. We shall refer to estimators based on (10) as integrated conditional quantile function (ICQF) estimators and to estimators based on (11) as integrated conditional distribution function (ICDF) estimators. Unlike the unconditional case, one cannot generally guarantee that $\hat{Q}(p \,|\, x) = \inf\{y : \hat{F}(y \,|\, x) \geqslant p\}$, $p \in (0, 1)$. Hence, the two classes of estimators need not coincide.

In the remainder of this paper, we propose specific versions of these two classes of estimators, corresponding to specific choices of $\hat{Q}(p \,|\, x)$ and $\hat{F}(y \,|\, x)$.

3.2.2. ICQF estimators. Conditional quantiles are often assumed to be linear in parameters, that is, of the form

$$Q(p \,|\, x) = \beta(p)^{\top} x$$

This is in fact the case originally considered by Koenker and Bassett [12], who proposed estimating $\beta(p)$ by solving

$$\min_{\beta} \sum_{t=1}^{n} \ell_p(Y_t - \beta^{\top} X_t)$$

where

$$\ell_p(u) = u(p - \mathbb{1}\{u < 0\}), \quad 0 < p < 1$$

is the asymmetric absolute loss function. Given a linear regression quantile estimator $\hat{\beta}(p)$, an estimator of $Q(p \,|\, x)$ is easily obtained as $\hat{Q}(p \,|\, x) = \hat{\beta}(p)^{\top} x$. Under general conditions, $\hat{\beta}(p)$

and $\hat{Q}(p|x)$ can be shown to be consistent provided that $Q(p|x)$ is linear in parameters. These estimators can also be shown to be asymptotically normal irrespective of whether the linear conditional quantile model is correctly specified (see Angrist *et al.* [13]). These results generalize to any fixed collection $\hat{\beta}(p_1), \ldots, \hat{\beta}(p_I)$ of linear regression quantile estimators.

Based on these results, a simple class of ICQF estimators of $\tau(\alpha|x)$ consists of weighted sums of linear regression quantile estimators, namely

$$\hat{\tau}(\alpha|x) = \sum_{i=1}^{I} w_i \hat{Q}(p_i|x) = \hat{\beta}^*(\alpha)^\top x$$

where $w_1, \ldots, w_I$ is a set of weights, the number I of terms in the weighted sum may depend on the sample size, and

$$\hat{\beta}^*(\alpha) = \sum_{i=1}^{I} w_i \hat{\beta}(p_i)$$

with $0 < p_1 < \cdots < p_I \leqslant \alpha$. To guarantee consistency of this estimator, I should be required to increase with the sample size T. Automatic choice of I is a topic for future research.

As for the asymptotic behavior of estimators of this type, we conjecture that, being linear combinations of asymptotically normal estimators, they are also asymptotically normal. The Monte Carlo evidence in Section 4 provides strong support for this conjecture. A formal proof of the asymptotic properties of $\hat{\tau}(\alpha|x)$ requires some care because $\hat{\beta}(p_1), \ldots, \hat{\beta}(p_I)$ is a collection of dependent and not identically distributed random vectors, and one needs to control the rate at which the number of terms I in the collection grows with the sample size T.

A drawback of the class of ICQF estimators is that linear regression quantile estimators may cross each other, that is, we may have $\hat{Q}(p|x) < \hat{Q}(p'|x)$ for $p > p'$ at some x value. This problem does not occur at $x = \bar{X}$, where $\bar{X}$ is the sample average of the X_t [see Dodge and Jurečková [14], pp. 127–128], but may occur at x values in the tails of the distribution of X_t, especially when Y_t is conditionally heteroskedastic, that is, its conditional variance is not constant but depends on X_t. How to impose monotonicity on estimating a family of conditional quantiles is an important but still largely unresolved issue.

3.2.3. ICDF estimators. In order to estimate $F(y|x)$, we follow the approach in Peracchi [6]. We select J distinct values $y_1, \ldots, y_J$ such that $Y_{(1)} < y_1 < \cdots < y_J < Y_{(T)}$ and define the log-odds

$$\eta_j(x) = \ln \frac{F_j(x)}{1 - F_j(x)}, \quad j = 1, \ldots, J$$

where $F_j(x) = F(y_j|x) = \Pr\{Y_t \leqslant y_j|x\}$. Because each rv $\mathbb{1}\{Y_t \leqslant y_j\}$ has a Bernoulli distribution with parameter $F_j(x)$, we can estimate each $\eta_j(x)$ by a separate logistic regression. Given an estimator $\tilde{\eta}_j(x)$ of $\eta_j(x)$, we can then estimate $F_j(x)$ by

$$\tilde{F}_j(x) = \frac{\exp \tilde{\eta}_j(x)}{1 + \exp \tilde{\eta}_j(x)}$$

After putting $y_0 = Y_{(1)}$ and $y_{J+1} = Y_{(T)}$, linear interpolation between thresholds gives the following estimate of the cdf:

$$\tilde{F}(y\,|\,x) = \begin{cases} \tilde{F}_0(x) = 0 & \text{if } y \leqslant y_0 \\ (1 - \varepsilon_j)\tilde{F}_j(x) + \varepsilon_j \tilde{F}_{j+1}(x) & \text{if } y_j \leqslant y < y_{j+1} \text{ and } j = 1, \ldots, J \\ \tilde{F}_{J+1}(x) = 1 & \text{if } y \geqslant y_{J+1} \end{cases}$$

where $\varepsilon_j = (y - y_j)/(y_{j+1} - y_j)$.

Given $\tilde{F}(y\,|\,x)$ and an estimator $\hat{Q}(\alpha\,|\,x)$ of $Q(\alpha\,|\,x)$, we obtain the following analog estimator of $\tau(\alpha)$:

$$\tilde{\tau}(\alpha\,|\,x) = \hat{Q}(\alpha\,|\,x) - \frac{1}{\alpha}\int_{-\infty}^{\hat{Q}(\alpha\,|\,x)} \tilde{F}(y\,|\,x)\,\mathrm{d}y = \hat{Q}(\alpha\,|\,x) - \frac{1}{\alpha}\sum_{i=1}^{I} \omega_i \tilde{F}_i(x)$$

where $\omega_1, \ldots, \omega_I$ is a set of weights, the number of terms I in the weighted sum is required to increase with the sample size T, and $\tilde{F}_I(x) = \max\{\alpha, \tilde{F}_{I-1}(x)\}$.

Automatic choice of I is again a topic for future research. Linear interpolation of the cdf corresponds to

$$\omega_i = \begin{cases} (y_{i+1} - y_{i-1})/2 & \text{if } i = 1, \ldots, I-1 \\ (y_I - y_{I-1})/2 & \text{if } i = I \end{cases}$$

with $y_I = \hat{Q}(\alpha\,|\,x)$, but other choices of weights are possible.

