EIEF WORKING PAPER SERIES

Einaudi Institute for Economics and Finance

EIEF Working Paper 13/10
June 2010

Asymptotically Efficient Estimation of the Conditional Expected Shortfall

by

Samantha Leorato
(University of Rome "Tor Vergata")
Franco Peracchi
(University of Rome "Tor Vergata" and EIEF)
Andrei V. Tanase
(National Bank of Romania)

Asymptotically efficient estimation of the conditional expected shortfall[☆]

Samantha Leorato[a,*], Franco Peracchi[a,b], Andrei V. Tanase[c]

[a] *University of Rome Tor Vergata, Italy*
[b] *EIEF, Italy*
[c] *National Bank of Romania*

Abstract

This paper proposes a procedure for efficient estimation of the trimmed mean of a random variable conditional on a set of covariates. For concreteness, the paper focuses on a financial application where the trimmed mean of interest corresponds to the conditional expected shortfall, which is known to be a coherent risk measure. The proposed class of estimators is based on representing the estimand as an integral of the conditional quantile function. Relative to the simple analog estimator that weights all conditional quantiles equally, asymptotic efficiency gains may be attained by giving different weights to the different conditional quantiles while penalizing excessive departures from uniform weighting. The approach presented here allows for either parametric or nonparametric modeling of the conditional quantiles and the weights, but is essentially nonparametric in spirit. The paper establishes the asymptotic properties of the proposed class of estimators. Their finite sample properties are illustrated through a set of Monte Carlo experiments and an empirical application.

Keywords: expected shortfall, quantile regression, asymptotic efficiency.

1. Introduction

Quantile regression, introduced by Roger Koenker and Gib Bassett (Koenker and Bassett [19]), has gradually evolved from a robust alternative to least squares to a way of summarizing the conditional distribution of a random variable given a set of covariates. As such, it can be used in a large variety of situations. In this paper we employ quantile regression methods to estimate the trimmed mean of a random variable of interest conditional on a set of covariates. Trimmed means are widely used as alternative location parameters to the ordinary mean because of their robustness and their superior properties under certain types of censoring. They are usually not

[☆] We thank Iván Fernández-Val, Michael Lechner, the Associate Editor and two anonymous referees for useful suggestions. We also thank Zongwu Cai and Xian Wang for sharing with us their Matlab code.

[*] Corresponding author

Email addresses: samantha.leorato@uniroma2.it (Samantha Leorato), franco.peracchi@uniroma2.it (Franco Peracchi), andrei.valentin.tanase@uniroma2.it (Andrei V. Tanase)

Preprint submitted to Elsevier *December 1, 2010*

of direct interest, however, in the sense that, absent other considerations such as robustness or censoring, one would be perfectly happy with the ordinary mean. Here we focus instead on a financial application where the trimmed mean is of substantive interest in itself as a coherent measure of risk.

Specifically, let Y_t be a continuous random variable that represents the uncertain return on a single asset or a portfolio of assets between time t and time $t+1$, and let X_t be a set of covariates that represent the relevant information available up to time t. This information typically consists of lagged values of other financial or nonfinancial variables, possibly including lagged values of Y_t itself. Let $f(y \,|\, x)$ and $Q(\alpha \,|\, x)$, with $0 < \alpha < 1$, respectively denote the conditional density and the αth conditional quantile of Y_t given $X_t = x$. Then the trimmed mean of interest is

$$\tau^0(\alpha \,|\, x) = \frac{1}{\alpha} \int_{-\infty}^{Q(\alpha \,|\, x)} y\, f(y \,|\, x)\, \mathrm{d}y, \tag{1.1}$$

namely the mean over the left tail of the conditional distribution of Y_t up to the αth quantile. In the financial literature, this is known as the α-level conditional expected shortfall (CES) of Y_t, with α typically set to .05 or .10. The negative CES corresponds to the loss expected when $X_t = x$ from holding the asset or the portfolio, given that the loss exceeds the αth conditional quantile of Y_t, a quantity known in the financial literature as the $(1-\alpha)$-percent conditional Value-at-Risk (VaR). The CES provides a natural way of incorporating information on economic and market conditions into a measure of potential loss that is continuous in α and, unlike the VaR, is always coherent, that is, it simultaneously satisfies sub-additivity, monotonicity, positive homogeneity and translation invariance (Artzer et al. [4]). For further references, see Acerbi and Tasche [1], Delbaen [13], and Bertsimas et al. [6], among others.

Most existing estimators of the CES are plug-in estimators based either on (1.1) or on alternative characterizations. Cai and Wang [8] proposed a class of nonparametric estimators obtained by replacing the density in (1.1) with a kernel based estimate. Their estimators, called weighted double kernel local linear (WDKLL) estimators, combine the attractive features of the double-kernel local linear estimator of Fan and Gijbels [14] with those of the weighted Nadaraya-Watson estimators of Hall et al. [17], especially its monotonicity and good boundary behavior. Their main drawback is computational complexity and the fact that their rate of convergence is slow and decays rapidly with the number of covariates reflecting the curse-of-dimensionality problem.

Peracchi and Tanase [25] proposed a class of semiparametric estimators based on the following equivalent representation of the CES

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

where $F(y \,|\, x) = \int_{-\infty}^{y} f(u \,|\, x) \mathrm{d}u$ is the conditional distribution function (CDF) of Y_t. These estimators, called integrated conditional distribution function (ICDF) estimators, combine a parametric or semi-parametric estimator of the CDF with a quantile regression estimator of the $(1-\alpha)$-percent conditional VaR.

Another class of semi-parametric estimators, also proposed by Peracchi and Tanase [25], is based on the equivalent representation of the CES as an integral of the conditional quantile function (CQF), that is,

$$\tau^0(\alpha \,|\, x) = \frac{1}{\alpha} \int_0^{\alpha} Q(p \,|\, x) \, \mathrm{d}p.$$

These estimators, called integrated conditional quantile function (ICQF) estimators, are of the form

$$\hat{\tau}^0_J(\alpha \,|\, x) = \alpha^{-1} \sum_{j=1}^{J} (p_j - p_{j-1}) \hat{Q}_j(x),$$

where $\hat{Q}_j(x)$ is an estimator of the conditional quantile $Q(p_j \,|\, x)$ and p_j is a point in the interval $(0, \alpha]$. Notice that, unlike the L-estimators analyzed by Koenker [20], which are based on a fixed grid of p_j points, an ICQF estimator is based on a grid of points whose number J and location is allowed to depend on the data. A closely related estimator has recently been suggested by Wang and Zhou (2010) for estimating the conditional mean of a monotone transformation of a random variable Y_t. They assume a heteroskedastic regression model for Y_t and exploit the property of equivariance to monotone transformations of the quantile function. Their estimator of the conditional mean is based on integrating the estimated CQF over a trimmed interval, where the trimming proportion vanishes as the sample size increases. They assume independent and identically distributed data, an assumption that we weaken here in order to deal with financial applications.

This paper generalizes the ICQF estimator by introducing a weighting scheme that weighs the J conditional quantile estimates differently. Estimators of this type are called weighted integrated conditional quantile function (WICQF) estimators. Intuitively, introducing nonuniform weights enables one to compensate the inefficiency of quantile estimators at extreme quantiles by giving more weight to the quantiles near α, which are more precisely estimated. The idea of introducing a set of weights to increase asymptotic efficiency when estimating a population parameter of interest is widely used in parametric and nonparametric statistics, and is a key feature of generalized method of moments and minimum distance methods. Of course, weighting may also introduce bias, which we control by penalizing excessive departures from uniform weighting.

Koenker (2005) proposed a weighted version of the linear quantile regression estimator with the aim of improving efficiency. The difference with respect to our approach is substantial: in his approach the weights enter the minimization problem that defines the estimator. We instead proceed on a two-step basis: first we estimate all the necessary quantiles (not necessarily via linear quantile regression), then we average them using a set of weights chosen via a minimum penalized variance criterion. Although we focus on estimating the CES, our method applies with minor changes to more general trimmed means, for example two-sided trimmed means with limits defined by conditional quantiles or other functions of X_t.

Asymptotically, a WICQF estimator corresponds to replacing the CDF in the definition of the CES by a transformation $W(F(y\,|\,x)\,|\,x)$, which is itself a CDF if the function $W(\cdot\,|\,x)$ is nondecreasing on $(0, \alpha]$. The use of a transformed version of the CDF in the definition of the CES may be related to the theory of non-expected utility of Yaari [29] and Prelec [26], where modifying the distribution of the returns accommodates risk aversion of the investor. We do not pursue this subjective interpretation and confine ourselves to weighting as a way of improving asymptotic efficiency of estimation.

ICQF and WICQF estimators depend crucially on the underlying estimates of the CQF. An important drawback of conventional quantile regression estimators is the fact that they do not guarantee monotonicity. When using linear quantile regression estimators, the linearity assumption is an additional problem because its failure may lead to bias. Despite this problem, linear quantile regression estimators are widely used because of parsimony, computational convenience, and the fact that they remain asymptotically normal under model misspecification (Angrist et al. [3]). They can also be used as preliminary nonmonotonic curves to be rearranged according to the method recently proposed by Chernozhukov et al. [11]. For these reasons, although presenting the asymptotic results for arbitrary estimators of the CQF, in the Monte Carlo and in the empirical exercise, we focus on the case when the quantile regression model is linear, or at least linear in the parameters.

The remainder of the paper is organized as follows. Section 2 formally defines the class of WICQF estimators. Section 3 analyzes their asymptotic properties. Section 4 discusses how to choose an optimal estimator. Section 5 presents the results of a set of Monte Carlo experiments. Section 6 presents an application to real data to highlight the potentials of our procedure. Finally, Section 7 concludes. All proofs are collected in the appendix.

2. Definition of WICQF estimators

Let the uncertain return on a given asset or portfolio between time t and time $t+1$ be represented by a continuous random variable Y_t with values in $\mathcal{Y}$, and let the information about Y_t available up to time t be represented by a K-dimensional random vector $X_t = (X_{t1}, \ldots, X_{tK})$ with values in $\mathcal{X}$. We assume that the data $\{(X_t, Y_t), t = 1, \ldots, T\}$ is a sample from a stationary strongly-mixing time series. This assumptions, which covers the case when the data are independent and identically distributed (iid), is relevant for the Monte Carlo simulations in Section 5 and the financial application in Section 6, where X_t includes lagged values of Y_t. Let $F(y \,|\, x) = \Pr\{Y_t \leq y \,|\, X_t = x\}$, $f(y \,|\, x) = (\partial/\partial y)\, F(y \,|\, x)$ and $Q(\alpha \,|\, x) = \inf\{y\colon F(y \,|\, x) \geq \alpha\}$, $\alpha \in (0, 1)$, respectively denote the CDF, the conditional density and the CQF of Y_t given $X_t = x$.

A WICQF estimator of the CES is any estimator of the form

$$\hat{\tau}_J(\alpha \,|\, x) = \sum_{j=1}^{J} w_j(x) \hat{Q}_j(x), \tag{2.1}$$

where $w_j(x)$ is the weight assigned to an estimate $\hat{Q}_j(x)$ of the p_jth conditional quantile of Y_t given $X_t = x$ and the p_j are grid points such that $0 < p_1 < \cdots < p_J = \alpha$. The weights $w_j(x)$ may be negative but must add up to one. An ICQF estimator is a special case of (2.1) corresponding to uniform weights $w_j(x) = (p_j - p_{j-1})/\alpha$. Both the weights $w_j(x)$ and the number and location of the grid points may depend on the data. To keep things simple, this dependence is momentarily ignored. From now on, we also drop the explicit reference to α and simply write a WICQF estimator as $\tau_J(x)$.

