PhD dissertation: Expected Shortfall estimators and their use in asset allocation

Andrei Valentin Tanase
Tor Vergata University, Rome

June 16, 2010

Abstract

The Expected Shortfall (ES) is a risk measure that averages out all losses more severe than the Value at Risk (VaR). As the ES shares the properties of coherent risk measures, its use as risk constraint in asset allocation has became relevant. First of all, we propose estimators for ES, considering the important case when additional information as some set of regressors is available. The estimators are based on the equivalent representation of ES in terms of the conditional distribution function and the conditional quantile function. Within the estimation framework, departing from a generalized weighted representation of ES, we work on improving the statistical and forecasting properties of the weighted estimators. In the first case, we derive the weighting that minimizes the asymptotic variance of the estimators, while, in the second case, the weighting minimizes some suitably defined forecast error. Nevertheless we are concerned with the use of these estimators in financial applications and construct a simple asset allocation model that maximizes expected return with a loss constraint on ES.

Acknowledgements

Acknowledgements go to my PhD supervisors Professor Franco Peracchi and Dr. Samantha Leorato. Moreover, I received valuable suggestions from Professors Alessandro Penati, Frank Vella, Karim Abadir, Rafaelo Seri, Jana Jurečková, Ioanid Rosu, Stefano Herzel and participants at seminars and workshops where parts of the thesis have been presented (the SMAP Workshop 2006 - Cassino, IX Workshop in Quantitative Finance 2008 - Rome and the Risk panel at EAA-ESEM 2009 - Barcelona. Nevertheless, I thank the other PhD Professors for sharing their knowledge and my colleagues for their continuous support.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 The Expected Shortfall (ES)

Next to the Value at Risk (VaR), the expected value of the returns on the left tail of the distribution has been proposed as an alternative measure of risk in financial applications. This quantity, known as the expected shortfall (ES) 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 Y_t be a real-valued random variable (rv) that represents the returns on a given asset during some period. Assume that it has a continuous and strictly increasing distribution function (df) $F(y) = \Pr\{Y_t \leq y\}$. Its quantile function (qf) is defined as $Q(p) = \inf\{y \colon F(y) \geq p\}$, 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 $F(y) = Q^{-1}(y)$. Then the VaR at level α is equal to the αth quantile of Y_t, $Q(\alpha)$, and the α-level expected shortfall is $\tau(\alpha) = \mathrm{E}(Y_t \,|\, Y_t \leq Q(\alpha))$, where $\mathrm{E}(Y_t \,|\, A_t)$ denotes the conditional expectation of Y_t given the event A_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.

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

$$\mathrm{E}(Y_t \,|\, Y_t \leq c) = \frac{1}{F(c)} \int_{-\infty}^{c} y \, \mathrm{d}F(y). \tag{1.1}$$

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

$$\tau(\alpha) = \mathrm{E}(Y_t \,|\, Y_t \leq Q(\alpha)) = \frac{1}{\alpha} \int_{-\infty}^{Q(\alpha)} y \, \mathrm{d}F(y), \tag{1.2}$$

where $Q(\alpha)$ is the αth quantile of Y_t.

Unlike the VaR, the expected shortfall takes into account all possible losses that exceed the severity level corresponding to the VaR. As shown by Acerbi and Tasche (1), this enables ES to satisfy the properties of a coherent risk measure: sub-additivity, monotonicity, positive homogeneity, and translation invariance (see Artzner et al. (7) and Delbaen (16)). Formally, 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) $\tau_X(\alpha) \leq 0$, if $X \leq 0$ (monotonicity), (ii) $\tau_{X+Y}(\alpha) \geq \tau_X(\alpha) + \tau_Y(\alpha)$ (sub-additivity), (iii) $\tau_{bX} = b\,\tau_X(\alpha)$, for $b \geq 0$ (positive homogeneity), and (iv) $\tau_{X+c}(\alpha) = \tau_X(\alpha) + c$, for $c \in \mathbb{R}$ (translation invariance). Besides being a coherent risk measure, the expected shortfall is continuous with respect to α regardless of the underlying distribution of Y_t, and therefore it is not too sensitive to small changes in α.

Acerbi and Tasche (1) also show that sub-additivity may be violated by the Value at Risk (VaR), therefore VaR is not a coherent risk measure. However, it is still widely used in financial applications, because it gives a lower bound on the loss made in the worst α percent of the cases during some period. The practical usefulness of the VaR is however limited by the fact that it does not account for the magnitude of the worst-case scenario, giving the same importance to all losses that are more severe than itself.

Alternative representations of ES

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

$$\mathrm{E}(Y_t \,|\, Y_t \leq c) = \frac{1}{F(c)} \int_{F(-\infty)}^{F(c)} F^{-1}(p)\,\mathrm{d}p = \frac{1}{F(c)} \int_0^{F(c)} Q(p)\,\mathrm{d}p.$$

Thus we have the equivalent representation

$$\tau(\alpha) = \frac{1}{\alpha} \int_0^{\alpha} Q(p)\,\mathrm{d}p. \tag{1.3}$$

This representation is particularly convenient when the quantiles of a 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 $\mu \in \mathbb{R}$ and $\sigma > 0$, where U_t is a 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

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

where $\tau^*(\alpha) = \alpha^{-1} \int_0^{\alpha} \zeta(p)\,\mathrm{d}p$ 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 e.g. Hirsch and Smale (22)) 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 (1.1) by parts, we get

$$\mathrm{E}(Y_t \,|\, Y_t \leq c) = c - \int_{-\infty}^{c} \frac{F(y)}{F(c)}\,\mathrm{d}y.$$

Therefore

$$\tau(\alpha) = Q(\alpha) - \frac{1}{\alpha} \int_{-\infty}^{Q(\alpha)} F(y)\,\mathrm{d}y. \tag{1.4}$$

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

Examples

We illustrate computation of the expected shortfall for 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

$$\mathrm{E}(Y_t \,|\, Y_t \leq c) = \mu + \sigma\,\mathrm{E}(U_t \,|\, U_t \leq c^*) = \mu - \sigma\,\frac{\phi(c^*)}{\Phi(c^*)},$$

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

$$\tau(\alpha) = \mu - \frac{\sigma}{\alpha}\,\phi(\Phi^{-1}(\alpha)). \tag{1.5}$$

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 (1.4), 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. Any continuous distribution may be approximated arbitrarily well by a mixture of J normal distributions (see e.g. McLachlan and Peel (29)). 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 [1]. In this case, the df of 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 (1.1) becomes

$$\mathrm{E}(Y_t \,|\, Y_t \leq c) = \theta(c)\,\mu_1(c) + [1 - \theta(c)]\,\mu_2(c),$$

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.

[1] The result presented here is easily generalized to the case when Y_t is a mixture of $J \geq 2$ normals with mixing probabilities $\pi_1, \ldots, \pi_J$ that are positive and add up to one.

Relationship between ES and 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

$$e(c) = \mathrm{E}(Y_t - c \,|\, Y_t \leq c) = \mathrm{E}(Y_t \,|\, Y_t \leq c) - c.$$

This quantity is an important tool in financial risk management and in various other fields, such as medicine (see Embrechts et al (18), 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, nonzero mean μ. The Lorenz curve is defined as

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

and so, from (1.3),

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

The generalized Lorenz curve (Shorrocks (38)) is the Lorenz curve scaled up by the mean and is equal to

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

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

1.2 Economic motivation

We address the important problem of estimation of the expected shortfall when auxiliary information about asset returns Y_t is provided by a set of predictors X_t that represents the information available at time $t-1$. Lags of Y_t might also be included in X_t.

Assume that the distribution of Y_t conditional on the set of predictors X_t is continuous and strictly increasing. Moreover, let $F(y \,|\, x) = \Pr\{Y_t \leq y \,|\, X_t = x\}$ and $Q(p \,|\, x) = \inf\{y \colon F(y \,|\, x) \geq p\}$ be, respectively, the conditional distribution function (cdf) and the conditional quantile function (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 α-level conditional expected shortfall of Y_t given $X_t = x$ is

$$\begin{aligned} \tau(\alpha \,|\, x) &= \frac{1}{\alpha} \int_{-\infty}^{Q(\alpha \,|\, x)} y\,\mathrm{d}F(y \,|\, x) \\ &= \frac{1}{\alpha} \int_0^\alpha Q(p \,|\, x)\,\mathrm{d}p \\ &= Q(\alpha \,|\, x) - \frac{1}{\alpha} \int_{-\infty}^{Q(\alpha \,|\, x)} F(y \,|\, x)\,\mathrm{d}y. \end{aligned} \tag{1.6}$$

In some cases, if the cdf is known, one may compute the expected shortfall analytically. 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 (1.5), $\tau(\alpha \,|\, x) = \mu(x) - \frac{\sigma(x)}{\alpha}\,\phi(\Phi^{-1}(\alpha))$. Still, we can incur situations when either the integral that defines the expected shortfall is hard to compute or we do not even know the shape of the conditional distribution function of Y_t.

In Chapter 2, that is mainly based on the work of Peracchi and Tanase (2008), we introduce two classes of analog estimators based on two alternative representations of the conditional ES, either as an integral of the cqf of Y_t given X_t or as an integral of the cdf of Y_t given X_t.

Then, departing from a generalized weighted representation of the ES, we propose weighted versions of estimators for ES and aim to improve their statistical and forecasting properties. Following Leorato et al (2009), we work out the asymptotic distribution of two of the conditional estimators and analytically derive the weighting that minimizes their asymptotic variance. As for the second objective, the weighting is numerically derived such as to minimize a suitably defined forecast error of the weighted estimators for ES. The methodologies are detailed in Chapter 3 and Chapter 4 respectively. Our results are supported by sets of Monte Carlo experiments and illustrated in empirical applications on real data.

Moreover, in Chapter 4, we construct an asset allocation model that maximizes expected return with a constraint on either of the risk measures that we have discussed: ES, weighted ES and VaR. The model does not fully exploit the convexity of the optimization problem, but it is solved by a numerical algorithm. Using real daily data, we develop an empirical application and compare various performance indicators and weights stability measures of the optimal portfolios.

Chapter 2

Estimation of the expected shortfall

This chapter is based on the work of Peracchi and Tanase (36) that consider alternative approaches to estimation of the α-level ES $\tau(\alpha)$ and the α-level conditional expected shortfall $\tau(\alpha \,|\, x)$. We basically exploit the equivalence between (1.2), (1.3) and (1.4) in the unconditional case, and the equivalence (1.6) in the conditional case. The equivalent representations open the way to estimation of the expected shortfall by replacing the population conditional quantile function (cqf) and the population conditional distribution function (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 conditional quantile function is assumed to be linear in parameters. Here we may face the problem of quantile crossing, i.e. 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, we have an approach that naturally impose monotonicity of the estimated cdf and avoid the quantile crossing problem that might arise with the first representation.

2.1 Unconditional estimators

Given a random sample $Y_1, \ldots, Y_T$ from a distribution with df F and qf Q, the ES $\tau(\alpha)$ may simply be estimated by replacing F and Q in (1.2) and (1.4) by their empirical counterparts, namely the empirical distribution function (edf) $\hat{F}$, defined on the real line by

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

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

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

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

$$\hat{\tau}_I^{(Q)}(\alpha) = \frac{1}{\alpha} \int_0^{\alpha} \hat{Q}(p)\, \mathrm{d}p = \hat{Q}(\alpha) - \frac{1}{\alpha} \int_{-\infty}^{\hat{Q}(\alpha)} \hat{F}(y)\, \mathrm{d}y = \frac{1}{\alpha T} \sum_{t=1}^{[\alpha T]} Y_{(t)} + \left(1 - \frac{[\alpha T]}{\alpha T}\right) Y_{([\alpha T]+1)}, \tag{2.1}$$

where $[\alpha T]$ denotes the integer part of αT. Thus, $\hat{\tau}_I^{(Q)}(\alpha)$ 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 $\tau(\alpha) = (\alpha T)^{-1} \sum_{t=1}^{[\alpha T]} Y_{(t)}$. Notice that this estimator coincides with the maximum likelihood estimator under the assumption that the distribution of Y_t conditional on $Y_t \leq Q(\alpha)$ is exponential on the negative half-line.

The study of the asymptotic distribution of estimators of the form (2.1) has been carried out by Csörgö et al. (14). In particular, they provide necessary and sufficient conditions for $\hat{\tau}_I^{(Q)}(\alpha)$ 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 (2.1) are satisfied (see their Corollary 1), then

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

as $T \to \infty$, where

$$AV(\alpha) = \int_0^\alpha \int_0^\alpha [\min(s,t) - st] \, \mathrm{d}Q(s)\mathrm{d}Q(t). \tag{2.2}$$

In fact, Csörgö et al. (14) establish, more generally, the asymptotic properties of weighted sums of extreme order statistics of the form

$$\hat{\tau}_I^{(D)}(\alpha) = \sum_{i=1}^{I} w_i Y_{(i)}, \tag{2.3}$$

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 $I/T \to \alpha$.

2.2 Conditional estimators

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 $F(y \,|\, x)$.

2.2.1 Non-parametric estimators

A simple class of fully non-parametric estimators of $\tau(\alpha \,|\, x)$ are local versions of (2.1), that is, averages of the smallest order statistics over a neighborhood of x defined by a suitably defined kernel function $K(\cdot)$. This corresponds to the class of estimators of the form

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

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 (1.6) suggests two classes of analog estimators of $\tau(\alpha \,|\, x)$, namely

$$\hat{\tau}_I^{(Q)}(\alpha \,|\, x) = \frac{1}{\alpha} \int_0^\alpha \hat{Q}(p \,|\, x) \, \mathrm{d}p \tag{2.4}$$

and

$$\hat{\tau}_I^{(D)}(\alpha \,|\, x) = \hat{Q}(\alpha \,|\, x) - \frac{1}{\alpha} \int_{-\infty}^{\hat{Q}(\alpha \,|\, x)} \hat{F}(y \,|\, x) \, \mathrm{d}y, \tag{2.5}$$

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 refer to estimators based on (2.4) as integrated cqf (ICQF) estimators and to estimators based on (2.5) as integrated cdf (ICDF) estimators. Unlike the unconditional case, one cannot generally guarantee that $\hat{Q}(p \,|\, x) = \inf\{y \colon \hat{F}(y \,|\, x) \geq p\}$, $p \in (0,1)$. Hence, the two classes of estimators need not coincide.

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

2.2.2 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 (24), who proposed estimating $\beta(p)$ by solving

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

where

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

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. (6)). 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

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

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 $0 < p_1 < \cdots < p_I \leq \alpha$. To guarantee consistency of this estimator, I should be required to increase with the sample size T.

As for the asymptotic behavior of estimators of this type, as we show in Chapter 3, being linear combinations of asymptotically normal estimators, they are also asymptotically normal. The Monte Carlo evidence in Section 2.3 provides support for our results.

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 (Dodge and Jurečková (17), 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.

2.2.3 ICDF estimators

In order to estimate $F(y \,|\, x)$, we follow the approach in Peracchi (35). We select J distinct values $y_1, \dots, y_J$ such that $Y_{(1)} < y_1 < \cdots < y_J < Y_{(T)}$, and define the log-odds

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

where $F_j(x) = F(y_j \,|\, x) = \Pr\{Y_t \le y_j \,|\, x\}$. Because each rv $\mathbb{1}\{Y_t \le y_j\}$ 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

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

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

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

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

where $\omega_1, \dots, \omega_I$ 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

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

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 satisfy the condition that $\hat{F}_j(x) \geq \hat{F}_{j-1}(x)$ for all x. A simple way of imposing monotonicity is to exploit the fact that

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

where

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

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 $Y_t \geq y_{h-1}$, $h = 2, \ldots, J$. Given $\tilde{F}_1(x)$ 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

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

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

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

As for the sampling properties of these two classes of ICDF estimators, in this Chapter we again confine ourselves to the evidence presented in the next section. Later on, in Chapter 3, we work out the asymptotic distribution of the non-monotonic class of ICDF estimators.

2.3 Monte Carlo evidence

We present some Monte Carlo evidence on the sampling properties of the unconditional estimator $\hat{\tau}_I^{(Q)}(\alpha)$, the fully non-parametric estimator $\bar{\tau}(\alpha \,|\, x)$, the ICQF estimator $\hat{\tau}_I^{(Q)}(\alpha \,|\, x)$, and the two ICDF estimators $\hat{\tau}_I^{(D)}(\alpha \,|\, x)$ and $\hat{\tau}_I^{(D*)}(\alpha \,|\, x)$. We consider sample 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.

2.3.1 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 $\hat{\tau}_I^{(Q)}(\alpha)$ are given in Table 2.1. 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 $\hat{\tau}_I^{(Q)}(\alpha)$, 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). We assume the following distributions: $Y_t \sim \mathcal{N}(0,1)$, $Y_t \sim 0.8 * \mathcal{N}(0,1) + 0.2 * \mathcal{N}(0,2)$, $Y_t \sim t[2]$ and $Y_t \sim t[4]$ respectively, for α =1%, 5%, 10%. The Monte Carlo distributions are based on 1000 samples of size T = 250, 500, 1000.

Figures 2.1 and 2.2 plot, for each set of parameters, kernel estimates of the Monte Carlo densities, respectively for the case when the parent distribution is normal $Y_t \sim \mathcal{N}(0,1)$ and a normal mixture $Y_t \sim 0.8 * \mathcal{N}(0,1) + 0.2 * \mathcal{N}(0,2)$. The panels in each figure correspond to different values of α and present Monte Carlo densities corresponding to the various sample sizes ($\alpha = 1\%, 5\%, 10\%$, and T = 250, 500, 1000). The densities are based on 1000 samples. 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.

Table 2.1: Summary statistics of the Monte Carlo distribution of $\hat{\tau}_I^{(Q)}(\alpha)$

			$\mathcal{N}(0,1)$							$0.8 * \mathcal{N}(0,1) + 0.2 * \mathcal{N}(0,4)$						
α	T	I	$\tau(\alpha)$	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

			$t[2]$							$t[4]$						
α	T	I	$\tau(\alpha)$	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

Figure 2.1: Monte Carlo densities of the unconditional estimator $\hat{\tau}_I^{(Q)}(\alpha)$ for the standard normal distribution

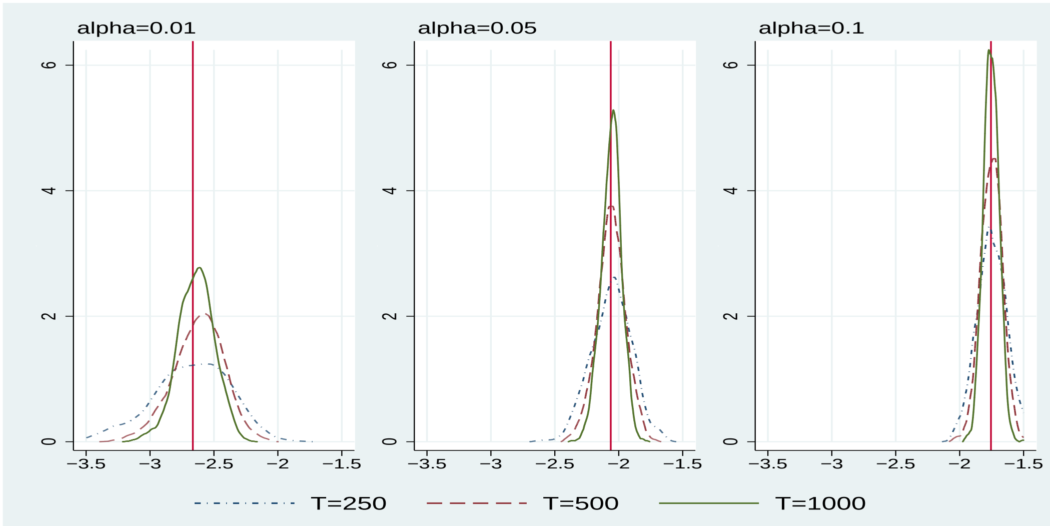
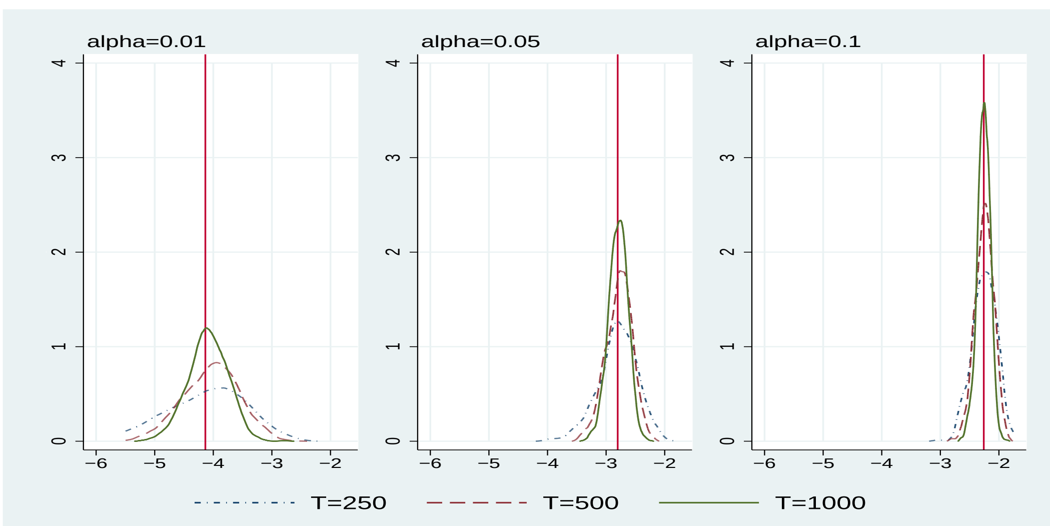


Figure 2.2: Monte Carlo densities of the unconditional estimator $\hat{\tau}_I^{(Q)}(\alpha)$ for the mixture of normals distributions



2.3.2 Estimation of the conditional expected shortfall

For the comparison of alternative estimators of the conditional ES, 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 $\bar{\tau}(\alpha\,|\,x)$, denoted by NP, (ii) the ICQF $\hat{\tau}_I^{(Q)}(\alpha\,|\,x)$, (iii) the ICDF $\hat{\tau}_I^{(D)}(\alpha\,|\,x)$ based on the non-monotonic cdf estimate $\hat{F}(y\,|\,x)$, denoted by ICDF1, and (iv) the ICDF $\hat{\tau}_I^{(D*)}(\alpha\,|\,x)$ based on the monotonic cdf estimate $\tilde{F}(y\,|\,x)$, denoted by ICDF2.

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 $\alpha = .01$. 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 between 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 drop the "failed" experiments (those where the condition is not met) and draw Monte Carlo samples until a predetermined number of 1000 "successful" experiments is 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 2.2 and 2.3, separately 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 2.3 and 2.5 (for the homoskedastic case) and Figures 2.4 and 2.6 (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. Overall, the fully nonparametric 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 are 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.

Table 2.2: Summary statistics of the Monte Carlo distribution of alternative estimators of $\tau(\alpha|x)$, $x = -1.282$ (the 10th percentile of X_t)

Homoskedastic model, $x = -1.282$

			NP			ICQF				ICDF1				ICDF2			
α	T	$\tau(\alpha\|x)$	Bias	SD	RMSE	I	Bias	SD	RMSE	I	Bias	SD	RMSE	I	Bias	SD	RMSE
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, $x = -1.282$

			NP			ICQF				ICDF1				ICDF2			
α	T	$\tau(\alpha\|x)$	Bias	SD	RMSE	I	Bias	SD	RMSE	I	Bias	SD	RMSE	I	Bias	SD	RMSE
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

Table 2.3: Summary statistics of the Monte Carlo distribution of alternative estimators of $\tau(\alpha|x)$, $x = 0$ (the median of X_t)

Homoskedastic model, $x = 0$

			NP			ICQF				ICDF1				ICDF2			
α	T	$\tau(\alpha\|x)$	Bias	SD	RMSE	I	Bias	SD	RMSE	I	Bias	SD	RMSE	I	Bias	SD	RMSE
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, $x = 0$

			NP			ICQF				ICDF1				ICDF2			
α	T	$\tau(\alpha\|x)$	Bias	SD	RMSE	I	Bias	SD	RMSE	I	Bias	SD	RMSE	I	Bias	SD	RMSE
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

Figure 2.3: Monte Carlo densities of alternative estimators of $\tau(\alpha \mid x)$, homoskedastic model, $x = -1.282$ (the 10th percentile of X_t)

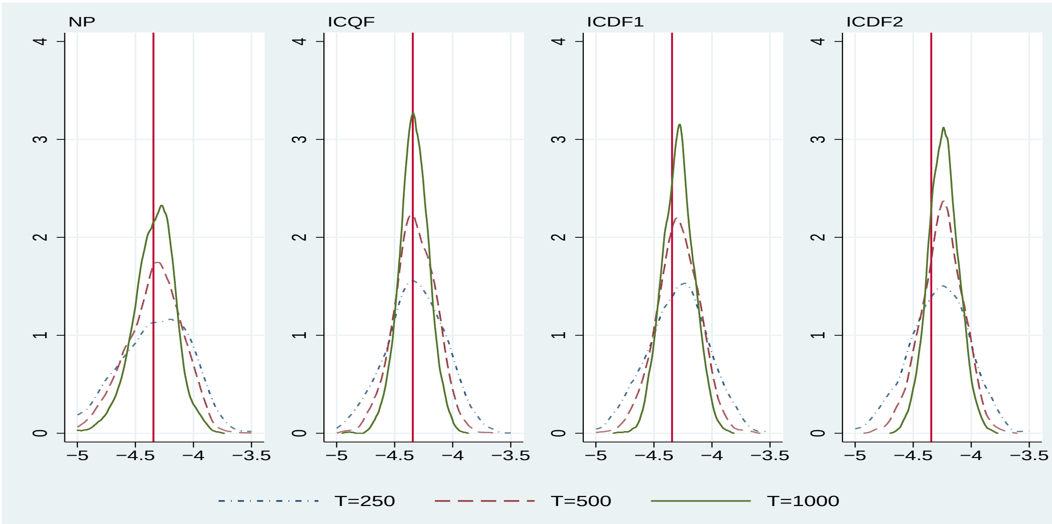


Figure 2.4: Monte Carlo densities of alternative estimators of $\tau(\alpha \mid x)$, heteroskedastic model, $x = -1.282$ (the 10th percentile of X_t)

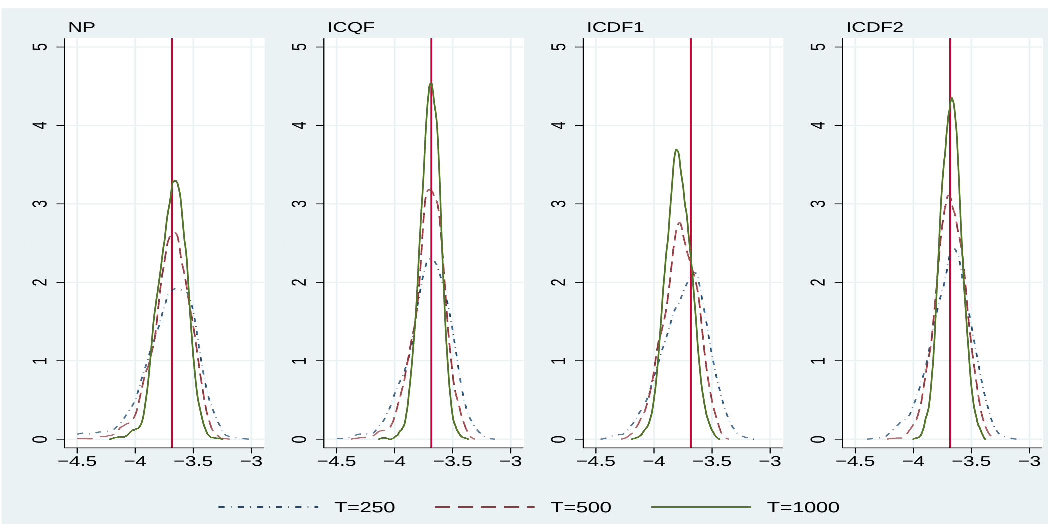


Figure 2.5: Monte Carlo densities of alternative estimators of $\tau(\alpha \mid x)$, homoskedastic model, $x = 0$ (the median of X_t).