One drawback of this class of ICDF estimators is that the estimated cdf $\hat{F}$ need not be monotonic, that is, it need not satisfy the condition that $\hat{F}_j(x) \geqslant \hat{F}_{j-1}(x)$ for all x. A simple way of imposing monotonicity is to exploit the fact that

$$F_j(x) = 1 - [1 - F_1(x)]\prod_{h=2}^{j}[1 - \lambda_h(x)], \quad h = 2, \ldots, J$$

where

$$\lambda_h(x) = \frac{F_h(x) - F_{h-1}(x)}{1 - F_{h-1}(x)} = \Pr\{y_{h-1} \leqslant Y_t < y_h \,|\, Y_t \geqslant y_{h-1}, x\}$$

Estimators for the $\lambda_h(x)$ may be obtained by fitting $J-1$ separate logistic regressions, one for each binary rv $\mathbb{1}\{Y_t < y_h\}$ conditional on $Y_t \geqslant y_{h-1}$, $h = 2, \ldots, J$. Given an estimator $\tilde{F}_1(x)$ of $F_1(x)$ and estimators $\tilde{\lambda}_h(x)$ of the $\lambda_h(x)$, we can then estimate $F_j(x)$ by the monotone estimator

$$\tilde{F}_j^*(x) = 1 - [1 - \tilde{F}_1(x)]\prod_{h=2}^{j}[1 - \tilde{\lambda}_h(x)], \quad j = 2, \ldots, J$$

and obtain a monotone estimate $\tilde{F}^*(y\,|\,x)$ of the cdf by linear interpolation. Replacing the non-monotone estimate $\tilde{F}$ by the monotone estimate $\tilde{F}^*$ gives another class of ICDF estimators of $\tau(\alpha)$, namely

$$\tilde{\tau}^*(\alpha\,|\,x) = \hat{Q}(\alpha\,|\,x) - \frac{1}{\alpha}\int_{-\infty}^{\hat{Q}(\alpha\,|\,x)} \tilde{F}^*(y\,|\,x)\,\mathrm{d}y = \hat{Q}(\alpha\,|\,x) - \frac{1}{\alpha}\sum_{i=1}^{I} \omega_i^* \tilde{F}_i^*(x)$$

As for the sampling properties of these two classes of ICDF estimators, in this paper we again confine ourselves to the Monte Carlo evidence presented in Section 4.

4. MONTE CARLO EVIDENCE

In this section we present some Monte Carlo evidence on the sampling properties of the unconditional estimator $\hat{\tau}(\alpha)$, the fully non-parametric estimator $\bar{\tau}(\alpha|x)$, the ICQF estimator $\hat{\tau}(\alpha|x)$, and the two ICDF estimators $\tilde{\tau}(\alpha|x)$ and $\hat{\tau}^*(\alpha|x)$. In our Monte Carlo experiments, we consider varying samples sizes of 250, 500 and 1000 observations. For each sample size, the number of Monte Carlo replications is set equal to 1000. As for the level α, we consider three typical values, namely 1, 5 and 10%. The Monte Carlo experiments were carried out using the statistical package Stata, version 9.1.

4.1. Estimation of the unconditional expected shortfall

We consider the case when the data are a random sample from four alternative distributions, all with a finite variance and symmetric about a mean of zero. The first distribution is the standard normal, the second is the mixture of a $\mathcal{N}(0, 1)$ with probability 80% and a $\mathcal{N}(0, 2)$ with probability 20%, the third and the fourth are Student's t distributions, with 2 and 4 degrees of freedom, respectively. Details on the Monte Carlo distribution of $\hat{\tau}(\alpha)$ are given in Table I. In addition to the number $I = [\alpha T]$ of extreme-order statistics that enter the estimation, the tables report the values of $\tau(\alpha)$ for each parent distribution and summaries of the Monte Carlo distribution of the estimator $\hat{\tau}(\alpha)$, namely the mean bias (Bias), the median bias (MBias), the standard deviation (SD), the root mean squared error (RMSE), and the coefficients of skewness (Skew) and kurtosis (Kurt).

Figures 1 and 2 plot, for each set of parameters, kernel estimates of the Monte Carlo densities, respectively, for the case when the parent distribution is normal and a normal mixture. The panels in each figure correspond to different values of α and present Monte Carlo densities corresponding to the various sample sizes. A vertical line in each panel marks the value of $\tau(\alpha)$.

The bias of the unconditional estimator tends to be small, except for small values of α in the case of the t distributions (especially the t distribution with 2 degrees of freedom). In small samples and for a small values of α, the estimator is not very precise. However, its precision increases with the sample size T and the level α.

4.2. Estimation of the conditional expected shortfall

To facilitate comparisons of alternative estimators of the conditional expected shortfall, we consider cases when the conditional mean of the outcome Y_t depends linearly on a constant and a single regressor X_t. Both homoskedastic and heteroskedastic versions of the model are considered. In the homoskedastic version, the conditional distribution of Y_t given $X_t = x$ is $\mathcal{N}(-1+x, 1)$. In the heteroskedastic version it is $\mathcal{N}(-1+x, (1+0.25x)^2)$. In either case, X_t is distributed as $\mathcal{N}(0, 1)$, and the population regression R^2 is about 50%.

We compare the Monte Carlo behavior of four estimators: (i) the fully non-parametric estimator $\bar{\tau}(\alpha|x)$, denoted by NP, (ii) the ICQF estimator $\hat{\tau}(\alpha|x)$, (iii) the ICDF estimator $\tilde{\tau}(\alpha|x)$ based on the non-monotonic cdf estimate $\hat{F}(y|x)$, denoted by ICDF1, and (iv) the ICDF estimator $\tilde{\tau}^*(\alpha|x)$ based on the monotonic cdf estimate $\tilde{F}(y|x)$, denoted by ICDF2.

Table I. Summary statistics of the Monte Carlo distribution of $\hat{\tau}(\alpha)$ when $Y_t \sim \mathcal{N}(0,1)$, $Y_t \sim 0.8*\mathcal{N}(0,1)+0.2*\mathcal{N}(0,2)$, $Y_t \sim t[2]$ and $Y_t \sim t[4]$, respectively, for $\alpha = 1, 5, 10\%$.

| α | T | l | $\mathcal{N}(0,1)$ | | | | | | | $0.8*\mathcal{N}(0,1)+0.2*\mathcal{N}(0,4)$ | | | | | | |
			$\tau(\alpha)$	Bias	MBias	SD	RMSE	Skew	Kurt	$\tau(\alpha)$	Bias	MBias	SD	RMSE	Skew	Kurt
0.01	250	2	−2.665	−0.016	0.003	0.304	0.304	−0.386	3.040	−4.135	−0.014	0.074	0.723	0.723	−0.580	3.676
0.01	500	5	−2.665	0.043	0.057	0.198	0.202	−0.317	3.130	−4.135	0.093	0.120	0.483	0.491	−0.223	2.935
0.01	1000	10	−2.665	0.025	0.029	0.142	0.145	−0.217	3.173	−4.135	0.044	0.050	0.339	0.342	−0.120	3.083
0.05	250	12	−2.063	0.005	0.012	0.158	0.158	−0.156	3.013	−2.802	−0.002	0.019	0.336	0.336	−0.454	3.391
0.05	500	25	−2.063	0.004	0.003	0.111	0.111	−0.105	3.125	−2.802	0.022	0.040	0.224	0.225	−0.305	2.976
0.05	1000	50	−2.063	0.004	0.010	0.080	0.080	−0.204	3.164	−2.802	0.008	0.011	0.165	0.165	−0.204	3.107
0.1	250	25	−1.755	0.007	0.007	0.117	0.117	0.045	2.910	−2.259	0.011	0.018	0.210	0.210	−0.235	3.019
0.1	500	50	−1.755	0.004	0.005	0.084	0.085	−0.116	3.115	−2.259	0.011	0.018	0.158	0.158	−0.225	2.994
0.1	1000	100	−1.755	0.002	0.000	0.061	0.061	0.023	3.129	−2.259	0.001	0.003	0.110	0.110	−0.237	3.330