Of particular interest are WICQF estimators based on linear quantile regression estimators of the form $\hat{Q}_j(x) = \hat{\beta}_j^\top x$, where

$$\hat{\beta}_j = \arg\min_{\beta} \sum_{t=1}^{T} \ell_{p_j}(Y_t - \beta^\top X_t)$$

and $\ell_p(u) = u(p - \mathbb{1}\{u < 0\})$ is the asymmetric absolute loss function (see Koenker [20]). If the true conditional quantiles are not linear in x, a linear quantile regression estimator only gives the best linear approximation to the CQF relative to a particular measure of deviation (Angrist et al. [3]). The resulting ICQF estimator takes the particularly simple form $\hat{\tau}_J(x) = \bar{\beta}_J(x)^\top x$, where $\bar{\beta}_J(x) = \sum_{j=1}^{J} w_j(x) \hat{\beta}_j$.

3. Asymptotic properties

Construction of a WICQF estimator requires the choice of J grid points $p_1, \ldots, p_J$ in the interval $(0, \alpha]$, estimators $\hat{Q}_1(x), \ldots, \hat{Q}_J(x)$ of the J conditional quantiles, and a set of weights $w_1(x), \ldots, w_J(x)$. All these choices affect the asymptotic properties of an estimator.

Let $\hat{Q}(p \,|\, x)$ be any function defined on $(0,1) \times \mathcal{X}$, that coincides with $\hat{Q}_j(x)$ when $p = p_j$. As in Angrist et al. [3], we assume that, for all $p \in (0,1)$, the pth estimated conditional quantile $\hat{Q}(p \,|\, x)$ converges in probability as $T \to \infty$, uniformly in x, to a function $Q^*(p \,|\, x)$ that may or may not coincide with the pth population conditional quantile $Q(p \,|\, x)$, that is, the difference $Q^*(p \,|\, X_t) - Q(p \,|\, X_t)$ may be nonzero with positive probability.

As for the weights, we assume that $w_j(x) = W(p_j \,|\, x) - W(p_{j-1} \,|\, x)$ for al j, where $W(p \,|\, x)$ is a continuously differentiable function on $(0, \alpha) \times \mathcal{X}$, with $W(0 \,|\, x) = 0$ and $W(\alpha \,|\, x) = 1$ for all $x \in \mathcal{X}$. We say that weights are uniform if they do not depend on x and are proportional to the distance between two consecutive grid points. This implies that uniform weights are of the form $w_j = (p_j - p_{j-1})/\alpha$, $j = 1, \ldots, J$, with $p_0 = 0$. A special case of uniform weights are the constant weights corresponding to $w_j = J^{-1}$ for all j.

In order to study the asymptotic properties of WICQF estimator, it is useful to decompose the estimation error $\hat{\tau}_J(x) - \tau^0(x)$ as follows

$$\hat{\tau}_J(x) - \tau^0(x) = [\hat{\tau}_J(x) - \tau^*_J(x)] + [\tau^*_J(x) - \tau_J(x)] + [\tau_J(x) - \tau^0_J(x)] + [\tau^0_J(x) - \tau^0(x)], \qquad (3.1)$$

with

$$\tau^*_J(x) = \sum_{j=1}^{J} w_j(x) Q^*_j(x),$$

$$\tau_J(x) = \sum_{j=1}^{J} w_j(x) Q_j(x),$$

and

$$\tau^0_J(x) = \sum_{j=1}^{J} u_j Q_j(x),$$

where $u_1, \ldots, u_J$ denotes the set of uniform weights. Consistency of $\hat{\tau}_J(x)$ for the CES $\tau^0(x)$ requires all four terms on the right-hand side of (3.1) to be negligible. Let us examine each of these terms in turn.

The first component in (3.1),

$$\hat{\tau}_J(x) - \tau^*_J(x) = \sum_{j=1}^{J} w_j(x) [\hat{Q}_j(x) - Q^*_j(x)],$$

reflects the sampling error. The next result implies that, in general, this component is negligible for large T.

Theorem 1. *Let $\hat{Q}(p \mid x)$ be an estimator of $Q^*(p \mid x)$ that is r_T-consistent for all $p \in (0, \alpha]$, where r_T is a divergent sequence, and assume that for every J-tuple $(p_1, \ldots, p_J)$ the random vector $\{r_T[\hat{Q}_j(x) - Q_j^*(x)], j = 1, \ldots, J\}$, converges in distribution to a multivariate Gaussian vector with mean zero and covariance matrix $\boldsymbol{V}(x)$. Then*

$$r_T[\hat{\tau}_J(x) - \tau_J^*(x)] \xrightarrow{d} \mathcal{N}(0, \boldsymbol{w}(x)^\top \boldsymbol{V}(x) \boldsymbol{w}(x)),$$

where $\boldsymbol{w}(x)^\top = (w_1(x), \ldots, w_J(x))$.

The second component in (3.1),

$$\tau_J^*(x) - \tau_J(x) = \sum_{j=1}^{J} w_j(x)[Q_j^*(x) - Q_j(x)], \tag{3.2}$$

reflects the bias that arises when the assumed model for the CQF is incorrectly specified There is clearly a trade-off between simplicity and tractability on the one hand, and bias on the other hand. A linear model for the CQF is particularly simple and tractable, but is likely to be misspecified. One way to overcome this problem is to consider more flexible estimators of $Q_j(x)$. For example, a semiparametric estimator may be obtained by inverting the monotone CDF estimator proposed by Foresi and Peracchi [16] and further discussed in Peracchi [24]. Nonparametric estimators based on inversion of kernel based estimators of the CDF may also be used. Examples include the weighted Nadaraya-Watson estimator of Cai (2002), the double-kernel local linear estimator of Yu and Jones (1998), and the WDKLL estimator of Cai and Wang (2008). In general, semiparametric estimators of the CQF represent a reasonable compromise between flexibility and tractability, while nonparametric methods suffer dramatically the increase of the number of regressors due to the curse-of-dimensionality problem. From now on we assume that (3.2) is either zero or negligible relative to the other terms in (3.1).

The third component in (3.1),

$$\tau_J(x) - \tau_J^0(x) = \sum_{j=1}^{J} [w_j(x) - u_j] Q_j(x),$$

reflects the bias due to the use of nonuniform weights. This component is negligible if the weight function $w_j(x)$ is not far from uniform weighting. The penalization criterion described in Section 4 aims at controlling this term.

Finally, the last component in (3.1),

$$\tau_J^0(x) - \tau^0(x) = \sum_{j=1}^{J} u_j Q_j(x) - \frac{1}{\alpha}\int_0^{\alpha} Q(p \mid x)\, dp,$$

reflects the bias due to approximating the integral $\tau^0(x)$ by the finite sum $\tau_J^0(x)$. This component is nonzero in general, unless J is allowed to grow with the sample size in such a way that the length of all intervals $p_j - p_{j-1}$ goes to zero at a proper rate. The next theorem gives the order of magnitude of the approximation error $\tau_J^0(x) - \tau^0(x)$ for a specific choice of weights, namely

$$u_j = \begin{cases} p_1/\alpha, & \text{if } j = 1, \\ (1-u_1)/(J-1), & \text{if } j = 2, \dots, J. \end{cases} \tag{3.3}$$

In this case

$$\tau_J^0(x) = \frac{1-u_1}{J-1}\sum_{j=2}^{J} Q_j(x) + u_1 Q_1(x).$$

Constant weights are a special case corresponding to the choice $u_1 = J^{-1}$.

Theorem 2. *Assume that the function $Q(p \mid x)$ is continuously differentiable in p for all x with derivative $q(p \mid x) = (\partial/\partial p)\, Q(p \mid x)$. Also assume that $Q(p \mid x) \leq c|x|^{\gamma}[p(1-p)]^{-a}$ for some $0 < a < 1-\varepsilon$, $\gamma > 0$ and $q(p \mid x) \leq c|x|^{\gamma}[p(1-p)]^{-a-1}$ for all $p \in (0, p_1]$, where $|x| = \max_{1\leq k\leq K}|x_k|$. Then, $[\tau_J^0(x) - \tau^0(x)] \leq O(|x|^{\gamma}\,(p_1^{1-a} + J^{a-1}))$.*

When the dimension of the grid increases with the sample size then, under appropriate regularity conditions, the limiting distribution of $r_T[\hat\tau_J(x) - \tau_J(x)]$ is still Gaussian with asymptotic variance

$$\sigma^2 = \int_0^{\alpha}\int_0^{\alpha} w(p \mid x) w(s \mid x) V(p, s \mid x) \mathrm{d}p\, \mathrm{d}s, \tag{3.4}$$

where $V(p, s \mid x)$ is the asymptotic covariance between $\hat Q(p \mid x)$ and $\hat Q(s \mid x)$, with $p, s \in (0, \alpha]$. The necessary regularity conditions depend on the nature of the estimator $\hat Q_j(x)$ and the behavior of the CQF and its derivative near zero. In particular, both $Q(p \mid x)$ and $q(p \mid x)$ should not grow too fast in absolute value as p approaches zero. We also assume that there is no misspecification, namely that for all p, $|Q(p \mid x) - Q^*(p \mid x)|$ is zero, or at least negligible compared to the other terms in (3.1).

The following theorem establishes asymptotic normality of rescaled difference $r_T[\hat\tau_J(x) - \tau^0(x)]$. The proof relies on the existence of a Bahadur representation for the estimator $\hat Q(p \mid x)$. This representation is available for many families of estimators, such as nonparametric M-estimators of the unknown function $m(x) = \arg\min_{\theta} \mathbb{E}[\rho(Y, \theta) \mid X = x]$ (Cheng and De Gooijer [9]). The special case $\rho(Y,\theta) = \ell_p(Y-\theta)$, corresponding to the p-th conditional quantile, is studied by Honda [18]. Although nonstandard convergence rates could also be considered, here we confine ourselves to the standard rate $r_T = \sqrt{T}$.

Theorem 3. *Assume that the following conditions hold.*

(i) The sequence $\{p_1, \ldots, p_J\}$ of grid points is such that, for some $0 \le b \le 1/4$,

$$p_1 = O\left(T^{-1/(1+4b)}\right) \qquad \text{and} \qquad J \ge O(p_1^{-1}). \tag{3.5}$$

(ii) There exist $\gamma > 0$, $c > 0$, $0 < a < 1/2 - 2b + \varepsilon$ and $\alpha_0 \le \alpha$ such that, for all $p \in (0, \alpha_0]$, $|Q(p\,|\,x)| \le c|x|^\gamma [p(1-p)]^{-a}$ and $q(p\,|\,x) \le c|x|^\gamma [p(1-p)]^{-a-1}$.

