Conditional Value-at-Risk and Average Value-at-Risk: Estimation and Asymptotics

Chun, So Yeon and Shapiro, Alexander and Uryasev, Stan

School of Industrial and Systems Engineering, Georgia Institute of Technology, epartment of Industrial and Systems Engineering, University of Florida

3 April 2011

Conditional Value-at-Risk and Average Value-at-Risk: Estimation and Asymptotics

So Yeon Chun*
School of Industrial and Systems Engineering, Georgia Institute of Technology, schun@isye.gatech.edu

Alexander Shapiro†
School of Industrial and Systems Engineering, Georgia Institute of Technology, ashapiro@isye.gatech.edu

Stan Uryasev
Department of Industrial and Systems Engineering, University of Florida, uryasev@ufl.edu

We discuss linear regression approaches to estimation of law invariant conditional risk measures. Two estimation procedures are considered and compared; one is based on residual analysis of the standard least squares method and the other is in the spirit of the M-estimation approach used in robust statistics. In particular, Value-at-Risk and Average Value-at-Risk measures are discussed in details. Large sample statistical inference of the estimators is derived. Furthermore, finite sample properties of the proposed estimators are investigated and compared with theoretical derivations in an extensive Monte Carlo study. Empirical results on the real-data (different financial asset classes) are also provided to illustrate the performance of the estimators.

Key words: Value-at-Risk, Average Value-at-Risk, linear regression, least squares residuals, M-estimators, quantile regression, conditional risk measures, law invariant risk measures, statistical inference

1. Introduction

In financial industry, sell-side analysts periodically publish recommendations of underlying securities with target prices (e.g., Goldman Sach's Conviction Buy List). Those recommendations reflect specific economic conditions and influence investors' decisions and thus price movements. However, this type of analysis does not provide risk measures associated with underlying companies. We see similar phenomena in the buy-side analysis as well. Each analyst or team covers different sectors (e.g., Airlines VS Semi-conductors) and typically makes separate recommendations for the portfolio managers without associated risk measures. However, risk measures of the covering companies

* Research of this author was partly supported by the NSF award DMS-0914785.

† Research of this author was partly supported by the NSF award DMS-0914785 and ONR award N000140811104.

are one of the most important factors for investment decision making. In this paper, we consider ways to estimate risk measures for a single asset at given market conditions. These information could be useful for investors and portfolio managers to compare prospective securities and to pick the best. For example, when portfolio managers expect the crude oil price to hike (due to inflation or geo-political conflicts), they could select securities less sensitive to oil price movements in the airline industry.

In order to formalize our discussion, let us introduce the following setting. Let $(\Omega, \mathcal{F})$ be a measurable space equipped with probability measure P. A measurable function $Y : \Omega \to \mathbb{R}$ is called a random variable. With random variable Y, we associate a number $\rho(Y)$ to which we refer as *risk measure*. We assume that "smaller is better", i.e., between two possible realizations of random data, we prefer the one with smaller value of $\rho(\cdot)$. The term "risk measure" is somewhat unfortunate since it can be confused with the probability measure. Moreover, in applications one often tries to reach a compromise between minimizing the expectation (i.e., minimizing on average) and controlling the associated risk. Thus, some authors use the term "mean-risk measure", or "acceptability functional" (e.g. Pflug and Römisch 2007). For historical reasons, we use here the "risk measure" terminology. Formally, risk measure is a function $\rho : \mathcal{Y} \to \mathbb{R}$ defined on an appropriate space $\mathcal{Y}$ of random variables. For example, in some applications it is natural to use the space $\mathcal{Y} = L_p(\Omega, \mathcal{F}, P)$, with $p \in [1, \infty)$, of random variables having finite p-th order moments.

It was suggested in Artzner et al. (1999) that a "good" risk measure should have the following properties (axioms), and such risk measures were called *coherent*.

(A1) *M*onotonicity: If $Y, Y' \in \mathcal{Y}$ and $Y \succeq Y'$, then $\rho(Y) \geq \rho(Y')$.

(A2) *C*onvexity:

$$\rho(tY + (1-t)Y') \leq t\rho(Y) + (1-t)\rho(Y')$$

for all $Y, Y' \in \mathcal{Y}$ and all $t \in [0, 1]$.

(A3) *T*ranslation Equivariance: If $a \in \mathbb{R}$ and $Y \in \mathcal{Y}$, then $\rho(Y + a) = \rho(Y) + a$.