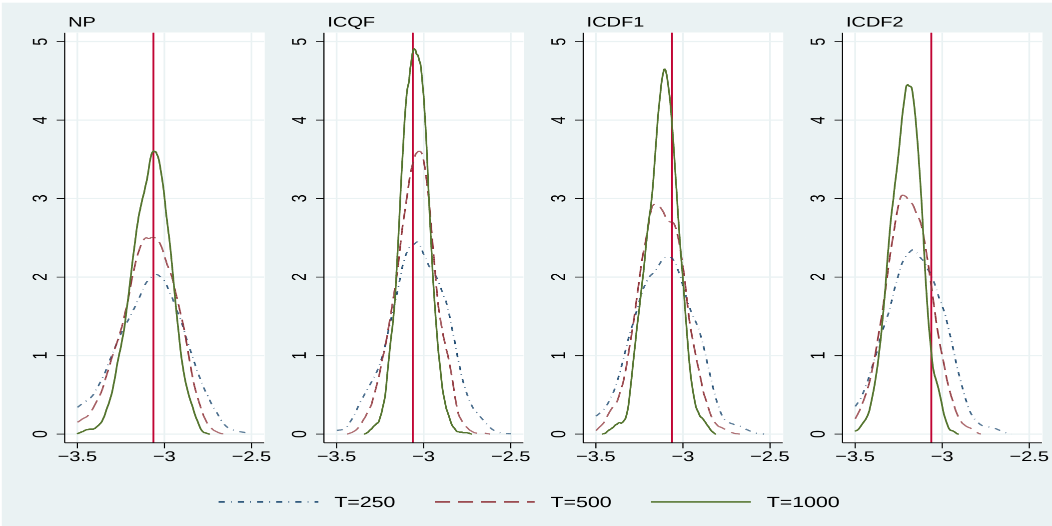
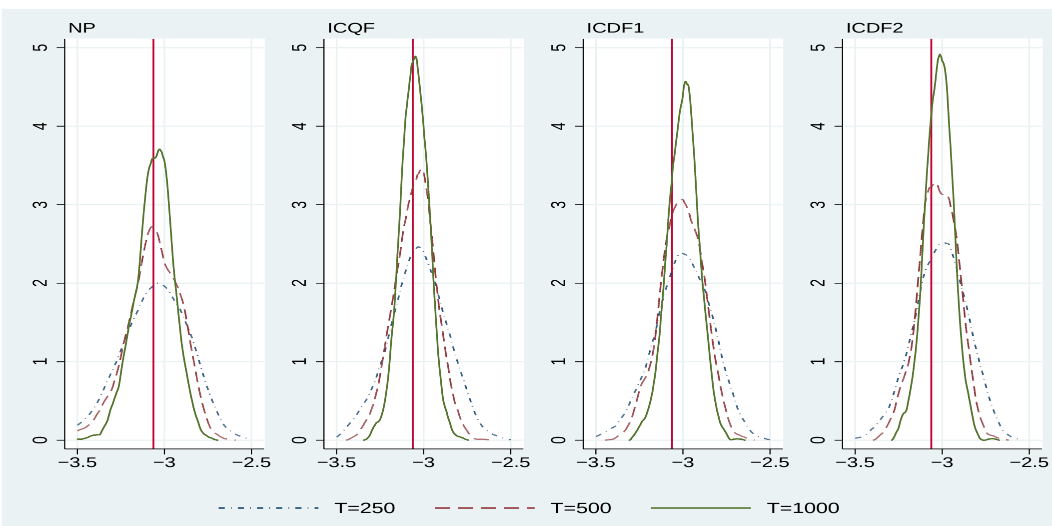


Figure 2.6: Monte Carlo densities of alternative estimators of $\tau(\alpha \mid x)$, heteroskedastic model, $x = 0$ (the median of X_t).



2.4 Empirical application

In this section we present an application on real data. The dependent variable is the daily excess return on the S&P 500 index, while the regressors are real and financial variables. We consider the unconditional estimator $\hat{\tau}_I^{(Q)}(\alpha)$, the ICQF estimator $\hat{\tau}_I^{(Q)}(\alpha \mid x)$ and the ICDF estimators $\hat{\tau}_I^{(D)}(\alpha \mid x)$ and $\hat{\tau}_I^{(D*)}(\alpha \mid x)$, excluding the fully non-parametric estimator $\bar{\tau}(\alpha \mid x)$ because of the curse-of-dimensionality problems due to the high number of predictors.

2.4.1 The data

Our raw data are daily from December 30, 1994, to December 31, 2004. The dependent variable is the daily excess return on the US stock market, defined as the difference between the return on the S&P 500 (the logarithmic difference in the index plus dividend payments) and the return on a 3-month US money market instrument issued by JPMorgan (the logarithmic difference in its price), and computed excluding weekends and holidays. The set of predictors includes both 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 US money market instrument issued by JPMorgan and the Lehman U.S. aggregate Baa bond, we use the Thomson Datastream series which have been concatenated backwards starting with December 31, 1996, and June 30, 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 2.4 and 2.5.

Figure 2.7 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.

Table 2.4: Data sources.

Code	Description	Source
MKU	S&P500 total return index	Bloomberg
RUS	3m US cash total return	Thomson Datastream, Bloomberg
OIL	Oil Nymex future price (1^{st} contract)	Bloomberg
COM	Goldman Sachs non-energy index	Bloomberg
DYUS	S&P500 equity dividend yield	Bloomberg
RYUS	Lehman US aggregate Baa yield	Thomson Datastream, Bloomberg
TB10Y	US generic govt 10 year yield	Bloomberg
TB3M	US treasury bill 90 days yield	Bloomberg

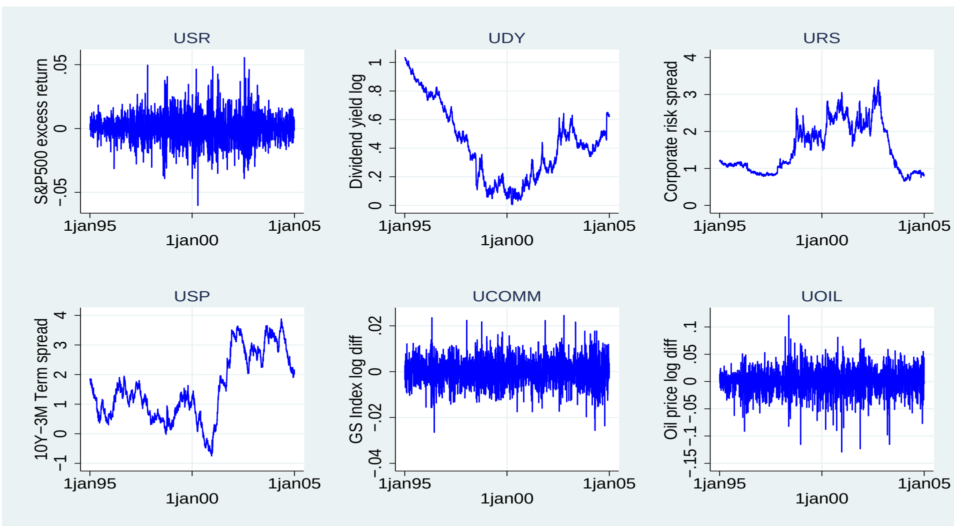
2.4.2 Empirical results

We estimate the model repeatedly using rolling windows of $T = 499$ observations. The first of these windows goes from January 4, 1995, when all transformed variables are available, to July 10, 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

Table 2.5: Transformations and summary statistics of the data.

Variable	Description	Transformation	Mean	SD	$Q_{.01}$	$Q_{.99}$
USR	S&P500 excess return	$\ln(MKU_t/MKU_{t-1}) - \ln(RUS_t/RUS_{t-1})$	$2.3 \cdot 10^{-4}$	0.011	-0.029	0.033
UOIL	Oil price log diff	$\ln(OIL_t/OIL_{t-1})$	$5.6 \cdot 10^{-4}$	0.023	-0.059	0.053
UCOMM	Commodity price log diff	$\ln(COM_t/COM_{t-1})$	$0.8 \cdot 10^{-4}$	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 2.7: Transformed daily data between January 4, 1995, and December 31, 2004.



estimate the expected shortfall over the next day. We call this estimate the one-step ahead predicted shortfall.

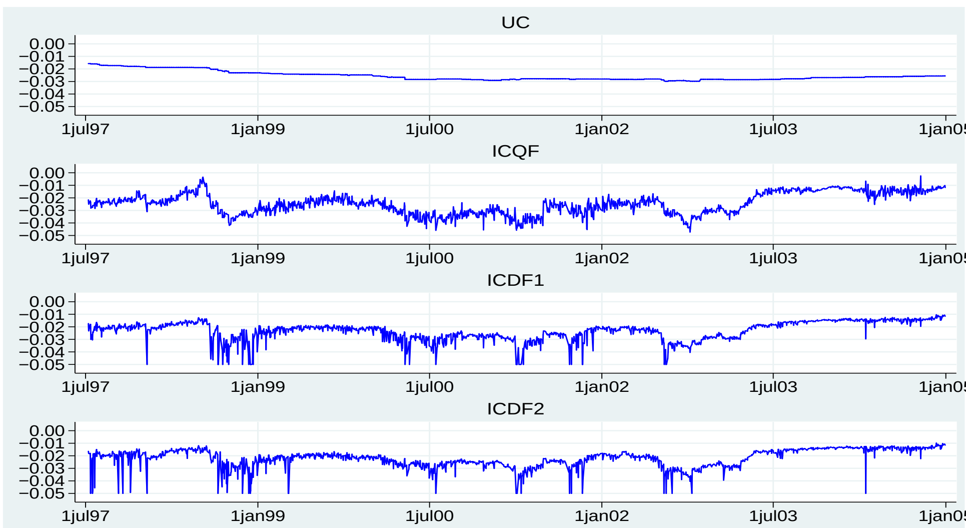
Table 2.6 summarizes the results obtained using the unconditional estimator UC, and our three conditional estimators, namely ICQF, ICDF1 and ICDF2. It presents the mean, the standard deviation, and the 1st and 99th percentiles of the empirical distribution over our 1460 rolling windows of the one-step 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 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 2.8 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 2.8, 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.

Table 2.6: Summary statistics of the empirical distribution of the one-step ahead predicted 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

Figure 2.8: One-step ahead predicted shortfall based on the unconditional and the conditional estimators



To assess the predictive accuracy of our estimators, we follow McNeil and Frey (30) 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 (30) 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 conditionally using the linear quantile regression estimator of Koenker and Bassett (24). Table 2.7 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. Notice that the mean prediction error is negative (under-prediction of the loss) for the ICQF and the ICDF2 estimators, and positive (over-prediction of the loss) for the ICDF1 estimator. Also notice 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.

Table 2.7: Summary statistics of the empirical distribution of the one-step ahead prediction error

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

2.5 Conclusions

We have extended the concept of ES to the important case when auxiliary information about the outcome of interest is available. 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 ICQF and 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 consists of local versions of the unconditional estimator.

The properties of the proposed estimators are studied 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. 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. Overall, the conditional estimators, and especially the ICQF estimator, tend to have a better performance than the unconditional estimator.

Chapter 3

Efficient expected shortfall estimators

3.1 Introduction

In the first chapter we have introduced the Expected Shortfall (ES) that, differently from the Value at Risk (VaR), is a coherent risk measure, meaning that it simultaneously satisfies sub-additivity, monotonicity, positive homogeneity and translation invariance (see Artzer et al (7)). Moreover, the ES is continuous with respect to α irrespective of the distribution of returns. References can be found in Acerbi and Tasche (2), Delbaen (16) and Bertsimas *et al.* (11), among the others.

Following Peracchi and Tanase (36), in Chapter 2, we proposed estimators of ES, both unconditional and conditional on some available set of regressors X_t. The unconditional estimator is a linear combination of extreme order statistics. For the conditional case, a nonparametric and two linear versions that are plug in estimators based on the representation of ES in terms of the conditional distribution function and the conditional quantile function.

In this Chapter, that is mainly based on Leorato et al (26), we focus on generalized analog estimators for ES and improve on their efficiency in terms of asymptotic variance (AV). Our approach consists in generalizing the ES to the Weighted Expected Shortfall (WES) by using a weighting function $W : [0, \alpha] \mapsto [0, 1]$ that maps the original distribution function $F(y) = \Pr\{Y_t \leq y\}$ onto $W(F(y)) = W(\Pr\{Y_t \leq y\})$, derive the asymptotic properties of the weighted estimators and optimize over W such as to minimize the AV of the estimators.

The weighted ES estimators are defined in Section 3.3 as analog consistent estimators of the corresponding Weighted ES (WES). This is obtained as the expectation of the returns, up to a given quantile, according to a distribution modified by some weights. The introduction of weighted quantile risk measures is not new in the literature, they are considered for instance by Acerbi (1), who studies spectral measures of risk and conditions on $W(\cdot)$ that guarantee coherence of WES, and Cherny (13) and Mansini *et al.* (28) that propose portfolio allocation problems with constraints on such measures. Rather than focusing on a new coherent risk measure, alternative to the ES, the purpose of this chapter consists in improving on the asymptotic efficiency of the conditional ES estimators proposed in (36). Moreover, the analysis of the asymptotic properties of the ICDF and ICQF estimators, that is not pursued in the previous chapter is obtained here.

The chapter is structured as follows. In the next Section, we define the so called Weighed Expected Shortfall and also the main conditions on the weighting functions are defined. In

Section 3.3, we propose the class of weighted ES estimators, as a generalization of those introduced in the previous chapter, and we study their asymptotic properties in Section 3.4. Since the WES coincides with the ES for $W(u) = u/\alpha$, and so do the corresponding estimators, then the asymptotic properties of the ICDF and ICQF estimators proposed in the previous chapter are embedded into the results of this section. Section 3.5 and 3.6 present the efficient WES estimators and a study of the gain of asymptotic efficiency that can be attained for different distributions of the returns (heavy tailed or low tailed). Finally, a Monte Carlo study and an application to real data are performed. Some of the proofs are postponed to the Appendix. The full set of proofs can be found in Leorato et al (26).

3.2 The Weighted ES: definition

The Weighted ES can be defined departing from the definition of ES, by modifying the distribution of the returns of an asset/portfolio Y_t according to a given weighting function W. As we shall see in the next section, the definition of the WES allows to construct the weighted ES estimators, as a weighted version of the ES estimators.

Let Y_t be the random variable representing the returns of a given asset or portfolio at time t and let the $k-$dimensional random vector X_t represent a given set of covariates for the rv Y_t. We shall assume that all Y_t's are conditionally independent and identically distributed with d.f. $F_x(y) = F_t(y \,|\, x)$. This is a rather limiting assumptions and an extension to the case where dependence across time is allowed is a priority for further work.

We start by defining the WES, for a given weighting function $W : [0, \alpha] \mapsto [0, 1]$. We consider the following assumptions on the weights:

(A1) $W : [0, \alpha] \mapsto [0, 1]$ is a function in $\mathcal{C}^1[0, \alpha]$ (the class of all functions with continuous first derivative) and satisfies: $W(0) = 0$ and $W(\alpha) = 1$

(A2) $W' = w$ is non-negative everywhere in $[0, \alpha]$.

Definition 1 *Let $F_x(y)$ be the distribution function of the returns Y_t conditional on a set of k covariates $X_t = x$. Then, for every W satisfying (A1) and (A2), the $\alpha-$level weighted ES is defined as*

$$\tau_w(\alpha \,|\, x) = \int_0^{Q_x(\alpha)} y \mathrm{d}W(F_x(y)) \tag{3.1}$$

where $Q_x(\alpha)$ is the conditional αth quantile.

As for the expected shortfall (see (36)), by simply integrating by parts or performing a change of variable, one can write the following chain of equivalent representations for the weighted ES:

$$\tau_w(\alpha \,|\, x) = \int_{-\infty}^{Q_x(\alpha)} y \mathrm{d}W(F_x(y) = \int_0^{\alpha} Q_x(p) w(p) \mathrm{d}p = Q_x(\alpha) - \int_{-\infty}^{Q_x(\alpha)} W(F_x(y)) \mathrm{d}y \tag{3.2}$$

Clearly, the unconditional weighted ES as well as the ES, that we denote by $\tau(\alpha \,|\, x)$, are embedded in Definition 1, corresponding to the cases $X_t = x$ with probability 1 and $w(p) = \frac{1}{\alpha}$ respectively.

Remark 1 *Assumptions (A1) and (A2) are sufficient for (3.1) to be a risk measure satisfying monotonicity, positive homogeneity and translation invariance. The following third condition (see Theorem 4.1 in Acerbi (1)) would moreover guarantee coherency:*

(A3') W is concave on $[0, \alpha]$ (w is a decreasing density function if also (A2) holds).

Since under (A1) and (A2) the cdf $F_x(y)$ is mapped into a new cdf $W(F_x(y))$ the weighted ES can be interpreted as the expected loss one would suffer in the worst $\alpha-$percent cases, if the distribution of the returns changed from F_x into $W(F_x)$. This interpretation of the weighted ES is related to the theory of non-expected utility of Yaari (40), where modifying the distribution of returns accommodates risk aversion of the investor. However, this subjective interpretation of weights goes beyond the scope of the present work and, in fact, assumption (A3') is neglected.

As stated above, every continuous cumulative df $W(\cdot)$ with support $[0, \alpha]$ is suitable to define a (non necessarily coherent) risk measure WES. We conclude this section by the specification of the class $\mathcal{W}$ of *admissible* weighting functions, among which we choose the optimal one, according to the minimum AV criterion.

We can follow two different approaches in order to identify the class $\mathcal{W}$: a parametric or a nonparametric specification.

Because of the form of the objective functions that, as will be shown below, is convex in w, the nonparametric representation for $\mathcal{W}$ corresponds to a convex optimization problem, and therefore in principle outdoes the parametric approach. Nonetheless, the solutions found when the optimization problems are restricted to a parametric class $\mathcal{W}_P$ are almost as good as in the nonparametric case. At the same time the parametric approach has the advantage of incorporating the inequality constraint $w \geq 0$ in the functional form.

For this reasons, although focusing in the nonparametric specification we will also present the parametric approach and make comparisons within the two.

3.2.1 Nonparametric specification

We here allow for $\mathcal{W}$ to contain all weighting functions, without restrictions to a given parametric form.

Besides restrictions (A1) and (A2), since the aim is to estimate ES by means of weighted expected shortfall estimators, then we must account for another constraint, also linear in w:

$$\int_0^\alpha Q_x(p) \left(w(p) - \frac{1}{\alpha} \right) \mathrm{d}p = 0 \tag{3.3}$$

The restriction (3.3), that corresponds to imposing $\tau_w(\alpha \,|\, x) = \tau(\alpha \,|\, x)$, is due to the fact that each weighted ES estimator $\hat{\tau}_w^{(\cdot)}(\alpha \,|\, x)$, introduced in Section 3.3, is be consistent for $\tau_w(\alpha \,|\, x)$ rather than for $\tau(\alpha \,|\, x)$.

Then, the set of weighting functions considered in the nonparametric case is:

$$\mathcal{W} = \left\{ w(p) \geq 0,\, p \in [0, \alpha],\, \int_0^\alpha w(p) dp = 1,\, \int_0^\alpha Q_x(p) \left(w(p) - \frac{1}{\alpha} \right) \mathrm{d}p = 0 \right\} \tag{3.4}$$

In the following, we will also consider the family

$$\mathcal{W}_U = \left\{ w = w(p) : \int_0^\alpha w(p) \mathrm{d}p = 1,\, \int_0^\alpha Q_x(p) \left(w(p) - \frac{1}{\alpha} \right) \mathrm{d}p = 0 \right\}$$

where the inequality constraints deriving from (A2) are ignored. This will be used to define the nonparametric unconstrained version of the efficient estimator. This simplification allows for the explicit derivation of the global optimum of the minimization problem (see Theorems 3 and 4). Anyway, in this case W is no longer a distribution function and therefore the WES associated to W can not be interpreted as a *proper* average of the quantiles up to level α.

3.2.2 Parametric specification

Alternatively to the nonparametric specification, we can choose to restrict the class $\mathcal{W}$ of admissible weighting functions to an appropriately chosen parametric class, depending on a finite number of parameters. The parametric class is embedded into $\mathcal{W}$ and the minimum asymptotic variance one obtains in this case can not be smaller than the minimum in the nonparametric specification.

Nonetheless, as already mentioned earlier, the parametric specification has some advantages. First, defining a parametric class of distribution functions allows us to get rid of the nonnegativity constraints, that are satisfied by construction. Second, although restricting to a subset of $\mathcal{W}$, if the parametric family is large enough we can easily find values of AV that almost achieve those obtained in the nonparametric specification. One important drawback of this approach is given by the fact that, in restricting to the parametric class $\mathcal{W}_P$, we loose the simple structure of the optimization problem, that, although depending on a limited number of parameters, can be solved only computationally.

As far as to the parametric class to choose, we remark that the families of weighting functions corresponding to an S-shaped W, that are usually modelled in the framework of non-expected utility or when accounting for subjective risk attitude, such as, for example, the classes of functions $W(p) = exp\{-\{-\ln p\}^{\beta}\}$, $0 < \beta < 1$ (Prelec (37)) or $W(p) = 1 - (1 - p/\alpha)^b$ (Bassett *et al.* (10)) are not wide enough for our purposes. In fact, all these classes involve a monotone density w either increasing (risk aversion) or decreasing (risk propension), while, in order to attain a significant gain in efficiency the densities w must be allowed to be multimodal. For example, the asymptotic variance of WICQF is a weighted average of the inverse of the density of the returns evaluated at different quantiles (up to level α). Whenever the density is increasing in the left tail, the larger the weights assigned to the highest quantiles, the lower is the AV. On the other hand, the condition $\tau_w(\alpha \,|\, x) = \tau(\alpha \,|\, x)$ requires the distribution of W to be centered around the value $F_x(\tau(\alpha \,|\, x)) \leq \alpha/2$. For all the above reasons, we specify the class $\mathcal{W}_P$ by a mixture of beta distributions, restricted to $[0, \alpha]$, namely

$$w(p) = \pi \frac{\frac{1}{\alpha}\left(\frac{p}{\alpha}\right)^{a_1-1}\left(1-\frac{p}{\alpha}\right)^{b_1-1}}{b(a_1,b_1)} + (1-\pi)\frac{\frac{1}{\alpha}\left(\frac{p}{\alpha}\right)^{a_2-1}\left(1-\frac{p}{\alpha}\right)^{b_2-1}}{b(a_2,b_2)}, \tag{3.5}$$

$(a_i, b_i) \in \Theta, \pi \in [0, 1/2]$, for all $0 \leq p \leq \alpha$, where $b(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ and where Θ is a compact subset of $\mathbb{R}^4$ satisfying $(1,1) \in int\{\Theta\}$.

For all 5−tuple of parameters $(\pi, a_1, b_1, a_2, b_2)$ the non-negativity constraints are satisfied by construction.

It now remains to consider the restriction (3.3), that implies that estimators consistent for the conditional weighted ES are consistent for $\tau(\alpha \,|\, x)$ as well.

Let w_1 and w_2 denote the two components of $w(p) = \pi w_1(p) + (1-\pi) w_2(p)$, and define $\tau_{w_1}(\alpha \,|\, x) = \int_0^\alpha Q_x(p) \mathrm{d}w_1(p)$ and $\tau_{w_2}(\alpha \,|\, x) = \int_0^\alpha Q_x(p) \mathrm{d}w_2(p)$. Once we set the parametrization (3.5), equation (3.3), solved with respect to π, gives the set of solutions:

$$\pi^* = \pi(a_1, b_1, a_2, b_2) = \frac{\tau_{w_2}(\alpha \,|\, x) - \tau(\alpha \,|\, x)}{\tau_{w_2}(\alpha \,|\, x) - \tau_{w_1}(\alpha \,|\, x)}. \tag{3.6}$$

Therefore, in the parametric specification, we will minimize the AV over the subset:

$$\mathcal{W}_P = \{w = w(\pi, a_1, b_1, a_2, b_2),\ (a_i, b_i) \in \Theta, \pi = \pi^*\}. \tag{3.7}$$

3.3 The Weighted ES: estimation

In this section, we present different WES estimators constructed on representation (3.2), with $w(\cdot)$ belonging to $\mathcal{W}$, $\mathcal{W}_U$, $\mathcal{W}_P$. These are generalized weighted versions of those presented in Section 2.2 for the ES. Since our interest is on the ES conditional on a set of regressors, we here focus only on the analog parametric estimators.

The two analog estimators proposed here rely on two different representations of $\tau_w(\alpha \,|\, x)$.

The first estimator is a plug in version of

$$\tau_w(\alpha \,|\, x) = \int_0^\alpha Q_x(p)\mathrm{d}W(p) = \int_0^\alpha Q_x(p)w(p)\mathrm{d}p \tag{3.8}$$

and for this reason, we call it the weighted integrated conditional quantile function (WICQF) estimator and denote it by $\hat{\tau}_w^{(Q)}$.

The second estimator departs from

$$\tau_w(\alpha \,|\, x) = Q_x(\alpha) - \int_{-\infty}^{Q_x(\alpha)} W(F_x(y))\mathrm{d}y \tag{3.9}$$

and we call it as the weighted integrated conditional distribution function (WICDF) estimator.

3.3.1 The WICQF estimator

For a given weighting function W, the WICQF is obtained from (3.8) by replacing $Q_x(p)$ with an estimate and by approximating the integral with an analog sum. Assuming that conditional quantiles are linear in parameters as in Koenker and Bassett (24), that is $Q_x(p) = \beta(p)^\top x$, the parameter $\beta(p)$ is estimated by solving

$$\min_\beta \sum_{t=1}^n \ell_p(Y_t - \beta^\top X_t) \tag{3.10}$$

where $\ell_p(u) = u(p - \mathbb{1}\{u < 0\})$, $0 < p < 1$, is the asymmetric absolute loss function (see also Koenker (25)).

Then, given I linear regression quantile coefficients estimates $\hat{\beta}(p_1), \ldots, \hat{\beta}(p_I)$, with $0 < p_1 < p_2 < \ldots p_I \leq \alpha$ each of them solving (3.10), we define the WICQF estimator as

$$\hat{\tau}_{w,I}^{(Q)}(\alpha \,|\, x) = \sum_{i=1}^I w_i \hat{Q}_x(p) \tag{3.11}$$

where $w_i = W(p_i) - W(p_{i-1})$. Because of the linearity of $\hat{Q}_x(p_i)$, $\hat{\tau}_{w,I}^{(Q)}(\alpha \,|\, x)$ also writes

$$\hat{\tau}_{w,I}^{(Q)}(\alpha \,|\, x) = \tilde{\beta}_T(\alpha)^\top x \tag{3.12}$$

where $\tilde{\beta}_T(\alpha) = \sum_{i=1}^I w_i \hat{\beta}_T(p_i)$.

The WICQF presents the inconvenience that it does not guarantee monotonicity of the estimated quantiles. When for some $0 < p_1 < p_2 < 1$, the estimated quantile cross each other, that is $\hat{Q}_x(p_1) > \hat{Q}_x(p_2)$, then monotonicity of $\hat{\tau}_{w,I}^{(Q)}(\alpha \,|\, x)$ is violated.

3.3.2 The WICDF estimators

The estimator based on (3.9) combines estimation of the α-level quantile with estimation of the cumulative distribution function at different grid points. Replacing the integral with the analog sum, the estimator WICDF takes the form

$$\hat{\tau}^{(D)}_{w,I}(\alpha\,|\,x) = \hat{Q}_x(\alpha) - \sum_{i=1}^{I}(y_i - y_{i-1})W(\hat{F}_i(x)) \tag{3.13}$$

where points $y_0, y_1, \ldots, y_I$ are arbitrarily chosen such that $-\infty < y_0 < y_1 < \cdots < y_{I-1} < y_I \leq Q_x(\alpha)$. Here $\hat{Q}_x(\alpha)$ is the quantile regression estimator described in Subsection 3.3.1, while $\hat{F}_i(x)$ is the estimator for $F_i(x) = F_x(y_i)$ described below.

According to different choices for the estimator $\hat{F}_i(x)$, one obtains different definitions for the WICDF estimator. We here follow Peracchi (35) and propose an estimator for $\hat{F}_i(x)$ based on logistic regression. Having defined the log-odds

$$\eta_i(x) = \ln\left[F_i(x)/(1 - F_i(x))\right], \;\; i = 1, \ldots, I$$

we fit I separate logistic regressions and, given estimators $\hat{\eta}_i(x)$, we estimate $F_i(x)$ by

$$\hat{F}_i(x) = \frac{\exp\hat{\eta}_i(x)}{1 + \exp\hat{\eta}_i(x)}. \tag{3.14}$$

The estimator (3.14) is not guaranteed to be monotonic, that is we might have $F_i(x) < F_{i-1}(x)$ for some $i \in [2, I]$ and, as in the case of the WICQF estimator, we loose monotonicity of $\hat{\tau}^{(D)}_{w,I}(\alpha\,|\,x)$. A monotone estimator $\tilde{F}_i(x)$ of the cdf $F_i(x)$ can also be obtained by fitting logistic regressions as presented in Section 2.2.3. Inserting $\tilde{F}_i(x)$ into (4.8) results in the monotone WICDF estimator, which we denote by $\tilde{\tau}^{(D)}_{w,I}(\alpha\,|\,x)$ in order to differentiate it from $\hat{\tau}^{(D)}_{w,I}$ that is based on (3.14). Here, we limit ourselves to mentioning $\tilde{\tau}^{(D)}_{w,I}(\alpha\,|\,x)$ as one of the possible ways to overcome the crossing problem. The study of the asymptotic properties is here limited to $\hat{\tau}^{(Q)}_{w,I}$ and $\hat{\tau}^{(D)}_{w,I}$ only.

3.4 Asymptotic properties

We make some initial remarks concerning the behavior of the estimators.

The WICQF estimator is a linear functional of the regression quantile estimator and this suggests that continuous mapping theorem, if applicable, will yield consistency and asymptotic normality. This property is preserved even if the linear conditional quantile model is misspecified (see Angrist et al. (6)).

Similar arguments can be used for the WICDF estimators that are transformations of both the regression quantile and the logit estimator $\hat{F}(x) = (\hat{F}_1(x), \ldots, \hat{F}_I(x))$. In this case, in order to be able to write down the asymptotic distribution one has to take into account not only convergence of the regression quantile and (integrated) logit term separately, but also their interaction.

We first focus, in Subsections 3.4.1 and 3.4.2, on the asymptotic behavior of WICQF and WICDF estimators, when the dimensions of the grids $(p_0, \ldots, p_I)$ or $(y_0, \ldots, y_I)$ are fixed and constant with T. Theorems 1 and 2 in particular prove that $\hat{\tau}^{(Q)}_{w,I}$ and $\hat{\tau}^{(D)}_{w,I}$ are both asymptotically normal estimators (under certain conditions) and are consistent for the approximations of $\tau_w(\alpha \,|\, x)$ resulting from the substitution of the integrals with analog sums in (3.8) and (3.9) respectively. Leorato et al (26) deals also with the case of sequences of grids whose dimension grows with T. Throughout all the Section, w (or its primitive W) is some fixed weighting function satisfying (A1) only, unless explicitly stated otherwise. Assumption (A2) is, in fact, not invoked in Theorems 1 and 2.

3.4.1 The WICQF estimator $\hat{\tau}^{(Q)}_{w,I}(\alpha \,|\, x)$

Let (X_1, Y_1),..., (X_T, Y_T) be a random sample of T pairs from the joint distribution of (X, Y), where X is a k-dimensional vector, and let the conditional distribution of Y_t given $X_t = x$ have strictly positive density $f_x(y)$ for every y and for all t. Consider a suitable weighting function $W : [0, \alpha] \mapsto [0, 1]$ and define the vector of I weights $\mathbf{w} = \{w_1, \dots w_I\}^\top$, $w_i = W(p_i) - W(p_{i-1})$ where $0 = p_0 < p_1 < \cdots < p_I \leq \alpha$.