(iii) The weighting function $W(p\,|\,x)$ is continuously differentiable in p with derivative $w(p\,|\,x)$. Moreover, the weights $w_j(x) = W(p_j\,|\,x) - W(p_{j-1}\,|\,x)$ are such that $\sum_{j=1}^J [w_j(x)/u_j - 1]^2 u_j \le |x|^2 h_T^2$, where the weights u_j satisfy (3.3) and $h_T = o(T^{-1/2-\varepsilon} J^{-a/2})$.

(iv) For all x, the estimator $\hat{Q}(p\,|\,x)$ is $\sqrt{T}$-consistent for $Q(p\,|\,x)$ uniformly on $(0, \alpha]$. In addition, for some sequence of positive numbers, $\delta_T \to 0$,

$$\sqrt{T}[\hat{Q}(p\,|\,x) - Q(p\,|\,x)] = \frac{1}{\sqrt{T}} \sum_{t=1}^T G_t(p\,|\,x) + O_P(\delta_T),$$

where, for all x, $G_t(\cdot\,|\,x)$ is a sequence of zero-mean random functions defined on $(0, \alpha]$ and the process $T^{-1/2} \sum_{t=1}^T G_t(\cdot\,|\,x)$ is asymptotically Gaussian with covariance function $V(p, s\,|\,x)$, with $V(p, p\,|\,x) \le c[p(1-p)]^{-2a-1} |x|^{2\gamma}$.

Then, for all x such that $|x| \le O(\log T)$, $\sqrt{T}(\hat{\tau}_J(x) - \tau^0(x)) \to_d \mathcal{N}(0, \sigma^2)$, with σ^2 given by (3.4).

Condition (i) allows to control the approximation error rate. Condition (ii) implies the main assumptions of Theorem 2. Condition (iii) requires the weights $w_j(x)$ to lie within a small distance from the uniform weights u_j, in order to control for the bias induced by weighting. Condition (iv) guarantees asymptotic normality of the dominating term in (3.1) and, because of the $\sqrt{T}$-consistency requirement, excludes the problem of misspecification. Throughout the paper, the orders of magnitude $o_P(\cdot)$ and $O_P(\cdot)$ are intended in the outer measure sense whenever measurability of the random elements involved is not guaranteed. The existence of an asymptotically Gaussian Bahadur representation for the quantile process has to be checked for specific choices of $\hat{Q}(p\,|\,x)$ and additional moment conditions may be needed for a central limit theorem to be applicable.

As an example, consider the case when $Q(p \mid x) = \beta(p)^\top x$. Assume that the data generating process is stationary and strongly mixing, which corresponds to the (possibly) heteroskedastic model

$$Y_t = \beta(p)^\top X_t + U_t(p),$$

where the error $U_t(p)$ has p-th quantile equal to zero conditional on X_t. Replace Conditions (ii) and (iv) of Theorem 3 by the following:

(ii.1) *Condition (ii) holds for $\gamma = 1$, $0 < a < 1/2 - 2b + \varepsilon$ and $\varepsilon/2 \le b < 1/4$. For all $\alpha_0 > 0$ and $p \in [\alpha_0, 1-\alpha_0]$, the coefficient $\beta(p)$ belongs to a compact set.*

(iv.1) *The mixing coefficient α_t is asymptotically decaying at the rate $\lambda < -2r/(r-2)$ for some $r > 2$. For all x, the conditional density of $U_t(p)$ given $X_t = x$ is absolutely continuous and bounded on a bounded interval. Further, $\mathbb{E}(p - \mathbb{1}\{Y_t - \beta(p)^\top X_t < 0\})X_t = 0$ for all $p \in (0, \alpha]$ and all t, and $\max_{1\le t\le T} \max_{1\le k\le K} \mathbb{E}\,|X_{tk}|^{r'} < \infty$, where $r' = \max\{r, 3+\eta\}$ for some $\eta > 0$.*

(iv.2) *The following matrices are positive definite for all T and all $p, s \in (0, \alpha]$:*

$$\boldsymbol{\Sigma}_T(p,s) = \mathbb{E}\left[T^{-1}\sum_{t=1}^T (p - \mathbb{1}\{Y_t < \beta(p)^\top X_t\})(s - \mathbb{1}\{Y_t < \beta(s)^\top X_t\}) X_t X_t^\top\right],$$

$$\boldsymbol{D}_T = \mathbb{E}\left[T^{-1}\sum_t X_t X_t^\top\right] \quad \textit{and} \quad \boldsymbol{J}_T(p) = \mathbb{E}\left[T^{-1}\sum_t f(\beta(p)^\top X_t \mid X_t) X_t X_t^\top\right].$$

As $T \to \infty$, the limits $\boldsymbol{\Sigma}_T(p,s) \to \boldsymbol{\Sigma}(p,s)$, $\boldsymbol{D}_T \to \boldsymbol{D}$ and $\boldsymbol{J}_T(p) \to \boldsymbol{J}(p)$ exist and are positive definite.

One can show, by adapting the argument in Wang and Zhou [28], that the conclusion of Theorem 3 holds with

$$V(p, s \mid x) = x^\top [\boldsymbol{J}^{-1}(p)\, \boldsymbol{\Sigma}(p,s)\, \boldsymbol{J}^{-1}(s)] x.$$

4. Optimal WICQF estimators

This section presents our proposal for the optimal choice of the weights defining an WICQF estimator. Although the number of grid points is now allowed to depend on the sample size T, we do not make explicit this dependence. For notational simplicity we also omit the dependence of the weights and the covariance matrix $\boldsymbol{V}$ on x. Thus, we denote by $\boldsymbol{w} = (w_1, \ldots, w_J)^\top$ the vector of weights, by $\boldsymbol{Q} = (Q_1(x), \ldots, Q_J(x))^\top$ the vector of population conditional quantiles, and by $\hat{\boldsymbol{Q}} = (\hat{Q}_1(x), \ldots, \hat{Q}_J(x))^\top$ the vector of estimated conditional quantiles. With this notation, we can write the WICQF estimator as $\hat{\tau}_J = \boldsymbol{w}^\top \hat{\boldsymbol{Q}}$ and its discrete population counterpart by $\tau_J = \boldsymbol{w}^\top \boldsymbol{Q}$.

An optimal set of weights may be obtained by minimizing the asymptotic variance of a WICQF estimator subject to two constraints. The first constraint is correct specification of the model for the CQF. This constraint may be approximately satisfied by choosing some flexible conditional quantile estimator. The second constraint, necessary for consistency of the WICQF estimator, requires that, asymptotically, $(\boldsymbol{w}-\boldsymbol{u})^\top \boldsymbol{Q} = 0$ for large enough J, where $\boldsymbol{u} = (u_1, \ldots, u_J)$ is the vector of uniform weights. Imposing this constraint directly would give the trivial solution $\boldsymbol{w}^* = \boldsymbol{u}$. To avoid this difficulty, we choose the optimal weights by minimizing a penalized asymptotic variance criterion, that is, we solve the problem

$$\min_{\boldsymbol{w}:\boldsymbol{w}^\top \boldsymbol{\imath}=1} \boldsymbol{w}^\top \boldsymbol{V} \boldsymbol{w} + c(\boldsymbol{w}-\boldsymbol{u})^\top \boldsymbol{A}(\boldsymbol{w}-\boldsymbol{u}), \tag{4.1}$$

where c is a nonnegative constant, $\boldsymbol{A}$ a symmetric positive definite matrix to be chosen by the investigator, and $\boldsymbol{u}$ is the J-dimensional vector of uniform weights.

The constant c captures the trade-off between asymptotic efficiency and consistency of the WICQF estimator. The smaller is c, the smaller is the importance that we attribute to consistency for the CES. Indeed, for $c = 0$ problem (4.1) coincides with the minimization of the asymptotic variance over all vectors satisfying $\boldsymbol{w}^\top \boldsymbol{\imath} = 1$. If instead $c \to \infty$, then only the penalization matters, so the optimal vector of weights is the vector of uniform weights and we obtain the ICQF estimator.

As for the matrix $\boldsymbol{A}$, three interesting special cases are $\boldsymbol{A} = \boldsymbol{I}$, $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{Q}^\top$, and $\boldsymbol{A} = \operatorname{diag}[u_j^{-1}]$. In the first case, the penalization acts on the Euclidean distance between the vector $\boldsymbol{w}$ and the vector $\boldsymbol{u}$ of uniform weights. The second case corresponds instead to penalizing directly large differences between the WICQF estimator and the unweighted ICQF estimator. The third case implies that $(\boldsymbol{w}-\boldsymbol{u})^\top \boldsymbol{A}(\boldsymbol{w}-\boldsymbol{u}) = \sum_j (w_j/u_j - 1)^2 u_j$, so the penalty term in (4.1) becomes a χ^2-type divergence between the probability mass functions $\boldsymbol{w}$ and $\boldsymbol{u}$.

The next result characterizes the solution to problem (4.1).

Theorem 4. *The vector of asymptotically optimal weights is*

$$\boldsymbol{w}_c^* = \frac{(\boldsymbol{V}+c\boldsymbol{A})^{-1}\boldsymbol{\imath}}{\boldsymbol{\imath}^\top(\boldsymbol{V}+c\boldsymbol{A})^{-1}\boldsymbol{\imath}} + c\left[\boldsymbol{I} - \frac{(\boldsymbol{V}+c\boldsymbol{A})^{-1}\boldsymbol{\imath}\boldsymbol{\imath}^\top}{\boldsymbol{\imath}^\top(\boldsymbol{V}+c\boldsymbol{A})^{-1}\boldsymbol{\imath}}\right](\boldsymbol{V}+c\boldsymbol{A})^{-1}\boldsymbol{A}\boldsymbol{u}. \tag{4.2}$$

When $c = 0$, the vector of asymptotically optimal weights is just $\boldsymbol{w}_0^* = (\boldsymbol{\imath}^\top \boldsymbol{V}^{-1}\boldsymbol{\imath})^{-1}\boldsymbol{V}^{-1}\boldsymbol{\imath}$, irrespective of $\boldsymbol{A}$. This corresponds to the unpenalized minimum asymptotic variance estimator, which is not in general consistent for the CES. Another interesting special case is when $\boldsymbol{u} = J^{-1}\boldsymbol{\imath}$, the vector of constant weights, and $\boldsymbol{A} = \boldsymbol{I}$. In this case, the vector of asymptotically optimal weights is

$$\boldsymbol{w}_c^* = \frac{(\boldsymbol{V}+c\boldsymbol{I})^{-1}\boldsymbol{\imath}}{\boldsymbol{\imath}^\top(\boldsymbol{V}+c\boldsymbol{I})^{-1}\boldsymbol{\imath}}. \tag{4.3}$$

In fact, there is no loss of generality in confining attention to the case of constant weights, because an analogous simplification of the formula of the vector of asymptotically optimal weights is obtained for an arbitrary set of uniform weights $\boldsymbol{u}$ by choosing $\boldsymbol{A} = \operatorname{diag}[u_j^{-1}]$. In this case, $\boldsymbol{Au} = \boldsymbol{\imath}$ and formula (4.2) simplifies to $\boldsymbol{w}_c^* = [\boldsymbol{\imath}^\top(\boldsymbol{V}+c\boldsymbol{A})^{-1}\boldsymbol{\imath}]^{-1}(\boldsymbol{V}+c\boldsymbol{A})^{-1}\boldsymbol{\imath}$. Thus, from now on, we set $\boldsymbol{A} = \boldsymbol{I}$ and $\boldsymbol{u} = J^{-1}\boldsymbol{\imath}$. Notice that the order of magnitude of c in the case of $\boldsymbol{A} = \operatorname{diag}[u_j^{-1}]$ is different: if in fact we chose $\boldsymbol{u} = J^{-1}\boldsymbol{\imath}$, then $\boldsymbol{A} = \operatorname{diag}[u_j^{-1}] = J\boldsymbol{I}$, so the above formula for $\boldsymbol{w}_c^*$ coincides with formula (4.3) for $\boldsymbol{w}_{c'}^*$, with $c' = J\,c$.