(A4) *P*ositive Homogeneity: If $t \geq 0$ and $Y \in \mathcal{Y}$, then $\rho(tY) = t\rho(Y)$.

The notation $Y \succeq Y'$ means that $Y(\omega) \geq Y'(\omega)$ for a.e. $\omega \in \Omega$. We may refer, e.g., to Detlefsen and Scandolo (2005), Weber (2006), Föllmer and Schied (2011) for a further discussion of mathematical properties of risk measures.

An important example of risk measures is the *Value-at-Risk* measure

$$\mathsf{V@R}_\alpha(Y) = \inf\{t : F_Y(t) \geq \alpha\}, \tag{1}$$

where $\alpha \in (0,1)$ and $F_Y(t) = \Pr(Y \leq t)$ is the cumulative distribution function (cdf) of Y, i.e., $\mathsf{V@R}_\alpha(Y) = F_Y^{-1}(\alpha)$ is the left side α-quantile of the distribution of Y. This risk measure satisfies axioms (A1),(A3) and (A4), but not (A2), and hence is not coherent. Another important example is the so-called *Average Value-at-Risk* measure, which can be defined as

$$\mathsf{AV@R}_\alpha(Y) = \inf_{t \in \mathbb{R}} \left\{ t + (1-\alpha)^{-1} \mathbb{E}[Y - t]_+ \right\} \tag{2}$$

(cf., Rockafellar and Uryasev 2002), or equivalently

$$\mathsf{AV@R}_\alpha(Y) = \frac{1}{1-\alpha} \int_\alpha^1 \mathsf{V@R}_\tau(Y) d\tau. \tag{3}$$

Note that $\mathsf{AV@R}_\alpha(Y)$ is finite iff $\mathbb{E}[Y]_+ < \infty$. Therefore, it is natural to use the space $\mathcal{Y} = L_1(\Omega, \mathcal{F}, P)$ of random variables having finite first order moment for the $\mathsf{AV@R}_\alpha$ risk measure. The Average Value-at-Risk measure is also called the Conditional Value-at-Risk or Expected Shortfall measure. (Since we discuss here "conditional" variants of risk measures, we use the Average Value-at-Risk rather than Conditional Value-at-Risk terminology.)

The Value-at-Risk and Average Value-at-Risk measures are widely used to measure and manage risk in the financial industry (see, e.g., Jorion 2003, Duffie and Singleton 2003, Gaglianone et al. 2011, for the financial background and various applications). Note that in the above two examples, risk measures are functions of the distribution of Y. Such risk measures are called *law invariant*. Law invariant risk measures have been studied extensively in the financial risk management literature (e.g., Acerbi 2002, Frey and McNeil 2002, Scaillet 2004, Fermanian and Scaillet 2005, Chen and Tang 2005, Zhu and Fukushima 2009, Jackson and Perraudin 2000, Berkowitz et al.

2002, Bluhm et al. 2002, and reference therein). Sometimes, we write a law invariant risk measure as a function $\rho(F)$ of cdf F.

Now let us consider a situation where there exists information composed of economic and market variables $X_1, ..., X_k$ which can be considered as a set of predictors for a variable of interest Y. In that case, one can be interested in estimation of a risk measure of Y conditional on observed values of predictors $X_1, ..., X_k$. For example, suppose we want to measure (predict) the risk of a single asset given specific economic conditions represented by market index and interest rates. Then, for a random vector $\boldsymbol{X} = (X_1, ..., X_k)^\top$ of relevant predictors, the conditional version of a law invariant risk measure ρ, denoted $\rho(Y|\boldsymbol{X})$ or $\rho_{|\boldsymbol{X}}(Y)$, is obtained by applying ρ to the conditional distribution of Y given $\boldsymbol{X}$. In particular, $\mathsf{V@R}_\alpha(Y|\boldsymbol{X})$ is the α-quantile of the conditional distribution of Y given $\boldsymbol{X}$, and

$$\mathsf{AV@R}_\alpha(Y|\boldsymbol{X}) = \frac{1}{1-\alpha}\int_\alpha^1 \mathsf{V@R}_\tau(Y|\boldsymbol{X})d\tau. \tag{4}$$

Recently, several researchers have paid attention to estimation of the conditional risk measures. For the conditional Value-at-Risk, Chernozhukov and Umantsev (2001) used a polynomial type regression quantile model and Engle and Manganelli (2004) proposed the model which specifies the evolution of the quantile over time using a special type of autoregressive processes. In both models, unknown parameters were estimated by minimizing the regression quantile loss function. For conditional Average Value-at-Risk, Scaillet (2005) and Cai and Wang (2008) utilized Nadaraya-Watson (NW) type nonparametric double kernel estimation while Peracchi and Tanase (2008) and Leorato et al. (2010) used the semiparametric method.