Let $\tau^{(Q)}_{w,I}(\alpha \,|\, x)$ denote the approximation of $\tau_w(\alpha \,|\, x) = \int_0^\alpha Q_x(p)w(p)dp$ by the analog sum:

$$\tau^{(Q)}_{w,I}(\alpha \,|\, x) = \sum_{i=1}^{I} w_i Q_x(p_i).$$

Theorem 1 *Under the set of conditions (see Theorem 4.1 in Koenker (25)):*

(i) The cdf F_x is absolutely continuous, with continuous density $f_x(\cdot)$, uniformly bounded away from zero at the points $y \in [\varepsilon, 1 - \varepsilon]$, for every $\varepsilon > 0$

(ii) There exist positive definite matrices $\mathbf{D}$ and $\mathbf{J_1(p)}$ such that

$$\lim_{T\to\infty} T^{-1} \sum X_t X_t' = \mathbf{D}$$

and

$$\lim_{T\to\infty} T^{-1} \sum f(Q_x(p) X_t X_t' = \mathbf{J_1(p)} \quad \text{and} \quad \max_{\mathbf{t=1\dots T}} ||\mathbf{X_t}||/\sqrt{\mathbf{T}} \to \mathbf{0},$$

where $\|\cdot\|$ is the Euclidean norm of a vector in $\mathbb{R}^k$,

the limiting distribution of the normalized difference $\sqrt{T}\,[\hat{\tau}^{(Q)}_{w,I}(\alpha \,|\, x) - \tau^{(Q)}_{w,I}(\alpha \,|\, x)]$ is Gaussian with zero mean and variance

$$AV(\hat{\tau}^{(Q)}_{w,I}(\alpha \,|\, x)) = x^\top (\mathbf{w}^\top \otimes \mathcal{I}_k)\, \mathbf{\Omega}_1 \,(\mathbf{w} \otimes \mathcal{I}_k)\, x \tag{3.15}$$

where $\mathcal{I}_k$ is the k-dimensional identity matrix and $\mathbf{\Omega}$ is an $Ik \times Ik$ block matrix with each $k \times k$ block

$$\Omega_{1;i,j} = \mathbf{J}_1{}^{-1}(p_i)\, \Sigma_{1;i,j}\, \mathbf{J}_1{}^{-1}(p_j) \tag{3.16}$$

where $\mathbf{J}_1(p) := \mathrm{E}\left[f(\beta(p)^\top X \,|\, X)\, X X^\top\right]$ and

$$\Sigma_{1;i,j} := \mathrm{E}[(p_i - \mathbb{1}\{Y < \beta(p_i)^\top X\})(p_j - \mathbb{1}\{Y < \beta(p_j)^\top X\}) X X^\top] \tag{3.17}$$

are positive definite $k \times k$ matrices. If the model is correctly specified, that is $Q_X(p) = \beta(p)^\top X$, then $\Sigma_{1;i,j} := [\min(p_i, p_j) - p_i p_j]\, \mathrm{E}[X X^\top]$.

In the following corollary, for inference purposes, we give a consistent estimator of the asymptotic variance $AV(\hat{\tau}^{(Q)}_{w,I}(\alpha \,|\, x))$.

Corollary 1 *If the model is correctly specified, a consistent estimator of the asymptotic variance* $AV(\hat{\tau}^{(Q)}_{w,I}(\alpha\,|\,x))$ *is*

$$\widehat{AV}(\hat{\tau}^{(Q)}_{w,I}(\alpha\,|\,x)) = x^\top \left\{ \sum_{i=1}^{I}\sum_{j=1}^{I} w_i w_j \left[\min(p_i,p_j) - p_i p_j\right] \hat{\mathbf{J}}_1^{-1}(p_i)\,\hat{D}\,\hat{\mathbf{J}}_1^{-1}(p_j) \right\} x \tag{3.18}$$

where $\hat{D} = T^{-1}\sum_{t=1}^T X_t X_t^\top$ *and* $\hat{\mathbf{J}}_1(p_i) = T^{-1}\sum_{t=1}^T \hat{f}(\hat{\beta}_T(p_i)^\top X_t\,|\,X_t)\,X_t X_t^\top$, *with* $\hat{f}$ *a consistent estimator of the conditional density* f.

This is immediate from consistency of $\hat{D}$ and $\hat{\mathbf{J}}_1(p)$ for $\mathrm{E}[XX^\top]$ and $\mathbf{J}_1(p)$ respectively.

3.4.2 The WICDF estimator $\hat{\tau}^{(D)}_{w,I}(\alpha\,|\,x)$

The WICDF estimator has the form

$$\hat{\tau}^{(D)}_{w,I}(\alpha\,|\,x) = \hat{Q}_x(\alpha) - \sum_{i=1}^{I}(y_i - y_{i-1})\,W(\hat{F}_i(x)) \tag{3.19}$$

where $\hat{Q}_x(p)$ is the quantile regression estimator of $Q_x(p)$ as proposed by Koenker and Bassett (24) and $\hat{F}_i(x)$ is the estimator of $F_i(x)$ defined by (3.14).

Under standard regularity conditions (see Theorem 3 in Angrist *et al.* (6)) $\sqrt{T}[\hat{Q}_x(\alpha) - Q_x(\alpha)]$ is asymptotically normal. The same is true for $\sqrt{T}[\hat{F}_i(x)-F_i(x)]$. Conditionally on the grid points and provided that some conditions on the weighting function W are encountered, the sum $\sum_{i=1}^I (y_i - y_{i-1})\,W(\hat{F}_i(x))$, suitably normalized, is approximately normal, too.

The aim of this section is to prove asymptotic Gaussianity of the estimator (3.19), which also writes as

$$\hat{\tau}^{(D)}_{w,I}(\alpha\,|\,x) = \hat{Q}_x(\alpha) - \hat{\delta}_w(\alpha\,|\,x,\mathbf{y})$$

where $\hat{\delta}_w(\alpha\,|\,x,\mathbf{y}) = \sum_{i=1}^I (y_i - y_{i-1})\,W(\hat{F}_i(x))$ and $\mathbf{y} = (y_0,\dots,y_I)$.

Consistency of $\hat{\tau}^{(D)}_{w,I}$, for I fixed, is proved with respect to the approximate WICDF:

$$\begin{aligned}\tau^{(D)}_{w,I} &= Q_x(\alpha) - \delta_w(\alpha\,|\,x,\mathbf{y}) \\ &= Q_x(\alpha) - \sum_{i=1}^{I}(y_i - y_{i-1})W(F_i(x)).\end{aligned} \tag{3.20}$$

To achieve this, we use standard arguments based on approximating both $\hat{Q}_x(\alpha)$ and $\hat{\delta}_w(\alpha\,|\,x,\mathbf{y})$ via functional empirical processes indexed by two different Donsker classes of functions. We are then able to write $\hat{\tau}^{(D)}_{w,I}(\alpha\,|\,x)$ as an empirical process indexed by sums $f+g$ of functions belonging to the two Donsker classes, that is also Donsker, thanks to the permanence property.

For (X,Y) and h a measurable function $h : \mathcal{X},\mathcal{Y} \to \mathbb{R}$, define the empirical expectations

$$\mathbb{E}_T[h(X,Y)] := T^{-1}\sum_{t=1}^{T} h(X_t,Y_t)$$

$$\mathbb{G}_T[h(X,Y))] := T^{-1/2}\sum_{t=1}^{T}(h(X_t,Y_t) - \mathrm{E}[h(X_t,Y_t)])$$

Moreover, for any square matrix Σ, with $\lambda_{\min}(\Sigma)$ we denote the minimum eigenvalue of Σ.

As far as to the term $\hat{Q}_x(\alpha)$, we already know, from Theorem 3 in Angrist *et al.* (6) that the estimated quantile coefficient $\hat{\beta}(\alpha)$ can be approximated as

$$\mathbf{J}_1(\alpha)\sqrt{T}[\hat{\beta}(\alpha) - \beta(\alpha)] = \mathbb{G}_T[(1\{Y \leq \beta(\alpha)^\top X\} - \alpha)X] + o_P(1) \tag{3.21}$$

where matrix $\mathbf{J}_1(\alpha)$ is defined as in Theorem 1, that is

$$\mathbf{J}_1(\alpha) = \mathrm{E}[f(\beta(\alpha)^\top X \,|\, X)\, XX^\top]$$

and $\mathbb{G}_T[(1\{Y \leq \beta(\alpha)^\top X\} - \alpha)X] \xrightarrow{d} Z_Q$, where Z_Q is a zero mean Gaussian random vector with covariance matrix $\Sigma_{1;i,j}$ specified in (3.17).

We now focus on the term $\hat{\delta}(\alpha \,|\, x, \mathbf{y})$, following closely the proof of Theorem 3 in Angrist *et al.* (6). The following theorem concerns the asymptotic behavior of the logit parameter $\hat{\theta} = \hat{\theta}(y)$, the k-dimension row-vector maximizing the log-likelihood

$$\hat{\theta}(y) = \arg\sup_{\theta} \frac{1}{T}\left[\sum_{t=1}^{T} \mathbb{1}\{Y_t \leq y\}\theta^\top X_t - \ln\left(1 + e^{\theta^\top X_t}\right)\right] \doteq \arg\sup \mathcal{L}_T(y, \theta) \tag{3.22}$$

The idea is to establish the uniform consistency and the asymptotic Gaussianity of the logistic regression process through an empirical process approximation for the function $\sqrt{T}[\hat{\theta}(y) - \theta(y)]$, where $\theta(y)$ is the value that maximizes

$$\mathcal{L}_\infty(y, \theta) := \mathrm{E}\left[\mathbb{1}\{Y \leq y\}\theta^\top X - \ln\left(1 + e^{\theta^\top X}\right)\right]$$

over $\Theta \subseteq \mathbb{R}^k$.

For simplicity, we consider an arbitrarily set non-stochastic grid of points $\mathbf{y} = \{y_0, \ldots, y_I\}$. Data-dependent choices of the grid $\{y_0, \ldots, y_I\}$ are discussed in Leorato et al (26).

Theorem 2 *Suppose that the following conditions are met:*

(i) $(X_1, Y_1), \ldots, (X_T, Y_T)$ *is a sample of* T *iid pairs from the joint distribution of* (X, Y) *defined on the probability space* $(\Omega, \mathcal{F}, \mathcal{P})$

(ii) The function $\hat{\theta} = \hat{\theta}(y)$ *maximizes* $\mathcal{L}_T(y, \theta)$ *over a compact set* $\Theta \subseteq \mathbb{R}^k$

(iii) The function $\mathcal{L}_\infty(y, \theta)$ *has a unique maximum at* $\theta(y) \in \Theta$ *for all values of* y *such that* $f_Y(y)$*, the marginal density of* Y*, is strictly positive*

(iv) $\mathrm{E}\,\|X\|^{2+\epsilon} < \infty$ *for some* $\epsilon > 0$*, where* $\|\cdot\|$ *is the Euclidean norm of a vector in* $\mathbb{R}^k$

(v) $\mathbf{J}_2(y) := \mathrm{E}[F_x(y)(1 - F_x(y))XX^\top]$ *is positive definite for all* $y \in \mathbb{R}$.

Then, the logit regression process is uniformly consistent, $\sup_{y\in\mathbb{R}} \|\hat{\theta}(y) - \theta(y)\| = o_P(1)$ *and* $\mathbf{J}_2(\cdot)\sqrt{T}[\hat{\theta}(\cdot) - \theta(\cdot)]$ *converges in distribution to a zero mean Gaussian process* $Z_\theta(\cdot)$ *defined by the covariance function*

$$\Sigma_{2;j,k} = \mathrm{Cov}[Z_\theta(y_j), Z_\theta(y_k)] = \mathrm{E}\left[\varphi(y_j, \theta(y_j))\varphi(y_k, \theta(y_k))XX^\top\right] \tag{3.23}$$

with

$$\varphi(y, \theta(y)) := \left[\mathbb{1}\{Y \leq y\} - \frac{e^{\theta(y)^\top X}}{1 + e^{\theta(y)^\top X}}\right]$$

If the log-odds ratio $\ln[F_x(y)(1 - F_x(y))]$ *is linear in* x *for any* $y \in \mathbb{R}$*, then* $\Sigma_{2;j,k}$ *simplifies to* $\Sigma_{2;j,k} = [\min(F_j, F_k) - F_jF_k)]\,\mathrm{E}[XX^\top]$.

Let

$$\tau_{w,I}^{(D)}(\alpha\,|\,x) = Q_x(\alpha) - \sum_{i=1}^{I}(y_i - y_{i-1})W(F_i(x)).$$

Corollary 2 *$\sqrt{T}[\hat{\tau}_{w,I}^{(D)}(\alpha\,|\,x) - \tau_{w,I}^{(D)}(\alpha\,|\,x)] \xrightarrow{d} N(0, AV(\hat{\tau}_{w,I}^{(D)}(\alpha\,|\,x)))$, where*

$$\begin{aligned} AV(\hat{\tau}_{w,I}^{(D)}(\alpha\,|\,x)) = {} & x^\top \alpha(1-\alpha)\mathbf{J}_1(\alpha)^{-1}\,\mathrm{E}[XX^\top]\mathbf{J}_1(\alpha)^{-1}x + \\ & + x^\top\left(\Delta_2^\top \otimes \mathcal{I}_k\right)\boldsymbol{\Omega}_2\left(\Delta_2 \otimes \mathcal{I}_k\right)x + \\ & - 2x^\top\left(\Delta_1^\top \otimes \mathcal{I}_k\right)\boldsymbol{\Omega}_3\left(\Delta_1 \otimes \mathcal{I}_k\right)x \end{aligned} \tag{3.24}$$

where $\Delta_1 = (\sqrt{\mathrm{d}_1^w}, \cdots \sqrt{\mathrm{d}_I^w})^\top$, $\Delta_2 = diag(\Delta_1\Delta_1^\top) = (\mathrm{d}_1^w, \cdots \mathrm{d}_I^w)^\top$ are I-dimensional column vectors and where

$$\mathrm{d}_i^w = [(y_i - y_{i-1})w(F(y_i))F(y_i)(1-F(y_i))]$$

and $\boldsymbol{\Omega}_2$ is an block matrix with each $k \times k$ block $\Omega_{2;j,k} = \mathbf{J}_2^{-1}(y_j)\,\Sigma_{2;j,k}\,\mathbf{J}_2^{-1}(y_k)$ with $\mathbf{J}_2(y_j)$ and $\Sigma_{2;j,k}$ defined as in Theorem 2. Moreover, $\boldsymbol{\Omega}_3$ is an $Ik \times Ik$ diagonal block matrix with each $k \times k$ block $\Omega_{3,j} = \mathbf{J}_2^{-1}(y_j)\,\Sigma_{3,j}\,\mathbf{J}_1^{-1}(\alpha)$ where

$$\Sigma_{3,j} = \mathrm{E}[(1\{Y \le \beta(\alpha)^\top X\} - \alpha)\varphi(y_j, \theta)XX^\top].$$

If both the conditional quantile $Q_x(\alpha)$ and the log odds ratio $\ln[F_x(y)(1-F_x(y))]$ are linear in x for any $y \in \mathbb{R}$, then $\Sigma_{3,j} = [\min(F_j, \alpha) - F_j\alpha]\,\mathrm{E}[XX^\top]$

For inference purposes, we also give an estimator of $AV(\hat{\tau}_{w,I}^{(D)}(\alpha\,|\,x))$ under hypothesis of well specified quantile function only.

Corollary 3 *Let $\hat{w}_i = [W(\hat{F}_i(x)) - W(\hat{F}_{i-1}(x))]$. Moreover, define*

$$\hat{\mathbf{J}}_1(p) = T^{-1}\sum_{t=1}^{T}\hat{f}(\hat{\beta}_T(p)^\top X_t\,|\,X_t)\,X_tX_t^\top,$$

$$\hat{\mathbf{J}}_2(y_i) = T^{-1}\sum_{t=1}^{T}\hat{F}_i(x)(1-\hat{F}_i(x))X_tX_t^\top$$

$$\hat{\Sigma}_{2;j,k} = T^{-1}\sum_{t=1}^{T}\varphi_t(y_j,\theta(y_j))\varphi_t(y_k,\theta(y_k))\,X_tX_t^\top$$

$$\hat{\Sigma}_{3,j} = T^{-1}\sum_{t=1}^{T}(\mathbb{1}\{Y_t \le \hat{\beta}(\alpha)^\top X_t\} - \alpha)\varphi_t(y_j,\theta(y_j))\,X_tX_t^\top$$

where $\hat{f}$ consistently estimates density f, $\hat{F}$ is the logistic regression estimator for the cdf F (3.14) and $\varphi_t(y,\theta(y)) = \left[\mathbb{1}\{Y_t \le y\} - \frac{e^{\hat{\theta}(y)^\top X_t}}{1+e^{\hat{\theta}(y)^\top X_t}}\right]$.

Under the hypothesis of well specified quantile function, the estimator $\widehat{AV}(\hat{\tau}^{(D)}_{w,I}(\alpha\,|\,x))$

$$\widehat{AV}(\hat{\tau}^{(D)}_{w,I}(\alpha\,|\,x)) = x^\top \alpha(1-\alpha)\hat{\mathbf{J}}_1^{-1}(\alpha)\,\hat{D}\,\hat{\mathbf{J}}_1^{-1}(\alpha)x +$$
$$+ x^\top \left\{ \sum_{i=1}^{I} \frac{\hat{w}_i[\hat{F}_i(x)(1-\hat{F}_i(x))](y_i - y_{i-1})}{\hat{F}_i(x) - \hat{F}_{i-1}(x))} \sum_{j=1}^{I} \frac{\hat{w}_j[\hat{F}_j(x)(1-\hat{F}_j(x))](y_j - y_{j-1})}{\hat{F}_j(x) - \hat{F}_{j-1}(x))} \cdot \right.$$
$$\left. \cdot \left[\hat{\Sigma}_{2;i,j}\right] \hat{\mathbf{J}}_2^{-1}(y_i)\hat{\mathbf{J}}_2^{-1}(y_j) \right\} x +$$
$$2 \cdot x^\top \left\{ \sum_{i=1}^{I} \frac{\hat{w}_i[\hat{F}_i(x)(1-\hat{F}_i(x))](y_i - y_{i-1})}{\hat{F}_i(x) - \hat{F}_{i-1}(x))} \cdot [\hat{\Sigma}_{3,i}]\,\hat{\mathbf{J}}_2^{-1}(y_i)\hat{\mathbf{J}}_1^{-1}(\alpha) \right\} x \tag{3.25}$$

is consistent for $AV(\hat{\tau}^{(D)}_{w,I}(\alpha\,|\,x))$.

This follows from consistency of $\hat{\mathbf{J}}_1(p)$, $\hat{\mathbf{J}}_2(y)$, $\hat{\Sigma}_{2;j,k}$ and $\hat{\Sigma}_{3;j}$ for $\mathbf{J}_1(p)$, $\mathbf{J}_2(y)$, $\Sigma_{2;j,k}$ and $\Sigma_{3;j}$ respectively.

Remark 2 (Linearity of log odds ratio) *If moreover the log odds ratio is linear in* X_t*, then, in the above corollary* $\hat{\Sigma}_{2;j,k}$ *and* $\hat{\Sigma}_{3;j}$ *boil down to*

$$\hat{\Sigma}_{2;j,k} = \left[\min(\hat{F}_i(x), \hat{F}_j(x)) - \hat{F}_i(x)\hat{F}_j(x)\right]\hat{D}$$

and

$$\hat{\Sigma}_{3;j} = [\min(\hat{F}_i(x), \alpha) - \hat{F}_i(x)\,\alpha]\hat{D}$$

where $\hat{D} = T^{-1}\sum_{t=1}^{T} X_t X_t^\top$..

3.5 Efficient weighted ES estimators

In this section we present our proposal for efficient estimation of ES via the minimization of the asymptotic variance with respect to the weights w.

We will consider separately the cases where the specification of the weights is fully nonparametric or parametric (following (3.5)). In either cases, the objective functions are the asymptotic variances of the WICQF and of the WICDF that are given in formulas (3.15) and (3.24).

In the whole section, as well as in Section 3.6, since the focus is now in the WES as a function of the weights, we refer to $\tau_{w,I}(\alpha\,|\,x)$ as $\tau(\mathbf{w})$ as well as $\hat{\tau}(\mathbf{w}) = \hat{\tau}_{w,I}(\alpha\,|\,x)$, where $\mathbf{w} = (w_1, \ldots, w_I)$ is a vector of weights (corresponding to function w). For the sake of simplicity, we also omit the explicit reference to the level α and to the covariate value x in $\tau := \tau(\alpha\,|\,x)$ and $\hat{\tau} := \hat{\tau}(\alpha\,|\,x)$.

3.5.1 Nonparametric specification

In order to write down the definition of the minimum asymptotic variance of the weighted ES estimators, we rewrite the classes $\mathcal{W}, \mathcal{W}_U$ in discrete form. So, corresponding to a grid of $I < \infty$ points $p_0 < \ldots < p_I = \alpha$, we denote $\bar{\mathbf{p}}_i = p_i - p_{i-1}$ and $\mathbf{w} = (w_1, \ldots, w_I) =$

$(W(p_1) - W(p_0), \ldots, W(p_I) - W(p_{I-1}))$. Then, for example $\mathcal{W}$ writes:

$$\mathcal{W}_I = \left\{ \mathbf{w} = (w_1, \ldots, w_I) : w_i \geq 0, \sum_{i \leq I} w_i = 1, \sum_{i \leq I} Q_x(p_i) w_i - \tau(\alpha \,|\, x) = 0 \right\}.$$

The efficient WICQF and WICDF estimators are then defined as the weighted ES estimators corresponding to $\mathbf{w}^* \in \mathcal{W}_I$, where $\mathbf{w}^*$ is such that

$$AV\left(\hat{\tau}_{\mathbf{w}^*}(\alpha \,|\, x)\right) \leq AV\left(\hat{\tau}_{\mathbf{w}}(\alpha \,|\, x)\right), \text{ for all } \mathbf{w} \in \mathcal{W}_I.$$

Definition 2 *The* minimum AV WICQF estimator *of the $\alpha-$level ES, $\hat{\tau}^{(Q)}(\mathbf{w}_Q^*)$, is given by the solution of:*

$$\begin{aligned} &\min_{\mathbf{w} \in \mathcal{W}_I} AV(\hat{\tau}^{(Q)}(\mathbf{w})) \\ &w_i \geq 0, \quad i = 1, \ldots, I \\ &\sum_{i=1}^{I} w_i = 1 \end{aligned} \tag{3.26}$$

The *minimum AV WICDF estimator* $\hat{\tau}^{(D)}(\mathbf{w}_D^*)$ is defined accordingly.

Definition 3 *The* unconstrained minimum AV WICQF estimator *(*unconstrained minimum AV WICDF estimator*) $\hat{\tau}^{(Q)}(\mathbf{w}_{u,Q}^*)$ (resp. $\hat{\tau}^{(D)}(\mathbf{w}_{u,D}^*)$) of the $\alpha-$level ES is given by the solution of:*

$$\begin{aligned} &\min_{\mathbf{w} \in \mathcal{W}_{U,I}} AV(\hat{\tau}^{(\cdot)}(\mathbf{w})) \\ &\sum_{i=1}^{I} w_i = 1 \end{aligned} \tag{3.27}$$

We underline that, the objective functions $\widehat{AV}(\hat{\tau}^{(\cdot)}(\mathbf{w}))$ are convex in the vector $\mathbf{w}$, while the inequality and equality constraints are linear. In other words, the efficient estimators $\hat{\tau}^{(\cdot)}(\mathbf{w}^*)$ and $\hat{\tau}^{(\cdot)}(\mathbf{w}_u^*)$ are obtained as solutions to standard convex optimization problems. As such, they take advantage of the following convenient properties:

(i) Any local optimum is necessarily a global optimum;

(ii) Duality theory can be used to infeasibility detection, hence algorithms are easy to initialize;

(iii) Efficient numerical solution methods are available.

The global optimum $\mathbf{w}^*$ of (3.26) may not be unique. We can define by convention the vector $\mathbf{w}^*$ to be any vector in the set $\mathcal{W}^*$ of the solutions to (3.26) satisfying

$$\sum_i (w_i^* - \bar{\mathbf{p}}_i)^2 \bar{\mathbf{p}}_i \leq \sum_i (w_i - \bar{\mathbf{p}}_i)^2 \bar{\mathbf{p}}_i, \; \mathbf{w} = (w_1, \ldots, w_I) \in \mathcal{W}^*.$$

The global optimum of (3.27) is instead unique because of strict convexity of the asymptotic variances. Moreover, $\mathbf{w}_{u,\cdot}^*$ can be easily found explicitly. As already noted, the price to be paid in this case is the loss of the interpretation of $W(F)$ as a biased probability distribution since $\mathbf{w}^*_{u,\cdot}$ is allowed to be a signed density.

On the other hand, allowing $\mathbf{w}$ to take negative entries yields a larger efficiency gain because of the inclusion $\mathcal{W} \subseteq \mathcal{W}_u$. This will be pointed out in the next Section.

Regarding the unconstrained minimization problem, we have the following result

Theorem 3 *Let $Y_t \,|\, X_t = x \sim F_x$, independent and identically distributed for all fixed x. The unconstrained minimum AV WICQF estimator of the $\alpha-$level ES is given by $\hat{\tau}^{(Q)}(\mathbf{w}^*_{u,Q})$, with:*

$$\begin{pmatrix} \mathbf{w}^*_{u,Q} \\ \boldsymbol{\lambda} \end{pmatrix} = \mathbb{C}^{-1}\boldsymbol{\tau} \tag{3.28}$$

where $\boldsymbol{\lambda}^\top = (\lambda_1, \lambda_2)$ is the vector of Lagrange multipliers, $\boldsymbol{\tau}^\top = (0, \ldots, 0, -1, -\tau)$ and $\mathbb{C}$ is the $(I+2)-$dimensional square matrix:

$$\mathbb{C} = \begin{pmatrix} 2x^\top \Omega_{1;1,1} x & 2x^\top \Omega_{1;1,2} x & \cdots & 2x^\top \Omega_{1;1,I} x & -1 & -Q_x(p_1) \\ \vdots & & \ddots & \vdots & & \vdots \\ 2x^\top \Omega_{1;I,1} x & 2x^\top \Omega_{1;I,2} x & \cdots & 2x^\top \Omega_{1;I,I} x & -1 & -Q_x(p_I) \\ -1 & -1 & \cdots & -1 & 0 & 0 \\ -Q_x(p_1) & -Q_x(p_2) & \cdots & -Q_x(p_I) & 0 & 0 \end{pmatrix}. \tag{3.29}$$

Proof. The Lagrangian function

$$\mathcal{L}^Q_u = x^\top(\mathbf{w}^\top \otimes I_k)\boldsymbol{\Omega}_1(\mathbf{w} \otimes I_k)x + \lambda_1(1 - \mathbf{w}^\top \boldsymbol{\imath}_I) + \lambda_2(\tau - \mathbf{w}^\top \mathbf{Q}) \tag{3.30}$$

where $\boldsymbol{\imath}_T$ is the $I-$dimensional vector of ones and $\mathbf{Q}^\top = (Q_x(p_1), \ldots, Q_x(p_I))$.

By solving the system of linear equations:

$$\begin{cases} \frac{\partial \mathcal{L}^Q_u}{\partial \mathbf{w}} = 0 \\ \frac{\partial \mathcal{L}^Q_u}{\partial \lambda_i} & i = 1, 2 \end{cases}$$

one obtains the result (3.28). The fact that $\boldsymbol{\Omega}_1$ is positive semidefinite ensures that $\hat{\tau}^{(Q)}(\mathbf{w}^*_{u,Q})$ is the minimum over $\mathcal{W}_I$. □

A similar result (and structure of the proof) holds for the unconstrained minimum AV WICDF estimator. This minimizes the Lagrange function

$$\begin{aligned} \mathcal{L}^D_u = {} & x^\top \left(\Delta_2^\top \otimes \mathcal{I}_k\right) \boldsymbol{\Omega}_2 \left(\Delta_2 \otimes \mathcal{I}_k\right) x + \\ & - 2x^\top \left(\Delta_1^\top \otimes \mathcal{I}_k\right) \boldsymbol{\Omega}_3 \left(\Delta_1 \otimes \mathcal{I}_k\right) x \\ & + \lambda_1(1 - \mathbf{w}^\top \boldsymbol{\imath}_I) + \lambda_2(\hat{\tau} - Q_x(\alpha) + \mathbf{w}^\top dY) \end{aligned} \tag{3.31}$$

where λ_1 and λ_2 are the Lagrange multipliers, dY and $\boldsymbol{\imath}_I$ are the $I-$dimensional column vectors with elements $y_I - y_{i-1}$ and 1 respectively. The other vectors and matrices are defined as in (3.24), while

$$\tau = \tau^{(D)}_{w,I} = Q_x(\alpha) - \frac{1}{\alpha}\sum_{i=1}^{I}(y_I - y_{i-1})(F_x(y_i) - F_x(y_{i-1})).$$

Theorem 4 *Let $Y_t \,|\, X_t = x \sim F_x$, iid for all fixed x. The unconstrained minimum AV WICDF estimator of the $\alpha-$level ES is $\hat{\tau}^{(D)}(\mathbf{w}^*_{u,D})$, with:*

$$\begin{pmatrix} \mathbf{w}^*_{u,D} \\ \boldsymbol{\lambda} \end{pmatrix} = \mathbb{D}^{-1}\boldsymbol{\upsilon} \tag{3.32}$$

where $\boldsymbol{\lambda}^\top = (\lambda_1, \lambda_2)$ is the vector of Lagrange multipliers,

$$\boldsymbol{\upsilon}^\top = \Big(2(y_1 - y_0)x^\top \Omega_{3;1}x, \ldots, 2(y_I - y_{I-1})x^\top \Omega_{3;I}x, -1, -\tau - Q_x(\alpha)\Big)$$

and $\mathbb{D}$ is the $(I+2)-$dimensional square matrix:

$$\mathbb{D} = \begin{pmatrix} \frac{2x^\top \Omega_{2;1,1}x}{(y_1-y_0)^{-2}} & \cdots & \frac{2x^\top \Omega_{2;1,I}x}{[(y_1-y_0)(y_I-y_{I-1})]^{-1}} & -1 & (y_I - y_0) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{2x^\top \Omega_{2;I,1}x}{[(y_1-y_0)(y_I-y_{I-1})]^{-1}} & \cdots & \frac{2x^\top \Omega_{2;I,I}x}{(y_I-y_{I-1})^{-2}} & -1 & (y_I - y_{I-1}) \\ -1 & \cdots & -1 & 0 & 0 \\ (y_I - y_0) & \cdots & (y_I - y_{I-1}) & 0 & 0 \end{pmatrix}, \tag{3.33}$$

where $\Omega_{2;j,k}$ and $\Omega_{3;j}$ are $k \times k$ matrices defined as in (3.24).