Since $\boldsymbol{V}$ and $\boldsymbol{Q}$ are unknown, in practice we replace them with consistent estimates $\hat{\boldsymbol{V}}$ and $\hat{\boldsymbol{Q}}$. The resulting vector of weights is denoted by $\hat{\boldsymbol{w}}_c^*$, so our feasible asymptotically optimal estimator is $\tilde{\tau}_J = \hat{\boldsymbol{w}}_c^{*\top}\hat{\boldsymbol{Q}} = \sum_{j=1}^J \hat{w}_j^*\hat{Q}_j$. The asymptotic error of this estimator is

$$\operatorname{plim}(\tilde{\tau}_J - \tau_J^0) = \operatorname{plim}\left[\frac{(\hat{\boldsymbol{V}}+c\boldsymbol{I})^{-1}\boldsymbol{\imath}}{\boldsymbol{\imath}^\top(\hat{\boldsymbol{V}}+c\boldsymbol{I})^{-1}\boldsymbol{\imath}} - \frac{\boldsymbol{\imath}}{J}\right]^\top \boldsymbol{Q},$$

which can be written

$$\operatorname{plim}(\tilde{\tau}_J - \tau_J^0) = \operatorname{plim}(\hat{\boldsymbol{w}}_c^* - \boldsymbol{w}_c^*)^\top\boldsymbol{Q} + (\boldsymbol{w}_c^* - J^{-1}\boldsymbol{\imath})^\top\boldsymbol{Q}. \tag{4.4}$$

The first term in (4.4) is asymptotically negligible if $\hat{\boldsymbol{V}}$ is consistent for $\boldsymbol{V}$ at a rate faster than $\|\boldsymbol{Q}\|$ or, if Condition (i) and (ii) of Theorem 3 are satisfied, faster than $O(J^a)$. This occurs independently of c, in fact even for $c = 0$. However, large values of c will increase the rate of convergence because, for large enough c, both $\hat{\boldsymbol{w}}_c^*$ and $\boldsymbol{w}_c^*$ approach the uniform weights $\boldsymbol{\imath}/J$ regardless of the distance between $\boldsymbol{V}$ and $\hat{\boldsymbol{V}}$. The second term of (4.4) depends instead crucially on c. The following theorem gives bounds on the rate of convergence of this term.

Theorem 5. *Under Conditions (i) and (ii) of Theorem 3,* $(\boldsymbol{w}_c^* - J^{-1}\boldsymbol{\imath})^\top\boldsymbol{Q} \le O\left(J^a\,|x|^\gamma\,[\underline{\lambda}(\boldsymbol{V}) + c]^{-1}\right)$*, where* $\underline{\lambda}(\boldsymbol{V})$ *is the minimum eigenvalue of the matrix* $\boldsymbol{V}$*.*

It follows immediately from Theorem 5 that condition (iii) of Theorem 3 is satisfied if c grows faster than $J^{-a}T^{-1/2}$.

5. Monte Carlo experiments

To ensure comparability with the results in Cai and Wang [8], the design of our set of Monte Carlo experiments follows closely their design. Thus, we consider two models that correspond to Model I and Model II of Cai and Wang [8].

The first model is an AR(1)-ARCH(1) model of the form $Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \sigma_t U_t$, where $\sigma_t^2 = \gamma_0 + \gamma_1 U_{t-1}^2$ and the U_t are iid $\mathcal{N}(0,1)$ random variables. The vector of model parameters is $\theta = (\alpha_0, \alpha_1, \gamma_0, \gamma_1)$. Since only Y_{t-1} and Y_{t-2} contain information useful to predict Y_t, we can set $X_t = (Y_{t-1}, Y_{t-2})$ and $x = (y_1, y_2)$. The pth conditional quantile of Y_t given $X_t = x$ for this model is of the form $Q(p \,|\, x) = \mu(x) + \Phi^{-1}(p)\,\sigma(x)$, where $\mu(x) = \alpha_0 + \alpha_1 y_1$ and $\sigma(x) = [\gamma_0 + \gamma_1(y_1 - \alpha_0 - \alpha_1 y_2)^2]^{1/2}$. Thus, the CES is of the form

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

where ϕ and Φ respectively denote the density and the distribution function of the standard normal distribution.

The second model is of the form $Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \sigma_t U_t$, as for an AR(2) process, where $\sigma_t^2 = \gamma_0 + \gamma_1 Y_{t-1}^2 + \gamma_2 Y_{t-2}^2$ and the U_t are iid $\mathcal{N}(0,1)$ random variables. We shall refer to this model as the heteroskedastic AR(2) model. The vector of model parameters is now $\theta = (\alpha_0, \alpha_1, \alpha_2, \gamma_0, \gamma_1, \gamma_2)$. The CQF and the CES for this model have the same general form as that for the AR(1)-ARCH(1) model, except that now $\mu(x) = \alpha_0 + \alpha_1 y_1 + \alpha_2 y_2$ and $\sigma(x) = (\gamma_0 + \gamma_1 y_1^2 + \gamma_2 y_2^2)^{1/2}$.

For each model, we consider three alternative parametric specifications for the conditional quantiles. The first is the "naive" specification $Q^*(p \,|\, x) = \beta_0 + \beta_1 y_1$. This specification is clearly incorrect since the conditional quantiles of Y_t depend nonlinearly on both Y_{t-1} and Y_{t-2}. The second is the "linear-with-interaction" specification $Q^*(p \,|\, x) = \beta_0 + \beta_1 y_1 + \beta_2 y_2 + \beta_3 y_1 y_2$. The third is the quadratic specification $Q^*(p \,|\, x) = \beta_0 + \beta_1 y_1 + \beta_2 y_2 + \beta_3 y_1^2 + \beta_4 y_2^2 + \beta_5 y_1 y_2$, which is likely to provide a better approximation to the conditional quantiles of Y_t.

Our Monte Carlo design is as follows:

1. We select the key parameters, namely the level α, the parameters in the two models, the number N and location of the points x at which the CES is evaluated, the sample size T, the number M of Monte Carlo replications, the size J and location $p_1, \ldots, p_J$ of the grid points in the interval $(0, \alpha]$, and the values of the penalty parameter c for WICQF estimators. The level α and the model parameters are as in Cai and Wang [8], namely $\alpha = .05$, $\theta = (.01, .62, .15, .65)$ for the AR(1)-ARCH(1) model and $\theta = (0, .63, -.47, .5, .23, .30)$ for the heteroskedastic AR(2) model. To evaluate the CES, we choose $n = 11$ equally spaced points y_1 and y_2 between -1 and 1 (extremes included), and consider all $N = n^2 = 121$ pairs of the form $x = (y_1, y_2)$. As for the sample size, we set $T = 500$. As for the number of grid points, we consider both $J = 5$ and $J = 10$. Following condition (3.5) in Theorem 3, we choose $p_1 = \alpha T^{-1/(1+4b)}$ and $p_j = p_1 + (\alpha - p_1)(j-1)/(J-1)$, $j = 2, \ldots, J$, with b equal to either .05 or .1. Finally,

after experimenting with the penalty parameter c, we select $c = 0.01$ for the AR(1)-ARCH(1) model and $c = 0.1$ for the heteroskedastic AR(2) model.

2. We recursively generate $T + 100$ observations $X_t = (Y_{t-1}, Y_{t-2})$ from each of the two models, starting from $\sigma_0^2 = U_0 = 0$ for the AR(1)-ARCH(1) model and $Y_0 = Y_{-1} = 0$ for the heteroskedastic AR(2) model. To avoid nonstationarity issues, we keep the last T of them.
3. For each model, we use the data $\{(X_t, Y_t)\}$ to construct the nonparametric WDKLL estimator of Cai and Wang [8]. We also estimate the coefficients $\hat{\beta}_j$ of the three alternative parametric specifications of the conditional quantiles and we construct an estimator $\hat{\mathbf{\Omega}}$ of the asymptotic variance of the estimated quantile regression coefficients by the moving-blocks bootstrap (Fitzenberger [15]), with blocks equal to 5% of the sample size T.
4. For each model, each specification of the conditional quantiles, and each evaluation point x, we form the vector $\hat{\boldsymbol{Q}}(x) = (\hat{Q}_1(x), \ldots, \hat{Q}_J(x))$ and use the matrix $\hat{\mathbf{\Omega}}$ to construct the matrix $\hat{\boldsymbol{V}}(x) = (\boldsymbol{I} \otimes x)^\top \hat{\mathbf{\Omega}} (\boldsymbol{I} \otimes x)$. We also construct the ICQF estimator $\hat{\tau}_J^0(x) = \boldsymbol{u}^\top \hat{\boldsymbol{Q}}(x)$, where $\boldsymbol{u}$ is the vector of uniform weights (3.3).
5. For each evaluation point x and each value of the penalty parameter c, we compute the asymptotically optimal weights $\hat{\boldsymbol{w}}_c^*(x)$ and the associated asymptotically optimal WICQF estimator $\tilde{\tau}_J(x) = \hat{\boldsymbol{w}}_c^*(x)^\top \hat{\boldsymbol{Q}}(x)$.
6. We repeat steps 2.–5. $M = 500$ times (as in Cai and Wang [8]), and save the results.

Figure 1 shows the α-level CES corresponding to our two models for $\alpha = .05$ and different pairs $x = (y_1, y_2)$ in the interval $[-1, 1] \times [-1, 1]$. Notice that the probability that the pair (Y_{t-1}, Y_{t-2}) belongs to this interval is about 85 percent for the AR(1)-ARCH(1) model and is about 65 percent for the heteroskedastic AR(2) model.

To avoid cluttering the paper with numerical tables, we present graphical summaries of our Monte Carlo experiments. Detailed tabulations of the results are available from the authors upon request. Figure 2 compares the nonparametric WDKLL estimator with our ICQF and WICQF estimators based on the "naive" specification of conditional quantiles. The left and right columns correspond, respectively, to the AR(1)-ARCH(1) model and the heteroskedastic AR(2) model, while the rows correspond to alternative choices of the parameters b and J, namely $(b = .1, J = 10)$, $(b = .05, J = 10)$ and $(b = .1, J = 5)$. The various estimators are compared based on their Monte Carlo mean absolute deviation error (MADE). Using the root mean squared error (RMSE) as performance criterion produces very similar results. In the case of the AR(1)-ARCH(1) model, the WDKLL is always the worst estimator, except when y_1 is close to zero. The ICQF estimator is the best when y_1 is large in absolute value, the WICQF estimator with $c = .01$ is the best when y_1 is close to zero, while the WICQF estimator with $c = 0$ is the best when y_1 is close to .50 in absolute value. Results are reversed in the case of the heteroskedastic AR(2) model, as the WDKLL estimator now dominates the ICQF estimator, does better than the WICQF with $c = .10$ over most of y_1 range, and does slightly better than the WICQF estimator with $c = 0$ for values of y_1 close to zero. Notice that, for large values of y_1, the performance of the WKDLL estimator tends to deteriorate while that of the ICQF and WICQF estimators tends to improve. Qualitatively, results do not change when we increase b from .05 to .10, or we reduce J from 10 to 5. In general, increasing b reduces the RMSE and the MADE, while reducing J increases them.