| α | T | l | $t[2]$ | | | | | | | $t[4]$ | | | | | | |
			$\tau(\alpha)$	Bias	MBias	SD	RMSE	Skew	Kurt	$\tau(\alpha)$	Bias	MBias	SD	RMSE	Skew	Kurt
0.01	250	2	−13.968	−1.153	2.137	11.996	12.045	−4.346	33.615	−5.217	−0.170	0.178	1.699	1.707	−1.697	7.450
0.01	500	5	−13.968	0.550	2.543	11.218	11.226	−14.401	308.183	−5.217	0.158	0.356	1.077	1.088	−2.193	14.894
0.01	1000	10	−13.968	0.228	1.658	6.319	6.320	−4.679	39.411	−5.217	0.083	0.197	0.771	0.775	−1.200	5.919
0.05	250	12	−6.118	0.043	0.552	2.130	2.129	−2.204	10.489	−3.201	0.008	0.072	0.488	0.488	−0.781	4.021
0.05	500	25	−6.118	−0.058	0.366	2.151	2.150	−6.300	69.543	−3.201	0.001	0.039	0.353	0.353	−0.794	4.627
0.05	1000	50	−6.118	−0.046	0.292	1.409	1.409	−2.715	16.501	−3.201	0.004	0.039	0.261	0.261	−0.722	3.937
0.1	250	25	−4.210	−0.064	0.241	2.108	2.108	−13.848	289.912	−2.498	0.006	0.035	0.307	0.307	−1.068	8.244
0.1	500	50	−4.210	−0.034	0.133	1.097	1.097	−6.692	90.494	−2.498	0.002	0.014	0.217	0.217	−0.543	3.797
0.1	1000	100	−4.210	−0.040	0.076	0.918	0.919	−10.293	171.314	−2.498	−0.001	0.001	0.149	0.149	−0.174	3.017

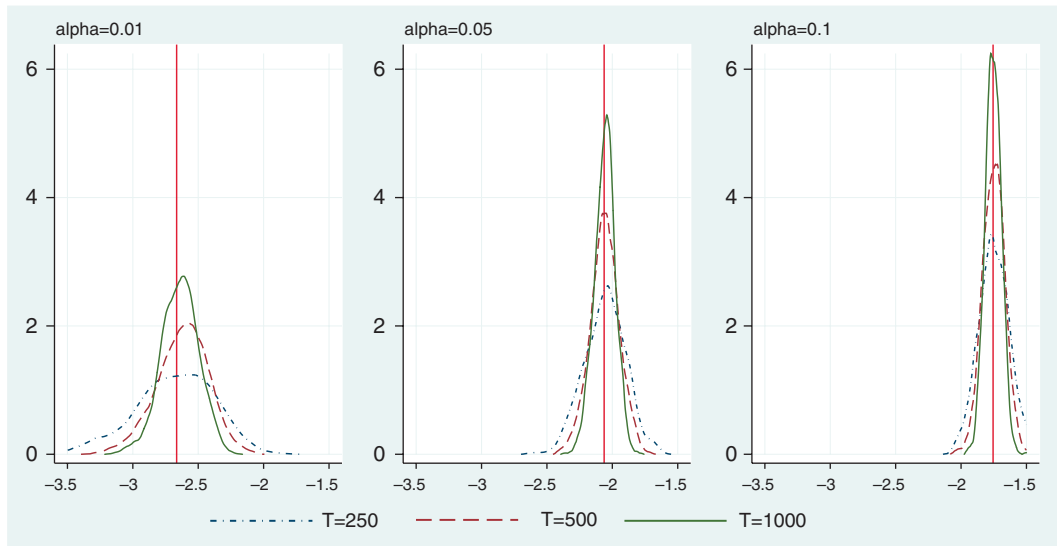The Monte Carlo distributions are based on 1000 samples of size $T = 250, 500, 1000$.

Figure 1. Monte Carlo densities of the unconditional estimator $\hat{\tau}(\alpha)$ when $Y_t \sim \mathcal{N}(0,1)$, for $\alpha = 1, 5, 10\%$. The densities are based on 1000 samples of size $T = 250, 500, 1000$.
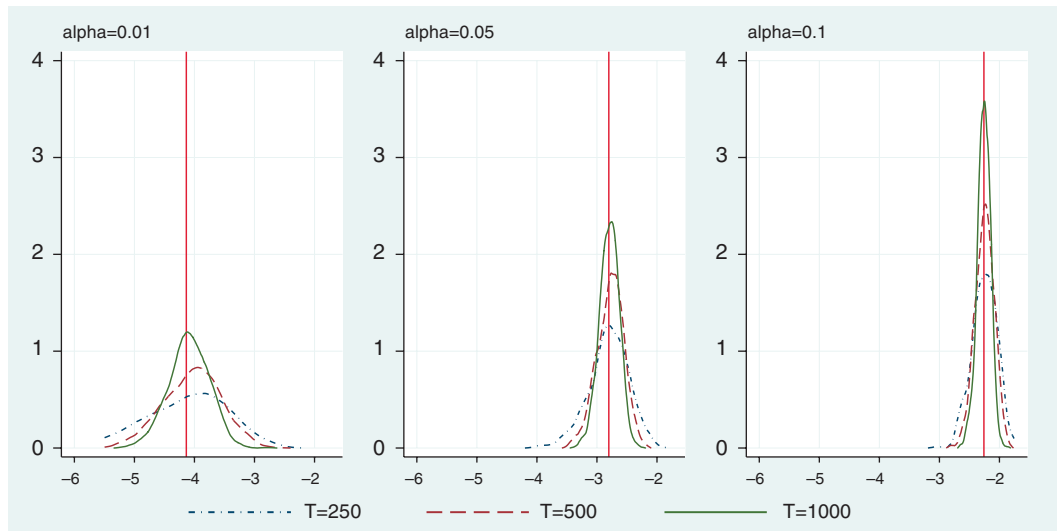

Figure 2. Monte Carlo densities of the unconditional estimator $\hat{\tau}(\alpha)$ when $Y_t \sim 0.8 * \mathcal{N}(0,1) + 0.2 * \mathcal{N}(0,2)$, for $\alpha = 1, 5, 10\%$. The densities are based on 1000 samples of size $T = 250, 500, 1000$.

For the non-parametric estimator NP, we use a Gaussian kernel with bandwidth $h = \sigma_X T^{-1/5}$, where σ_X is the standard deviation of X_t and T is the sample size.