In this paper, we discuss procedures for estimation of conditional risk measures. Especially, we will pay attention to estimation of conditional Value-at-Risk and Average Value-at-Risk measures. We assume the following linear model (linear regression)

$$Y = \beta_0 + \boldsymbol{\beta}^\top \boldsymbol{X} + \varepsilon, \tag{5}$$

where β_0 and $\boldsymbol{\beta} = (\beta_1, ..., \beta_k)^\top$ are (unknown) parameters of the model and the error (noise) random variable ε is assumed to be independent of random vector $\boldsymbol{X}$. Meaning of the model (5) is that there is a true (population) value $\beta_0^*, \boldsymbol{\beta}^*$ of the respective parameters for which (5) holds. Sometimes, we will write this explicitly and sometimes suppress this in the notation.

Let $\rho(\cdot)$ be a law invariant risk measure satisfying axiom (A3) (Translation Equivariance), and $\rho_{|\boldsymbol{X}}(\cdot)$ be its conditional analogue. Note that because of the independence of ε and $\boldsymbol{X}$, it follows that $\rho_{|\boldsymbol{X}}(\varepsilon) = \rho(\varepsilon)$. Together with axiom (A3), this implies

$$\rho_{|\boldsymbol{X}}(Y) = \rho_{|\boldsymbol{X}}(\beta_0 + \boldsymbol{\beta}^\top \boldsymbol{X} + \varepsilon) = \beta_0 + \boldsymbol{\beta}^\top \boldsymbol{X} + \rho_{|\boldsymbol{X}}(\varepsilon) = \beta_0 + \boldsymbol{\beta}^\top \boldsymbol{X} + \rho(\varepsilon). \tag{6}$$

Since $\beta_0 + \rho(\varepsilon) = \rho(\varepsilon + \beta_0)$, we can set $\rho(\varepsilon) = 0$ by adding a constant to the error term. In that case, for the true values of the parameters, we have $\rho_{|\boldsymbol{X}}(Y) = \beta_0^* + \boldsymbol{\beta}^{*\top} \boldsymbol{X}$. Hence, the question is how to estimate these (true) values $\beta_0^*, \boldsymbol{\beta}^*$ of the respective parameters.

This paper is organized as follows. In Section 2, we describe two different estimation procedures for the conditional risk measures; one is based on residuals of the least squares estimation procedure and the other is based on the M-estimation approach. Asymptotic properties of both estimators are provided in Section 3. In Section 4, we investigate the finite sample and asymptotic properties of the considered estimators. We present Monte Carlo simulation results under different distribution assumptions of the error term. Later, we illustrate the performance of different methods on the real data (different financial asset classes) in Section 5. Finally, Section 6 gives some conclusion remarks and suggestions for future research directions.

2. Basic Estimation Procedures

Suppose that we have N observations (data points) $(Y_i, \boldsymbol{X}_i)$, $i = 1, ..., N$, which satisfy the linear regression model (5), i.e.,

$$Y_i = \beta_0 + \boldsymbol{\beta}^\top \boldsymbol{X}_i + \varepsilon_i, \ \ i = 1, ..., N. \tag{7}$$

We assume that: (i) $\boldsymbol{X}_i$, $i = 1, ..., N$, are iid (independent identically distributed) random vectors, and write $\boldsymbol{X}$ for random vector having the same distribution as $\boldsymbol{X}_i$, (ii) the errors $\varepsilon_1, .., \varepsilon_N$ are iid with finite second order moments and independent of $\boldsymbol{X}_i$. We denote by $\sigma^2 = \mathrm{Var}[\varepsilon_i]$ the common variance of the error terms.