Figures 3–6 present the results for the the ICQF and WICQF estimators with "linear-with-interaction" and quadratic specification of the conditional quantiles. The format of each figure is the same. The columns plot the Monte Carlo MADE of three alternative estimators corresponding to each model. The alternative estimators are the WICQF estimator with $c = 0$ (first column), the WICQF estimator with $c = .01$ for the AR(1)-ARCH(1) model and $c = .10$ for the heteroskedastic AR(2) model (second column), and the ICQF estimator, corresponding to the WICQF estimator with $c = \infty$ (third column). The rows correspond instead to three alternative choices of the parameters b and J, namely $(b = .1, J = 10)$, $(b = .05, J = 10)$ and $(b = .1, J = 5)$.

Figure 3 shows the results for the AR(1)-ARCH(1) model when the conditional quantiles are specified as "linear-with-interaction". The ICQF estimator is the best when y_1 is large in absolute value, unless y_2 is also large in absolute value, in which case the WICQF estimator with $c = 0$ tends to be the best. The latter estimator also tends to be the best when y_1 is close to zero but y_2 is large in absolute value. The WICQF estimator with $c = .01$ is instead the best when either y_1 and y_2 are both large in absolute value, or y_2 is close to zero.

Figure 4 shows the results for the AR(1)-ARCH(1) model when the conditional quantiles are specified as quadratic. In this case, the performance of all estimators improves substantially. The IQCF estimator tends to be the best in general, unless y_1 and y_2 are both close to zero or y_2 is large in absolute value. The WICQF estimator with $c = 0$ is almost always the worst estimator.

Figure 5 shows the results for the heteroskedastic AR(2) model when the conditional quantiles are specified as "linear-with-interaction". In this case, the ICQF estimator is always dominated by the WICQF estimator with $c = .01$, which is itself always dominated by the WICQF estimator with $c = 0$.

Finally, Figure 6 shows the results for the heteroskedastic AR(2) model when the conditional quantiles are specified as quadratic. Again, the performance of all estimators improves substantially. The ICQF estimator is always the worst, although it does better than the WICQF estimator with $c = 0$ when y_1 and y_2 are both large in absolute value. The WICQF estimator with $c = 0$ is the best when either y_1 is close to zero, unless y_2 is large in absolute value, or y_1 is large in absolute value but y_2 is close to zero. The WICQF estimator with $c = .01$ is the best when either y_1 or y_2 are large in absolute value.

6. Empirical application

This section considers an application based on daily stock market data. Specifically, we use daily data on the returns on the Euro Stoxx 50, Europe's leading blue-chip index for the Eurozone, to construct one-day ahead forecasts of the CES at level $\alpha = .05$. Raw daily data range from December 30, 1994, to October 29, 2010.

The outcome variable Y_t is the daily return, defined as the logarithmic difference in the stock index between day $t+1$ and day t, computed excluding weekends and holidays, while the covariates in X_t consist of lagged daily returns. The length of the raw series allow us to include periods of exceptionally high and low returns, such as those that characterize the market swings of the last two years. Daily returns are plotted in Figure 7. In total, we have 3,225 observations. The mean and median of the returns are equal to $0.27\,10^{-4}$ and $4.6\,10^{-4}$ respectively, their standard deviation is equal to 1.4 percent, and their first and last percentiles are equal to -4.5 and 3.8 percent respectively.

We compare the results obtained for the nonparametric WDKLL estimator, the ICQF estimator and two WICQF estimators. The WKDLL estimator is computed using a grid of 251 equally-spaced points in the interval [-.125,.125], whereas for the ICQF and WICQF estimators we set $b = .10$ and $J = 10$. For the ICQF and WICQF estimators, we use the same specifications of the conditional quantiles that we used in our Monte Carlo experiments, namely "naive" (which includes only the first lag of the returns), "linear-with-interaction" (which adds as regressors the second lag of the returns and the cross product of the first and second lag) and quadratic (which includes as regressors the first two lags of the returns, their cross product and their squares). For the WICQF estimators, we show results using increasing values of the penalty parameter, namely $c = 0$, $c = 10^{-5}$, $c = 10^{-4}$, $c = 10^{-3}$ and $c = 10^{-2}$.

The conditional quantiles and the CES are estimated using rolling samples of size $T_0 = 500$, the same sample size as in our Monte Carlo experiments. For each $t = T_0, \ldots, T-1$, the estimated CES evaluated at the current value of the covariates is then used to form one-day-ahead forecasts of the shortfall. The forecast ability of the various estimators is compared by looking at the distribution of their forecast error for all quantile violation events. Following McNeil and Frey [23], the forecast error is defined as the difference between the return observed next day and the forecast of the shortfall, while a quantile violation event is a case when the realized return is lower than the corresponding predicted α-level quantile. McNeil and Frey [23] formally test for unbiasedness of forecasts by looking at how large is the average forecast error. We do not use their test, but simply present a few summaries of the distribution of the forecast error.

Table 1 presents, for each estimator, the number of quantile violation events along with the mean, the median (Med), the standard deviation (SD), the root mean square error (RMSE), the mean absolute deviation error (MADE), and the first (Q_1) and last (Q_{99}) percentiles of the forecast error computed for these quantile violation events. The WKDLL estimator produces very volatile estimates, with many outliers, which result in large outliers in the forecast error. For the ICQF and WICQF estimators, the best results are obtained under the quadratic specification of the conditional quantiles. The corresponding forecast error is always the closest to zero, has the smallest variability and, consequently, the smallest RMSE and MADE. Comparing WICQF estimators with different values of the penalty parameter c, the mean and the median forecast error decrease (in absolute value) as c increases, while the standard deviation of the forecast error increases, as one would expect based on our previous results. As a consequence, WICQF estimators with a finite positive value of c enjoy a slight advantage in terms of RMSE over the ICQF estimator and the WICQF estimator with $c = 0$.

7. Conclusions

In this paper we propose the general class of WICQF estimators of the α-level CES. These estimators are obtained by integrating the estimated CQF over a possibly data-dependent interval using different weights for different quantiles, thus attaining higher asymptotic efficiency relative to the case when no weighs are used while, at the same time, controlling for bias. We also provide asymptotic results that open the way to inference.

Our Monte Carlo evidence shows that it does matter how the conditional quantiles are specified. Even in the "naive" case, however, our WICQF estimators compare well with the nonparametric WDKLL estimator of Cai and Wang [8] despite their severe bias due to misspecification of the CQF. Substantial improvements in the properties of WICQF estimators are obtained by using more flexible parametric specifications of the conditional quantiles. This is because adding more flexibility reduces one of the terms entering their asymptotic bias, namely $\tau_J^*(x) - \tau_J(x)$ in the decomposition (3.1).

Our empirical application to daily stock returns confirms the good properties of our WICQF estimators in practice. The U-shaped pattern of their RMSE of forecast reflects the different behavior of its two components, one decreasing (the squared bias) and the other increasing (the variance) with the value of the penalization constant.

References

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

[2] Acerbi, C., 2002. Spectral measures of risk: A coherent representation of subjective risk aversion. *Journal of Banking and Finance* 26, 1505–1518.

[3] Angrist, J., Chernozhukov, V., Fernández-Val, I., 2006. Quantile regression under misspecification, with an application to the U.S. wage structure. *Econometrica* 74, 539–563.

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

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

[6] Bertsimas, D., Lauprete, G.J., Samarov, A., 2004. Shortfall as a risk measure: Properties, optimization and applications. *Journal of Economic Dynamics & Control* 28, 1353–1381.

[7] Cai, Z., 2002. Regression quantiles for time series. *Econometric Theory* 18, 169–192.

[8] Cai Z., Wang X., 2008. Nonparametric estimation of conditional VaR and expected shortfall. *Journal of Econometrics* 2008, 120–130.

[9] Cheng, Y., De Gooijer, J.G., 2009. Bahadur representation for the nonparametric M-estimator under alpha-mixing dependence. *Statistics* 43, 443–462.

[10] Chernozhukov, V., 2005. Extremal quantile regression. *Annals of Statistics*, 33, 806–839.

[11] Chernozhukov, V., Fernandez-Val, I., Galichon, A., 2010. Quantile and probability curves without crossing. *Econometrica*, 1093–1125.

[12] Csörgö, S., Haeusler, E., Mason, D.M., 1991. The asymptotic distribution of extreme sums. *Annals of Probability* 19, 783–811.

[13] Delbaen, F., 2000. Coherent risk measures on general probability spaces. ETH Zürich, mimeo.

[14] Fan, J., Gijbels, I., 1996. *Local Polynomial Modeling and Its Applications*. London: Chapman and Hall.

[15] Fitzenberger, B., 1998. The moving blocks bootstrap and robust inference for linear least squares and quantile regressions. *Journal of Econometrics* 82, 235–287.

[16] Foresi, S., Peracchi, F., 1995. The condition distribution of excess returns: An empirical analysis. *Journal of The American Statistical Association* 90, 451–466.

[17] Hall, P., Wolff, R.C.L., Yao, Q., 1999. Methods for estimating a conditional distribution function, *Journal of the American Statistical Association* 94, 154–163.

[18] Honda, T. 2000. Nonparametric estimation of a conditional quantile for α-mixing processes, *Annals of the Institute of Statistical Mathematics*, 52, 459–470.

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

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

[21] Mason, D.M., Shorack, G.R., 1992. Necessary and sufficient conditions for asymptotic normality of L-statistics. *Annals of Probability* 20, 1779–1804.

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

[23] McNeil, A.J., Frey, R., 2000. Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. *Journal of Empirical Finance* 7, 271–300.

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

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

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

[27] van der Vaart, A.W., Wellner, J.A., 1996. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer: Berlin.

[28] Wang, H.J., Zou, X.-H., 2010. Estimaton of the retransformed conditional mean in health care cost studies. *Biometrika* 97, 147–158.

[29] Yaari, M.E., 1987. The dual theory of choice under risk. *Econometrica* 55, 95–115.

[30] Yu, K., Jones, M.C., 1998. Local linear quantile regression. *Journal of the American Statistical Association* 93, 228–237.

Table 1: Summary statistics of the empirical distribution of the one-step-ahead forecast error (expressed in percentage points) for quantile violation events and alternative estimators of the CES over 2,724 rolling windows. The level α is equal to .05. For WICQF estimators, conditional quantiles are specified as (i) "naive", (ii) "linear-with-interaction", and (iii) quadratic.