One of the issues with the constrained nonparametric specification, is related with the dimension of the grid I. It is in fact clear that the finer the grid the larger the set of admissible values for w. On the other hand, the objective function depends on I also through the vector $\bar{\mathbf{p}}$. Moreover, since I coincides with the dimension of the convex problem (3.26), the computational burden gets heavier as I increases.

Since both $AV\,(\hat{\tau}(\mathbf{w}))$ (either for WICQF and WICDF) and the linear constraints in (3.4) depend on the distribution of the returns, we must replace it with consistent estimates. Formulas (3.18) and (3.25) are then used, while the condition (3.3) is replaced by

$$\left|\hat{\tau}^{(Q)}(\mathbf{w}) - \hat{\tau}^{(Q)}\right| \leq r_T \text{ or } \left|\hat{\tau}^{(D)}(\mathbf{w}) - \hat{\tau}^{(D)}\right| \leq r_T$$

respectively, where r_T is a sequence of positive numbers converging to zero at a proper rate, discussed in Theorem 5. The estimator we obtain will be based on feasible versions of $\mathbf{w}^*_Q, \mathbf{w}^*_D$ that we denote by $\hat{\mathbf{w}}^*_Q, \hat{\mathbf{w}}^*_D$.

The *feasible* optimal weights vector is defined by $\hat{\mathbf{w}}^*$ that satisfies the following condition:

$$\widehat{AV}\left(\hat{\tau}^{(\cdot)}(\hat{\mathbf{w}}^*)\right) \leq \widehat{AV}\left(\hat{\tau}^{(\cdot)}(\mathbf{w})\right), \text{ for all } \mathbf{w} \in \widehat{\mathcal{W}}_I.$$

where $\widehat{\mathcal{W}}_I$ is:

$$\widehat{\mathcal{W}}_I = \left\{ \mathbf{w} = (w_1, \ldots, w_I) : \sum_i w_i = 1,\ w_i \geq 0,\ \left|\hat{\tau}^{(\cdot)}(\mathbf{w}) - \hat{\tau}^{(\cdot)}\right| \leq r_T \right\}.$$

The feasible minimum AV (or minimum unconstrained AV) WICQF and WICDF estimators are then denoted by $\tilde{\tau}^{(\cdot)}$. For the minimum unconstrained AV WICQF, for example $\tilde{\tau}^{(Q)}_u := \hat{\tau}(\hat{\mathbf{w}}^*_{u,Q})$. The following result guarantees that the vectors $\hat{\mathbf{w}}^*_Q$ and $\hat{\mathbf{w}}^*_{u,Q}$ (resp. $\hat{\mathbf{w}}^*_D$ and $\hat{\mathbf{w}}^*_{u,D}$) are consistent for the optimal vectors $\mathbf{w}^*_Q$ and $\mathbf{w}^*_{u,Q}$ (resp. $\mathbf{w}^*_D$ and $\mathbf{w}^*_{u,D}$) that identify the efficient nonparametric estimators.

Theorem 5

(a) *Under the conditions of Theorem 1 and Corollary 1 and if moreover*

$$\lim_{T\to\infty} r_T = 0 \qquad \lim_{T\to\infty} r_T T^{1/2} = \infty, \tag{3.34}$$

then

(a.i) *If the vector* $\mathbf{w}_Q^* = \arg\min_{\mathbf{w}\in\mathcal{W}_I} AV(\hat{\tau}^{(Q)}(\mathbf{w}))$ *is the unique minimum point in* $\mathcal{W}_I$,

$$\|\mathbf{w}_Q^* - \hat{\mathbf{w}}_Q^*\| = \sup_{1\le i\le I} |w_i^* - \hat{w}_i^*| = o_P(1) \tag{3.35}$$

as $T \to \infty$.

(a.ii) *Moreover,*

$$\|\mathbf{w}_{u,Q}^* - \hat{\mathbf{w}}_{u,Q}^*\| \to 0 \tag{3.36}$$

(b) *Under the conditions of Theorem 2 and Corollary 3 and if moreover (3.34) holds*

(b.i) *If the vector* $\mathbf{w}_D^* = \arg\min_{\mathbf{w}\in\mathcal{W}_I} AV(\hat{\tau}^{(D)}(\mathbf{w}))$ *is the unique minimum point in* $\mathcal{W}_I$,

$$\|\mathbf{w}_D^* - \hat{\mathbf{w}}_D^*\| = o_P(1).$$

(b.ii) *Moreover, for the unconstrained optimal weights, it holds,*

$$\|\mathbf{w}_{u,D}^* - \hat{\mathbf{w}}_{u,D}^*\| = o_P(1)$$

Form Theorem 5, the following result is easily derived.

Corollary 4 *Under the conditions of Theorem 5,*

$$|\hat{\tau}(\hat{\mathbf{w}}^*) - \tilde{\tau}| = o_P(1) \tag{3.37}$$

where $\hat{\tau}(\mathbf{w}^*)$ *indicates either the minimum (minimum unconstrained) AV WICQF or WICDF estimators and* $\tilde{\tau}$ *is the corresponding feasible version.*

Proofs of Theorem 5 and Corollary 3.37 can be found in Leorato et al (26).

3.5.2 Parametric specification

For the parametric specification, we have to find the optimal value of w as a function of at most 5 parameters.

Definition 4 *The* minimum parametric AV WICQF estimator *(*minimum parametric AV WICDF estimator*) of the* $\alpha-$*level ES,* $\tau^{(Q)}(w_{P,Q}^*)$ $(\tau^{(D)}(w_{P,D}^*))$, *is given by the solution of:*

$$\min_{\{a_1,b_1,a_2,b_2,\pi=\pi^*\}} AV(\hat{\tau}^{(\cdot)}(w)) \tag{3.38}$$

where $w = w(a_1, b_1, a_2, b_2, \pi)$ *is given by (3.5) and* π^* *is defined by (4.14)*

Unfortunately, in this case the objective functions are not convex in the parameters. Moreover, differentiation of $AV(\hat{\tau}^{(\cdot)}(\mathbf{w}))$ with respect to (a_i, b_i) is not an easy task, especially for WICDF. This means that we can only proceed by numerical methods for deriving the optimal solution. Indeed, despite these limitations, the loss with respect to the nonparametric specification is rather limited. In order to show what is the efficiency gain that one can attain by our estimation method, we have considered some particular distributions, some of which satisfying the invariance property (3.40) described below.

Remark 3 (Invariance with respect to location and scale parameters) *Let us assume that the class of distribution functions F_x for the conditional returns is a parametric class such that*

$$Q_x(p) = \mu + \sigma\zeta(p), \tag{3.39}$$

for some $\mu \in \mathbb{R}$, $\sigma > 0$ and $\zeta(p)$ a continuous and strictly increasing quantile function. In this case, it is easy to see that the asymptotic bias is equal to

$$\tau(w) - \tau = \sigma \int_0^\alpha \zeta(p) \left[w(p) - \frac{1}{\alpha} \right] \mathrm{d}p \tag{3.40}$$

Moreover, under (3.39), the formula of the asymptotic variance of $\hat{\tau}^{(Q)}$ satisfies, for all $\mathbf{w}$*,*

$$AV_{\mu,\sigma}\left(\hat{\tau}^{(Q)}(\mathbf{w})\right) = \sigma^2 AV_{0,1}\left(\hat{\tau}^{(Q)}(\mathbf{w})\right) \tag{3.41}$$

where the subscript μ, σ in $AV_{\mu,\sigma}$ clearly refers to the location and scale parameters of F_x. In Section 3.6, we study the performances of the efficient estimators, for different distributions of the returns, in terms of the ratio:

$$eAV(\hat{\tau}^{(\cdot)}(\mathbf{w})) = 1 - \frac{AV\left(\hat{\tau}^{(\cdot)}(\mathbf{w})\right)}{AV\left(\hat{\tau}^{(\cdot)}\right)} \tag{3.42}$$

Following (3.41), for $\hat{\tau}^{(Q)}$, whenever the distribution of returns satisfies (3.39), we can limit ourselves, without loss of generality, to standardized values for the parameters of the distribution ($\mu = 0$, $\sigma = 1$), in view of

$$eAV_{\mu,\sigma}\left(\hat{\tau}^{(Q)}(\mathbf{w})\right) = eAV_{0,1}\left(\hat{\tau}^{(Q)}(\mathbf{w})\right).$$

It is clear that, although the above identity holds, the optimal weights $\mathbf{w}^*$ as well as the asymptotic variance of $\hat{\tau}^{(Q)}(\mathbf{w}^*)$ will in general depend on the location and scale parameters.

There are several examples of distributions for $Y \mid X = x \sim F$ that satisfy equation (3.39): first of all, if $F \sim N(\mu, \sigma^2)$, and Φ is the cumulative distribution function of a $N(0,1)$ rv, then equation (3.39) holds with $\zeta(p) = \Phi^{-1}(p)$. Other examples are given by the logistic(μ, σ) or the Gumbel(μ, β) distributions, where $Q(p) = \mu - \sigma \log\left(\frac{1}{p} - 1\right)$ and $Q(p) = \mu - \beta \log\left(\log \frac{1}{p}\right)$ respectively.

The Exponential(λ) distribution (f.i. with support in $(-\infty, 0)$) $f(y) = \lambda e^{\lambda y}$ also satisfies $Q(p) = log(p)/\lambda$ and equation (3.40) writes:

$$\tfrac{1}{\lambda}\left[\int_0^\alpha \log p \, w(p) dp - \log\alpha + 1\right] = 0$$

More generally, if $\zeta_k(p)$ is the quantile function of a Gamma$(k, 1)$ rv, and $F \sim$ Gamma(k, λ), then $Q(p) = \frac{\zeta_k(p)}{\lambda}$, where λ is the scale parameter (whereas k is the shape parameter).

For the generalized Pareto distribution,

$$Y \mid X \sim F(y) = \left(1 - \xi \frac{(y-\mu)}{\sigma}\right)^{-1/\xi}$$

with $\mu \in \mathbb{R}$, the location parameter, $\sigma > 0, \xi \in \mathbb{R}$ the scale and shape parameters respectively, we have $Q(p) = \mu + \sigma \frac{1-p^{-\xi}}{\xi}$ and the ratio eAV in this case depends on variations on the shape parameter ξ only.

3.6 Gain in asymptotic efficiency of WES estimators

In this section, we present a study on the asymptotic efficiency gain of the WICQF and WICDF estimators. The efficiency gain is defined in terms of asymptotic variance AV as

$$eAV = 1 - \frac{AV(\hat{\tau}^{(\cdot)}(\mathbf{w}))}{AV(\hat{\tau}^{(\cdot)})} \tag{3.43}$$

where $\hat{\tau}^{(\cdot)} = \hat{\tau}(\alpha \mid x)$ and $\hat{\tau}^{(\cdot)}(\mathbf{w}) = \hat{\tau}_{\mathbf{w}}(\alpha \mid x)$ denote the ordinary and the weighted versions of any of the WICQF or WICDF estimators. Formulas of AV are derived in (3.15) and (3.24) for the WICQF and the WICDF estimators respectively for $\mathbf{w}$ belonging to $\mathcal{W}_I, \mathcal{W}_{I,U}, \mathcal{W}_P$.

As we have already pointed out in Section 2, we are not able to give general results for arbitrary distributions of the conditional returns of the asset. Anyway, the analysis we present here, although limited to some specifications of the distribution of $Y_t \mid X_t$, permits us to give an idea of the potentiality of the weighted version of the ES estimators, showing that the gain in terms of AV is relevant, especially for distributions with heavy tails.

Specifications of $w(\cdot)$

We consider both a nonparametric and a parametric specifications of the weighting function $w(\cdot)$. The nonparametric specification is proposed in the unconstrained and constrained version (with non-negative weights), as described in Section 3.5.

In order to impose non-negativity of weights, we use a penalty method and change the loss functions into

$$\mathcal{L}_c^Q = \mathcal{L}_u^Q + \gamma_p \left\{ \sum_i (\tilde{\mathbf{w}}_u^\top - \mathbf{w}_u^\top) \boldsymbol{\imath}_I \right\}^{\lambda_p} \tag{3.44}$$

and

$$\mathcal{L}_c^D = \mathcal{L}_u^D + \gamma_p \left\{ \sum_i (\tilde{\mathbf{w}}_u^\top - \mathbf{w}_u^\top) \boldsymbol{\imath}_I \right\}^{\lambda_p} \tag{3.45}$$

Here $\tilde{\mathbf{w}}_u = (|w_1|, |w_2|, \ldots |w_I|)^\top$ is the vector of weights with absolute value operator. Parameters γ_p and λ_p have to be chosen such as to ensure derivability of the penalty function and smoothness of the optimization algorithm at boundaries of the feasible region (see Mulvey *et al.* (32) and Lillo *et al.* (27)). In particular, we set $\gamma_p = 10^{20}$ and $\lambda_p = 1.1$ in most of the cases. A further refining, that we do not apply here, is to set a small value of γ_p for the first iterations of the optimization algorithm and increase it successively.

As for the parametric specification, the weights are non-negative by definition, while the unconstrained minimum AV WICQF and minimum AV WICDF estimators are given by formulas (3.28) and (3.32).

Choice of distributions and computational aspects

Our examples follow distributional characteristics of returns of financial assets (see e.g McNeil and Frey (30)). More precisely, we consider a mixture of normal distributions, the t distribution (with a low number of degrees of freedom), the exponential distribution, the logistic distribution, the Gumbel distribution and the generalized Pareto distribution (GPD).

The first choice is motivated by the fact that any continuous distribution can be approximated arbitrarily well by a mixture of normals, allowing for asymmetry, skewness and heavy tails (see e.g. McLachlan and Peel (29)). The mixture parameters are such that the left tail of the dominating distribution is contaminated. The nested case is $Y_t|X_t \sim \mathcal{N}(0,1)$.

We consider the simple case when the rv $Y_t \,|\, X_t$ has a df that is a mixture of two components, the standard normal distribution $\mathcal{N}(0,1)$ and another $\mathcal{N}(\mu,\sigma^2)$ distribution, with mixing coefficient π_n. This setting can be further generalized to more than two components with any of the moments conditional on some set of k covariates. Here we only condition the mean of the second component on the value of the covariate X_t and specify it as $\mu = \mu(x) = x$. The df of $Y_t \,|\, X_t$ is therefore

$$F(y\,|\,x) = \pi_n\,\Phi(y) + (1-\pi_n)\,\Phi\left(\frac{y-x}{\sigma}\right)$$

where $\Phi(\cdot)$ is the df of the standard normal distribution $\mathcal{N}(0,1)$. Although the quantiles of the mixture do not have a closed-form expression, we evaluate them numerically[1]. As values for the x, we choose extreme left percentiles (the 1^{st}, the 2^{nd} and the 3^{rd}) of X_t. The standard deviation σ and the mixing parameter π_n take values (0.2, 0.3) and 95% respectively.

Besides the mixture of normals, we also study the following distributions: the $t-student$ with 2, 3 or 4 degrees of freedom, the exponential, the logistic, the Gumbel and the generalized Pareto with shape parameter ξ taking values 0.1, 0.2 and 0.3.

Except for the mixture of normals and the t distribution, in all other cases we exploit Remark 3 on invariance of the optimal weights $\mathbf{w}^*$ with respect to the location and scale parameters of the distribution of $Y\,|\,X$. For these examples, we only look at $\mu = 0, \sigma = 1$ or $\lambda = 1$. For WICDF, the above simplification is not possible as in formula (3.24) for the AV, the third term is linear in the standard deviation parameter σ of the density function, rather than the squared value σ^2, while for the second term we cannot factorize σ^2.

All other parameters of the study are set as follows. The level that we consider is $\alpha = 0.1$. For WICQF, I is equal to 25, with the sequence $\{0 = p_0 < p_1 \ldots < p_I\}$ given by $p_i = (\alpha i)/I$. For WICDF, we set the grid points $\{\tilde{y}_1 \leq, \ldots, \tilde{y}_{I_0}\}$ as follows: we set $I = 100$ and the initial grid is defined as $y_i = -10 + i\cdot 0.1$, then retain I_0 points that meet condition $Q_x(0.5\%) < y_i < Q_x(\alpha)$. Moreover, we fix $\tilde{y}_0 = Q_x(0.1\%)$. Remark that I_0 varies according to the distribution of $Y\,|\,X$.

For some distributional assumptions for the $Y_t\,|\,X_t$, we also make a sensitivity analysis of the eAV at parameter I varying from 2 to 200. In particular, we look at the standard normal, the exponential (with parameter $\lambda = 1$), the $t-student$ with 2, 3 and 4 degrees of freedom and the Generalized Pareto with shape parameter ξ set equal to 0.1, 0.2 and 0.3 respectively.

In the parametric specification, $W(\cdot)$ is a mixture of two beta distributions (see (3.5)). We work on subsets of parameters (a_i, b_i), $i = 1,2$ that take 12 equally distant values in the subset $[0.25, 2.5] \times [0.25, 2.5]$, while the mixing coefficient π is computed on basis of (4.14).

[1] Assuming that F is differentiable with strictly positive density $f = F'$, then we employ a Newton-Raphson algorithm, based on iterations of the form $Q^{(i+1)} = Q^{(i)} + [p - F(Q^{(i)})]/f(Q^{(i)}), i = 0,1,2,\ldots,$ where $p = \Pr\{Y_t \leq y|X_t = x\}$. As starting value, we consider $Q^{(0)} = \pi_n\Phi^{-1}(p) + (1-\pi_n)[x + \sigma\Phi^{-1}(p)]$

In both the nonparametric and the parametric specifications, we approximate $\tau_w(\alpha \,|\, x)$ by the sums

$$\tau_{w,I}^{(Q)} = \sum_{i=1}^{I} w_i Q(p_i) = \sum_{i=1}^{I} [W(p_i) - W(p_{i-1})] Q(p_i)$$

$$\tau_{w,I}^{(D)} = Q(\alpha) - \sum_{i=1}^{I_0} (\tilde{y}_{I_0} - \tilde{y}_{i-1})[W(F(\tilde{y}_i)) - W(F(\tilde{y}_{i-1}))]$$

General considerations on the study

Tables 3.1 and 3.2 show results on efficiency gain eAV for minimum AV WICQF and WICDF estimators, with different specifications of the weighting function $w(\cdot)$: (i) nonparametric unconstrained, (ii) nonparametric with non-negativity constraint and (iii) parameterized as a mixture of beta. For each distributional assumption in turn, results are grouped following this order.

The embedding relation between weights specifications is that eAV for the nonparametric unconstrained one is greater or equal to that of the nonparametric constrained which is itself grater or equal to eAV corresponding to the parametric case.

Table 3.1, corresponding to the WICQF estimator, is composed by two panels. The first panel is dedicated to distributions where $w(\cdot)$ and eAV varied with the location and scale parameters. Here enters the mixture of normals and the t distribution. In the second panel, invariance of the efficiency gain to location and scale applies; for WICQF, we look at the normal, the logistic, the Gumbel, the generalized Pareto and the exponential distributions, all with standard values $\mu = 0, \sigma = 1$ or $\lambda = 1$ accordingly. For each distribution, we increase the heaviness of the tail from left side to right side columns. This is verified on decreasing the mean of the contamination distribution in the mixture, decreasing the number of degrees of freedom for the t and increasing ξ, the shape parameter for the GPD.

For the WICDF estimator, the invariance to location and scale parameters does not work. Therefore, Table 3.2 is not divided in panels as the previous one. Moreover we show results for two values of the scale parameter for the Normal, the Logistic, the Gumbel and the exponential distributions.

In Figures 3.1 and 3.2, for both the WICQF and the WICDF estimators and for w having the unconstrained nonparametric specification, we analyze sensitivity of the eAV to parameter I. As range of interest, we look at 100 distinct values in the interval [2 and 200].

Results for the WICQF estimator

Table 3.1 shows results for the WICQF estimator. For the mixture of normals, in general, the higher is the contamination of the original distribution (lower σ_n and x), the higher is eAV. This ranges from 4% for $\sigma_n = 0.3$ and $x = -1.881$ to values around 30% for the parametric specification and more than 50% for the two nonparametric specifications, values that correspond to parameters $\sigma_n = 0.3$ and $x = -2.326$. The values of eAV are high as the constructed example is "ideal" in the sense that we have a local contamination at the quantiles of interest.

For the t-student, eAV ranges from around 2% for the $t[4]$ and approach 20%, for $t[2]$. When $Y_t|X_t$ has a normal distribution $\mathcal{N}(\mu, \sigma^2)$, eAV is very low, around 1%, for all specifications. The gain is also small if $Y_t \mid X_t$ has any of the logistic, the exponential distribution or the Gumbel distributions, with values less than 3%. For the GPD, the gain increases as ξ increases, arriving to 6% for ξ equal to 0.3.

We also make a remark on the effective values of the AV. Comparing across distributions, we see no connection between these values and eAV. We conjecture that what influences eAV is the local heaviness of the tail in the proximity of the quantiles of interest. That is the reason why for the mixture of normals, we have a combination of low AV and high eAV, whilst for the logistic and the exponential it is the other way round. Instead, for the t-student distribution, the high AV is matched with high eAV.

In Figure 3.1, we plot the efficiency gain for different values of parameter I and for different distributions. We remark that eAV in the case of the standard Normal and the Exponential distribution is a concave function of I, with maximum around $I = 10$. For the heavy tailed distribution, the curves also appear concave but in this case they look like logaritmic and bounded above for high values of I. This behavioral distinction among exponential and heavy tailed distributions is to be remarked for the WICDF estimators, too.

Results for the WICDF estimator

Passing to the WICDF estimator, in Table 3.2, for the mixture of normals we note sensitivity of eAV to variation in the conditional mean $\mu(x) = x$ and the standard deviation of the contaminating distribution. The gain has values that range from approximately 5% to 15%. For the t-student distribution, eAV is almost double if compared to the WICQF estimator, with values that are close to 5% for the $t[4]$ and 35% in the case of $t[2]$. When $Y_t|X_t$ has a normal distribution $\mathcal{N}(\mu, \sigma^2)$, eAV is again quite low, around 3% for all specifications. The gain is also small, within the $(0, 10\%]$ interval, if $Y_t \mid X_t$ has any of the logistic, the exponential distribution or the Gumbel distributions, with scale parameter $\sigma = 1$. If, on the other hand, we increase the scale parameter, we notice the increase of eAV, which is to be underlined for the logistic distribution.

As for the value of AV across estimators, we remark much similarity. Exceptions are for the distributions with very heavy tails, the t[2] and the GPD[0,1,0.3] where the AV of the WICDF estimator is higher.

In Figure 3.2, we show the sensitivity of the efficiency gain at different values of I. As in the case of the WICQF estimator, we also remark that for the standard Normal and the Exponential distribution the curves are concave with the maximum around $I = 100$, whilst for the t-student and the Generalized Pareto, the curves appear also concave with an increase in the efficiency gain that dies out as I increases.

Table 3.1: The estimated asymptotic variance of the ordinary WICQF estimator $\hat{\tau}^{(Q)}(\alpha \mid x)$ and the correspondent efficiency gain eAV of the weighted version

		$w(\cdot)$	AV	eAV	AV	eAV	AV	eAV
					$Y_t \mid X_t \sim \pi_n \mathcal{N}(0,1) + (1-\pi_n)\mathcal{N}(x, \sigma_n^2)$			
π_n	σ_n		$x = -1.881$		$x = -2.054$		$x = -2.326$	
.95	.2	(i)	2.147	25.7%	2.570	30.1%	3.995	57.3%
		(ii)		20.7%		20.1%		28.7%
		(iii)		16.7%		15.7%		16.7%
.95	.3	(i)	2.583	4.3%	2.979	5.9%	4.313	25.8%
		(ii)		4.2%		4.9%		12.6%
		(iii)		3.6%		3.2%		10.0%
					$Y_t \mid X_t \sim t[r]$			
			$r = 4$		$r = 3$		$r = 2$	
		(i)	17.319	2.7%	31.497	6.6%	113.249	17.5%
		(ii)		2.5%		5.7%		13.3%
		(iii)		2.4%		5.3%		11.9%
					$Y_t \mid X_t \sim GPD(\mu, \sigma, \xi)$			
μ	σ		$\xi = .1$		$\xi = .2$		$\xi = .3$	
0	1	(i)	35.943	.3%	75.817	2.2%	164.282	5.8%
		(ii)		.3%		1.9%		4.9%
		(iii)		.3%		1.8%		4.5%
μ	σ		$Y_t \mid X_t \sim \mathcal{N}(\mu, \sigma)$		$Y_t \mid X_t \sim Logistic(\mu, \sigma)$		$Y_t \mid X_t \sim Gumbel(\mu, \sigma)$	
0	1	(i)	3.601	1.4%	18.997	.1%	1.608	2.9%
		(ii)		1.4%		.1%		2.9%
		$(iiii)$		1.4%		.1%		2.9%
	λ				$Y_t \mid X_t \sim Exponential(\lambda)$			
	1	(i)	17.474	.2%				
		(ii)		.2%				
		(iii)		.2%				

Table 3.2: The estimated asymptotic variance of the ordinary WICDF estimator $\hat{\tau}^{(D)}(\alpha\,|\,x)$ and the correspondent efficiency gain eAV of the weighted version

		$w(\cdot)$	AV	eAV	AV	eAV	AV	eAV
					$Y_t\,|\,X_t \sim \pi_n\mathcal{N}(0,1)+(1-\pi_n)\mathcal{N}(x,\sigma_n^2)$			
π_n	σ_n		$x=-1.881$		$x=-2.054$		$x=-2.326$	
.95	.2	(i)	1.795	11.4%	2.557	12.3%	2.365	17.6%
		(ii)		8.5 %		11.2%		13.4%
		(iii)		2.4 %		5.6		10.6%
.95	.3	(i)	1.675	10.2%	1.937	9.9%	2.477	9.2%
		(ii)		8.7 %		8.5%		9.2%
		(iii)		3.7 %		4.3%		7.5%
					$Y_t\,|\,X_t \sim t[r]$			
			$r=4$		$r=3$		$r=2$	
		(i)	16.958	15.1%	38.629	21.7%	211.218	36.9%
		(ii)		13.2%		19.2%		27.7%
		(iii)		11.7%		15.1%		18.8%
					$Y_t\,|\,X_t \sim GPD(\mu,\sigma,\xi)$			
μ	σ		$\xi=.1$		$\xi=.2$		$\xi=.3$	
0	1	(i)	29.212	7.9%	75.532	15.6%	341.084	41.9%
		(ii)		6.4%		12.9%		22.5%
		(iii)		5.8%		11.6%		12.6%
			$Y_t\,|\,X_t \sim \mathcal{N}(\mu,\sigma)$		$Y_t\,|\,X_t \sim Logistic(\mu,\sigma)$		$Y_t\,|\,X_t \sim Gumbel(\mu,\sigma)$	
μ	σ							
0	1	(i)	2.079	3.3%	12.541	3.8%	.803	7.5%
		(ii)		3.3%		3.5%		7.0%
		(iii)		1.5%		1.9%		4.8%
0	2	(i)	7.797	4.9%	55.645	12.7%	3.202	8.6%
		(ii)		4.3%		6.8%		8.1%
		(iii)		2.3%		4.1%		6.9%
					$Y_t\,|\,X_t \sim Exponential(\lambda)$			
			$\lambda=1$		$\lambda=2$			
		(i)	12.064	3.8%	3.062	3.7%		
		(ii)		3.6%		1.5%		
		(iii)		2.0%		1.5%		

Figure 3.1: The efficiency gain eAV of the WICQF estimator

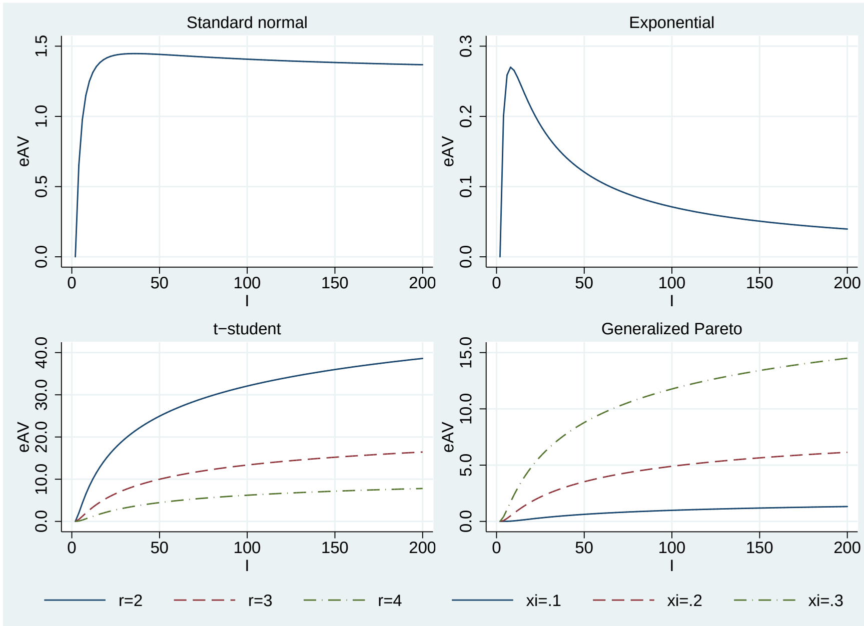
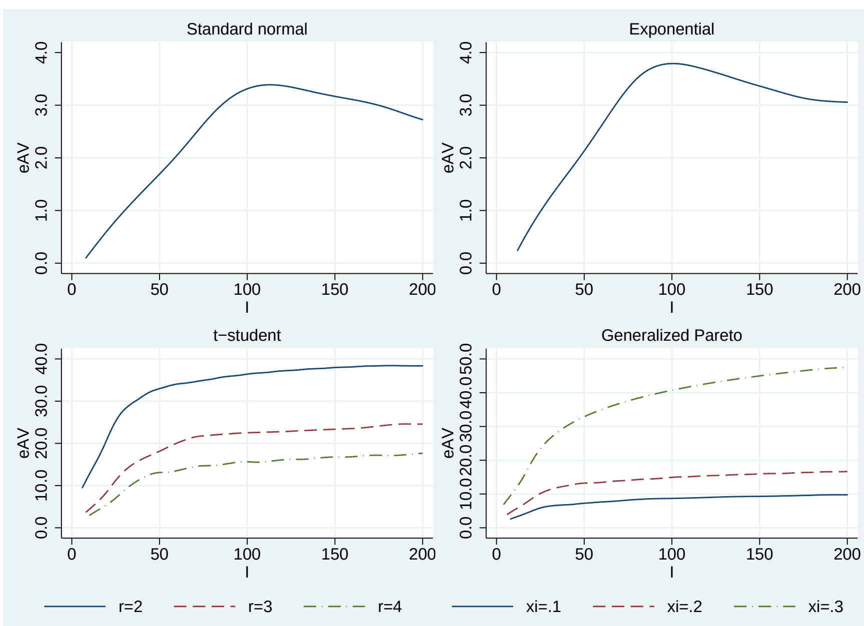


Figure 3.2: The efficiency gain eAV of the WICDF estimator



3.7 Monte Carlo evidence

In this section we present some Monte Carlo evidence on the sampling properties of the WICQF $\hat{\tau}^{(Q)}_{w,I}(\alpha \,|\, x)$ and the WICDF $\hat{\tau}^{(D)}_{w,I}(\alpha \,|\, x)$ estimators. We are interested wether the weighting function w that minimizes the AV has a positive impact on the precision of the estimator in finite samples and look at the standard deviation of the Monte Carlo distribution for both the uniformly and non-uniformly weighted estimators.