After some experimentation, the number of estimated regression quantiles for the ICQF estimator and the number of thresholds for the two ICDF estimators is set equal to $I = 1, 2, 4$, respectively,

for $T = 250, 500, 1000$ and $\alpha = 0.01$. For α equal to 5 and 10%, the value of I is scaled up proportionally (thus, $I = 5, 10, 20$, respectively, for $T = 250, 500, 1000$ and $\alpha = 5\%$). For the ICQF estimator, we also choose $p_i = \alpha(2i-1)/2I$, with $i = 1, \ldots, I$, and uniform weights $w_i = 1/I$. This corresponds to choosing the p_i equally spaced between $p_1 = \alpha/(2I)$ and $p_I = \alpha - \alpha/(2I)$, that is, $p_i = p_{i-1} + \delta$, with $\delta = \alpha/I$. For the ICDF estimators, the thresholds correspond to equally spaced order statistics between $y_1 = Y_{(1+\delta)}$ and $y_I = Y_{(1+\delta I)}$, where $\delta = [S/(I+1)]$ is the integer part of the ratio of the number S of data points to the left of $\hat{Q}(\alpha \,|\, x)$ and the number I of thresholds. In practice, construction of the ICDF estimator requires the estimated conditional quantile to be greater than the first sample-order statistic $Y_{(1)}$. This condition is not guaranteed, especially when $\alpha = 1\%$ and $T = 250$. In this case, we decided to drop the 'failed' experiments (those where the condition is not met) and draw Monte Carlo samples until a predetermined number of 1000 'successful' experiments are reached. For $\alpha = 1\%$, the ratio of 'failed' to 'successful' experiments is between 20 and 25% for $T = 250$, is between 5 and 7% for $T = 500$, drops to less than 1% for $T = 1000$, and is zero or negligible in all other cases.

Details on the Monte Carlo distribution of the four alternative estimators for different α-levels ($\alpha = 1, 5$ and 10%), sample sizes ($T = 250, 500, 1000$) and x-values (the 10th and 50th percentiles of X_t) are given in Tables II and III, respectively, for the homoskedastic and the heteroskedastic case. Each table reports the value of I for the ICDF and the ICQF estimators, and the mean bias (Bias), the standard deviation (SD) and the root mean squared error (RMSE) of all four estimators. In Figures 3 and 5 (for the homoskedastic case) and Figures 4 and 6 (for the heteroskedastic case), we plot kernel estimates of the Monte Carlo densities of the various estimators for $\alpha = 5\%$. In each graph, we keep the x-value fixed and increase the size T of the Monte Carlo sample.

The Monte Carlo results follow the same pattern as for the unconditional estimator, with the bias and the RMSE of all estimators falling with the sample size. In most cases, the coefficients of skewness and kurtosis (not reported to save space) range in the intervals $(-0.5, 0.5)$ and $(2.5, 3.5)$, respectively. For moderate and large sample sizes (500 or 1000 observations) and values of α equal to 5 and 10%, the Monte Carlo distribution of all estimators looks approximately normal, supporting the conjecture of asymptotic normality of these estimators. Overall, the fully non-parametric estimator NP has a smaller bias, a larger SD and a larger RMSE than the other three estimators. In general, the ICQF estimator performs better than the ICDF estimators in terms of RMSE except when either the level α or the sample size T is small. This is mostly due to its smaller bias which, for small α or small T, is offset by a larger variability. Of the two ICDF estimators, ICDF1 tends to have a smaller bias but a larger variability than ICDF2. In terms of RMSE, ICDF1 tends to do better in the homoskedastic case irrespective of α, x and T, whereas ICDF2 tends to do better in the heteroskedastic case.

5. EMPIRICAL ILLUSTRATION

In this section we present an empirical application using real data. The dependent variable is the daily excess return on the S&P 500 index, while the regressors include a number of real and financial variables. We report results for the unconditional estimator $\hat{\tau}(\alpha)$, the ICQF estimator $\tilde{\tau}(\alpha \,|\, x)$ and the two ICDF estimators $\tilde{\tau}(\alpha \,|\, x)$ and $\tilde{\tau}^*(\alpha \,|\, x)$. We exclude the fully non-parametric estimator $\bar{\tau}(\alpha \,|\, x)$ because it suffers of curse-of-dimensionality problems given the relatively high number of predictors that we consider.

Table II. Summary statistics of the Monte Carlo distribution of alternative estimators of $\tau(\alpha|x)$ in the homoskedastic model $Y_t|X_t \sim \mathcal{N}(-1+X_t, 1)$, $X_t \sim \mathcal{N}(0,1)$ and in the heteroskedastic model $Y_t|X_t \sim \mathcal{N}(-1+X_t, (1+0.25X_t)^2)$, $X_t \sim \mathcal{N}(0,1)$ for $x = -1.282$ (the 10th percentile of X_t) and $\alpha = 1, 5, 10\%$.

			NP			ICQF				ICDF1				ICDF2			
α	T	$\tau(\alpha|x)$	Bias	SD	RMSE	I	Bias	SD	RMSE	I	Bias	SD	RMSE	I	Bias	SD	RMSE
Homoskedastic model, $x = -1.282$																	
0.01	250	−4.947	0.027	0.546	0.547	1	0.163	0.483	0.510	1	0.149	0.443	0.467	1	0.149	0.443	0.467
0.01	500	−4.947	−0.018	0.448	0.448	2	0.095	0.333	0.346	2	0.042	0.336	0.338	2	0.040	0.334	0.336
0.01	1000	−4.947	0.010	0.339	0.339	4	0.043	0.249	0.253	4	0.018	0.250	0.251	4	0.022	0.248	0.249
0.05	250	−4.344	−0.028	0.362	0.363	5	0.024	0.255	0.256	5	0.045	0.274	0.277	5	0.061	0.271	0.278
0.05	500	−4.344	0.000	0.247	0.247	10	0.021	0.179	0.180	10	0.051	0.186	0.193	10	0.093	0.179	0.202
0.05	1000	−4.344	0.005	0.179	0.179	20	0.012	0.127	0.128	20	0.053	0.135	0.145	20	0.104	0.130	0.166
0.1	250	−4.037	0.003	0.250	0.250	10	0.020	0.197	0.198	10	0.062	0.206	0.215	10	0.118	0.197	0.229
0.1	500	−4.037	−0.007	0.195	0.195	20	0.011	0.146	0.147	20	0.062	0.148	0.161	20	0.135	0.142	0.195
0.1	1000	−4.037	−0.001	0.132	0.132	40	0.004	0.095	0.095	40	0.057	0.100	0.115	40	0.138	0.095	0.167
Heteroskedastic model, $x = -1.282$																	
0.01	250	−4.093	0.047	0.305	0.309	1	0.038	0.367	0.369	1	0.100	0.263	0.281	1	0.100	0.263	0.281
0.01	500	−4.093	0.013	0.257	0.257	2	−0.001	0.251	0.251	2	−0.004	0.225	0.225	2	0.016	0.214	0.215
0.01	1000	−4.093	−0.002	0.212	0.211	4	−0.019	0.164	0.165	4	−0.060	0.177	0.187	4	−0.019	0.158	0.159
0.05	250	−3.683	−0.013	0.220	0.220	5	−0.019	0.182	0.183	5	−0.053	0.191	0.198	5	0.001	0.175	0.175
0.05	500	−3.683	−0.005	0.161	0.161	10	−0.021	0.135	0.136	10	−0.091	0.144	0.170	10	0.000	0.126	0.126
0.05	1000	−3.683	0.002	0.118	0.118	20	−0.010	0.087	0.087	20	−0.106	0.106	0.150	20	0.007	0.088	0.089
0.1	250	−3.474	0.000	0.161	0.161	10	−0.008	0.137	0.137	10	−0.076	0.154	0.171	10	0.024	0.136	0.138
0.1	500	−3.474	0.005	0.124	0.124	20	−0.004	0.092	0.092	20	−0.095	0.104	0.140	20	0.033	0.091	0.097
0.1	1000	−3.474	0.002	0.087	0.087	40	−0.002	0.063	0.063	40	−0.101	0.079	0.128	40	0.036	0.068	0.077

The Monte Carlo distributions are based on 1000 samples of size $T = 250, 500, 1000$.