Estimator	Obs.	Mean	Med	SD	RMSE	MADE	Q_1	Q_{99}
WKDLL	155	1.96	-0.25	6.21	6.51	3.79	-5.15	27.47
(i)								
WICQF ($c = 0$)	187	-0.97	-0.64	1.16	1.51	1.07	-5.01	1.03
WICQF ($c = 10^{-5}$)	187	-0.57	-0.23	1.11	1.25	0.82	-4.54	1.09
WICQF ($c = 10^{-4}$)	187	-0.44	-0.15	1.09	1.18	0.79	-4.27	1.19
WICQF ($c = 10^{-3}$)	187	-0.41	-0.14	1.10	1.17	0.79	-4.23	1.30
WICQF ($c = 10^{-2}$)	187	-0.41	-0.14	1.10	1.17	0.79	-4.22	1.34
ICQF	187	-0.41	-0.14	1.10	1.17	0.79	-4.22	1.35
(ii)								
WICQF ($c = 0$)	193	-0.99	-0.68	1.09	1.47	1.06	-4.47	0.79
WICQF ($c = 10^{-5}$)	193	-0.68	-0.38	1.11	1.30	0.86	-4.73	0.78
WICQF ($c = 10^{-4}$)	193	-0.54	-0.22	1.11	1.23	0.81	-4.73	1.02
WICQF ($c = 10^{-3}$)	193	-0.48	-0.20	1.12	1.22	0.81	-4.74	1.17
WICQF ($c = 10^{-2}$)	193	-0.47	-0.17	1.13	1.22	0.81	-4.74	1.30
ICQF	193	-0.47	-0.17	1.13	1.22	0.81	-4.74	1.33
(iii)								
WICQF ($c = 0$)	189	-0.80	-0.64	0.94	1.24	0.91	-3.30	1.33
WICQF ($c = 10^{-5}$)	189	-0.51	-0.40	0.97	1.10	0.77	-3.17	1.11
WICQF ($c = 10^{-4}$)	189	-0.40	-0.27	1.00	1.11	0.75	-3.14	1.93
WICQF ($c = 10^{-3}$)	189	-0.36	-0.26	1.04	1.10	0.75	-3.11	2.47
WICQF ($c = 10^{-2}$)	189	-0.35	-0.26	1.06	1.11	0.75	-3.12	2.63
ICQF	189	-0.35	-0.26	1.06	1.11	0.75	-3.12	2.65

Figure 1: CES at level α for the AR(1)-ARCH(1) model (left panel) and the heteroskedastic AR(2) model (right panel) for $\alpha = .05$ and different values of (y_1, y_2).

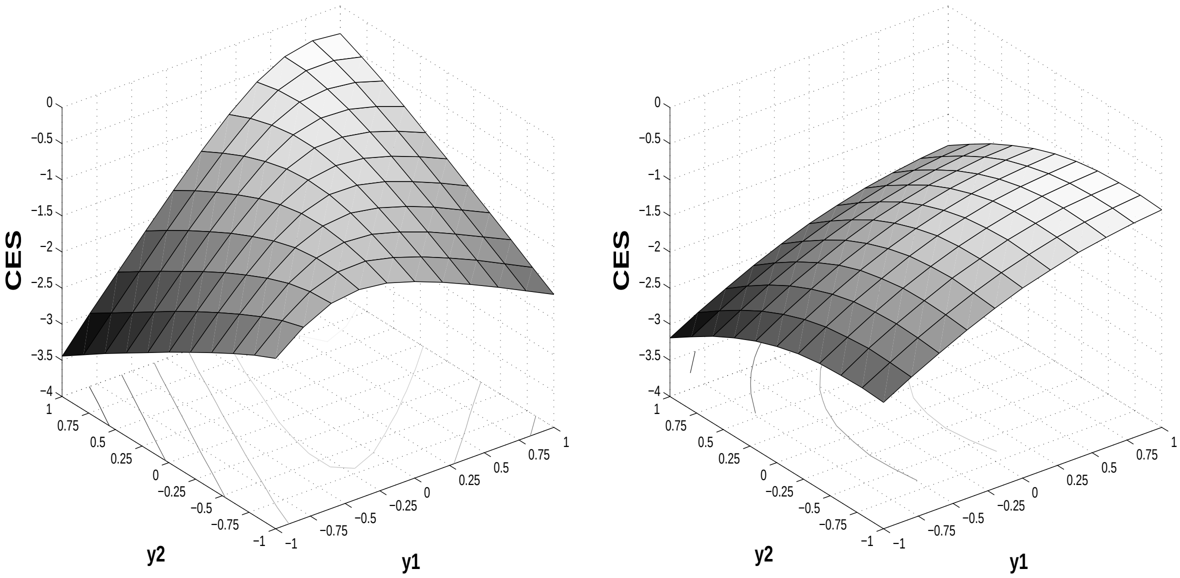


Figure 2: Monte Carlo MADE of alternative estimators of the α-level CES for the AR(1)-ARCH(1) model (left panel) and the heteroskedastic AR(2) model (right panel) with $\alpha = .05$, different values of y_1 and "naive" specification of the conditional quantiles. The rows of the table correspond, respectively, to $(b = .1, J = 10)$, $(b = .05, J = 10)$ and $(b = .1, J = 5)$.

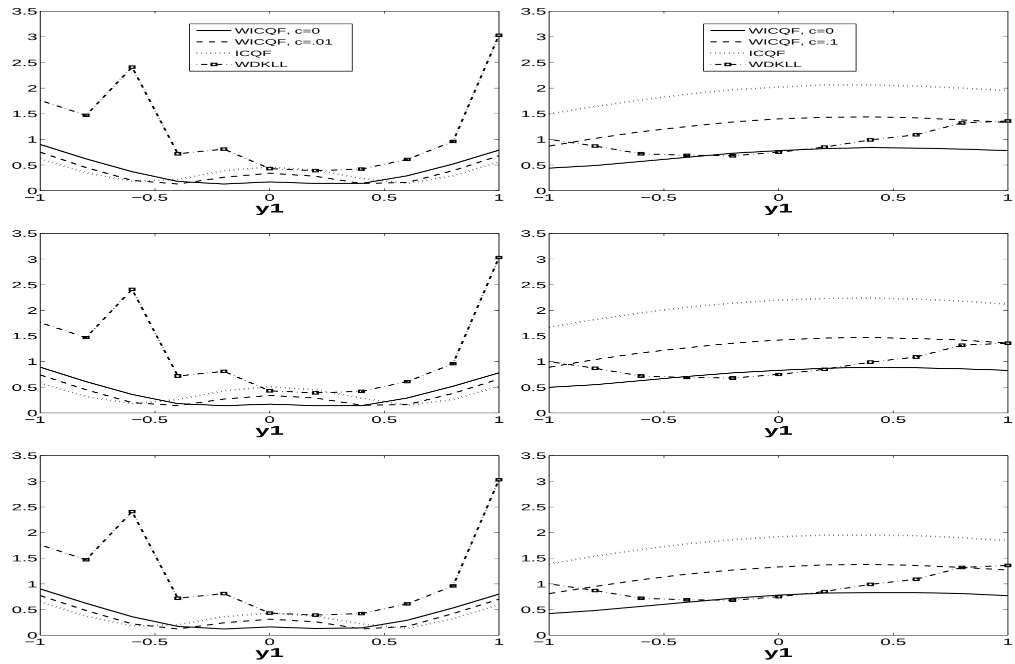


Figure 3: Monte Carlo MADE of alternative estimators of the α-level CES for the AR(1)-ARCH(1) model with $\alpha = .05$ and "linear-with-interaction" specification of the conditional quantiles. The rows of the table correspond, respectively, to $(b = .1, J = 10)$, $(b = .05, J = 10)$ and $(b = .1, J = 5)$.

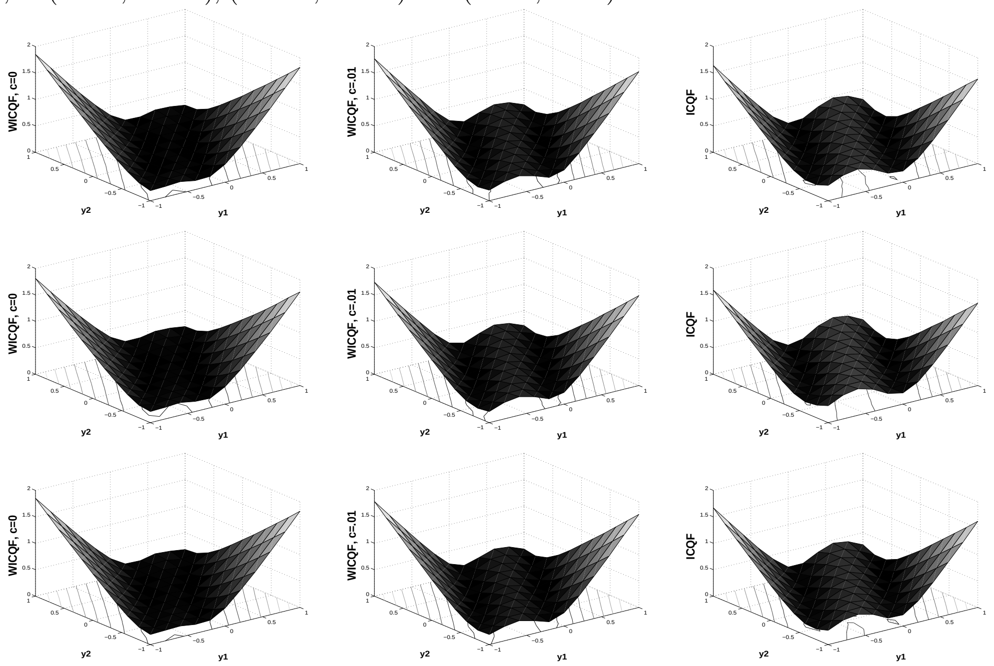


Figure 4: Monte Carlo MADE of alternative estimators of the α-level CES for the AR(1)-ARCH(1) model with $\alpha = .05$ and quadratic specification of the conditional quantiles. The rows of the table correspond, respectively, to $(b = .1, J = 10)$, $(b = .05, J = 10)$ and $(b = .1, J = 5)$.

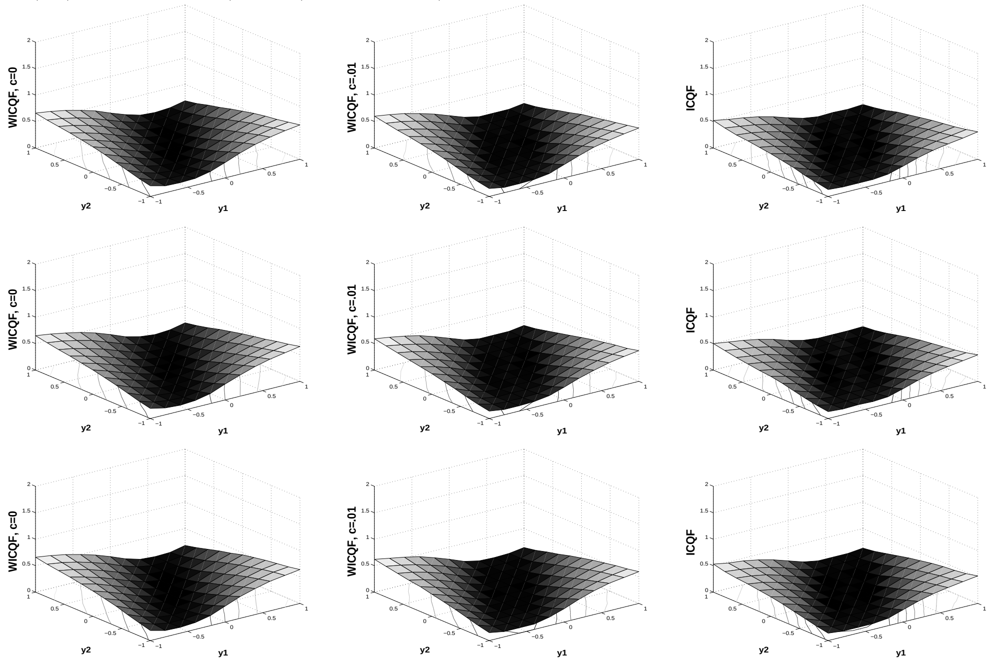


Figure 5: Monte Carlo MADE of alternative estimators of the α-level CES for the heteroskeastic AR(2) model with $\alpha = .05$ and "linear-with-interaction" specification of the conditional quantiles. The rows of the table correspond, respectively, to $(b = .1, J = 10)$, $(b = .05, J = 10)$ and $(b = .1, J = 5)$.