Firstly, we briefly recall the results of the Monte Carlo study performed by Peracchi and Tanase (36), although these are limited to the uniform weighted versions of the estimators. Compared by the RMSE of the Monte Carlo distributions, the conditional estimators are tested on both homoskedastic and heteroskedastic models and results support the asymptotic normality theory. Across the semiparametric estimators, the ICQF estimator (corresponding to (4.10) with uniform weights) tends to perform better.

In our Monte Carlo experiments, we simulate samples of size 250, 500 and 1000 and the number of Monte Carlo replications is set equal to 1000. The software that we use is the statistical package Stata, version 10.

We compare alternative estimators of the ES assuming that the conditional distribution of the rv $Y_t \,|\, X_t$ is either Normal or generalized Pareto, with the location and scale parameters depending on some vector of regressors X_t, with dimension k. We take the simple case $k = 1$. The mean of the cdf depends linearly on a constant and regressor X_t. Both a homoskedastic and heteroskedastic versions of the model are considered, in the latter case non only the mean but also the scale depending linearly on the regressor. For simplicity, the rv X_t is assumed to have a standard normal distribution.

Parameters of the conditional distribution functions are chosen such as to impose different degrees of heaviness of the distribution tails. In our first example, we assume that the rv $Y_t \,|\, X_t$ is distributed as

$$Y_t \,|\, X_t \sim \mathcal{N}(-2 + a \cdot X, (1 + b \cdot X)^2), X_t \sim \mathcal{N}(0, 1)$$

while in the second

$$Y_t \,|\, X_t \sim \text{GPD}(a \cdot X, 1 + b \cdot X, \xi), X_t \sim \mathcal{N}(0, 1)$$

We set $a = 0.5$, while b takes values 0 and 0.25 for the homoskedastic and the heteroskedastic models respectively. In the GPD example, the shape parameter ξ is set equal to 0.1, 0.2 and 0.3. This facilitates comparison with the study on the asymptotic efficiency gain that we presented in the previous section.

The other parameters of the estimating exercise are: level α is 10% and the number of points I is arbitrarily set equal to 10, 20 and 40 corresponding to sample sizes of 250, 500 and 1000 in the case of the WICQF estimator, while for the WICDF estimator, analogously to the previous section, we consider I points $y_i = -10 + i \cdot 0.1$ and then retain I_0 points that meet condition $Y_{(1)} < \tilde{y}_i < \hat{Q}_x(\alpha)$, that is between the first order statistic of Y_t and the estimated α level quantile of the rv $Y_t \,|\, X_t = x$. Moreover, we fix $\tilde{y}_0 = Y_{(1)}$ and $\tilde{y}_I = \hat{Q}_x(\alpha)$. Parameter I is set equal to 25, 50 and 100 corresponding to sample sizes of 250, 500 and 1000 and, within each simulation exercise, we double the value if necessary until $I_0 \geq 5$.

The weighting function is specified as nonparametric with unbounded domain for the weights, with a constraint on the bias of the estimator. They are derived maximizing the Lagrangean (3.44) and (3.45) for the WICQF and WICDF estimators respectively. As our interest is purely on the effect of the non-uniform weighting scheme avoiding to mingle results with estimation precision of the cdf parameters, we assume that for each simulated sample, a, b (and ξ) are known. Therefore, for the WICQF estimator, in (3.44) we use the true values of $Q_x(p_i), i = 1 \dots I$, while in $\Omega_{1;i,j}$, we estimate $\hat{\mathbf{J}}_1(p_i)$ as specified in Corollary 1. For the WICDF estimator, in (3.44), we also use the true values of $F_x(y_j), y_j = 1 \dots J$, while in $\Omega_{2;j,k}$ and $\Omega_{3;j}$, we estimate $\hat{\mathbf{J}}_2(y_i)$, $\Sigma_{2;j,k}$ and $\Sigma_{3,j}$ with the sample counterparts specified in Corollary 3. Estimator $\widehat{AV}(\hat{\tau}_w^{(D)}(\alpha \,|\, x))$ is therefore robust to non-linearity of log odds in the regressor.

Details on the sampling efficiency gain for different sample sizes, parameters of the cdf and covariate values are given in Tables 3.3 and 3.4. Each table reports a, b and ξ and the efficiency gain $eAV(\hat{\tau}^{(\cdot)}(\mathbf{w}))_{MC}$, that is the sampling analog of the asymptotic efficiency gain $eAV(\hat{\tau}^{(\cdot)}(\mathbf{w}))$ defined in (3.43). The sampling efficiency gain is measured in terms of the variances $\sigma^2_{MC}(\hat{\tau}(\mathbf{w}))$ and $\sigma^2_{MC}(\hat{\tau})$ of the Monte Carlo distributions of the non-uniformly and uniformly weighted estimators:

$$eAV(\hat{\tau}^{(\cdot)}(\mathbf{w}))_{MC} = 1 - \frac{\sigma^2_{MC}(\hat{\tau}(\mathbf{w}))}{\sigma^2_{MC}(\hat{\tau})} \tag{3.46}$$

An overall look show discrepancies between behavior of the two estimators in terms of sampling efficiency gain for different distributional assumptions and covariate values. Results on the WICQF estimator follow those on the asymptotic efficiency gain study, with positive $eAV(\hat{\tau}^{(Q)}(\mathbf{w}))_{MC}$ for almost all sets of parameters, while for WICDF, we encounter cases, even corresponding to heavy tailed distributions, where $eAV(\hat{\tau}^{(D)}(\mathbf{w}))_{MC}$ is negative.

As mentioned before, for the WICQF estimator, $eAV(\hat{\tau}^{(Q)}(\mathbf{w}))_{MC}$ is positive except for the heteroskedastic version of the normal distribution example. In absolute values, $eAV(\hat{\tau}^{(Q)}(\mathbf{w}))_{MC}$ is close to $eAV(\hat{\tau}^{(Q)}(\mathbf{w}))$, for $I = 20$. In general, the higher the heaviness of the tail, the higher is eAV_{MC}. In the heteroskedastic model, if compared to the homoskedastic one, the weighting function has a larger positive impact in terms of efficiency gain. Comparing across different values of x, for the GPD distribution, the heteroskedastic example point an efficiency gain that is decreasing in x, while for the homoskedastic case, this is not verified.

For the WICDF, results do not show a clear pattern on the sensitivity of $eAV(\hat{\tau}^{(D)}(\mathbf{w}))_{MC}$ at different degrees of tail heaviness of the cdf. If the cdf is normal, $eAV(\hat{\tau}^{(D)}(\mathbf{w}))_{MC}$ is high for all values of covariate x and sample sizes T. Moreover, $eAV(\hat{\tau}^{(D)}(\mathbf{w}))_{MC}$ appears to be convex in the covariate value. Also, for the GPD example, we note negative sampling efficiency gains when x take as values the first quartile or the median of the distribution of X_t and the shape parameter ξ is low. Depending on the value of the covariate, the sampling efficiency gain can be monotone with respect to the sample size T and the shape parameter ξ. For example, when $x = -.674$, for all distributional assumptions, we see that $eAV(\hat{\tau}^{(D)}(\mathbf{w}))_{MC}$ increases with T and ξ. This is paired with high standard deviation of the Monte Carlo distribution.

Table 3.3: Efficiency gain in terms of standard deviation $\sigma_{MC}(\hat{\tau})$ of Monte Carlo distributions for the WICQF estimator $\hat{\tau}^{(Q)}_{w,I}(\alpha \,|\, X_t = x)$

b	ξ	T	I	$\sigma_{MC}(\hat{\tau})$	eAV_{MC}	$\sigma_{MC}(\hat{\tau})$	eAV_{MC}	$\sigma_{MC}(\hat{\tau})$	eAV_{MC}
				$x = -.674$		$x = 0$		$x = .674$	
				$Y_t \,|\, X_t \sim \mathcal{N}(-2 + .5 \cdot X, (1 + b \cdot X)^2), X_t \sim \mathcal{N}(0,1)$					
0		250	10	.138	.7%	.114	1.7%	.142	1.2%
0		500	20	.101	1.3%	.083	1.0%	.101	1.3%
0		1000	40	.074	.7%	.060	.6%	.070	.5%
.25		250	10	.106	-3.4%	.136	-.2%	.191	- .4%
.25		500	20	.080	-5.8%	.097	-2.9%	.134	- .4%
.25		1000	40	.058	-6.3%	.068	-3.4%	.092	- .9%
				$Y_t \,|\, X_t \sim \text{GPD}(.5 \cdot X, 1 + b \cdot X, \xi), X_t \sim \mathcal{N}(0,1)$					
0	.1	250	10	.421	.1%	.352	- .1%	.426	- .1%
0	.1	500	20	.313	.4%	.264	.5%	.316	.2%
0	.1	1000	40	.238	.7%	.194	1.1%	.227	1.2%
0	.2	250	10	.604	1.6%	.509	1.1%	.606	.5%
0	.2	500	20	.454	2.3%	.387	3.2%	.460	2.1%
0	.2	1000	40	.351	3.6%	.289	5.2%	.335	5.1%
0	.3	250	10	.879	4.8%	.746	4.3%	.871	2.4%
0	.3	500	20	.670	6.4%	.578	8.4%	.680	6.0%
0	.3	1000	40	.527	8.9%	.439	12.1%	.504	11.5%
.25	.1	250	10	.326	.3%	.364	- .1%	.496	- .2%
.25	.1	500	20	.235	.8%	.271	.4%	.367	.1%
.25	.1	1000	40	.178	1.2%	.196	1.1%	.260	.7%
.25	.2	250	10	.473	2.5%	.524	1.0%	.702	0.0%
.25	.2	500	20	.347	4.2%	.397	3.1%	.530	1.7%
.25	.2	1000	40	.266	5.6%	.290	5.1%	.381	3.6%
.25	.3	250	10	.701	7.0%	.767	4.4%	1.007	2.0%
.25	.3	500	20	.522	10.6%	.594	8.5%	.780	5.5%
.25	.3	1000	40	.406	13.1%	.441	12.2%	.570	9.2%

Table 3.4: Efficiency gain in terms of standard deviation $\sigma_{MC}(\hat{\tau})$ of Monte Carlo distributions for the WICDF estimator $\hat{\tau}^{(D)}_{w,I}(\alpha \mid X_t = x)$

b	ξ	T	I	$\sigma_{MC}(\hat{\tau})$	eAV_{MC}	$\sigma_{MC}(\hat{\tau})$	eAV_{MC}	$\sigma_{MC}(\hat{\tau})$	eAV_{MC}
				$x = -.674$		$x = 0$		$x = .674$	
				$Y_t \mid X_t \sim \mathcal{N}(-2 + .5 \cdot X, (1 + b \cdot X)^2), X_t \sim \mathcal{N}(0, 1)$					
0		250	25	.387	19.1%	.063	18.7%	.028	20.3%
0		500	50	.322	17.7%	.048	18.3%	.021	19.9%
0		1000	100	.257	16.9%	.034	18.2%	.014	19.4%
.25		250	25	.394	24.9%	.062	19.7%	.018	21.1%
.25		500	50	.350	26.5%	.045	19.3%	.013	20.9%
.25		1000	100	.304	28.9%	.032	19.1%	.009	20.8%
				$Y_t \mid X_t \sim \text{GPD}(a \cdot X, 1 + b \cdot X, \xi), X_t \sim \mathcal{N}(0, 1)$					
0	.1	250	25	2.269	-2.1%	.247	-12.7%	.056	29.1%
0	.1	500	50	2.212	-5.8%	.187	-31.7%	.039	28.9%
0	.1	1000	100	1.953	-9.7%	.133	-19.2%	.027	29.7%
0	.2	250	25	3.600	10.3%	.375	0.9%	.073	29.2%
0	.2	500	50	3.215	13.5%	.282	-2.2%	.051	28.7%
0	.2	1000	100	2.600	20.9%	.202	2.9%	.035	29.0%
0	.3	250	25	3.651	16.7%	.541	9.4%	.100	26.8%
0	.3	500	50	2.630	22.6%	.403	6.8%	.070	26.1%
0	.3	1000	100	1.579	30.6%	.289	5.7%	.049	26.2%
.25	.1	250	25	2.402	-3.9%	.243	-18.4%	.036	30.0%
.25	.1	500	50	2.298	-2.7%	.178	-51.9%	.024	29.7%
.25	.1	1000	100	2.075	2.0%	.129	-37.4%	.017	30.6%
.25	.2	250	25	3.638	12.3%	.368	-1.2%	.052	28.2%
.25	.2	500	50	3.182	19.2%	.274	-5.2%	.036	27.9%
.25	.2	1000	100	2.530	27.4%	.199	-2.6%	.026	28.6%
.25	.3	250	25	3.606	18.7%	.531	9.2%	.078	25.6%
.25	.3	500	50	2.531	26.8%	.391	6.3%	.053	25.0%
.25	.3	1000	100	1.519	33.4%	.285	6.6%	.039	25.4%

3.8 Empirical application

The empirical application we develop is based on daily data on excess returns of 6 European stock indexes and a set of financial and macro variables. For each of the stock indexes, we estimate the conditional WES, both with uniformly distributed weights (ordinary ES) and with non-uniform weights, for a level α equal to 10%.

We conditionally estimate WES with the semiparametric estimators WICQF $\hat{\tau}_w^{(Q)}(\alpha \,|\, x)$ and WICDF $\hat{\tau}_w^{(D)}(\alpha \,|\, x)$. We follow McNeil and Frey (30) and compare the predictive capacity of the weighted and the ordinary ES estimators, not formally constructing a test but only looking at the distributions of the forecast error (FE) of the ES estimators for all quantile violation events. We define the FE as the absolute difference between the excess return and the estimated ES, using information up to period T, and the quantile violation as the event when the excess return from T to $T+1$ is lower than the predicted VaR of level (1-α) for period $T+1$.

Our raw data are daily from December 30, 1994, to December 31, 2007. As dependent variables we have the excess returns on the national composite indexes Xetra Dax 30 Frankfurt, CAC 40 Paris, S&P MIB 30 Milan, IBEX 35 Madrid and AEX Amsterdam. The set of predictors are real and financial variables that include micro and macroeconomic information. We estimate the ES of the composite stock indexes across companies by conditioning on prices of inputs (raw materials, credit conditions through short-term interest rate), on a balance sheets indicator (the dividend yield) and on general conditions regarding the financial markets and the economic cycle (exchange rate, government bond yields and credit risk). All data is measured as closing price.

Table 3.5: Description and sources of the composite indexes and covariates

Code	Description	Source	Provider
dax	Xetra Dax 30 Frankfurt Price Index	Frankfurt exchange	Bloomberg
cac40	CAC 40 Paris Price Index	Paris exchange	Bloomberg
mib30	S&P MIB 30 Milan Price Index	Milan exchange	Bloomberg
ibex35	IBEX 35 Madrid Price Index	Madrid exchange	Bloomberg
aex	AEX Amsterdam Price Index	Amsterdam exchange	Bloomberg
stoxx50	DJ Euro Stoxx 50 Price Index	DJ Eurostoxx	Bloomberg
comm	Goldman Sachs non-energy index	Goldman Sachs	Bloomberg
dyeu	DJ Euro Stoxx equity dividend yield	DJ Eurostoxx	Datastream, Bloomberg
gb5y	German Govt Bond Yield	Bloomberg	Bloomberg
fx	EUR/USD exchange rate	ECB	Bloomberg
oil	Oil Nymex future price (in US $)	Nymex	Bloomberg
re3m	Euribor 3M	European Banking Fdr.	Datastream, Bloomberg
ryeu	Lehman Euro Corp BBB Yield	Lehman Brothers	Datastream, Bloomberg
yd10e	10y Treasury Bond Yield Germany	Bloomberg	Bloomberg

The weighting function $w(\cdot)$ associated to any WES estimator is such that it minimizes the AV of the estimator the assumption that $Y_t \,|\, X_t \sim GPD(0, 1, \xi)$. We make this assumption as the GPD is often used as characterizing financial assets with fat tails and moreover, in the study, we obtained good results in terms of asymptotic efficiency gain. We assume the simple model with zero mean and the shape parameter not depending on the covariates.

Table 3.6: Transformations and summary statistics of the composite indexes and covariates

Variable	Description	Transformation	Mean	SD	$Q_{99} - Q_{01}$
DAX	Xetra Dax 30 return	$\ln(dax_t/dax_{t-1})$	$2.6 \cdot 10^{-4}$	$1.4 \cdot 10^{-2}$	$7.8 \cdot 10^{-2}$
CAC40	CAC 40 return	$\ln(cac40_t/cac40_{t-1})$	$3.5 \cdot 10^{-4}$	$1.3 \cdot 10^{-2}$	$7.4 \cdot 10^{-2}$
MIB30	S&P MIB 30 return	$\ln(mib30_t/mib30_{t-1})$	$3.5 \cdot 10^{-4}$	$1.3 \cdot 10^{-2}$	$6.8 \cdot 10^{-2}$
IBEX35	IBEX 35return	$\ln(ibex35_t/ibex35_{t-1})$	$5.8 \cdot 10^{-4}$	$1.3 \cdot 10^{-2}$	$6.9 \cdot 10^{-2}$
AEX	AEX return	$\ln(aex_t/aex_{t-1})$	$1.9 \cdot 10^{-4}$	$1.3 \cdot 10^{-2}$	$7.6 \cdot 10^{-2}$
STOXX50	DJ Euro Stoxx 50 return	$\ln(stoxx50_t/stoxx50_{t-1})$	$3.0 \cdot 10^{-4}$	$1.3 \cdot 10^{-2}$	$7.4 \cdot 10^{-2}$
ECOMM	Commodity price log diff	$\ln(comm_t/comm_{t-1})$	$2.3 \cdot 10^{-4}$	$0.7 \cdot 10^{-2}$	$3.4 \cdot 10^{-2}$
EDY	Dividend yield	$\ln(dyeu_t)$	0.79	0.27	1.01
EFX	EUR/USD log diff	$\ln(fx_t/fx_{t-1})$	$-0.5 \cdot 10^{-4}$	$0.6 \cdot 10^{-2}$	$3.1 \cdot 10^{-2}$
EOIL	Oil price log diff	$\ln(oil_t/oil_{t-1})$	$7.5 \cdot 10^{-4}$	$2.2 \cdot 10^{-2}$	$11.0 \cdot 10^{-2}$
ERSP	Risk spread	$ryeu_t - gb5y_t$	1.15	0.45	1.91
ESP	Term Spread	$yd10e_t - re3m_t$	1.46	0.90	3.75

This facilitates maximum likelihood estimation of ξ, fitted only on the negative values of Y_t, previously standardized by dividing them to the sample standard deviation. More precisely, let $y_t^\star$ denote the negative observations of Y_t. Then

$$\hat{\xi} = \max_{\mathbb{R}^+} \sum_{t=1}^{T} [\ln\left(1 + \xi \cdot (-y_t^\star)^{(\xi^{-1}-1)}\right) \cdot \mathbb{1}\{y_t < 0\}]$$

The optimal weighting function $\mathbf{w}$ is derived in a similar manner as in the Monte Carlo study. For each rolling window sample, we solve (3.28) for the WICQF and (3.32) for the WICDF respectively. For the WICQF estimator (4.10), we set $p_i = \alpha \cdot i/I$, with $i = 1, \ldots, I = 20$, therefore the p_i's are equally spaced between $p_1 = \alpha/I$ and $p_I = \alpha$. For the WICDF estimator (4.8), we set $I = 50$ points $y_i = -10 + i \cdot 0.1$ and successively retain I_0 of them that fall in the interval $(Y_{(1)}, \hat{Q}_x(\alpha))$, where $Y_{(1)}$ is the first order statistic of Y_t and $\hat{Q}_x(\alpha)$ is the estimated α level quantile of the rv $Y_t \,|\, X_t = x$. We also define $\tilde{y}_0 = Y_{(1)}$ and set $\tilde{y}_{I_0} = \hat{Q}_x(\alpha)$.

All predictors are lagged one period. After deriving $\mathbf{w}$, the model is fitted repeatedly for each of the 2147 rolling windows. The estimated model and the last available value of the covariates are then used to predict the one-step ahead WES over the next day. The first rolling window goes from January 3, 1995, to June 7, 1996 and the first estimate is the predicted WES between June 10, 1996 and June 11, 1997. The $k-$dimensional vector x is on day June 10, 1997.

In Tables 3.7 and 3.8, for all indexes, we report summary statistics of the empirical distribution of the WICQF and the WICDF estimators over the rolling windows and also of the empirical distribution of the FE corresponding to each estimator. We show the mean, the standard deviation and the range between the 1st and the 99th percentiles, all expressed in percentage points. The percentage of quantile violation cases, for each index and estimator is approximatively 12%.

The summary statistics between the uniformly weighted and the non-uniformly weighted estimators are very similar as the "estimated" weighting functions are close to the uniform scheme. Across indexes, the standard deviation of the empirical distribution of both the WICQF and WICDF estimators varies within approximatively 0.9% and 1.1%. The quantile range has values above 4% and it is smaller for the weighted estimator if compared to the ordinary estimator. Regarding the FE, we notice that the mean value is negative for the WICQF estimators while it is positive (and much larger, in absolute value) for the WICDF estimators. Moreover, the mean corresponding to the weighted estimator is lower in absolute value with around 0.1% for the WICQF, whilst it is lower with around 0.3% for the WICDF. In other words, even if the weighted estimator is not characterized by a lower standard deviation of the empirical distribution, it appears to have a better forecast precision translated into less severe underprediction for the WICQF and less severe overprediction for the WICDF estimator.

Table 3.7: Summary statistics for the one-step ahead predicted shortfall

Estimator	Mean	SD	$Q_{.99}$-$Q_{.01}$	Mean	SD	$Q_{.99}$-$Q_{.01}$
		$XetraDax30$			$CAC40$	
$\hat{\tau}^{(Q)}(\alpha \mid x)$	-2.359	1.072	4.833	-2.164	0.899	4.189
$\hat{\tau}_w^{(Q)}(\alpha \mid x)$	-2.365	1.079	4.864	-2.157	0.896	4.178
$\hat{\tau}^{(D)}(\alpha \mid x)$	-2.578	1.066	4.891	-2.349	0.858	3.987
$\hat{\tau}_w^{(D)}(\alpha \mid x)$	-2.553	1.098	5.028	-2.321	0.870	4.019
		$S\&PMIB30$			$IBEX35$	
$\hat{\tau}^{(Q)}(\alpha \mid x)$	-2.090	1.082	4.666	-2.082	1.021	4.596
$\hat{\tau}_w^{(Q)}(\alpha \mid x)$	-2.090	1.081	4.653	-2.078	1.019	4.613
$\hat{\tau}^{(D)}(\alpha \mid x)$	-2.282	1.070	4.854	-2.275	0.998	4.963
$\hat{\tau}_w^{(D)}(\alpha \mid x)$	-2.270	1.106	4.980	-2.248	1.029	5.176
		AEX			$DJEuroStoxx$	
$\hat{\tau}^{(Q)}(\alpha \mid x)$	-2.136	1.000	5.043	-2.128	0.936	4.474
$\hat{\tau}_w^{(Q)}(\alpha \mid x)$	-2.135	1.001	5.057	-2.125	0.934	4.468
$\hat{\tau}^{(D)}(\alpha \mid x)$	-2.370	0.983	4.807	-2.337	0.923	4.268
$\hat{\tau}_w^{(D)}(\alpha \mid x)$	-2.346	1.005	4.746	-2.312	0.948	4.313

Table 3.8: Summary statistics for the one-step ahead forecast error

Estimator	Mean	SD	$Q_{.99}$-$Q_{.01}$	Mean	SD	$Q_{.99}$-$Q_{.01}$
		$XetraDax30$			$CAC40$	
$\hat{\tau}^{(Q)}(\alpha \mid x)$	-0.072	0.754	3.479	-0.045	0.722	3.444
$\hat{\tau}_w^{(Q)}(\alpha \mid x)$	-0.067	0.753	3.460	-0.052	0.721	3.409
$\hat{\tau}^{(D)}(\alpha \mid x)$	0.229	0.788	4.354	0.173	0.700	3.421
$\hat{\tau}_w^{(D)}(\alpha \mid x)$	0.207	0.832	4.828	0.140	0.697	3.418
		$S\&PMIB30$			$IBEX35$	
$\hat{\tau}^{(Q)}(\alpha \mid x)$	-0.071	0.706	4.219	-0.052	0.783	4.323
$\hat{\tau}_w^{(Q)}(\alpha \mid x)$	-0.072	0.706	4.147	-0.058	0.783	4.330
$\hat{\tau}^{(D)}(\alpha \mid x)$	0.196	0.734	3.982	0.180	0.786	4.930
$\hat{\tau}_w^{(D)}(\alpha \mid x)$	0.173	0.741	4.020	0.152	0.785	4.825
		AEX			$DJEuroStoxx$	
$\hat{\tau}^{(Q)}(\alpha \mid x)$	-0.163	0.769	3.956	-0.093	0.695	3.416
$\hat{\tau}_w^{(Q)}(\alpha \mid x)$	-0.166	0.769	3.898	-0.099	0.692	3.387
$\hat{\tau}^{(D)}(\alpha \mid x)$	0.123	0.819	4.826	0.178	0.697	3.539
$\hat{\tau}_w^{(D)}(\alpha \mid x)$	0.093	0.820	4.911	0.154	0.695	3.783

3.9 Conclusions

We obtain an asymptotic efficiency gain of ES conditional estimators proposed in the previous Chapter. Our approach counts on introducing a weighting function that maps the cdf or equivalently the cqf. The efficiency appears to be significant for heavy tailed distributions and this opens way to applications on risk estimation for asset returns characterized by such distributions. The Monte Carlo exercise shows that the weighting function has an overall positive effect on estimation precision mostly in the case of the WICQF estimator. We also develop an application on daily financial data. As the weighting function is close to the uniform, we obtain quite similar estimates for both the ordinary and the weighted versions of the ES estimator, with marginally better forecast precision for the latter class.

Chapter 4

Asset allocation with expected shortfall risk constraint

4.1 Introduction

The $\alpha-$level expected shortfall (ES) of an asset Y is the expected loss that can be incurred on holding the asset, conditional on the loss being more severe than the α level quantile $Q(\alpha)$, a measure known as the (1-α)- percent Value at Risk (VaR)

$$\tau(\alpha) = \frac{1}{\alpha} \int_{-\infty}^{Q(\alpha)} y \, \mathrm{d}F(y) = \frac{1}{\alpha} \int_{0}^{\alpha} Q(p) \, \mathrm{d}p \tag{4.1}$$

where $F(y)$ and $Q(p)$ are respectively the distribution function (df) and the quantile function (qf) of the random variable (rv) Y. The expected value of a random variable can be written in terms of the ES by setting $\alpha = 100\%$, that is $\mu = \int_0^1 Q(p) \, \mathrm{d}p = \tau(1)$. This last observation allows us to adapt the estimators for ES and use them as estimators for the mean.

Unlike VaR, the ES is a coherent risk measure (see Artzer et al (7) and Acerbi and Tasche (2)). However, both ES and VaR are used as risk constraints in asset allocation models. We recall here Basak and Shapiro (9), Alexander and Baptista (3), Cherny (13), Mansini et al (28), Bassett et al (10), Cuoco et al (15) and Gundel and Weber (21) that solve portfolio optimization problems minimizing a well defined loss function written in terms of the mean return and a constraint on either ES or VaR.