Table III. Summary statistics of the Monte Carlo distribution of alternative estimators of $\tau(\alpha|x)$ in the homoskedastic model $Y_t|X_t \sim \mathcal{N}(-1+X_t, 1)$, $X_t \sim \mathcal{N}(0,1)$ and in the heteroskedastic model $Y_t|X_t \sim \mathcal{N}(-1+X_t, (1+0.25X_t)^2)$, $X_t \sim \mathcal{N}(0,1)$, for $x=0$ (the median of X_t) and $\alpha=1, 5, 10\%$.

α	T	$\tau(\alpha|x)$	NP Bias	NP SD	NP RMSE	ICQF I	ICQF Bias	ICQF SD	ICQF RMSE	ICDF1 I	ICDF1 Bias	ICDF1 SD	ICDF1 RMSE	ICDF2 I	ICDF2 Bias	ICDF2 SD	ICDF2 RMSE
Homoskedastic model, x=0																	
0.01	250	-3.665	-0.030	0.372	0.373	1	0.124	0.298	0.322	1	0.015	0.251	0.251	1	0.015	0.251	0.251
0.01	500	-3.665	-0.018	0.273	0.274	2	0.073	0.209	0.221	2	-0.079	0.182	0.199	2	-0.109	0.181	0.212
0.01	1000	-3.665	-0.004	0.201	0.201	4	0.042	0.148	0.154	4	-0.131	0.145	0.195	4	-0.197	0.143	0.243
0.05	250	-3.063	-0.034	0.207	0.209	5	0.030	0.160	0.162	5	-0.050	0.169	0.176	5	-0.099	0.164	0.192
0.05	500	-3.063	-0.028	0.151	0.154	10	0.014	0.112	0.113	10	-0.062	0.128	0.142	10	-0.133	0.125	0.183
0.05	1000	-3.063	-0.017	0.109	0.110	20	0.010	0.077	0.078	20	-0.055	0.089	0.105	20	-0.145	0.089	0.170
0.1	250	-2.755	-0.026	0.165	0.167	10	0.019	0.125	0.127	10	-0.010	0.146	0.147	10	-0.060	0.143	0.155
0.1	500	-2.755	-0.020	0.112	0.114	20	0.015	0.086	0.088	20	-0.017	0.101	0.103	20	-0.082	0.098	0.128
0.1	1000	-2.755	-0.021	0.085	0.088	40	0.007	0.061	0.061	40	-0.022	0.072	0.075	40	-0.099	0.071	0.122
Heteroskedastic model, x=0																	
0.01	250	-3.665	0.040	0.342	0.344	1	0.114	0.308	0.328	1	0.167	0.242	0.294	1	0.167	0.242	0.294
0.01	500	-3.665	0.034	0.266	0.268	2	0.070	0.217	0.228	2	0.137	0.181	0.227	2	0.129	0.178	0.220
0.01	1000	-3.665	0.021	0.208	0.209	4	0.039	0.152	0.157	4	0.131	0.136	0.189	4	0.116	0.133	0.176
0.05	250	-3.063	0.005	0.198	0.198	5	0.029	0.166	0.168	5	0.064	0.164	0.176	5	0.052	0.152	0.161
0.05	500	-3.063	0.002	0.150	0.149	10	0.014	0.115	0.116	10	0.056	0.122	0.134	10	0.040	0.111	0.118
0.05	1000	-3.063	0.003	0.110	0.110	20	0.009	0.080	0.080	20	0.063	0.088	0.109	20	0.044	0.079	0.091
0.1	250	-2.755	0.006	0.163	0.163	10	0.018	0.129	0.130	10	0.040	0.146	0.151	10	0.038	0.132	0.137
0.1	500	-2.755	0.003	0.112	0.112	20	0.014	0.090	0.091	20	0.035	0.105	0.111	20	0.032	0.093	0.098
0.1	1000	-2.755	-0.002	0.084	0.084	40	0.006	0.063	0.063	40	0.027	0.074	0.079	40	0.023	0.065	0.069

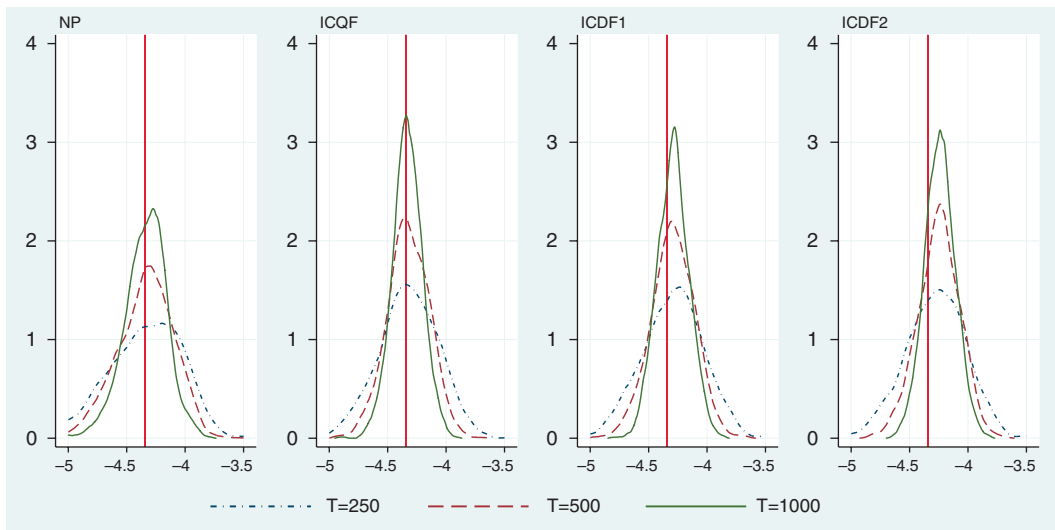The Monte Carlo distributions are based on 1000 samples of size $T=250, 500, 1000$.

Figure 3. Monte Carlo densities of alternative estimators of $\tau(\alpha|x)$ in the homoskedastic model $Y_t|X_t \sim \mathcal{N}(-1+X_t, 1)$ and $X_t \sim \mathcal{N}(0,1)$, for $\alpha = 5\%$ and $x = -1.282$ (the 10th percentile of X_t). The densities are based on 1000 samples of size $T = 250, 500, 1000$.
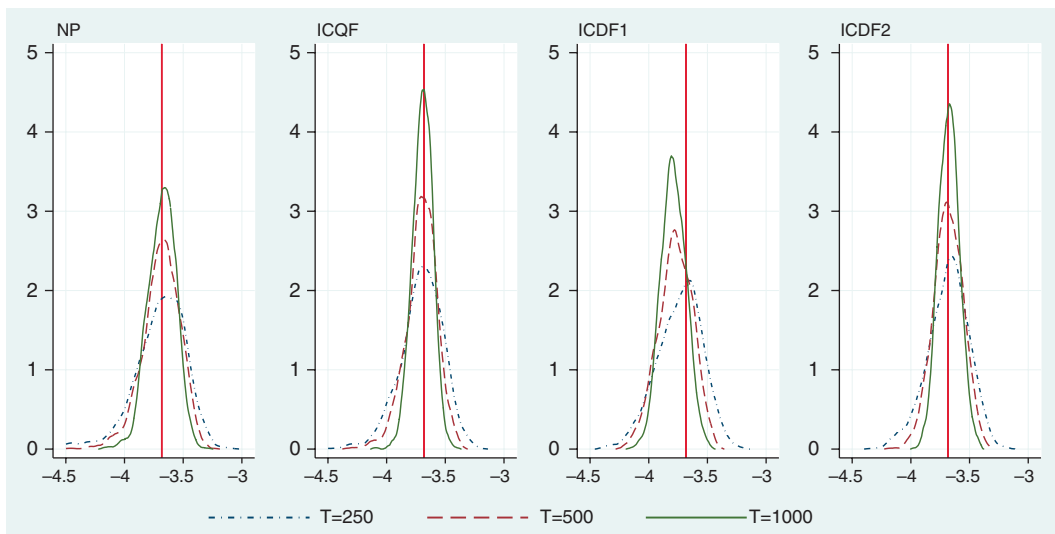