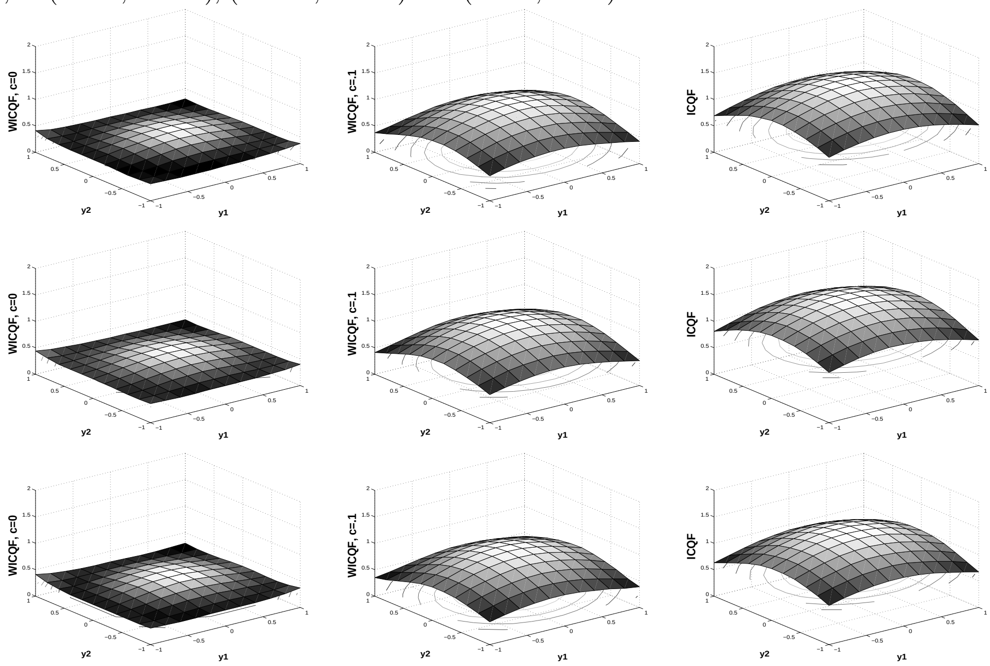


Figure 6: Monte Carlo MADE of alternative estimators of the α-level CES for the heteroskedastic AR(2) model with $\alpha = .05$ and quadratic specification of the conditional quantiles. The rows of the table correspond, respectively, to $(b = .1, J = 10)$, $(b = .05, J = 10)$ and $(b = .1, J = 5)$.

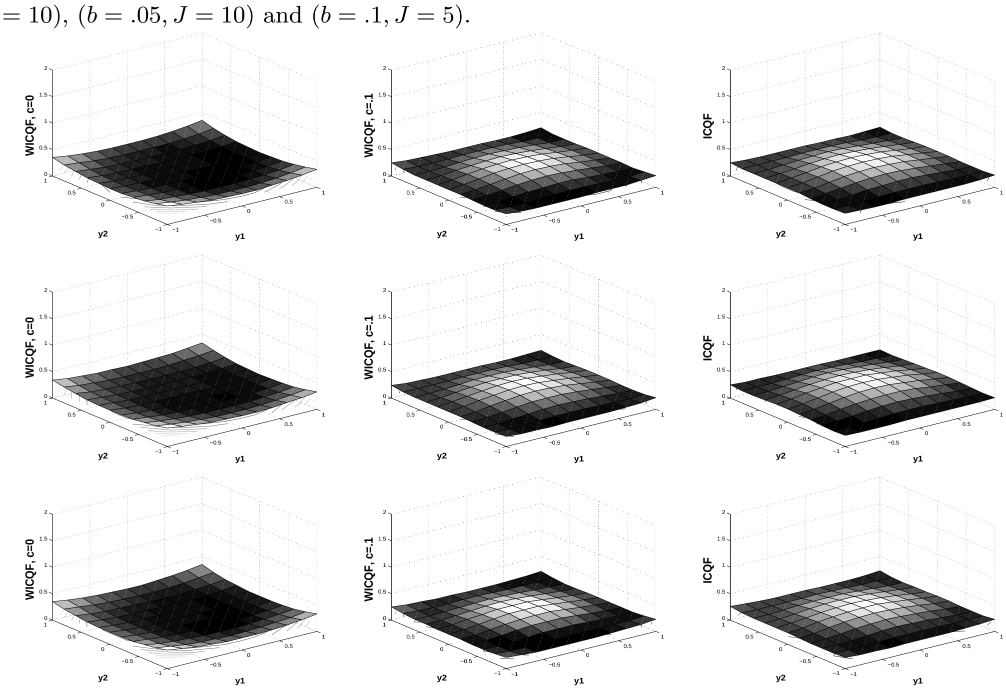
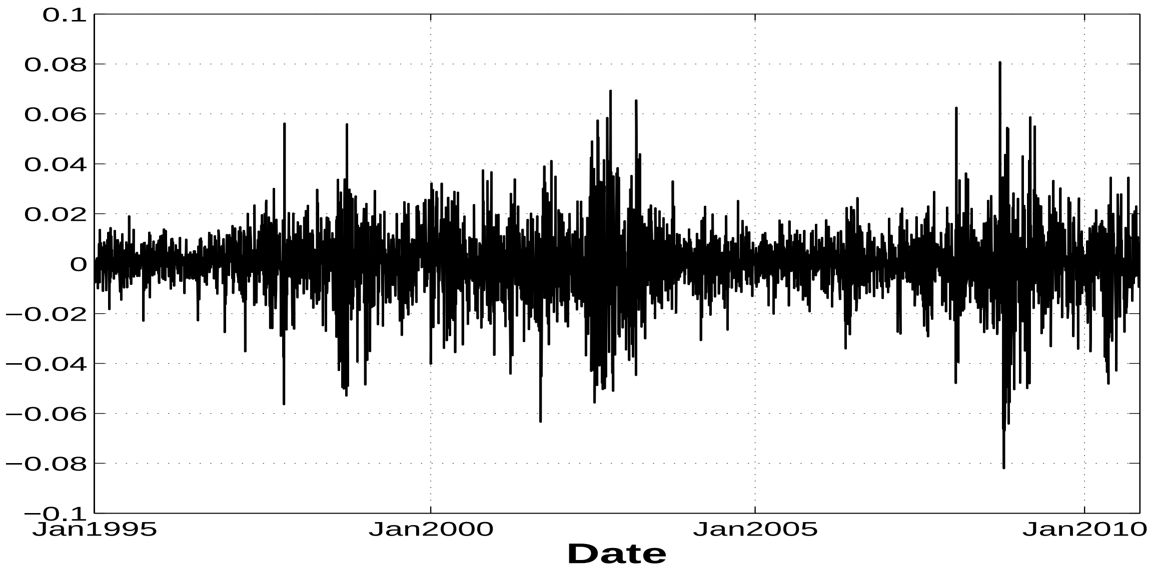


Figure 7: Daily returns on the Euro Stoxx 50 index between January 3, 1995, and October 29, 2010.



Appendix A.

Proof of Theorem 2

From $u_j = (p_j - p_{j-1})/\alpha$, we have

$$\tau_J^0(x) = \sum_{j=1}^{J} u_j Q_j(x) = \frac{1}{\alpha} \sum_{j=1}^{J} \int_{p_{j-1}}^{p_j} Q_j(x)\,\mathrm{d}p.$$

Thus, by a Taylor expansion of $Q(p \,|\, x)$ around $Q_j(x)$,

$$\begin{aligned}
|\tau_J^0 - \tau^0| &= \alpha^{-1} \left| \sum_{j=1}^{J} \int_{p_{j-1}}^{p_j} [Q_j(x) - Q(p \,|\, x)]\,\mathrm{d}p \right| \\
&= \alpha^{-1} \left| \sum_{j=1}^{J} \int_{p_{j-1}}^{p_j} [Q_j(x) - Q_j(x) + (p_j - p) q_j(x) + R(p \,|\, x)]\,\mathrm{d}p \right| \\
&\leq \alpha^{-1} \left| \sum_{j=1}^{J} \frac{(p_j - p_{j-1})^2}{2} q_j(x)(1 + O(1)) \right|,
\end{aligned}$$

where $R(p \,|\, x)$ is a remainder term such that $R(p \,|\, x) \leq O((p_j - p_{j-1}) q_j(p))$ for all $p \in (p_{j-1}, p_j]$. Because $|q_j(x)| \leq c|x|^\gamma (p_j(1-p_j))^{-a-1}$, we have

$$\begin{aligned}
|\tau_J^0 - \tau^0| &\leq c(2\alpha)^{-1} \sum_{j=1}^{J} (p_j - p_{j-1})^2 |x|^\gamma \,[p_j(1-p_j)]^{-a-1}(1 + O(1)) \\
&\leq \frac{c|x|^\gamma\,\alpha}{2(1-\alpha)^{a+1}} \left[u_1^2 (\alpha u_1)^{-a-1} + \sum_{j=2}^{J} u_j^2 \left(u_1 + \frac{j-1}{J-1}(1-u_1) \right)^{-a-1} \right] (1 + O(1)) \\
&\leq c_1 |x|^\gamma \left[u_1^{1-a} + \left(\frac{1-u_1}{J-1} \right)^2 \sum_{j=2}^{J} \left(u_1 + \frac{j-1}{J-1}(1-u_1) \right)^{-a-1} \right].
\end{aligned}$$

From convexity of the function $(x)^{-a-1}$ on $(0, \infty)$,

$$\begin{aligned}
|\tau_J^0 - \tau^0| &\leq c_1 |x|^\gamma \left[u_1^{1-a} + \left(\frac{1-u_1}{J-1} \right)^2 \sum_{j=2}^{J} \left(u_1 + (1-u_1) \left(\frac{j-1}{J-1} \right)^{-a-1} \right) \right] \\
&\leq O(|x|^\gamma\, u_1^{1-a}) + O(|x|^\gamma\, u_1 J^{-1}) + O\left(|x|^\gamma\, J^{a-1} \sum_{j=1}^{J-1} j^{-a-1} \right),
\end{aligned}$$

where the second term on the right-hand side is dominated by the others. The result then follows from the fact that $\sum_{j=1}^{\infty} j^{-a-1}$ is a convergent series for all $a > 1$.