In Chapter 2, we discuss basic form of estimators for ES proposed by Peracchi and Tanase (36). Then, in Chapter 3, following Leorato et al (26), we write a generalized weighted representation of ES and analytically derive the weighting function such as to obtain estimators of ES with minimum asymptotic variance.

In this Chapter we accomplish a broad study of conditional estimation of ES by focusing on the forecasting properties of the estimators. We achieve this by means of a generalization of the ES estimators. The weighting function that modifies the cdf of returns is numerically derived such as to minimize the forecast error of weighted estimators. This is defined, following McNeil and Frey (30), as a measure of deviation from the ES in the case of quantile violation events, that is, when the observed return is less than the estimated VaR.

The second objective of this Chapter is to compare results of a simple asset allocation model that maximizes expected return with a loss constraint on either ES or VaR. For example, we are interested in stability of the portfolio weights over time, or equivalently the degree of variability. This is of concern from the point of view of transaction and forecasting applications. We give a stylized example proving that the allocation that uses ES rather that VaR as risk constraint is characterized by higher stability. The asset allocation model is solved by means of a numerical algorithm. This does not fully exploit the convexity property of the optimization problem of the underlying model, but it is a simple iterative algorithm that generates random sets of portfolio weights and chooses the optimal one according to a well defined loss function. The loss function that is minimized within the asset allocation model is written in terms of the mean and the risk measure. In solving the model, the risk measures are estimated by analog estimators constructed assuming linearity in the covariates, either of the conditional quantiles or of the log-odds.

In the empirical section of the paper, we use real financial data and construct optimal portfolios that solve the asset allocation problem where the mean return and the risk constraint are estimated with the estimators introduced in Chapter 2 and generalized here. We compare the return-to-risk profile and the stability of the portfolio weights with either ES, WES or VaR risk constraints.

The rest of the chapter is organized as follows. After defining the ES and mean return, we present their estimators. Then we generalize the ES to the weighted expected shortfall (WES) and their correspondent estimators. The fourth section is dedicated to the definition of the forecast error of the weighted estimator of ES and derivation of the weighting function that minimizes the forecast error. Then we formulate the asset allocation problem, present the numerical algorithm of the constrained optimization problem and define our measures of allocation stability. The last section is dedicated to the empirical application.

4.2 The weighted mean and the weighted ES: definition and estimation

In this section we recall the definition of the conditional expected shortfall and the integrated cdf and integrated cqf estimators proposed by Peracchi and Tanase (36). We also show how these estimators can be used to estimate the mean. Moreover we generalize the expected shortfall and present a weighted version of the estimators for ES. The generalization is achieved by means of a well defined weighting function $w(\cdot)$ that modifies the distribution of the random variable without affecting consistency of the weighted estimators for ES. The estimators of WES and of the weighted mean are simply weighted versions of the integrated cdf and cqf estimators.

Assume that the rv Y_t has continuous and strictly increasing distribution function $F(y) = \Pr\{Y_t \leq y\}$ and quantile function $Q(p) = \inf\{y\colon F(y) \geq p\}$, with $p \in (0,1)$. Moreover, let

$$F(y \,|\, x) = \Pr\{Y_t \leq y \,|\, X_t = x\} \quad \text{and} \quad Q(p \,|\, x) = \inf\{y\colon F(y \,|\, x) \geq p\}$$

be the conditional distribution function (cdf) and the conditional quantile function (cqf) of Y_t given a $k-$value random vector of regressors X_t

The α-level conditional ES of Y_t given $X_t = x$ is

$$\tau(\alpha \,|\, x) = \frac{1}{\alpha} \int_0^{\alpha} Q(p \,|\, x)\,\mathrm{d}p = Q(\alpha \,|\, x) - \frac{1}{\alpha} \int_{-\infty}^{Q(\alpha \,|\, x)} F(y \,|\, x)\,\mathrm{d}y. \tag{4.2}$$

where $Q(\alpha \,|\, x)$ is the α-th quantile of Y_t conditional on $X_t = x$.

The conditional mean of Y_t given $X_t = x$ is

$$\mu(x) = \tau(1 \,|\, x) = \int_0^1 Q(p \,|\, x)\mathrm{d}p \tag{4.3}$$

Definition 1 *For any $w : [0,1] \mapsto [0,1]$ continuous, non-decreasing and differentiable function satisfying $w(0) = 0$, $w(\alpha) = \alpha$ and $w(1) = 1$, the α-level WES of Y_t conditional on $X_t = x$ is*

$$\tau_w(\alpha \,|\, x) = \frac{1}{\alpha} \int_0^{\alpha} Q(p \,|\, x)\mathrm{d}w(p) = Q(\alpha \,|\, x) - \frac{1}{\alpha} \int_{-\infty}^{Q(\alpha \,|\, x)} w(F_x(y))\mathrm{d}y \tag{4.4}$$

and the weighted mean of Y_t conditional on $X_t = x$ is

$$\mu_w(x) = \int_0^1 Q(p \,|\, x)\mathrm{d}w(p) \tag{4.5}$$

Remark 4 *In Chapter 3, we define the weighting function as $w : [0,\alpha] \mapsto [0,1]$ as we focus only on the left tail of the distribution. Here, w enters both the definition of the WES and the expected value of Y_t conditional on X_t, therefore we define it on the whole domain $[0,1]$.*

Remark 5 *The $\alpha-$level WES is itself a coherent risk measure if and only if w is concave on $[0,\alpha]$ (see Th. 4.1 in Acerbi (1)[1]).*

Remark 6 *In the definition of the weighting function, it is crucial to impose the constraint*

$$\int_0^{\alpha} Q(p \,|\, x)[w(p) - 1]\mathrm{d}p = 0 \tag{4.6}$$

which corresponds to imposing $\tau_w(\alpha \,|\, x) = \tau(\alpha \,|\, x)$. This ensures consistency of the weighted estimators $\hat{\tau}_w^{(Q)}(\alpha \,|\, x)$ for $\tau(\alpha \,|\, x)$. This is also true for the estimator $\hat{\tau}_w^{(D)}(\alpha \,|\, x)$.

In (4.4), the original cdf $F_x(y) = F(y \,|\, x)$ is mapped into a new cdf $w(F_x(y))$. The WES is the expected loss one would suffer in the worst $\alpha-$percent cases and $\mu_w(x)$ is the expected return if the distribution of returns were $w(F_x)$ instead of F_x.

In the rest of the section, the pairs $\{(Y_1, X_1) \ldots, (Y_T, X_T)\}$ are assumed to be a random sample drawn from the bivariate distribution of (Y, X), with X_t a $k-$valued random vector of regressors.

Integrated cdf estimators

Without loss of generality, we consider the set of J points $\tilde{Y} = \{y_1 < y_2 < \ldots < y_J\}$. Alternative estimators for $F_j(x) = F_j(y_j \,|\, x)$, $y_j \in \mathbb{R}$, either monotonic or not, were proposed in Section 2.2 following Peracchi (35). Given estimates of the cdf at the grid points $\{y_1, y_2, \ldots, y_J\}$, we approximate the second integral in (4.2) by a sum. Hence, the integrated cdf estimator takes the form

$$\hat{\tau}^{(D)}(\alpha \,|\, x) = \hat{Q}(\alpha \,|\, x) - \frac{1}{\alpha} \sum_{j=1}^{J'} (y_j - y_{j-1})\, \hat{F}_j(x) \tag{4.7}$$

Here $\hat{Q}(\alpha \,|\, x)$ is an estimator of the conditional quantile $Q(\alpha \,|\, x)$, for example a linear quantile regression estimate as in Koenker (25) (this will present in the next subsection), while J' is the index of the grid point $y_{J'} = \sup\{y \in \tilde{Y} \,|\, y \leq Q(\alpha \,|\, x)\}$, that is the largest grid point less or equal than the $\alpha-$level conditional quantile.

An analog estimator $\hat{\mu}^{(D)}(x))$ of the conditional mean can be constructed using estimates of the cdf at grid points $\{y_1, y_2, \ldots, y_J\}$ and setting $\alpha = 1 - \varepsilon$, with $\varepsilon > 0$.

[1]the spectral function $\phi(p)$ is the first derivative $w'(p)$ of the weighting function

Remark 7 *Leorato et al (26) work out the asymptotic variance of the $\hat{\tau}^{(D)}(\alpha \,|\, x)$ estimator. Here we do not make a rigorous analysis of monotonicity of the variance with respect to α. However, as $\hat{\tau}^{(D)}(\alpha \,|\, x)$ is the difference of two increasing quantities in α, we expect the variance of $\hat{\tau}^{(D)}(\alpha \,|\, x)$ to increase in α, too, rendering the $\hat{\mu}^{(D)}(x)$ estimator unstable as α approaches 1 and therefore problematic. In fact, in Theorem 3 in Angrist et al (6), the quantile regression process is defined over $[\varepsilon, 1-\varepsilon]$, for $\varepsilon > 0$. In the empirical exercise, after some trials, we estimate the mean with $\hat{\tau}^{(D)}(\alpha^* \,|\, x)$, where $\alpha^* = 1 - 10^{-3}$.*

According to different choices for $\hat{F}_j(x)$, one obtains different estimators for the mean and the ES. However, in the empirical application we use only the first estimator, the one obtained by inserting (3.14) into (4.2) and denoted with $\hat{\tau}^{(D)}(\alpha \,|\, x)$.

Passing to the weighted versions, we approximate the integrals in (4.4) and (4.5) with averages and have the following estimators for WES

$$\hat{\tau}_w^{(D)}(\alpha \,|\, x) = \hat{Q}(\alpha \,|\, x) - \frac{1}{\alpha}\sum_{j=1}^{J'}(y_j - y_{j-1})\, w(\hat{F}_j(x)) \tag{4.8}$$

We refer to the weighted estimators as the weighted integrated cdf estimator (WICDF) of ES, obtained by inserting (3.14) into (4.8) and denoted with $\hat{\tau}_w^{(D)}(\alpha \,|\, x)$. The weighted integrated cdf estimator of the mean (MICDF) is denoted with $\hat{\mu}_w^{(D)}(x)$ and is obtained by setting α very close to 100% (see Remark 7).

Integrated cqf estimators

As in Section 2.2, let $0 = p_0 < p_1 < \cdots < p_J \le 1$ be a set of J positive real numbers and define the set of weights $\omega_j = \alpha^{-1}(p_j - p_{j-1})$. Assuming, as in Koenker and Bassett (24) and Koenker (25), that for $p \in [0,1]$ we have linearity of the conditional quantile, that is $Q(p \,|\, x) = \beta(p)^\top x$, then parameter $\beta(p)$ is estimated by solving $\min_\beta \sum_{t=1}^T \ell_p(Y_t - \beta^\top X_t)$, where $\ell_p(u) = u(p - \mathbb{1}\{u < 0\})$, $0 < p < 1$, is the asymmetric absolute loss function.

Let $J = [\alpha T]$, then the integrated cqf estimator of the ES is obtained by replacing $Q(\alpha \,|\, x)$ in (4.2) with an estimate and approximating the second integral by a sum

$$\hat{\tau}^{(Q)}(\alpha \,|\, x) = \sum_{j=1}^{J} \omega_j \hat{Q}(p_j \,|\, x) = \tilde{\beta}_T(\alpha)^\top x \tag{4.9}$$

where $\tilde{\beta}_T(\alpha) = \sum_{j=1}^J \omega_j \hat{\beta}_T(p_j)$. As for the integrated cqf estimator of the mean, $\hat{\mu}^{(Q)}(\alpha \,|\, x)$, it is analog to the ES estimator with $\alpha = 1$.

Turning to the weighted versions of the integrated cqf estimators, define the set of weights $\omega_j = \alpha^{-1}[w(p_j) - w(p_{j-1})]$. The integrated cqf weighted estimator(WICQF) of ES

$$\hat{\tau}_w^{(Q)}(\alpha \,|\, x) = \sum_{j=1}^{J=[\alpha T]} \omega_j \hat{Q}(p_j \,|\, x) = \tilde{\beta}_T(\alpha)^\top x, \tag{4.10}$$

where $\tilde{\beta}_T(\alpha) = \sum_{j=1}^J \omega_j \hat{\beta}_T(p_j)$ is a weighted average of coefficient estimates $\hat{\beta}(p_1), \ldots, \hat{\beta}(p_J)$. The weighted integrated cqf estimator of the mean (MICQF) $\hat{\mu}_w^{(Q)}(x)$ is obtained by setting α equal to 1, which is a plug in approximation of the second integral in (4.5).

For details on the asymptotic properties of estimators of $\tau_w(\alpha \,|\, x)$ with a nonparametric specification of $w(\cdot)$, see Chapter 3.

4.3 Minimizing the forecast error of weighted estimators of ES

In Chapter 3, we depart from the generalized weighted representation of ES (3.2) and derive the optimal weighting function that minimizes the asymptotic variance of WICDF and WICQF estimators of ES. Here, departing from the same representation, we derive the optimal weighting function that minimizes the forecast error of the WICDF and WICQF estimators of ES. The forecast error is defined using a test for the prediction accuracy of ES introduced by McNeil and Frey (30). As explained in Section 2.4.2, McNeil and Frey (30) propose a formal test of the hypothesis that the bias of some ES estimator is zero. Define the quantile violation event as the case case when the observed excess return is lower than the VaR at level α. Under the null hypothesis, the forecast error, computed as the difference between the observed excess return and the corresponding one-step ahead predicted shortfall under quantile violation event should have mean zero. The test rejects the null hypothesis if the average forecast error is large. Here we do not test the null hypothesis but, for different weighted estimators, compute the average of the empirical distribution of the forecast error and select the optimal weighting function as the one corresponding to the estimator with minimum average forecast error.

Now we define the forecast error $\mathbf{f}e$ for any weighted estimator $\hat{\tau}_w(\alpha)$ for the ES of level α of the rv Y. As mentioned before, the $\mathbf{f}e$ is computed only for the quantile violation events, that is when the observed return is less than the (1-α) VaR

$$\mathbf{f}e_{Y,\hat{\tau}_w(\alpha)} = \begin{cases} \mathbf{g}(Y - \hat{\tau}_w(\alpha)) & , \quad \text{if} \quad Y \le \hat{Q}(\alpha \,|\, x) \\ \text{not defined} & , \quad \text{otherwise} \end{cases} \tag{4.11}$$

where $g : \mathbb{R} \mapsto \mathbb{R}^+$ is an arbitrary loss function defined, for example quadratic $\mathbf{g}(u) = u^2$ or absolute value $\mathbf{g}(u) = |\,u\,|$. Note that the forecast error $\mathbf{f}e$ depends on the estimates $\hat{\tau}_w(\alpha)$ and $\hat{Q}(\alpha \,|\, x)$.

We search for the weighting function that minimizes the mean forecast error over a well specified set $\mathcal{W}$ of *admissible* weighting functions

$$w^{\star}_{\hat{\tau}_w(\alpha)} = \arg\min_{\mathcal{W}}\{\mathrm{E}\left(\mathbf{g}(Y - \hat{\tau}_w(\alpha) \mid Y \le \hat{Q}(\alpha \,|\, x)\right)\} \tag{4.12}$$

where

$$\mathcal{W} = \{w = w(\pi, b_1, b_2), b_1 \in (0,1], b_2 \in [1, +\infty], \pi = \pi^*\}$$

As in Chapter 3, we give the function $w(\cdot)$ a parametric specification as a mixture of beta distributions

$$w(p) = \begin{cases} \alpha\left(\pi\frac{\left(1-\frac{p}{\alpha}\right)^{b_1-1}}{b(1,b_1)} + (1-\pi)\frac{\left(1-\frac{p}{\alpha}\right)^{b_2-1}}{b(1,b_2)}\right), & p \in [0,\alpha] \\ p, & p \in (\alpha, 1] \end{cases} \tag{4.13}$$

where $b(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$, $b_1 \in (0,1], b_2 \in [1,+\infty]$ and $\pi \in [0,1]$ and

$$\pi^* = \pi(b_1 b_2) = \frac{\tau_{w_2}(\alpha \,|\, x) - \tau(\alpha \,|\, x)}{\tau_{w_2}(\alpha \,|\, x) - \tau_{w_1}(\alpha \,|\, x)}. \tag{4.14}$$

that solves equation (4.6). Here, $w_1(p)$ and $w_2(p)$ are the components of $w(p)$, that is

$$w(p) = \pi w_1(p) + (1-\pi) w_2(p), p \le \alpha$$

while $\tau_{w_1}(\alpha\,|\,x) = \int_0^\alpha Q(p\,|\,x)\mathrm{d}w_1(p)$ and $\tau_{w_2}(\alpha\,|\,x) = \int_0^\alpha Q(p\,|\,x)\mathrm{d}w_2(p)$. Note that the weighting is non-uniform exclusively on the left tail of the distribution, changing only probabilities assigned to losses more severe than the estimated quantiles $\hat{Q}(\alpha\,|\,x)$.

We do not give an analytical solution for the optimal weighting function $w^\star$, but only an numerical one. We replace the expected value operator E() in the minimization problem (4.12) with its empirical counterpart. Given the set of observations $(Y_1, X_1) \ldots (Y_T, X_T)$ from the bivariate distribution of (Y, X), with X a $k-$valued random vector of regressors, the optimal estimated $\hat{w}^\star$ solves the following minimization problem

$$\hat{w}^\star_{\hat{\tau}_w(\alpha)} := \arg\min_{w\in\mathcal{W}} \left\{ \frac{1}{R} \sum_{t=T-t_w}^{T} \mathbb{1}\{Y_t < \hat{Q}_t(\alpha\,|\,x)\} \cdot \mathbf{g}(Y_t - \hat{\tau}_w(\alpha\,|\,x)) \right\} \tag{4.15}$$

where $t_w > 1$. Here $\hat{Q}_t(\alpha\,|\,x)$ and $\hat{\tau}_w(\alpha\,|\,x)$ are the estimates of the $\alpha-$VaR and the $\alpha-$ level weighted ES at time t using information up to time $(t-1)$ and $R = \sum_{t=t_0}^{T} \mathbb{1}\{Y_t < \hat{Q}_t(\alpha\,|\,x)\}$.

4.4 The asset allocation model

4.4.1 The general setting

The Markovitz asset allocation model maximizes expected return with a risk constraint on the standard deviation. Alternative formulations propose as risk constraints the VaR (see Alexander and Baptista (3)) and, more recently, the expected shortfall (see Basak and Shapiro(9)).

Here we construct an asset allocation model that maximizes the expected return imposing an upper bound on the risk that the portfolio might suffer, writing the model in terms of a Lagrange function with some general risk measure. Either the ES, the WES or the VaR can play the role of the risk measure.

We recall the literature that supports convexity of optimization problems similar to ours. However, we do not take advantage of the full potentiality of the convexity of the asset allocation problem and derive the solution by means of a numerical algorithm that searches for the optimal allocation out of a set of alternative portfolio weights.

Let

$$\mathcal{Q} = \{q \in \mathcal{R}^M : \sum_{m=1}^{M} q_m = 1, 0 \le q_m < 1\} \tag{4.16}$$

denote the set of available alternative components weights vectors q, each of them defining unique portfolios P_q with returns Y_q. The optimization problem that we solve is

$$q^\star = \arg\max_{q\in\mathcal{Q}} \mathcal{L} \tag{4.17}$$

where

$$\mathcal{L} = \mu(Y_q\,|\,x) - \lambda_1 \cdot (\psi(Y_q\,|\,x) - \bar{\delta}^\psi) - \lambda_2 \cdot \left(\sum_{m=1}^{M} q_m - 1 \right) \tag{4.18}$$

Here $\psi(Y_q\,|\,x)$ is some risk measure of portfolio P_q, conditional on available regressors $X = x$. Moreover, quantity $\bar{\delta}^\psi$ is the risk threshold and λ_1 and λ_2 are the Lagrange multipliers. The threshold $\bar{\delta}^\psi$ is equivalent to the portion of wealth that the investor is ready to lose.

In general, such portfolio optimization problems cannot be solved analytically and, provided that some conditions are met, they are dealt with numerical solutions via convex programming techniques. For example, if in (4.17), we replace $\psi(Y_q \,|\, x)$ with the standard deviation, then the problem can be solved analytically (see Meucci (31), chapter 6). If, on the other hand, we consider the VaR of level α as risk constraint, that is $\psi(Y_q \,|\, x) = Q(\alpha \,|\, x)$, then problem (4.17) enter the class of convex programming problems and admits a numerical solution (see Ghaoui (19)).

Remark 8 *As $\tau(\alpha \,|\, x)$ is an average of the up to the α-level, the convexity conditions hold if we replace ψ with $\tau(\alpha \,|\, x)$, too. Therefore, the optimization problem admits a numerical solution (see Cuoco et al (28) and Alexander and Baptista (3)).*

In our case, the optimality conditions that guarantee the existence of a unique solution (see Barbu and Precupanu (8), chapter 3), can be written as:

(i) Function $\mathcal{L}$ is convex

(ii) The set $\{\psi(Y_q \,|\, x) \le \bar{\delta}^{\psi}\}$ is convex

(iii) $\mathcal{Q} \cap \{\psi(Y_q \,|\, x) \le \bar{\delta}^{\psi}\} \neq \varnothing$

We propose a similar setting of the model (4.17) and (4.18), in which we replace the mean with its weighted version $\mu_w(x)$ and the risk measures with weighted ES $\tau_w(\alpha \,|\, x)$. Therefore, the problem is

$$q^\star = \arg\max_{q \in \mathcal{Q}} \mathcal{L} \quad \text{s.t.} \quad \tau_w(Y_q \,|\, x) \le \bar{\delta}^{\tau_w}, \tag{4.19}$$

where

$$\mathcal{L} = \mu_w(Y_q \,|\, x) - \lambda_1 \cdot (\tau_w(Y_q \,|\, x) - \bar{\delta}^{\tau_w}) - \lambda_2 \cdot \left(\sum_{m=1}^{M} q_m - 1 \right) \tag{4.20}$$

Moreover, each portfolio P_q is associated to a unique weighting function $w^\star(q)$ obtained by the solution of (4.15), the minimization of the forecast error of weighted estimators of ES. The weighted estimators (WES) are consistent for the ES (see Remark 4.6 on the definition of the weighting function). Although not proved here, we assume, in base of Remark 8 that optimization problem that controls for the weighted mean and the weighted ES is convex and has a unique solution when the true quantities are replaced by their weighted consistent estimates. This allows us to jointly derive the vector of optimal portfolio weights $q^\star$ and the weighting function $w^\star$ using the numerical algorithm that we present in the next subsection.

4.4.2 The numerical algorithm

The algorithm is constructed following the basic sequence of iterative algorithms (see for example Brooks (12)): random generation of a finite number of subsets of weights and optimization over some well defined function up to a threshold level for the tolerance level. However, it is rather limited as it does not take into account the convexity property of the optimization problem. We generate a finite number N of random subsets $\mathcal{Q}^N$ of portfolio weights from the Dirichlet distribution and choose the optimal vector that by solving (4.19). Moreover, for each set of portfolio weights q, we derive the optimal weighting function $w^\star$ as a function of q solving (4.15). As we use $\mathcal{Q}^N$ instead of the set $\mathcal{Q}$, we only end up with an approximation $\tilde{q}^\star$ to the optimal solution $q^\star$.

We assume that, under some smoothness properties of the objective function $\mathcal{L}$, in particular convexity and continuity, the approximation $\tilde{q}^\star$ converges to $q^\star$

$$\forall \epsilon \in (0,1) \text{ and } \nu > 0, \quad \exists \bar{N} \text{ s.t. } \forall N > \bar{N}, \Pr\{\, |\, \tilde{q}^\star - q^\star \,| \, < \nu\} > 1 - \epsilon. \tag{4.21}$$

The sequential algorithm is implemented on time rolling windows indexed by period t. Let us denote by $q^\star_{(i)}(t)$ the vector found after the $i-$th iteration, $i \geq 1$, at period t and with $\mathcal{L}(q^\star_{(i)}(t))$ the value of the Lagrangian (4.20) computed on the cdf $F(y\,|\,x)$ of returns of the portfolio with weights $q^\star_{(i)}(t)$. The steps are

(i) Select a random sample of N_0 independent $M-$ dimensional vectors $(q_1, \ldots, q_{N_0})$ from the Dirichlet distribution with parameters $(\alpha_0 + M)q^\star_{(i)}(t) + 1$, with $\alpha_0 > 0$. Each portfolio P_{q_n} with returns Y_{q_n} has an associated $w^\star$ derived by solving the optimization problem 4.15.

(ii) Define $\tilde{q}^\star_{(i+1)}$ as the vector that maximizes equation (4.20) among the set of vectors $(q_1, \ldots, q_{N_0})$. Then if $\hat{\mathcal{L}}(\tilde{q}^\star_{(i+1)}) < \hat{\mathcal{L}}(q^\star_{(i)})$ set $q^\star_{(i+1)} = \tilde{q}^\star_{(i+1)}$, otherwise $q^\star_{(i+1)} = q^\star_{(i)}$

where $\hat{\mathcal{L}}(q)$ is computed on the estimated cdf $\hat{F}(y\,|\,x)$ and approximates the true value $\mathcal{L}(q)$.

As components weights vector for the initial time $t = t_0$, we set $q^\star_0(t_0) = (1/M, \ldots, 1/M)$, which corresponds to the equally weighted portfolio. In order to smoothen the portfolio weights, we impose some dependency in the parameters of the Dirichlet distribution. Therefore, for the successive time periods, in the first iteration, the mode of the Dirichlet distribution is set equal to $q^\star_0(t) = \tilde{q}^\star(t-1)$, where $\tilde{q}^\star$ is the approximated value of the true optimum $q^\star(t)$. This mitigates the problem of significant variations in the components weights across small time intervals. The scale parameter α_0 is used to allow for higher or lower dispersion of the sampled observations and can also be allowed to vary across iterations. As a stopping rule for the algorithm, the following conditions must be satisfied, for each time t:

$$\frac{|\,\mathcal{L}(q^\star_{(i)}(t)) - \mathcal{L}(q^\star_{(i-1)}(t))\,|}{\mathcal{L}(q^\star_{(i-1)}(t))} \leq s, \tag{4.22}$$

for some tolerance level s positive and close to 0. The scale of s is dependent on the degree of convexity of function $\mathcal{L}(\cdot)$. In practice, choosing s equal to a sufficiently low number (in our case, we set $s = 0.05$) ensures convergence and stability of the optimization exercise.

4.5 Allocation stability

When we derive optimal portfolio weights by replacing the unknown quantities with estimates, the estimation errors are passed as errors in the optimal allocation (see Meucci (31)). The notion of stable allocation is associated to sensitivity of the optimal allocation to the estimated market parameters. Some authors analyze stable optimal allocations in models that use quantile risk constraints (VaR, in particular). We recall here Ghaoui et al (19) and Natarajan et al (33).

We define stability of the portfolio weights in terms of the mean "distance" between components weights distribution over some period of time. In other words, we look at variability of portfolio components weights. Large variations of portfolio weights are of concern if one considers issues like transaction costs and forecasting of portfolio performance indicators. Transaction costs associated to re-balancing of portfolio weights are significant if such operations are performed with relatively high frequency and for a large number of components. Moreover, stable portfolio weights are easier to monitor and forecasting applications involves less uncertainty regarding future composition of portfolios.

Following the above mentioned works, we are interested in how the use of different risk constraints plugged in the asset allocation model affect stability of the optimal portfolio from the point of view of sensitivity of the allocation to the estimated market parameters. We refer to the optimal allocation as the vector of weights obtained by solving the portfolio optimization problem as specified in (4.19). Two alternative measures of portfolio weights stability over some time interval are proposed and a stylized example is constructed where we use VaR or the ES risk measures. In order to simplify the example, we only consider the simple case of normal returns and a time interval made up of two periods. In the empirical application, the two proposed stability measures are computed for the optimal portfolios.

4.5.1 Measures of allocation stability

Let Y_q denote the return of a portfolio constructed of M assets with weights $q \in \mathcal{Q}$, with $\mathcal{Q}$ specified in (4.16). Consider the following time interval made up of $S > 0$ periods prior to t

$$(t, S) = \{(t - S), (t - S + 1), \ldots, (t - 1)\}$$

Definition 2 *Stability $\rho_{q,\delta}(t, S)$ of the portfolio weights over the time interval (t, S) is defined as the reciprocal of the mean distance δ for all pairs (t, t_0)*

$$\rho_{q,\delta}(t, S) = \left\{ \frac{1}{S} \sum_{(t-S) \leq t_0 < t} \delta(q(t), q(t_0)) \right\}^{-1}, \quad \textit{with} \quad q(\cdot) \in \mathcal{Q} \tag{4.23}$$

where $\delta(q(t), q(t_0))$ is the distance between the distributions of portfolio weights $q(t)$ and $q(t_0)$, with $t_0 \in (t, S)$.

We interpret $\rho_{q,\delta}(t, S)$ as an absolute measure of stability of portfolio weights. The smaller is the variation in portfolio weights, the higher is $\rho_{q,\delta}(t, S)$. In Definition 2, if $S = 1$, we only consider two consecutive periods, t and $(t - 1)$. Otherwise, if we opt for monitoring the performance indicators of portfolios and re-balancing operations on a weekly or monthly basis (as usually encountered in practice), then S would take values 5 and 20, respectively. This is our choice in the empirical section that we develop further on.

As measures of distance δ between distributions, we have alternative proposals. Our first choice is the Kullback-Leibler "distance" (also called information divergence or relative entropy)

$$\delta^{KL}(q(t), q(t_0)) = \sum_{m=1}^{M} q_m(t) \log \frac{q_m(t)}{q_m(t_0)} \tag{4.24}$$

where $q_m(t)$ denotes the weight corresponding to the m^{th} portfolio component at time t. The alternative measure that we propose is the L^2-distance, that does not penalize large weights, as the KL distance does

$$\delta^{L^2}(q(t), q(t_0)) = \sqrt[2]{\sum_{m=1}^{M} |\, q_m(t) - q_m(t_0) \,|^2} \tag{4.25}$$

Plugging (4.24) and (4.25) into (4.23) results in the two alternative measures of portfolio weights stability, which we denote with $\rho^{KL}(t, S)$ and $\rho^{L^2}(t, S)$ respectively.