Figure 4. Monte Carlo densities of alternative estimators of $\tau(\alpha|x)$ in the heteroskedastic model $Y_t|X_t \sim \mathcal{N}(-1+X_t, (1+0.25X_t)^2)$ and $X_t \sim \mathcal{N}(0,1)$, for $\alpha = 5\%$ and $x = -1.282$ (the 10th percentile of X_t). The densities are based on 1000 samples of size $T = 250, 500, 1000$.

5.1. The data

Our raw data are daily from 30 December 1994 to 31 December 2004. The dependent variable is the daily excess return on the U.S. stock market, defined as the difference between the return
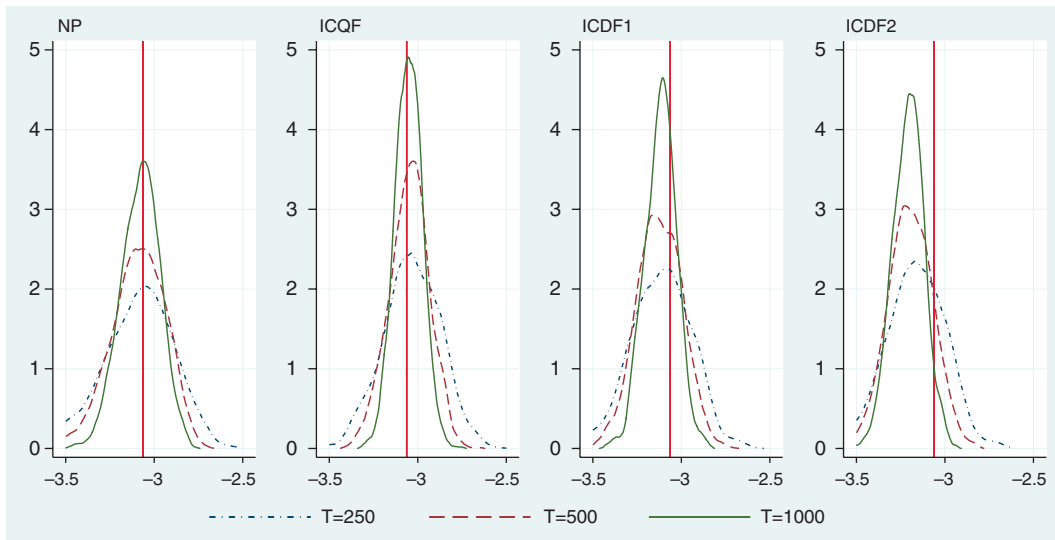
Figure 5. Monte Carlo densities of alternative estimators of $\tau(\alpha|x)$ in the homoskedastic model $Y_t|X \sim \mathcal{N}(-1+X_t, 1)$ and $X_t \sim \mathcal{N}(0, 1)$, for $\alpha = 5\%$ and $x = 0$ (the median of X_t). The densities are based on 1000 samples of size $T = 250, 500, 1000$.
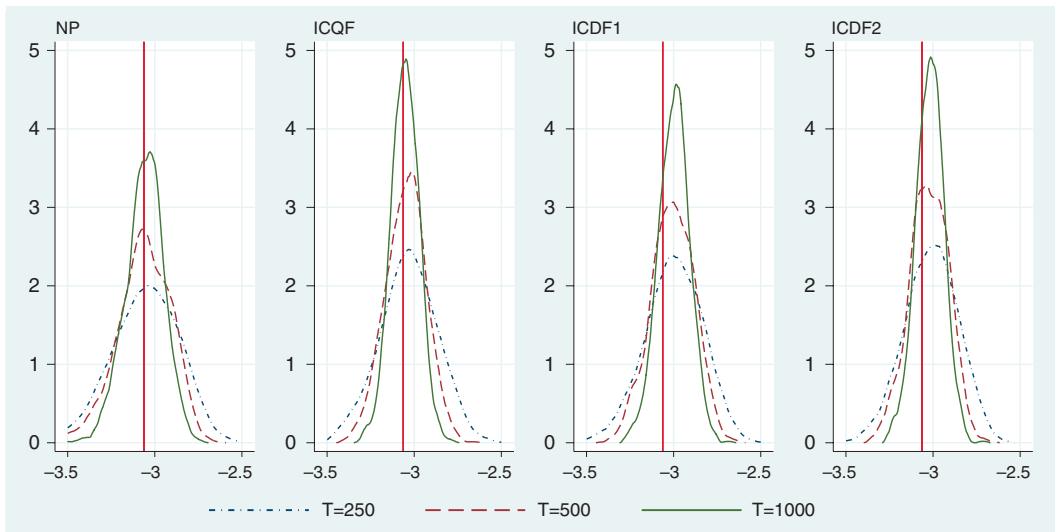

Figure 6. Monte Carlo densities of alternative estimators of $\tau(\alpha|x)$ in the heteroskedastic model $Y_t|X_t \sim \mathcal{N}(-1+X_t, (1+0.25X_t)^2)$ and $X_t \sim \mathcal{N}(0, 1)$, for $\alpha = 5\%$ and $x = 0$ (the median of X_t). The densities are based on 1000 samples of size $T = 250, 500, 1000$.

on the S&P 500 (the logarithmic difference in the index plus dividend payments) and the return on a 3-month U.S. money market instrument issued by JPMorgan (the logarithmic difference in its price), and computed excluding weekends and holidays. The set of predictors includes both

Table IV. Data sources.

Code	Description	Source
MKUTR	S&P500 total return index	Bloomberg
RUSTR	3m U.S. cash total return	Thomson Datastream, Bloomberg
OIL	Oil Nymex future price (first contract)	Bloomberg
COMMDTY	Goldman Sachs non-energy index	Bloomberg
DYUS	S&P500 equity dividend yield	Bloomberg
RYUS	Lehman U.S. aggregate Baa yield	Thomson Datastream, Bloomberg
TB10Y	U.S. generic govt 10 year yield	Bloomberg
TB3M	U.S. treasury bill 90 days yield	Bloomberg

real and financial variables. The real variables include the price of oil futures and a price index of non-energy commodities. The financial variables include the risk spread (the yield difference between a Lehman U.S. aggregate Baa bond and a U.S. Government 10-year bond), the term spread (the yield difference between a U.S. Government 10-year bond and a U.S. Treasury 90-day bill), and the dividend yield (the weighted average of the dividend per share on the stocks entering the S&P 500). This set of predictors has been chosen to include a broad mix of macro and micro indicators. The prices of basic materials carry information on the cost of industrial inputs, the risk spread and the dividend yield carry information on the risk premium and companies' profitability, whereas the term spread embodies expectations on short- and long-term inflation. All predictors are measured as of the end of the day. For the 3-month U.S. money market instrument issued by JPMorgan and the Lehman U.S. aggregate Baa bond, we use the Thomson Datastream series that have been concatenated backwards starting with 31 December 1996 and 30 June 1998, respectively. All other data are from Bloomberg.