Proof of Theorem 3

Write

$$\sqrt{T}[\hat{\tau}_J(x) - \tau^0(x)] = \int_{p_1}^{\alpha} \sqrt{T}[\hat{Q}(p\,|\,x) - Q(p\,|\,x)]w(p\,|\,x)\mathrm{d}p + A_1 + A_2 + A_3, \tag{A.1}$$

where

$$A_1 = \sqrt{T}\left[\sum_{j=1}^{J} \hat{Q}_j(x)w_j(x) - \int_{p_1}^{\alpha} \hat{Q}(p\,|\,x)w(p\,|\,x)\mathrm{d}p - \sum_{j=1}^{J} Q_j(x)w_j(x) + \int_{p_1}^{\alpha} Q(p\,|\,x)w(p\,|\,x)\mathrm{d}p\right],$$

$$A_2 = \sqrt{T}\sum_{j=1}^{J} Q_j(x)[w_j(x) - u_j],$$

$$A_3 = \sqrt{T}\left(\sum_{j=1}^{J} Q_j(x)u_j - \int_0^{\alpha} Q(p\,|\,x)\mathrm{d}p\right).$$

Because of Condition (iv), the first term of (A.1), namely

$$\sqrt{T}\int_{p_1}^{\alpha}[\hat{Q}(p\,|\,x) - Q(p\,|\,x)]w(p\,|\,x)\mathrm{d}p = T^{-1/2}\sum_{t=1}^{T}\int_{p_1}^{\alpha} G_t(p\,|\,x)w(p\,|\,x)\mathrm{d}p + O_P(\delta_T),$$

is asymptotically normal with mean zero and variance σ^2. Thus, it is then enough to prove that $A_i = o_P(1)$ for $i = 1, 2, 3$. Conditions (i)–(ii) and Theorem 2 imply that

$$A_3 = O(|x|^{\gamma}\, T^{1/2} p_1^{1-a}) = O((\log(T))^{\gamma}\, T^{(a+2b-1/2)/(1+4b)}) = o(1).$$

Because of Condition (ii), we also have

$$\begin{aligned}
|A_2| &\le \sqrt{T}\sum_{j=1}^{J}(\log(T))^{\gamma}[p_j(1-p_j)]^{-a}\left|\frac{w_j(x)}{u_j} - 1\right| u_j \\
&\le \sqrt{T}(\log(T))^{\gamma}\left[\sum_{j=1}^{J}\left(\frac{w_j(x)}{u_j} - 1\right)^2 u_j\right]^{1/2}\left[\sum_{j=1}^{J}(p_j(1-p_j))^{-2a}u_j\right]^{1/2} \\
&\le \sqrt{T}(\log(T))^{\gamma+1}\, h_T(1-\alpha)^{-2}\left(\sum_{j=1}^{J} u_j p_j^{-2}\right)^{a/2},
\end{aligned}$$

where we applied Hölder inequality and the trivial bound $(1 - p_j) \ge (1 - \alpha)$ in the last line. Recalling that $p_1 = \alpha u_1$ and $p_j = \alpha[u_1 + (j-1)(1-u_1)/(J-1)]$ for $j \ge 2$, we have

$$\begin{aligned}
|A_2| &\le c_1\sqrt{T}(\log(T))^{\gamma+1}\, h_T\left[\sum_{j=1}^{J} u_j\left(u_1 + \frac{j-1}{J-1}(1-u_1)\right)^{-2}\right]^{a/2} \\
&\le c_1\sqrt{T}(\log(T))^{\gamma+1}\, h_T\left\{u_1^{-1} + \frac{1-u_1}{J-1}\sum_{j=2}^{J}\left[u_1 + (1-u_1)\left(\frac{J-1}{j-1}\right)^2\right]\right\}^{a/2},
\end{aligned}$$

where we exploited the fact that the function $f(x) = x^{-2}$ is convex and where c_1 is a constant independent on T. Further,

$$\sum_{j=2}^{J}\left[u_1 + (1-u_1)\left(\frac{J-1}{j-1}\right)^2\right] = (J-1)u_1 + (1-u_1)(J-1)^2\sum_{j=1}^{J-1} j^{-2} = O(J^2),$$

because $\sum_{j=1}^{\infty} j^{-2}$ is a convergent series. Therefore, by Condition (i) and (iii),

$$|A_2| \le O((\log(T))^{\gamma+1}T^{1/2}h_T J^{a/2}) = o(1).$$

The term A_1 can be divided into two parts

$$\begin{aligned} A_1 = & \sum_{j=2}^{J}\int_{p_{j-1}}^{p_j}\sqrt{T}\left[\left(\hat{Q}_j(x) - Q_j(x)\right) - \left(\hat{Q}(p\,|\,x) - Q(p\,|\,x)\right)\right]w(p\,|\,x)\mathrm{d}p + \\ & + \sqrt{T}(\hat{Q}_1(x) - Q_1(x))w_1(x), \end{aligned}$$

which we denote by A_{11} and A_{12} respectively. Clearly $A_{11} = o(A_3) = o(1)$. Using Condition (iv),

$$\begin{aligned} |A_{12}| &= \sqrt{T}(\hat{Q}_1(x) - Q_1(x))w_1(x) \\ &= T^{-1/2}\sum_t G_t(p_1\,|\,x)w_1(x)(1 + o_P(1)). \end{aligned}$$

By Chebyshev's inequality, we have that for sufficiently large T and all $\delta > 0$

$$\begin{aligned} \Pr\{|A_{12}| > \delta\} &\le \frac{\mathrm{var}\left(T^{-1/2}\sum_t G_t(p_1\,|\,x)w_1(x)\right)}{\delta^2} \\ &\approx \frac{V(p_1, p_1\,|\,x)w_1^2(x)}{\delta^2} \\ &\le (\delta)^{-2}|x|^{2\gamma}(1-p_1)^{-1-2a}p_1^{-1-2a}(u_1 + w_1(x) - u_1)^2 \\ &\le \frac{(1-\alpha)^{-1-2a}}{\delta^2}|x|^{2\gamma}p_1^{-1-2a}p_1^2\frac{2}{\alpha}(1 + h_T^2|x|^2) \\ &\le 2\alpha^{-1}(1-\alpha)^{-1-2a}\delta^{-2}(\log T)^{2(\gamma+1)}(p_1^{1-2a} + h_T^2). \end{aligned}$$

Then, for $\delta^2 = p_1^{1-a}T^{1/2} = o(1)$,

$$\Pr\{|A_{12}| > \delta\} \le O\left((\log T)^{2(\gamma+1)}(T^{-1/2}p_1^{-a} + T^{-1/2}p_1^{a-1}h_T^2)\right) = o(1)$$

because of Conditions (i)–(iii) on the rates of p_1 and h_T.

Proof of Theorem 4

From the first-order conditions we get

$$\begin{pmatrix} \boldsymbol{V} + c\boldsymbol{A} & -\boldsymbol{\imath} \\ -\boldsymbol{\imath} & 0 \end{pmatrix}\begin{pmatrix} \boldsymbol{w} \\ \lambda \end{pmatrix} = \begin{pmatrix} c\boldsymbol{A}\boldsymbol{u} \\ 1 \end{pmatrix},$$

where λ is the Lagrange multiplier associate with the constraint $\boldsymbol{w}^\top\boldsymbol{\imath} = 1$. The result then follows form the inversion formulae for block matrices.

Proof of Theorem 5

First notice that $(\boldsymbol{w}_c^* - \boldsymbol{u})^\top \boldsymbol{Q} \leq \|\boldsymbol{w}_c^* - \boldsymbol{u}\| \|\boldsymbol{Q}\|$. Because $\|\boldsymbol{Q}\|$ is bounded under Condition (ii) of Theorem 3, to study the order of magnitude of the above term it is enough to focus on the distance $\|\boldsymbol{w}_c^* - \boldsymbol{u}\|$. We have

$$\begin{aligned}
\|\boldsymbol{w}_c^* - \boldsymbol{u}\| &= \left\| \frac{(\boldsymbol{V} + c\boldsymbol{I})^{-1}\boldsymbol{\imath}}{\boldsymbol{\imath}^\top (\boldsymbol{V} + c\boldsymbol{I})^{-1}\boldsymbol{\imath}} - \frac{\boldsymbol{\imath}}{J} \right\| \\
&= \left\| \frac{(\boldsymbol{V} + c\boldsymbol{I})^{-1}\boldsymbol{\imath} - \boldsymbol{\imath}\boldsymbol{\imath}^\top (\boldsymbol{V} + c\boldsymbol{I})^{-1}\boldsymbol{\imath}/J}{\boldsymbol{\imath}^\top (\boldsymbol{V} + c\boldsymbol{I})^{-1}\boldsymbol{\imath}} \right\| \\
&= \|(\boldsymbol{I} - J^{-1}\boldsymbol{\imath}\boldsymbol{\imath}^\top)\boldsymbol{w}_c^*\| \\
&= \|\boldsymbol{L}_J\, \boldsymbol{w}_c^*\| \\
&\leq \|\boldsymbol{L}_J\| \|\boldsymbol{w}_c^*\| \\
&= \|\boldsymbol{w}_c^*\|,
\end{aligned}$$

where $\boldsymbol{L}_J = \boldsymbol{I} - J^{-1}\boldsymbol{\imath}\boldsymbol{\imath}^\top$ is a symmetric idempotent matrix. Recall that the induced norm of a symmetric matrix $\boldsymbol{M}$ is $\|\boldsymbol{M}\| = \sup_{\boldsymbol{z}} |\boldsymbol{M}z|/||\boldsymbol{z}||$, which coincides with the largest eigenvalue $\overline{\lambda}(\boldsymbol{M})$ of $\boldsymbol{M}$. Thus,

$$\begin{aligned}
\|\boldsymbol{w}_c^*\| &= \left| \frac{(\boldsymbol{V} + c\boldsymbol{I})^{-1}\boldsymbol{\imath}}{\boldsymbol{\imath}^\top (\boldsymbol{V} + c\boldsymbol{I})^{-1}\boldsymbol{\imath}} \right| \\
&= \frac{\boldsymbol{\imath}^\top (\boldsymbol{V} + c\boldsymbol{I})^{-1/2} (\boldsymbol{V} + c\boldsymbol{I})^{-1} (\boldsymbol{V} + c\boldsymbol{I})^{-1/2}\boldsymbol{\imath}}{\boldsymbol{\imath}^\top (\boldsymbol{V} + c\boldsymbol{I})^{-1}\boldsymbol{\imath}} \\
&\leq \sup_{\boldsymbol{z}} \frac{\boldsymbol{z}^\top (\boldsymbol{V} + c\boldsymbol{I})^{-1}\boldsymbol{z}}{\boldsymbol{z}^\top \boldsymbol{z}} \\
&= \overline{\lambda}\left((\boldsymbol{V} + c\boldsymbol{I})^{-1}\right) \\
&= \frac{1}{\underline{\lambda}(\boldsymbol{V} + c\boldsymbol{I})} \\
&= \frac{1}{\underline{\lambda}(\boldsymbol{V}) + c}
\end{aligned}$$

where $\underline{\lambda}(\boldsymbol{V})$ is the smallest eigenvalue of $\boldsymbol{V}$. Condition (ii) of Theorem 3 then completes the proof.