There are two basic approaches to estimation of the true values of β_0 and $\boldsymbol{\beta}$. One approach is to apply the standard Least Squares (LS) estimation procedure and then to make an adjustment of the estimate of the intercept parameter β_0. That is, let $\tilde{\beta}_0$ and $\tilde{\boldsymbol{\beta}}$ be the least squares estimators of the respective parameters of the linear model (7) and

$$e_i := Y_i - \tilde{\beta}_0 - \tilde{\boldsymbol{\beta}}^\top \boldsymbol{X}_i,\ \ i = 1, ..., N, \tag{8}$$

be the corresponding residuals. By the standard theory of the LS method, we have that $\tilde{\beta}_0$ and $\tilde{\boldsymbol{\beta}}$ are unbiased estimators of the respective parameters of the linear model (5) provided $\mathbb{E}[\varepsilon] = 0$. Therefore, we need to make the correction $\tilde{\beta}_0 + \rho(\varepsilon)$ of the intercept estimator. If we knew the true values $\varepsilon_1, ..., \varepsilon_N$ of the error term, we could estimate $\rho(\varepsilon)$ by replacing the cdf F_ε of ε by its empirical estimate $\hat{F}_{\varepsilon,N}$ associated with $\varepsilon_1, ..., \varepsilon_N$, i.e., to estimate $\rho(F_\varepsilon)$ by $\rho(\hat{F}_{\varepsilon,N})$. Since true values of the error term are unknown, it is a natural idea to replace $\varepsilon_1, ..., \varepsilon_N$ by the residual values $e_1, ..., e_N$. Hence, we use the estimator $\tilde{\beta}_0 + \rho(\hat{F}_{e,N})$, where $\hat{F}_{e,N}$ is the empirical cdf of the residual values, i.e., $\hat{F}_{e,N}$ is the cdf of the probability distribution assigning mass $1/N$ to each point e_i, $i = 1, ..., N$ (see section 3.1 for further discussion). We refer to this estimation approach as the *Least Squares Residuals* (LSR) method.

An alternative approach is based on the following idea. Suppose that we can construct a function $h(y, \theta)$ of $y \in \mathbb{R}$ and $\theta \in \mathbb{R}$, convex in θ, such that the minimizer of $\mathbb{E}_F[h(Y, \theta)]$ will be equal to $\rho(F)$, i.e., $\rho(F) = \arg\min_\theta \mathbb{E}_F[h(Y, \theta)]$. Since $\rho(Y + a) = \rho(Y) + a$ for any $a \in \mathbb{R}$, it follows that the function $h(y, \theta)$ should be of the form $h(y, \theta) = \psi(y - \theta)$ for some convex function $\psi : \mathbb{R} \to \mathbb{R}$. We refer to $\psi(\cdot)$ as the *error function*. Therefore, we need to construct an error function such that

$$\rho(F) = \arg\min_\theta \mathbb{E}_F[\psi(Y - \theta)]. \tag{9}$$

This is equivalent to solving the equation

$$\mathbb{E}_F[\phi(Y - \theta)] = 0, \tag{10}$$

where $\phi(t) := \psi'(t)$. Note that the error function $\psi(\cdot)$ could be nondifferentiable, in which case the corresponding derivative function $\phi(\cdot)$ is discontinuous. That is, the function $\phi(\cdot)$ is monotonically nondecreasing.

The corresponding estimators $\hat{\beta}_0$ and $\hat{\boldsymbol{\beta}}$ are taken as solutions of the optimization problem

$$\operatorname*{Min}_{\beta_0,\boldsymbol{\beta}} \sum_{i=1}^{N} \psi\left(Y_i - \beta_0 - \boldsymbol{\beta}^{\mathsf{T}} \boldsymbol{X}_i\right). \tag{11}$$

In the statistics literature, such estimators are called M-estimators (the terminology which we will follow) and for an appropriate choice of the error function, this is the approach of robust regression (Huber 1981). For the $\mathsf{V@R}_\alpha$ risk measure, the error function is readily available (recall that $[t]_+ = \max\{0,t\}$):

$$\psi(t) := \alpha[t]_+ + (1-\alpha)[-t]_+. \tag{12}$$

The corresponding robust regression approach is known as the *quantile regression* method (cf. Koenker 2005).