Data sources, variable transformations, and summary statistics of the transformed data, namely the mean, the standard deviation (SD), the 1st percentile ($Q_{.01}$) and the 99th percentiles ($Q_{.99}$), are presented in Tables IV and V.

Figure 7 presents the transformed data. The upper left panel plots the dependent variable, the daily excess return on the US stock market. Note that the sample period is long enough to include the bull market of the second half of the 1990s, the bear market between 2000 and 2003, and the post-2003 period.

5.2. Empirical results

We estimate the model repeatedly using rolling windows of $T=499$ observations. The first of these windows goes from 4 January 1995, when all transformed variables are available, to 10 July 1997. In total, we have 1460 windows. All predictors are lagged one period. For each rolling window, we use the estimated model and the last available value of the predictors to estimate the expected shortfall over the next day. We call this estimate the one-step ahead predicted shortfall.

Table VI summarizes the results obtained using the unconditional estimator UC, and our three conditional estimators, namely ICQF, ICDF1 and ICDF2. It presents the mean, the standard deviation, and the 1st and 99th percentiles of the empirical distribution over our 1460 rolling windows of the one-step ahead predicted shortfall (expressed in percentage points) for $\alpha=5\%$. The standard deviation is significantly lower for the unconditional estimator than for the conditional estimators. This reflects the smoothness of the unconditional estimator, whose value is affected

Table V. Transformations and summary statistics of the data.

| Variable	Description	Transformation	Mean	SD	$Q_{.01}$	$Q_{.99}$
USR	S&P500 excess return	$\ln(MKUTR_t/MKUTR_{t-1}) - \ln(RUSTR_t/RUSTR_{t-1})$	2.3×10^{-4}	0.011	-0.029	0.033
UOIL	Oil price log diff	$\ln(OIL_t/OIL_{t-1})$	5.6×10^{-4}	0.023	-0.059	0.053
UCOMM	Commodity price log diff	$\ln(COMMDTY_t/COMMDTY_{t-1})$	0.8×10^{-4}	0.006	-0.014	0.013
URS	Risk spread	$RYUS_t - TB10Y_t$	1.556	0.667	0.707	3.054
UDY	Dividend yield	$\ln(DYUS_t)$	0.427	0.259	0.044	1.008
USP	Term Spread	$TB10Y_t - TB3M_t$	1.544	1.120	-0.572	3.612
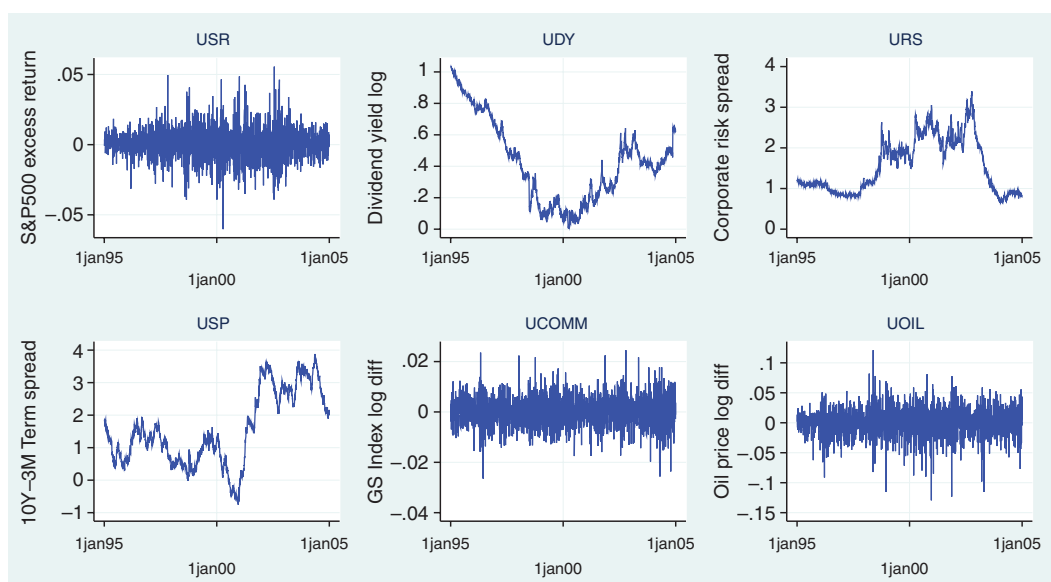
Figure 7. Transformed daily data between January 4, 1995, and December 31, 2004.

Table VI. Summary statistics of the empirical distribution over 1460 rolling windows of the one-step ahead predicted shortfall (expressed in percentage points) for alternative estimators of the expected shortfall.

Estimator	Mean	SD	$Q_{.01}$	$Q_{.99}$
UC	-2.560	0.360	-2.979	-1.603
ICQF	-2.481	0.844	-4.387	-0.978
ICDF1	-2.448	1.764	-6.540	-1.202
ICDF2	-2.302	1.284	-5.984	-1.199

only marginally by the accrual of new information as the rolling window changes. On the other hand, by construction, the conditional estimators are much more sensitive to short-run variations in the predictors.
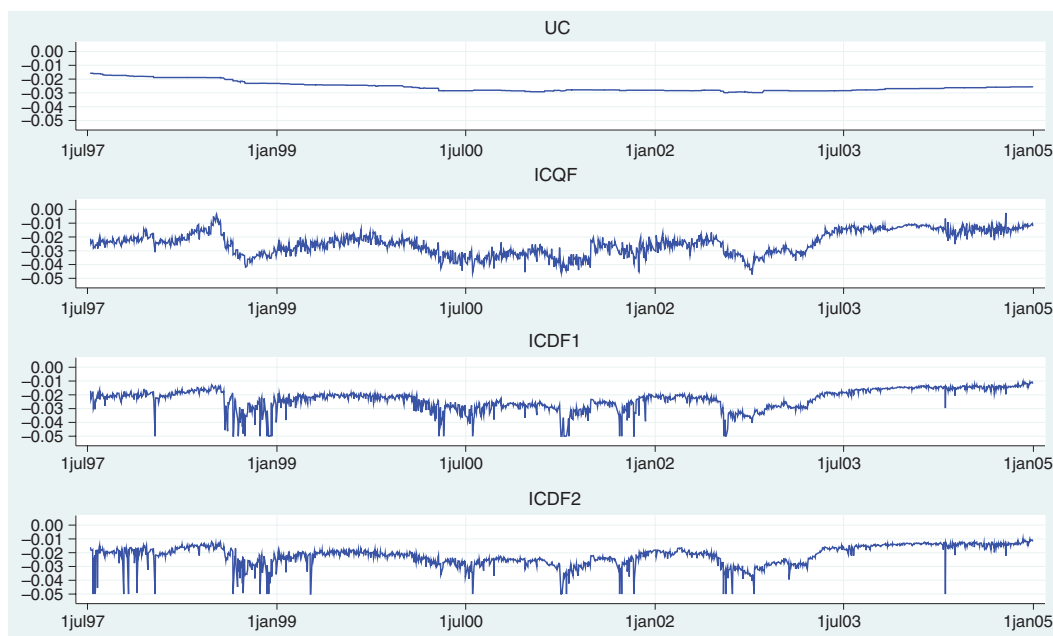
Figure 8. One-step ahead predicted shortfall based on the unconditional estimator $\hat{\tau}(\alpha)$ (denoted by UC) and three conditional estimators, ICDF, ICQF1 and ICQF2, with $\alpha = 5\%$.