4.5.2 A stylized example

We construct one stylized example using normal returns and prove that the allocation that uses ES rather that VaR as risk constraint is characterized by higher stability according to the measures proposed in the previous sub-section. This analysis should be furtherly extended by considering other classes of distributions, like mixtures of normals or fat tailed distributions for which we have analytical results regarding computation and estimation of the expected shortfall (see, for example, Peracchi and Tanase (36)).

Suppose that we observe two assets with the following bivariate normal distribution at t_0

$$\left(\begin{array}{c} Y_1 \\ Y_2 \end{array} \right) = \mathcal{N}\left(\left(\begin{array}{c} \mu \\ \mu \end{array} \right), \left[\begin{array}{cc} \sigma^2 & 0 \\ 0 & \sigma^2 \end{array} \right] \right). \tag{4.26}$$

We are interested in how the optimal allocation changes at $t = t_0 + 1$, assuming that the mean of the two assets change by some amount $\epsilon > 0$ but the variance remains constant, i.e. under the alternative distribution

$$\left(\begin{array}{c} Y_1 \\ Y_2 \end{array} \right) = \mathcal{N}\left(\left(\begin{array}{c} \mu + \epsilon \\ \mu - \epsilon \end{array} \right), \left[\begin{array}{cc} \sigma^2 & 0 \\ 0 & \sigma^2 \end{array} \right] \right) \tag{4.27}$$

We then prove that stability measures ρ^{KL} and ρ^{L^2} are indeed lower for the optimal portfolio that solves the ES constrained allocation model as compared to the VaR constrained allocation model, considering that the weights vector change from $q(t_0)$ to $q(t)$.

Let $q = (q_1, q_2)'$ denote the vector of portfolio weights. Then the return on portfolio $Y = q_1 * Y_1 + q_2 * Y_2$ is normally distributed

$$Y \sim N(\mu + \epsilon(q_1 - q_2), (q_1^2 + q_2^2) \cdot \sigma^2).$$

Weights (q_1, q_2) are derived by solving the optimization problem (4.19), or equivalently by finding a stationary point of the Lagrange function, with generic risk constraint ψ bounded by $\bar{\delta}^\psi$

$$\mathcal{L}^\psi = \mu + \epsilon(q_1 - q_2) - \lambda_1^\psi \cdot (\psi - \bar{\delta}^\psi) - \lambda_2^\psi \cdot (q_1 + q_2 - 1) \tag{4.28}$$

The risk constraint are either $VaR(\alpha)$ or the expected shortfall $\tau(\alpha)$. Both the VaR and the ES have close form expressions for a normal random variable with mean μ and variance σ^2. More precisely, $VaR_{N(\mu,\sigma)}(\alpha) = \mu + \sigma \cdot \Phi^{-1}(\alpha)$ and $\tau_{N(\mu,\sigma)}(\alpha) = \mu - \sigma\alpha^{-1} \cdot \phi(\Phi^{-1}(\alpha))$, where $\phi(y), y \in \mathbb{R}$ and $\Phi^{-1}(p) = Q(p), p \in (0,1)$ are the density function and the quantile function respectively of the standard normal distribution $\mathcal{N}(0,1)$. These two quantities enter the condition under which the allocation that uses ES rather that VaR as risk constraint is characterized by higher stability.

Replacing the risk constraint ψ with either $VaR_Y(\alpha)$ or $\tau_Y(\alpha)$ in (4.28), it comes out that the equally weighted portfolio $q_1 = q_2 = 1/2$ is the optimal allocation when the means of the distributions of the two assets are equal (model (4.26)). This result can be easily generalized to M assets. In this case the optimal allocation is $q_m = 1/M, m = 1 \ldots M$.

Now consider the case of returns with normal distributions but different means (model (4.27)) and denote with (q_1^{VaR}, q_2^{VaR}) and (q_1^τ, q_2^τ) the optimal allocations with $VaR_Y(\alpha)$ and $\tau_Y(\alpha)$ as risk constraints.

Let $VaR_\sigma = VaR_{N(0,\sigma)}(\alpha)$ and $\tau_\sigma = \tau_{N(0,\sigma)}(\alpha)$. Writing the first order conditions $\mathcal{L}^\psi \, \mathrm{d}q_i = 0, i = 1,2$ and $\mathcal{L}^\psi \, \mathrm{d}\lambda_2^\psi = 0$, with $\psi = VaR_Y(\alpha)$ we get that

$$q_1^{VaR} = \frac{1}{2} + \left(\frac{1 - \lambda_1^{VaR}}{\lambda_1^{VaR}} \right) \frac{\epsilon}{2VaR_\sigma}, \quad q_2^{VaR} = \frac{1}{2} - \left(\frac{1 - \lambda_1^{VaR}}{\lambda_1^{VaR}} \right) \frac{\epsilon}{2VaR_\sigma} \tag{4.29}$$

Analogously, writing the first order conditions for $\psi = \tau_Y(\alpha)$ we get

$$q_1^\tau = \frac{1}{2} + \left(\frac{1-\lambda_1^\tau}{\lambda_1^\tau}\right)\frac{\epsilon}{2\tau_\sigma}, \quad q_2^\tau = \frac{1}{2} - \left(\frac{1-\lambda_1^\tau}{\lambda_1^\tau}\right)\frac{\epsilon}{2\tau_\sigma} \tag{4.30}$$

Let ρ_{VaR}^{KL} and ρ_τ^{KL} denote the stability measures for the optimal portfolios with $\tau(\alpha)$ and $VaR(\alpha)$ as risk measures, with $q^{VaR}(t) = (q_1^{VaR}, q_2^{VaR})$, $q^\tau(t) = (q_1^\tau, q_2^\tau)$ and $q(t_0) = (1/2, 1/2)$. We prove that the optimal portfolio weights with $\psi = \tau(\alpha)$ as risk constraint are more stable that the optimal weights with $\psi = VaR$, which is equivalent to $(\rho_{VaR}^{KL})^{-1} - (\rho_\tau^{KL})^{-1} > 0$. Writing down the KL distance, we get that

$$(\rho_\tau^{KL})^{-1} - (\rho_{VaR}^{KL})^{-1} = \log 2\cdot(q_1^\tau \log(2q_1^\tau) + q_2^\tau \log(2q_2^\tau)) - \log 2\cdot(q_1^{VaR}\log(2q_1^{VaR}) + q_2^{VaR}\log(2q_2^{VaR}))$$

The above quantity is positive if and only if

$$1/2 < q_1^\tau < q_1^{VaR} \quad \text{and} \quad q_2^{VaR} < q_2^\tau < 1/2 \tag{4.31}$$

This condition is also necessary and sufficient for the alternative stability measures $\rho_{VaR}^{L^2}$ and $\rho_\tau^{L^2}$ and for inequality $(\rho_{VaR}^{L^2})^{-1} - (\rho_\tau^{L^2})^{-1} > 0$ to hold. Using expressions of q_1^{VaR} and q_1^τ from (4.29) and (4.30), inequality (4.31) is equivalent to

$$\left(\frac{1-\lambda_1^{VaR}}{\lambda_1^{VaR}}\right)\frac{1}{VaR_\sigma} > \left(\frac{1-\lambda_1^\tau}{\lambda_1^\tau}\right)\frac{1}{\tau_\sigma} \tag{4.32}$$

Note that, in the above expression, both VaR_σ and τ_σ enter in absolute value as measures of losses that the portfolio incurs. The multipliers λ_1^{VaR} and λ_1^τ represent the fraction of initial wealth that the investor is ready to loose (see, for example Alexander (4)). As $\tau(\alpha) > VaR(\alpha)$ by definition, it comes out that $\lambda_1^\tau > \lambda_1^{VaR}$, therefore inequality (4.32) is satisfied.

4.6 Empirical application

We develop an empirical application using daily data on European financial and non-financial stocks. The portfolio weights are chosen solving the asset allocation model (4.18) by maximizing expected return while controlling for either VaR, ES or the weighted ES (according to model (4.20)). The optimal portfolios are compared in terms of a set of performance indicators and measures of stability of components weights. Besides the stability measures presented in the previous section, we also show the standard deviation of portfolio weights.

The expected return and the expected shortfall are estimated either with the WICDF $\hat{\tau}_w^{(D)}(\alpha \mid x)$ and the MICDF $\hat{\mu}_w^{(D)}(\alpha \mid x)$ estimators, or with the WICQF $\hat{\tau}_w^{(Q)}(\alpha \mid x)$ and MICQF $\hat{\tau}_w^{(Q)}(\alpha \mid x)$ estimators. The VaR is estimated via linear quantile regression.

The regressors are a set of financial and macro variables and they are chosen such as to make up a broad enough spectrum of conditioning information. Regarding composition of the portfolios, we consider portfolios made of either financial or non-financial stocks. This allows us to compare the asset allocation model applied across the two main sectors of activity. All experiments are carried out using the statistical package Stata, version 10.

4.6.1 The data

Our raw data are daily from December 30, 1994, to December 31, 2007, and the sources are Bloomberg and Thomson Datastream. The dependent variable, in each estimation exercise in turn, is the daily return of alternative portfolios. The daily return is computed as the logarithmic difference in the price level and excluding weekends and holidays.

The set of predictors includes both real and financial variables, making up a balanced mix of macro and micro data. The real variables are the price of oil as the value of the 1^{st} month future contract, and a price index of non-energy commodities, both expressed in US dollars. The financial variables include the risk spread (the yield difference between an aggregate BBB bond and a German Government 5-year bond), the term spread (the yield difference between a German Government 10-year bond and the Euribor 3M), the dividend yield (on the DJ Eurostoxx 50 Index) and the EUR/USD exchange rate.

In order to reconstruct the series for the money market instruments as well as for the common currency and the benchmark for the computation of the spreads, we had to choose a representative national market before 1998. Macro convergence within the Euro area and trading volumes pointed Germany as the benchmark economy. Thus, the Euribor 3M and the Lehman Euro aggregate BBB bond yield were concatenated backwards using the dividend yield of the Datastream Totmarket Euro index before 31 January 1998, the German interbank 3 month interest rate before December 31, 1998 and the Lehman US aggregate BBB bond Yield before June 30, 1998 respectively.

In Tables 4.1 and 4.2, for the broad Eurostoxx index and the set of regressors, we give the data sources, variable transformations and summary statistics of the transformed data: the mean, the standard deviation (SD) and also the difference between the 99^{th} percentile (Q_{99}) and the 1^{st} percentiles (Q_{01}). The sample includes changes in market regimes, departing from the bull market of the second half of the 1990s, followed by the bear market between 2000 and 2003, and the post-2003 period. The number of observations is equal to 2646, starting from January 3, 1995, until December 28, 2007. All predictors are measured as of the end of the day.

Table 4.1: Description and sources of the Eurostoxx index and the covariates.

Code	Description	Source	Provider
stoxx50	DJ Euro Stoxx 50 Price Index	DJ Eurostoxx	Blommberg
comm	Goldman Sachs non-energy index	Goldman Sachs	Bloomberg
dyeu	DJ Euro Stoxx equity dividend yield	DJ Eurostoxx	Datastream, Bloomberg
gb5y	German Govt Bond Yield	Bloomberg	Bloomberg
fx	EUR/USD exchange rate	ECB	Bloomberg
oil	Oil Nymex future price (in US $)	Nymex	Bloomberg
re3m	Euribor 3M	European Banking Fdr.	Datastream, Bloomberg
ryeu	Lehman Euro Corp BBB Yield	Lehman Brothers	Datastream, Bloomberg
yd10e	10y Treasury Bond Yield Germany	Blommberg	Bloomberg

In Table 4.3, we give details on the raw and transformed variables. The series, provided by Thomson Datastream, were chosen among the components of the Eurostoxx50 index as of 24^{th} of June, 2008. We selected the stocks according to their availability but we also aimed at covering all Industry Classification Benchmark (ICB, according to primary revenue

Table 4.2: Transformations and summary statistics of the variables.

Variable	Description	Transformation	Mean	SD	$Q_{99} - Q_{01}$
STOXX50	DJ Euro Stoxx 50 return	$\ln(stoxx50_t/stoxx50_{t-1})$	$3.0 \cdot 10^{-4}$	$1.3 \cdot 10^{-2}$	$7.4 \cdot 10^{-2}$
ECOMM	Commodity price log diff	$\ln(comm_t/comm_{t-1})$	$2.3 \cdot 10^{-4}$	$0.7 \cdot 10^{-2}$	$3.4 \cdot 10^{-2}$
EDY	Dividend yield	$\ln(dyeu_t)$	0.79	0.27	1.01
EFX	EUR/USD log diff	$\ln(fx_t/fx_{t-1})$	$-0.5 \cdot 10^{-4}$	$0.6 \cdot 10^{-2}$	$3.1 \cdot 10^{-2}$
EOIL	Oil price log diff	$\ln(oil_t/oil_{t-1})$	$7.5 \cdot 10^{-4}$	$2.2 \cdot 10^{-2}$	$11.0 \cdot 10^{-2}$
ERSP	Risk spread	$ryeu_t - gb5y_t$	1.15	0.45	1.91
ESP	Term Spread	$yd10e_t - re3m_t$	1.46	0.90	3.75

Table 4.3: Description and summary statistics of the stock indexes

Variable	Description	Mean	SD	$Q_{99} - Q_{01}$
ALL	ALLIANZ (Fin.,Germany)	0.020	1.6	8.6
BCO	BCO SANTANDER (Fin., Spain)	0.065	2.0	11.2
BNP	BNP PARIBAS (Fin., France)	0.063	2.1	11.4
GEN	ASSIC. GENERALI (Fin., Italy)	0.024	1.8	9.8
INTS	INTESA SANPAOLO (Fin., Italy)	0.041	4.4	21.0
UNIC	UNICREDIT (Fin., Italy)	0.014	4.0	25.2
AEG	AEGON (Fin., Netherlands)	0.003	3.1	18.7
ING	ING GRP (Fin., Netherlands)	0.036	2.1	12.3
BASF	BASF (Basic Materials, Germany)	0.027	1.9	10.2
VOL	VOLKSWAGEN (Consumer Goods, Germany)	0.036	2.2	11.7
SAN	SANOFI-AVENTIS (Health Care, France)	0.083	2.1	11.6
SIE	SIEMENS (Industrials, Germany)	0.048	2.1	11.6
TOT	TOTAL (Oil&Gas, France)	0.031	1.8	9.5
SAP	SAP (Technology, Germany)	0.056	3.1	17.2
TELEF	TELEFONICA (Telecomm., Spain)	0.081	1.9	10.4
IBE	IBERDROLA (Utilities, Spain)	0.080	1.4	7.6

source) sectors. Half of them (9 companies) are classified as financial and the other half are non-financial stocks. The summary statistics of the indexes show a large range for the mean daily return, from $0.3 \cdot 10^{-4}$ for Aegon to $8.3 \cdot 10^{-4}$ for Sanofi Aventis. Looking at dispersion, we note that, compared to the Eurostoxx index, the stock indexes are much more volatile, with standard deviations from $1.4 \cdot 10^{-2}$ for Iberdrola to $4.4 \cdot 10^{-2}$ for Intesa Sanpaolo and interquantile ($Q_{99} - Q_{01}$) ranges almost all above $10.0 \cdot 10^{-2}$.

4.6.2 The estimation exercise

We estimate the mean, the ES and the WES using the analog estimators presented in the previous sections. The VaR is estimated via linear quantile regression. For simplicity, the WICDF and MICDF estimators are jointly referred to ICDF estimators and the same is for the ICQF estimators. We consider α equal to 10% (the largest of the three typical values, together with 1% and 5%) allowing us to capture regressors effect on the left tail of the distribution up to a sufficiently large threshold. For the estimation of the mean with $\hat{\mu}^{(D)}(x)$, we set $\alpha^\star = 0.999$ (see Remark 7). The weighting function is specified as in model (4.13) with b_1, b_2 taking as values all possible pairs from the sets $\{0.25, 0.50, 0.75, 1\}$ and $\{1, 2, 3, 4\}$ while the mixing parameter is derived according to (4.14). The length of the time interval for the weighing function derivation in (4.15) is set $t_w = 20$.

The model is fitted repeatedly using rolling windows of size $T = 199$. All predictors are used with a lag equal to one. For each rolling window sample, we estimate the model and use the last available value of the predictors to forecast the one-step ahead WES over the next day. The first rolling window includes observations from January 3, 1995, to June 7, 1996 and the first estimate is the predicted WES between June 10, 1996 and June 11, 1997. The first $k-$dimensional vector x is equal therefore to the covariates on day June 10, 1997. In total, we have 2346 windows and the estimating exercise is repeated for each window.

For the ICDF estimators (4.8), we set the thresholds $\tilde{y}_j, j = 1 \ldots J$ as equally spaced order statistics $\tilde{y}_1 = Y_{(1+\delta)}$ and $\tilde{y}_J = Y_{(1+\delta J)}$, where $\delta = [S/(J+1)]$ is the integer part of the ratio between the number S of data points to the left of $\hat{Q}(\alpha \,|\, x)$ and J is the number of thresholds. If S is very small, in particular if $S \leq 2J$, then the grid points are not optimally chosen as the largest one is equal to Y_{J+1}. However, in practice, J is chosen to be small enough both to avoid such situations and also considering computational burdening. For the ICQF estimators (4.10), we set $p_j = \alpha \cdot j/J$, with $j = 1, \ldots, J$, that is $p_j = p_{j-1} + \delta$, with $\delta = \alpha/J$. The p_j's are therefore equally spaced between $p_1 = \alpha/(2J)$ and $p_J = \alpha$. In the study, given the large number of rolling windows, we set a small value for J, equal to 5.

4.6.3 Asset allocation

We implement the algorithm presented in Section 4.4.2 on the data presented above. At each estimation exercise and for each iteration we generate $N_0 = 5$ alternative portfolio weights. Then, for each set of portfolio weights, we derive the weighting function that minimizes the forecast error of the weighted estimator of ES. Eventually, we choose among the alternative sets the one that minimizes the Lagrangean (4.20). The cdf $\hat{F}(y \,|\, x)$ that enters $\mathcal{L}(q)$ is estimated on a rolling window sample of 20 observations. The multiplier λ_1 is set equal to 0.1 assuming that the investor accepts losses no higher than 10% of the initial wealth. For simplicity, the same value of the multiplier is used for all allocation problems, irrespective of the risk constraint that we consider. The scale parameter α_0 of the Dirichlet distribution is 10^3. As for the loss function $g(\cdot)$ in the definition of the forecast error (see equation 4.11), we specify it as quadratic loss.

In Tables 4.4, 4.5 and 4.6 we show the performance indicators of the benchmark and the optimal portfolios, for both financial and non-financial stocks. As risk constraints, we use either the VaR or the expected shortfall. The WES is estimated either with ICDF or ICQF estimators. The number of iterations of the algorithm is set equal to 3.

In table 4.4 we list the performance indicators for different benchmark portfolios. We also show summary statistics characterizing the returns distribution, namely the mean, the median and the standard deviation. Then we present some loss measures (the Value at Risk and the ES, both at 95% level), computed both on empirical distribution and under normal assumption. Lastly, we compute indicators that characterize the Return to Risk profile (on daily return basis), namely the Sharpe ratio and the Return to ES at 95% level (both empirically and under normality assumption). The benchmarks that we consider are 100% investment in the Eurostoxx 50 index and the equally weighted portfolios.

Tables 4.5 and 4.6 show the performance indicators and the stability measures of portfolios

Table 4.4: Performance indicators for the benchmark index EuroStoxx50 and equally weighted portfolios

Indicator*	Description	EuroStoxx50	Equally weighted Fin.	Equally weighted Non-fin.
(1)	Mean daily return	0.025	0.025	0.044
(2)	StDev daily return	1.3	1.5	1.3
(3)	Median daily return	0.074	0.052	0.097
(4)	Empirical VaR at 95% level	2.2	2.3	2.2
(5)	Empirical ES at 95% level	3.2	3.7	3.0
(6)	Var at 95% level under normality (= -(1)+1.645*(2))	3.1	3.5	3.0
(7)	ES at 95% level under normality ($= -[(1) - 1/0.05 * (2) * \phi(\Phi^{-1}(0.05))]$)	2.7	3.1	2.7
(8)	Sharpe ratio (=(1)/(2))	29	26	53
(9)	Empirical Return to ES at 95% level (=(1)/(5))	0.8	0.7	1.4
(10)	Return to ES at 95% level under normality (=(1)/(7))	0.9	0.8	1.6

* The indicators are expressed in percentage points

that use either VaR, ES or WES risk constraints. The tables correspond to each type of components, financial or non-financial. As compared to the EuroStoxx50 benchmark, the strategies show a better return-to-risk profile in terms of Sharpe ratio. The other return-to-risk indicators also bring out an improvement in the profile of the strategies. On changing the benchmark and considering the equally weighted portfolios, the improvement is significant for portfolios that use both ICDF and ICQF estimators. As for the losses measures, in most cases, the optimal portfolios show less severe empirical VaR or ES at 95% level.

The non-financial portfolios distinguish themselves by very high mean returns, with the highest value registered with ES risk constraint for the ICQF estimator (0.070%). Across estimators, the results are similar for both ICDF and ICQF estimators, and the non-financial vs. financial ranking is preserved in all cases. As for the other statistics, we note that the median daily returns are generally lower than the benchmark Eurostoxx50, mainly for the non-financial portfolios. Moreover, most of the weighted portfolios are not characterized by severe estimated risk, irrespective of wether we assume normality of the returns distribution. Again, we do not observe high discrepancies across estimators used in the algorithms, all of them showing a better loss protection than the benchmark.

As for differences between VaR and either the ES or WES risk measures, the return-to-risk profile (in terms of Sharpe ratio and also of the empirical return to ES at 95% level) improves in most cases, both across type of components and estimators. The standard deviation of returns is similar, with values that are higher only for the ICQF estimator and non-uniform weighting, in the non-financial components case. Regarding the loss measures (empirical ES and VaR at 95% level), they are less severe when using the ES rather than VaR risk constraint, while if we use the WES, the loss measures are more severe.

For each type of components, we report the stability measures $\rho^{KL}(t, S)$ and $\rho^{L^2}(t, S)$ for $S = 5$ and $S = 20$, which roughly correspond to the number of working days within one week and one month respectively. The stability of portfolio weights is similar across different risk

Table 4.5: Performance indicators and alternative stability indicators for the optimal portfolios with financial stocks

Indicator		ICDF		ICQF	
	VaR	ES	WES	ES	WES
(1)	0.033	0.033	0.030	0.035	0.029
(2)	1.4	1.2	1.4	1.3	1.5
(3)	0.049	0.052	0.066	0.042	0.061
(4)	2.1	1.9	2.1	1.9	2.4
(5)	3.4	2.9	3.5	3.1	3.6
(6)	3.2	2.9	3.3	3.0	3.5
(7)	2.8	2.5	2.9	2.7	3.16
(8)	38	43	34	42	31
(9)	1.0	1.2	0.9	1.1	0.8
(10)	1.2	1.3	1.0	1.3	0.9
$\rho^{KL}(t,5)$	80.7	81.1	78.2	89.0	78.2
$\rho^{L^2}(t,5)$	20.3	21.0	19.8	21.8	19.5
$\rho^{KL}(t,20)$	24.5	20.4	23.5	22.3	22.3
$\rho^{L^2}(t,20)$	11.2	10.7	11.0	10.8	10.5

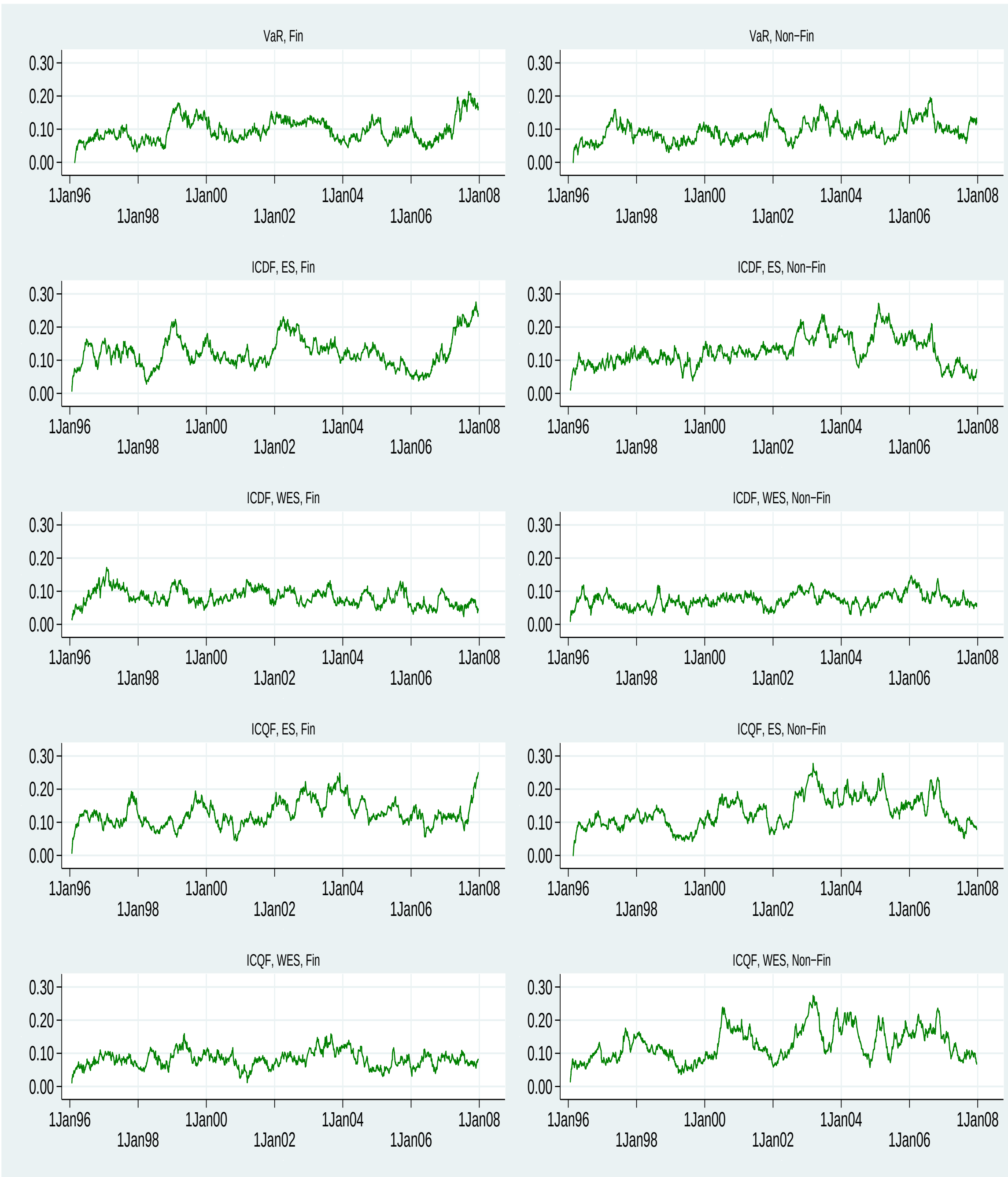
Table 4.6: Performance indicators and alternative stability indicators for the optimal portfolios with non-financial stocks

Indicator		ICDF		ICQF	
	VaR	ES	WES	ES	WES
(1)	0.051	0.056	0.059	0.070	0.055
(2)	1.2	1.1	1.3	1.2	1.4
(3)	0.075	0.062	0.075	0.065	0.075
(4)	2.0	1.7	2.2	1.8	2.1
(5)	2.8	2.5	3.1	2.6	3.2
(6)	2.8	2.6	3.1	2.7	3.2
(7)	2.5	2.3	2.7	2.4	2.8
(8)	65	78	70	94	63
(9)	1.8	2.2	1.9	2.7	1.7
(10)	2.0	2.4	2.2	3.0	2.0
$\rho^{KL}(t,5)$	82.2	77.7	79.6	90.0	78.7
$\rho^{L^2}(t,5)$	20.0	20.9	19.7	22.2	19.8
$\rho^{KL}(t,20)$	23.7	20.4	24.0	21.3	21.5
$\rho^{L^2}(t,20)$	10.9	10.7	11.1	11.0	10.5

constraints and measures that we propose. However, if we consider the weekly evaluation, portfolios that control for ES have a higher stability. For the monthly evaluation, portfolios that control for VaR have a higher $\rho^{KL}(t,20)$ measure in most cases. As for $\rho^{L^2}(t,20)$, it is also higher, except for the non-financial stocks case.

In Figure 4.1, we plot the standard deviation of the optimal portfolio weights. Lower standard deviation of the portfolio weights means higher diversification. This can also be regarded as an indicator of stability of the portfolio weights. Except for the case with non-financial components and ICQF estimator, the portfolios with WES risk constraint show a standard deviation of weights that is more stable with respect to portfolios that use either VaR or ES risk constraints.

Figure 4.1: Standard deviation of the optimal weights with different risk constraints



4.7 Conclusions

In this Chapter, we focus on the forecasting properties of the weighted estimators of ES and minimize some suitably defined forecast error. The we construct an asset allocation model that maximizes expected return with a constraint on some risk measure and solve it via a numerical algorithm. Using European daily data on financial and non-financial stocks, we develop an empirical application and compare the optimal portfolios that have either VaR, ES or WES risk constraints in terms of performance indicators. Moreover, we look at two measures of portfolio weights stability. Results show that, in general, the return-to-risk profile and the incurred losses of the optimal portfolios that use ES or WES as risk constraints are better, whilst the stability of components weights are lower mostly for shorter periods of time. This is in accordance with our stylized example.

Chapter 5

Summary and conclusions

Recently, financial applications have considered alternative risk measures of the $\alpha-$ VaR. One of them is the $\alpha-$level expected shortfall that represent the average of losses that are more severe than the VaR. The advantages of ES come from the coherence properties that VaR does not satisfy entirely (failing subadditivity).

We have considered the case when auxiliary information about the outcome of interest is available and extended the concept of ES accordingly. Our conditional estimators depart from two equivalent representations of the α-level expected shortfall in terms of the conditional distribution function and the conditional quantile function. The Monte Carlo experiments show that accuracy of the estimators increases rapidly with the level α and the sample size. Then, in the first empirical application, the predictive performance of the various estimators is assessed and the integrated conditional quantile function estimators spur as having a better performance than the unconditional estimator.

Then we generalize the estimators by means of a suitably defined weighting function. We do this as we aim to work on the statistical and forecasting properties of the weighted versions of the estimators of ES. Firstly, we analytically derive the weighting function such as to minimize their asymptotic variance. The efficiency gain that measures the difference in asymptotic variance appears to be significant for heavy tailed distributions. In order to achieve the second objective, the weighting is derived numerically such as to minimize some suitably defined forecast error of the estimators.

Lastly, the estimators are used within a simple asset allocation model that maximizes expected return with a constraint on either VaR, ES or weighted ES. The weighting is such that we minimize the forecast error of the estimator. The optimal portfolios are assessed in terms of performance indicators and also two measures of weights stability. The first set of indicators tend to improve when using ES or weighted ES rather than VaR, while stability of portfolio components weights is higher mostly when considering shorter periods of time.

Chapter 6

Appendix: Proofs for Chapter 3

Proofs of Theorems and Corollaries for Chapter 3 can be found in Leorato et al (26). Here we report proofs of Theorems 1 and 2 and Corollary 2.

Proof of Theorem 1

Under regularity conditions and assuming linearity in parameters, we have uniform consistency of the quantile regression process (see Koenker (25) and Angrist et al. (6)). More precisely,

$$\mathbf{J}_1(\cdot)\sqrt{T}[\hat{\beta}(\cdot) - \beta(\cdot)] \xrightarrow{d} Z_Q(\cdot), \tag{6.1}$$

where $\mathbf{J}_1(p) = \mathrm{E}\left[f(\beta(p)^\top X|X)XX^\top\right]$ is positive definite for all $p \in (0,1)$ and $z_Q(\cdot)$ is a zero mean Gaussian process defined by the covariance function $\Sigma_{(p_i,p_j)}$.

Now let $\tilde{\beta}_T(\alpha) = \sum_{i=1}^{I} w_i \hat{\beta}_T(p_i)$ be a linear combination of the I estimates $\hat{\beta}(p_1)\ldots\hat{\beta}(p_I)$ of the true population quantile coefficients $\beta(p_1)\ldots\beta(p_I)$. The weights vector $\mathbf{w} = (w_1,\ldots w_I)^\top$ is deterministic. Therefore, as matrix $\mathbf{J}_1(p)$ is positive definite for any $p \in (0,1)$, from (6.1) we have

$$\sqrt{T}\left(\tilde{\beta}_T(\alpha) - \sum_{i=1}^{I} w_i\beta(p_i)\right) \sim \mathcal{N}(0, AV(\tilde{\beta}_T(\alpha))) \tag{6.2}$$

with

$$AV(\tilde{\beta}_T(\alpha)) = (\mathbf{w}^\top \otimes \mathcal{I}_k)\,\Omega\,(\mathbf{w} \otimes \mathcal{I}_k)$$

and W, Ω and $\mathcal{I}_k$ defined as in (3.15).

It follows that

$$\sqrt{T}\left(\hat{\tau}_{w,I}^{(Q)}(\alpha\,|\,x) - \tau_{w,I}^{(Q)}(\alpha\,|\,x)\right) \sim \mathcal{N}(0, AV(\hat{\tau}_{w,I}^{(Q)}(\alpha\,|\,x))) \tag{6.3}$$

Proof of Theorem 2

The scheme of the proof is the following. We first establish uniform consistency of the logit regression process $y \mapsto \hat{\theta}(y)$ to $\theta(y)$. Finally, we prove asymptotic Gaussianity of the process $\sqrt{T}[\hat{\theta} - \theta]$ by an empirical process based approximation of $\hat{\theta}$.
Uniform consistency of $\hat{\theta}(\cdot)$

$\mathrm{E}\,\|X\| < \infty$ implies that $\mathrm{E}_\infty\left|\left(\mathbb{1}\{Y \leq y\}\theta^\top X - \ln(1 + e^{\theta^\top X})\right)\right|$ is finite and uniquely minimized at $\theta(y)$, $\forall y \in \mathbb{R}$.

We first show uniform convergence. This is equivalent to: for any compact set Θ,

$$\mathcal{L}_T(y,\theta) = \mathcal{L}_\infty(y,\theta) + o_{p^*}(1)$$

uniformly in $(y,\theta) \in (\mathbb{R}, \mathbb{R}^k)$, which follows from the Khinchine law of large numbers and also from stochastically equicontinuity of the empirical process $(y,\theta) \mapsto \mathcal{L}_T(y,\theta)$ as for any pairs (y',θ'), (y'',θ'') we have

$$|\,\mathcal{L}_T(y',\theta') - \mathcal{L}_T(y'',\theta'')\,| \leq C_{1T}^\top\,|\,\theta' - \theta''\,| + C_{2T}\,|\,|\,(F_Y(y'') - F_Y(y')\,|\,|$$

where $C_{1T} = 2\mathbb{E}_T\|X\| = O_p(1)$ and $C_{2T} = \mathbb{E}_T\|X\|\sup_{\theta\in\Theta}\|\theta\|(1 + O_p(T^{-1/2})) = O_p(1)$.

Now consider a collection of closed balls $\mathbf{B}_M(\theta(y))$ of radius M and center $\theta(y)$ and let $\theta_M(y) = \theta(y) + \delta_M(y)\nu(y)$, where $\nu(y)$ is a direction vector with unity norm $\|\nu(y)\| = 1$ and $\delta_M(y)$ is a positive scalar such that $\delta_M(y) \geq M$. Let $\theta_M^\star$ be the point of the boundary of $\mathbf{B}_M(\theta(y))$ on the line connecting $\theta_M(y)$ and $\theta(y)$. From concavity of $\mathcal{L}_T(y,\cdot)$

$$\mathcal{L}_T(y,\theta_M^\star) \geq \frac{M}{\delta_M}\mathcal{L}_T(y,\theta_M) + \left(1 - \frac{M}{\delta_M}\right)\mathcal{L}_T(y,\theta)$$

we get

$$\frac{M}{\delta_M}\left(\mathcal{L}_T(y,\theta) - \mathcal{L}_T(y,\theta_M)\right) \geq \mathcal{L}_T(y,\theta) - \mathcal{L}_T(y,\theta_M^\star)$$

and thus, because of the uniform convergence proved above.

$$\frac{M}{\delta_M}\left(\mathcal{L}_T(y,\theta) - \mathcal{L}_T(y,\theta_M)\right) \geq \mathcal{L}_\infty(y,\theta) - \mathcal{L}_\infty(y,\theta_M^\star) + o_{p^\star}(1) > \epsilon_M + o_{p^\star}(1).$$

The last inequality follows from the fact that $\theta(y)$ is the unique maximizer of $\mathcal{L}_\infty(y,\cdot)$. Since this holds for every $M > 0$, the estimate $\hat\theta(y)$ must lie in a radius$-M$ ball centered at $\theta(y)$ uniformly for all y and with probability approaching 1. Otherwise we would find $\mathcal{L}_T(y,\theta) - \mathcal{L}_T(y,\hat\theta) \geq \frac{\delta_M}{M}\epsilon_M + o_p^\star(\delta_M/M)$ for an arbitrarily large value of δ_M/M and $\forall T$ which contradicts the fact that $\hat\theta$ minimizes $\mathcal{L}_T(y,\cdot)$.

Asymptotic Gaussianity of $\sqrt{T}(\hat\theta - \theta)$

We remark that the class

$$\mathcal{F} = \left\{f(x) = \theta^\top x, \theta \in \Theta\right\} \tag{6.4}$$

is a VC-class of functions. In fact, the *subgraphs* are given by the class

$$\{x \in \mathbb{R}^k :\ \theta^\top x \leq t\}$$

for $\theta^T \in \mathbb{R}^k$ and $t \in \mathbb{R}$. In particular, if $\Theta = \mathbb{R}^k$, then the class is the class of halfspaces and has $VC-$index equal to $k+2$ (see Van der Vaart and Wellner (39), Ch. 2.6 Problems and Complements No.14).

Then, by Theorem 2.6.7 in Van der Vaart and Wellner (39), if $F(x)$ is a square integrable (with respect to a probability measure Q) envelope function for $\mathcal{F}$, we find the following bound for the covering numbers of the class $\mathcal{F}$:

$$N\left(\varepsilon\|F\|_{Q,2}, \mathcal{F}, L_2(Q)\right) \leq K\frac{1}{\varepsilon^{2(k+1)}} \tag{6.5}$$

where K is a constant depending on the dimension k.

This implies that

$$\int_0^\delta \sqrt{\log N\left(\varepsilon\|F\|_{Q,2}, \mathcal{F}, L_2(Q)\right)}d\varepsilon < \infty \tag{6.6}$$

because of $\int_0^1 \log(1/\varepsilon)d\varepsilon < \infty$ and this yelds that $\mathcal{F}$ is a Donsker class of functions.

Now, we can use Theorem 2.10.20 in Van der Vaart and Wellner (39), to conclude that the class of functions

$$\left\{\frac{e^{\theta^\top x}}{1+e^{\theta^\top x}},\ \theta^\top \in \Theta\right\}$$

is also Donsker. We state a simplified version of the theorem for convenience reasons.

Let $\mathcal{F}$ be a class of measurable real functions with a measurable envelope function F. Let $\phi : \mathbb{R} \mapsto \mathbb{R}$ be a map satisfying

$$|\phi \circ f(x) - \phi \circ g(x)|^2 \leq L_\alpha^2(x)|f-g|^{2\alpha}(x)$$

for a constant L_α, $\alpha \in (0,1]$ for every $f, g \in \mathcal{F}$ and for all $x \in \mathbb{R}$.

Then, for every $\delta > 0$,

$$\begin{aligned}
&\int_0^\delta \sup_Q \sqrt{\log N\left(\varepsilon\|L_\alpha F^\alpha\|_{Q,2}, \phi(\mathcal{F}), L_2(Q)\right)}d\varepsilon \\
&\leq \int_0^{\delta^{1/\alpha}} \sup_Q \sqrt{\log N\left(\varepsilon\|F\|_{Q,2\alpha}, \mathcal{F}, L_{2\alpha}(Q)\right)}\frac{d\varepsilon}{\varepsilon^{1-\alpha}}
\end{aligned} \tag{6.7}$$

where the supremum is taken over all finitely discrete probability measures Q. Then, if the right hand side is finite and $P^*(L_{\alpha f^\alpha}) < \infty$, $\phi(\mathcal{F})$ is a Donsker class, provided of course that its members are square integrable and that the class is measurable.

We can apply the above theorem to our context by defining $\phi(u) = \frac{e^u}{1+e^u}$ and by observing that

$$\|\phi \circ f - \phi \circ g\|^2(x) = \left|\frac{e^f}{1+e^f} - \frac{e^g}{1+e^g}\right|^2(x) \leq |f-g|^2(x)$$

Thus, in our case the condition of Theorem 2.10.20 is satisfyed with $\alpha = 1$, $L_1 = 1$. Then, for every $\delta > 0$,

$$\begin{aligned}
&\int_0^\delta \sup_Q \sqrt{\log N\left(\varepsilon\|F\|_{Q,2}, \phi(\mathcal{F}), L_2(Q)\right)}d\varepsilon \\
&\leq \int_0^\delta \sup_Q \sqrt{\log N\left(\varepsilon\|F\|_{Q,2}, \mathcal{F}, L_2(Q)\right)}d\varepsilon
\end{aligned} \tag{6.8}$$

where $\mathcal{F}$ is the class defined by (6.4). We have discussed and proved above that the integral in the right hand side of (6.8) is finite, and thus the class of logit functions

$$\phi(\mathcal{F}) = \left\{\frac{e^{\theta^\top x}}{1+e^{\theta^\top x}},\ \theta^\top \in \Theta\right\}$$

is a Donsker class.

The functional class $\mathcal{I}$=$\{\mathbb{1}\{Y \leq y\}, y \in R\}$ is a VC subgraph class and therefore a bounded Donsker class. Consequently, the functional classes $\varphi = \mathcal{I} - \phi(\mathcal{F})$ and φX are bounded

Donsker, the latter with square integrable envelope $2\max_{j=1\dots k}|X_j|$ (Theorem 2.10.6 Van der Vaart and Wellner (39)).

We are now able to find an approximation for $\mathbb{G}_T[\varphi(y,\hat{\theta})X]$ that is a functional of $\varphi(y,\hat{\theta})$. First, we remark that the mapping $(y,\theta)\mapsto\mathbb{G}_T[\varphi(y,\theta)X]$ is stochastically equicontinuous over $\mathbb{R}\times\mathbb{R}^k$ with respect to the $L_2(P)$ pseudometric

$$\rho\left((y',\theta'),(y'',\theta'')\right)^2=\max_{j=1\dots k}\mathrm{E}\left[\left(\varphi(y',\theta')X_j-\varphi(y'',\theta'')X_j\right)^2\right]$$

where X_j are the components of X. Moreover, as $\sup_{y\in\mathbb{R}}\|\hat{\theta}(y)-\theta(y)\|=o_{p^\star}(1)$, that follows from convergence w.r.t. pseudometric

$$\sup_{y\in\mathbb{R}}\rho\left((y,\hat{\theta}(y)),(y,\theta(y))\right)^2=o_p(1)$$

(here, the boundness condition $\mathrm{E}\,\|X\|^{2+\epsilon}<\infty$ is used), we conclude that

$$\mathbb{G}_T[\varphi(y,\hat{\theta})X]=\mathbb{G}_T[\varphi(y,\theta)X]+o_{p^\star}(1)\quad\text{in}\quad l^\infty(\mathbb{R}) \tag{6.9}$$

With a Taylor expansion of $\mathrm{E}[\varphi(y,\theta)X]\,|_{\,\theta=\hat{\theta}(y)}$ around θ and using uniform consistency of $\hat{\theta}(y)$ and the assumed uniform continuity and boundness of the mapping $y\mapsto(F(y)(1-F(y)))$, we have that uniformly in $y\in\mathbb{R}$,

$$\mathrm{E}[\varphi(y,\theta)X]\,|_{\,\theta=\hat{\theta}(y)}=[\mathbf{J}_2(\cdot)+o_p(1)]\,[\hat{\theta}(y)-\theta(y)] \tag{6.10}$$

From first order condition of (3.22) and from $\mathrm{E}\,\|X\|^{2+\epsilon}<\infty$ we obtain that $\sqrt{T}\,\mathbb{E}_T[\varphi(y,\hat{\theta})X]=o_p(1)$ and, as

$$\sqrt{T}\,\mathbb{E}_T[\varphi(y,\hat{\theta})X]=\sqrt{T}\,\mathrm{E}[\varphi(y,\theta)X]\,|_{\,\theta=\hat{\theta}(y)}+\mathbb{G}_T[\varphi(y,\hat{\theta})X]$$

it follows (using (6.9) and (6.10)) that, uniformly in $y\in\mathbb{R}$,

$$[\mathbf{J}_2(y)+o_p(1)]\,\sqrt{T}[\hat{\theta}(y)-\theta(y)]+\mathbb{G}_T[\varphi(y,\theta)X]=o_p(1) \tag{6.11}$$

Moreover, as $\mathbf{J}_2(y)$ is positive definite, then mineig$[\mathbf{J}_2(y)]\geq\lambda>0$ and

$$(\sqrt{\lambda}+o_p(1))\sqrt{T}\sup_{y\in\mathbb{R}}\|\hat{\theta}(y)-\theta(y)\|\leq\sup_{y\in\mathbb{R}}\|\mathbb{G}_T[\varphi(y,\theta)X]+o_p(1)\| \tag{6.12}$$

The mapping $y\mapsto\theta(y)$ is continuous and $y\mapsto\mathbb{G}_T[\varphi(y,\theta)X]$ is stochastically equicontinuous over $\mathbb{R}$. A multivariate central limit theorem imply that, in $l^\infty(\mathbb{R})$, $\mathbb{G}_T[\varphi(y,\theta)X]\xrightarrow{d}Z_\theta(\cdot)$, a zero mean Gaussian process defined by the covariance matrix $\Sigma_{2;j,k}$. Hence, (6.12) yields $\sup_{y\in\mathbb{R}}\sqrt{T}\|\hat{\theta}(y)-\theta(y)\|=O_{p^\star}(1)$ and taking into account (6.11) we conclude that, in $l^\infty(\mathbb{R})$

$$\mathbf{J}_2(y)\sqrt{T}[\hat{\theta}(y)-\theta(y)]=-\mathbb{G}_T[\varphi(y,\theta)X]+o_{p^\star}(1)\xrightarrow{d}Z_\theta(\cdot) \tag{6.13}$$

□

Proof of Corollary 2

As $\mathbf{J}_2(y)$ is positive definite, (6.13) writes as

$$\sqrt{T}[\hat{\theta}(y) - \theta(y)] = \mathbb{G}_T\left[-\mathbf{J}_2^{-1}(y)\varphi(y,\theta)X\right] + o_p(1)$$

Hence, for any $x \in \mathbb{R}^k$,

$$\sqrt{T}[x^\top\hat{\theta}(y) - x^\top\theta(y)] = \mathbb{G}_T\left[-x^\top\mathbf{J}_2^{-1}(y)\varphi(y,\theta)X\right] + o_p(1) \tag{6.14}$$

By a Taylor expansion of $F(y)$ around $(x^\top\theta(y))$, we obtain

$$\hat{F}(y) - F(y) = [x^\top\hat{\theta}(y) - x^\top\theta(y)][F(y)(1-F(y))] + o_p(1)$$

and, from (6.14) we get an empirical process approximation for the function $\sqrt{T}[\hat{F}(y)-F(y)]$ as

$$\sqrt{T}[\hat{F}(y) - F(y)] = \mathbb{G}_T\left[-[F(y)(1-F(y))]x^\top\mathbf{J}_2^{-1}(y)\varphi(y,\theta)X\right] + o_p(1) \tag{6.15}$$

Consider now a continuous and differentiable function $w(\cdot)$ with first derivative w' strictly positive. Then, under the condition that for all values of y we have that $f_Y(y)$, the marginal density of Y, is strictly positive, we can expand $w(F(y))$ around $F(y)$ and, from (6.15), we get

$$\sqrt{T}[w(\hat{F}(y)) - w(F(y))] = \mathbb{G}_T\left[-w'(F(y))[F(y)(1-F(y))]x^\top\mathbf{J}_2^{-1}(y)\varphi(y,\theta)X\right] + o_p(1) \tag{6.16}$$

In order to derive the empirical process that approximates the estimator

$$\hat{\delta}_w(x,y) = \sum_{i=1}^{I}(y_i - y_{i-1})\, w(\hat{F}_i(x))$$

of $\delta_w(x,y) = \sum_{i=1}^{I}(y_i - y_{i-1})\, w(\hat{F}_i(x))$, we first simplify notation and define

$$\gamma_\theta(x,y) = w'(F(y))[F(y)(1-F(y))]x^\top\mathbf{J}_2^{-1}(y)\varphi(y,\theta)X$$

Moreover, departing from the grid of points $\{y_0,\ldots,y_I\}$, we define the I-dimensional column vector of differences $D = (dy_1,\ldots,dy_I)^\top$, with $dy_i = y_i - y_{i-1}$ for $i = 1\ldots I$, and also the scalar

$$\gamma_{\delta_w}(y_1,\ldots,y_I,x) = D^\top\left(\gamma_\theta(x,y_1),\ldots,\gamma_\theta(x,y_I)\right)^\top.$$

Then (6.16) yields

$$\sqrt{T}[\hat{\delta}_w(x,y) - \delta_w(x,y)] = \mathbb{G}_T\left[-\gamma_{\delta_w}(y_2,\ldots,y_I,x)\right] + o_p(1) \tag{6.17}$$

Now we can write down the empirical process approximation for the WICDF estimator

$$\hat{\tau}_{w,I}^{(D)}(\alpha\,|\,x) = \hat{Q}(\alpha\,|\,x) - \sum_{i=1}^{I}(y_i - y_{i-1})\, w(\hat{F}_i(x))$$

and for this purpose we rewrite result (3.21) as

$$\sqrt{T}[\hat{Q}(\alpha\,|\,x) - Q(\alpha\,|\,x)] = \mathbb{G}_T\left[\gamma_Q(\alpha\,|\,x)\right] + o_p(1) \tag{6.18}$$

where $\gamma_Q(\alpha\,|\,x)$ is a scalar defined as

$$\gamma_Q(\alpha\,|\,x) = x^\top \mathbf{J}_1^{-1}(\alpha)(1\{Y \leq \beta(\alpha)^\top X\} - \alpha)X$$

The functional classes $\{\gamma_Q(\alpha\,|\,x)\}$ and $\{\gamma_{\delta_w}(y_2,\ldots,y_I,x)\}$ are bounded Donsker classes with square integrable envelopes equal to $(2\max_{y\in R} f(y))$ and

$$(32\max_{y\in R, j=1\ldots k} w'(F(y))|X_j|)$$

respectively. Consequently, the functional class $\{\gamma_Q(\alpha\,|\,x)+\gamma_{\delta_w}(y_2,\ldots,y_I,x)\}$ is also bounded Donsker, with square integrable $(\max_{y\in R, j=1\ldots k} 2\cdot f(y) + 32\cdot w'(F(y))|X_j|)$ (Theorem 2.10.6 Van der Vaart and Wellener). We can therefore write

$$\sqrt{T}[\hat{\tau}^{(D)}_{w,I}(\alpha\,|\,x) - \tau^{(D)}_{w,I}(\alpha\,|\,x)] = \mathbb{G}_T\left[\gamma_Q(\alpha\,|\,x) + \gamma_{\delta_w}(y_2,\ldots,y_I,x)\right] + o_p(1) \tag{6.19}$$

and then $\sqrt{T}[\hat{\tau}^{(D)}_{w,I}(\alpha\,|\,x) - \tau^{(D)}_{w,I}(\alpha\,|\,x)] \xrightarrow{d} Z_D(\cdot)$, where $Z_D(\cdot)$ is a zero mean Gaussian process defined by the covariance function

$$\begin{aligned} &AV(\hat{\tau}^{(D)}_{w,I}(\alpha\,|\,x)) = \mathrm{E}[(\gamma_Q(\alpha\,|\,x) + \gamma_{\delta_w}(y_2,\ldots,y_I,x))^2] - \mathrm{E}^2[\gamma_Q(\alpha\,|\,x) + \gamma_{\delta_w}(y_2,\ldots,y_I,x)] \\ &= T\,AV(\hat{Q}(\alpha\,|\,x)) + T\,AV(\delta_w(y_2,\ldots,y_I,x)) - 2T\,Acov(Q(\alpha\,|\,x),\delta_w(y_2,\ldots,y_I,x)) \end{aligned} \tag{6.20}$$

$AV(Q(\alpha\,|\,x))$ and $AV(\delta_w(y_2,\ldots,y_I,x))$ are straightforward, while for the last term we need $\mathrm{E}[Q(\alpha\,|\,x)\delta_w(y_1,\ldots,y_I,x)]$ that is equal to

$$\begin{aligned} &\sum_{i=1}^{I} x^\top dy_i W'(F_i(x))F_i(x)(1-F_i(x))\cdot \\ &\mathrm{E}\left[J_Q(\alpha)^{-1}\mathbf{J}_2(y_i)^{-1}\left(\mathbb{1}\{Y\leq X\beta(\alpha)\}\cdot\mathbb{1}\{Y\leq y_i\} - \mathbb{1}\{Y\leq X\beta(\alpha)\}\cdot\frac{e^{\theta_i^\top X}}{1+\theta_i^\top X}\right.\right. \\ &\left.\left.-\alpha\cdot\mathbb{1}\{Y\leq y_i\} + \alpha\cdot\frac{e^{\theta_i^\top X}}{1+\theta_i^\top X}\right)XX^\top\right]x \\ &= \sum_{i=1}^{I} x^\top dy_i W'(F_i(x))F_i(x)(1-F_i(x))\cdot\mathrm{E}\left[\mathbf{J}_1(\alpha)^{-1}\mathbf{J}_2(y_i)^{-1}\left(\min(\alpha,F_i(x)) - \alpha F_i(x)\right)XX^\top\right]x \end{aligned}$$

Bibliography

[1] Acerbi C. Spectral Measures of risk: a coherent represenatation of subjective risk aversion. *Journal of Banking and Finance* 2002; **26**: 1505–1518.

[2] Acerbi C, Tasche D. On the coherence of expected shortfall. *Journal of Banking and Finance* 2002; **26**: 1487–1503.

[3] Alexander GJ, Baptista AM. Conditional expected loss as a measure of risk: Implications for Portfolio Selection. *Management Science* 2004; **50**: 1261–1273.

[4] Alexander C. *Quantitative Methods in Finance, vol I.* Wiley, 2008.

[5] Allais M. Le Comportement de l'Homme Rationnel devant le Risque: Critique des Postulats et Axiomes de l'Ecole Americaine *Econometrica* 1953; **21**: 503–546.

[6] 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.

[7] Artzner P, Delbaen F, Eber J-M, Heath D. Coherent measures of risk. *Mathematical Finance* 1999; **9**: 203–228.

[8] Barbu V, Precupanu T. *Convexity and optimization in Banach spaces.* Sÿthoff and Nordhoff International Publishers: Alphen aan den Rijn, 1978.

[9] Basak S, Shapiro A. Value-at-Risk-Based risk management: optimal policies and asset prices. *The Review of financial studies* 2001; **14**: 371–405.

[10] Bassett GW, Koenker R, Kordas G. Pessimistic portfolio allocation and Choquet expected utility. *Journal of Financial Econometrics* 2004; **2**: 477–492.

[11] Bertsimas D, Lauprete GJ, Samarov A. Shortfall as a risk measure: properties, optimization and applications. *Journal of Economic Dynamics & Control* 2004; **28**: 1353–1381.

[12] Brooks SH. A discussion for random methods for seeking maxima. *Operations Reasearch* 1958; **6**: 244–251.

[13] Cherny AS. Weighted VaR and its properties. *Finance and Stochastics* 2006; **10**: 367–393.

[14] Csörgö S, Haeusler E, Mason DM. The asymptotic distribution of extreme sums. *The Annals of Probability* 1991, **19**: 783–811.

[15] Cuoco D, He H, Issaenko S. Optimal dynamic trading strategies with risk limits. *Operational Research* 2007: online.

[16] Delbaen F. Coherent risk measures on general probability spaces. *preprint. ETH Zuřich* 2000.

[17] Dodge J, Jurečková J *Adaptive Regression*. Springer: Berlin, 2000.

[18] Embrechts P, Klüppelberg C, Mikosch T. *Modelling Extremal Events*. Springer: Berlin, 1997.

[19] Ghaoui L, Oks M, Oustry F. Worst-case Value-at-Risk and Robust Portfolio Optimization. *Operations Reasearch* 2003; **51**: 543–556.

[20] Gonzales R, Wu G. On the shape of the probability weighting function. *Cognitive Psychology* 1999; **38**: 129–166.

[21] Gundel A, Weber S. Utility maximization under a shortfall risk constraint. *Journal of Mathematical Economics* 2008; **44**: 1126–1151.

[22] Hirsch MW, Smale S. *Differential Equations, Dynamical Systems, and Linear Algebra*. Academic Press: New York, 1974

[23] Kahneman D, Tversky A. An analysis of decision under risk. *Econometrica* 1979, **47**: 263–291.

[24] Koenker R, Bassett G. Regression quantiles. *Econometrica* 1978, **46**: 33–50.

[25] Koenker R. *Quantile Regression*. Cambridge University Press: New York, 2005.

[26] Leorato S, Peracchi F, Tanase A.V. Efficient expected shortfall estimation. *Working paper* 2009.

[27] Lillo WE, Loh MH, Hui S, Zak SH. On solving constrained optimization problems with neural networks: a penalty method approach. *Transactions on Neural Networks* 1993; **4**.

[28] Mansini R, Ogryczak W, Speranza MG. Conditional value at risk and related linear programming models for portfolio optimization. *Annals of Operational Research* 2007, **152**: 227–256.

[29] McLachlan G, Peel D. *Finite Mixture Models*. Wiley: New York, 2000.

[30] 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.

[31] Meucci A. *Risk and asset allocation*. Springer-Verlag Berlin, 2005.

[32] Mulvey JM, Vanderbei RJ, Zenios SA. Robust optimizaion of large-scale systems. *Operations Research* 1995; **43**: 264–281.

[33] Natarajan K, Pachamanova D, Sim M. Incorporating Asymmetric Distributional Information in Robust Value-at-Risk Optimization. *Management Science* 2008, **54**: 573–585.

[34] Ortobelli L, Rachev ST. Safety-first analysis and stable Paretian approach to portfolio choice theory. *Mathematical and Computer Modelling* 2001, **34**: 1037–1072.

[35] Peracchi F. On estimating conditional quantiles and distribution functions. *Computational Statistics and Data Analysis* 2002, **38**: 433–477.

[36] Peracchi F, Tanase AV. On estimating the conditional expected shortfall. *Applied Stochastic Models in Business and Industry* 2008, **24**: 471–493.

[37] Prelec D. The probability weighting function. *Econometrica* 1998, **66**: 497–527.

[38] Shorrocks AF. Ranking income distributions. *Econometrica* 1983, **50**: 3–17.

[39] van der Vaart AW, Wellner Jon A *Weak convergence and empirical processes: with applications to statistics* . Springer-Verlag: Berlin, New York, 1996.

[40] Yaari ME. The dual theory of choice under risk. *Econometrica* 1987, **55**: 95–115.