Figure 8 presents the time-series plot of the one-step ahead predicted shortfall based on the unconditional estimator UC and the three conditional estimators ICQF, ICDF1 and ICDF2. In a few cases (29 for the ICDF1 estimator and 23 for the ICDF2 estimator out of a total of 1460), the one-step ahead predicted shortfall is negative and larger than 5% in absolute value. To avoid scale problems, in Figure 8, we censor these values at -5%. A pattern that is common to all conditional estimators is that they follow the observed volatility clustering of the financial returns and tend to exhibit persistence in sub-intervals. Periods when the one-step ahead predicted shortfall is particularly high are the second semester of 1998 and 2001, respectively, the last quarter of 2002 and the first quarter of 2003. Among the conditional estimators, the ICDF2 is particularly sensitive to the market index swings of the second semester of 1997.

To assess the predictive accuracy of our estimators, we follow McNeil and Frey [15] who propose a formal test of the hypothesis that a particular method provides an unbiased predictor of the expected shortfall. Their test is based on the idea that, under the null hypothesis, the one-step ahead prediction error (defined as the difference between the observed excess return between T and $T+1$ and the one-step ahead predicted shortfall) should have mean zero under quantile violation, that is, in cases when the observed excess return is lower than the VaR at level α. The test rejects the null hypothesis whenever the average one-step ahead prediction error is large, the average being taken over all quantile violation cases.

We depart from McNeil and Frey [15] because we do not formally test the null hypothesis, but simply compare summaries of the empirical distribution of the one-step ahead prediction error for the various estimators under quantile violation. We estimate the VaR unconditionally by the order statistic $Y_{[\alpha T]}$, where $T = 499$ is the number of observations in each rolling window, and

Table VII. Summary statistics of the empirical distribution of the one-step ahead prediction error (expressed in percentage points) in the quantile violation cases for alternative estimators of the expected shortfall.

Estimator	Np. obs.	Mean	SD	$Q_{.01}$	$Q_{.99}$
UC	77 (5.3%)	−0.052	0.670	−3.346	0.698
ICQF	89 (6.1%)	−0.033	0.718	−3.366	1.699
ICDF1	89 (6.1%)	0.022	0.776	−3.565	2.076
ICDF2	89 (6.1%)	−0.059	1.410	−3.805	8.601

conditionally using the linear quantile regression estimator of Koenker and Bassett [12]. Table VII shows, for each estimator considered, the number and percentage of quantile violations (out of 1460 cases) and summaries of the empirical distribution of the one-step ahead prediction error—the mean, the standard deviation, and the 1st and 99th percentiles—under quantile violation. The standard deviation and the difference between the 99th and the 1st percentile (another measure of variability) are smallest for the unconditional estimator. However, the mean prediction error for this estimator (under quantile violation) is larger in absolute value than for the conditional estimators, except possibly ICDF2. Among the conditional estimators, the mean prediction error is smallest in absolute value for the ICDF1 estimator, whereas the variability of the prediction error is smallest for the ICQF estimator. Note that the mean prediction error is negative (underprediction of the loss) for the ICQF and the ICDF2 estimators, and positive (overprediction of the loss) for the ICDF1 estimator. Also note that, for the two ICDF estimators, the difference between the 99th and the 1st percentile tends to be large due to some extreme negative estimates.

6. SUMMARY AND CONCLUSIONS

In this paper we have extended the concept of expected shortfall to the important case when auxiliary information about the outcome of interest is available in the form of a set of predictors. Our starting points are two equivalent representations of the α-level expected shortfall. In the unconditional case, the two representations lead to the same estimator, namely an average of the smallest sample-order statistics. In the conditional case, instead, they may lead to two alternative classes of estimators, labelled integrated conditional quantile function (ICQF) and integrated conditional distribution function (ICDF) estimators. One advantage of the class of ICDF estimators is that we can more easily impose monotonicity of the estimated cdf and therefore avoid the quantile crossing problem that one may encounter with the class of ICQF estimators. We also consider a simple class of fully non-parametric estimators that consist of local versions of the unconditional estimator. The main drawback of this class of estimators is that they suffer of the curse-of-dimensionality problem.

We study the properties of the proposed estimators through a set of Monte Carlo experiments and through an empirical application using financial data. The Monte Carlo experiments show that accuracy of the estimators increases rapidly with the level α and the sample size. For moderate and large sample sizes (500 or 1000 observations) and values of α equal to 5 and 10%, the Monte Carlo distribution of the estimators looks approximately normal, supporting the conjecture of asymptotic normality of these estimators. The behavior of the conditional (ICQF and ICDF) estimators is very similar for central values of the predictors, but tends to differ for extreme values, in a way that depends on the underlying model.

In our empirical application, the predictive performance of the various estimators is assessed by analyzing the distribution of the one-step ahead prediction errors based on a sequence of rolling samples. Overall, the conditional estimators, and especially the ICQF estimator, tend to have a better performance than the unconditional estimator.

REFERENCES

1. Artzner P, Delbaen F, Eber J-M, Heath D. Coherent measures of risk. *Mathematical Finance* 1999; **9**:203–228.
2. Delbaen F. Coherent risk measures on general probability spaces. *Advances in Finance and Stochastics*. Springer: Berlin, 2002; 1–37.
3. Acerbi C, Tasche D. On the coherence of expected shortfall. *Journal of Banking and Finance* 2002; **26**:1487–1503.
4. Bassett GW, Koenker R, Kordas G. Pessimistic portfolio allocation and Choquet expected utility. *Journal of Financial Econometrics* 2004; **2**:477–492.
5. Koenker R. *Quantile Regression*. Cambridge University Press: New York, 2005.
6. Peracchi F. On estimating conditional quantiles and distribution functions. *Computational Statistics and Data Analysis* 2002; **38**:433–477.
7. Hirsch MW, Smale S. *Differential Equations*, *Dynamical Systems*, *and Linear Algebra*. Academic Press: New York, 1794.
8. Embrechts P, Klüppelberg C, Mikosch T. *Modelling Extremal Events*. Springer: Berlin, 1997.
9. Shorrocks AF. Ranking income distributions. *Econometrica* 1983; **50**:3–17.
10. McLachlan G, Peel D. *Finite Mixture Models*. Wiley: New York, 2000.
11. Csörgö S, Haeusler E, Mason DM. The asymptotic distribution of extreme sums. *The Annals of Probability* 1991; **19**:783–811.
12. Koenker R, Bassett G. Regression quantiles. *Econometrica* 1978; **46**:33–50.
13. Angrist J, Chernozhukov V, Fernández-Val I. Quantile regression under misspecification, with an application to the U.S. wage structure. *Econometrica* 2006; **74**:539–563.
14. Dodge J, Jurečková J. *Adaptive Regression*. Springer: Berlin, 2000.
15. McNeil AJ, Frey R. Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance* 2000; **7**:271–300.