For coherent risk measures, the situation is more delicate. Let us make the following observations. Suppose that the representation (9) holds. Let F_1 and F_2 be two cdf such that $\rho(F_1) = \rho(F_2) = \theta$. Then it follows by (9) (by (10)) that $\rho(tF_1 + (1-t)F_2) = \theta$ for any $t \in [0,1]$. This is quite a strong necessary condition for existence of a representation of the form (9). It certainly doesn't hold for the $\mathsf{AV@R}_\alpha$, $\alpha \in (0,1)$, risk measure. Indeed, consider the following probability distributions $F_1 := \alpha\delta_a + \frac{1}{2}(1-\alpha)(\delta_b + \delta_d)$, $F_2 := \alpha\delta_c + (1-\alpha)\delta_{(b+d)/2}$ and

$$\tfrac{1}{2}(F_1 + F_2) = \tfrac{1}{2}\alpha\delta_a + \tfrac{1}{4}(1-\alpha)\delta_b + \tfrac{1}{2}\alpha\delta_c + \tfrac{1}{4}(1-\alpha)\delta_d + \tfrac{1}{2}(1-\alpha)\delta_{(b+d)/2},$$

where δ_x denotes measure of mass one at x, $\alpha \in (\frac{1}{2}, 1)$ and $a < b < c < d$ are such that $c < \frac{1}{2}(b+d)$. It is straightforward to calculate that $\mathsf{AV@R}_\alpha(F_1) = \mathsf{AV@R}_\alpha(F_2) = \frac{1}{2}(b+d)$. However, since $\alpha \in (\frac{1}{2}, 1)$,

$$\mathsf{AV@R}_\alpha\big(\tfrac{1}{2}(F_1+F_2)\big) = \tfrac{1}{4}\big(b + 2c\alpha(1-\alpha)^{-1} + 2d\big) > \tfrac{1}{4}(b+2c+2d) > \tfrac{1}{4}(2b+c+2d) > \tfrac{1}{2}(b+d).$$

These arguments are due to Gneiting (2009).

This shows that for general coherent risk measures, possibility of constructing the corresponding M-estimators is rather exceptional, and such estimators certainly do not exist for the $\mathsf{AV@R}_\alpha$ risk measure. Nevertheless, it is possible to construct the following approximations (this construction is essentially due to Rockafellar et al. (2008)).

PROPOSITION 1. *Let $\psi_j:\mathbb{R}\to\mathbb{R}$, $j=1,...,r$, be convex functions, $\lambda_j\in\mathbb{R}$ be such that $\sum_{j=1}^r\lambda_j=1$ and*

$$\mathcal{E}(Y):=\inf_{\boldsymbol{\tau}\in\mathbb{R}^r}\left\{\mathbb{E}\left[\sum_{j=1}^r\psi_j(Y-\tau_j)\right]:\sum_{j=1}^r\lambda_j\tau_j=0\right\}. \tag{13}$$

Moreover, let $S_j(Y)$ be a minimizer of $\mathbb{E}[\psi_j(Y-\theta)]$ over $\theta\in\mathbb{R}$. Then $S(Y):=\sum_{j=1}^r\lambda_jS_j(Y)$ is a minimizer of $\mathcal{E}(Y-\theta)$ over $\theta\in\mathbb{R}$.

Proof Consider the problem

$$\operatorname*{Min}_{\theta,\boldsymbol{\tau}}\mathbb{E}\left[\sum_{j=1}^r\psi_j(Y-\theta-\tau_j)\right]\ \text{s.t.}\ \sum_{j=1}^r\lambda_j\tau_j=0. \tag{14}$$

By making change of variables $\eta_j=\theta+\tau_j$, $j=1,...,r$, we can write this problem in the form

$$\operatorname*{Min}_{\theta,\boldsymbol{\eta}}\mathbb{E}\left[\sum_{j=1}^r\psi_j(Y-\eta_j)\right]\ \text{s.t.}\ \sum_{j=1}^r\lambda_j\eta_j=\theta. \tag{15}$$

Since $S_j(Y)$ is a minimizer of $\mathbb{E}[\psi_j(Y-\eta_j)]$, it follows that $\eta_j=S_j(Y)$, $i=1,..,r$, $\theta=S(Y)$, is an optimal solution of problem (15). This completes the prof. $\square$

In particular, we can consider functions $\psi_j(\cdot)$ of the form (12), i.e.,

$$\psi_j(t):=\alpha_j[t]_++(1-\alpha_j)[-t]_+, \tag{16}$$

for some $\alpha_j\in(0,1)$, $j=1,...,r$. Then $S_j(Y)=\mathsf{V@R}_{\alpha_j}(Y)$ and hence the risk measure $\sum_{j=1}^r\lambda_j\mathsf{V@R}_{\alpha_j}(Y)$ is a minimizer of $\mathcal{E}(Y-\theta)$. We can view $\sum_{j=1}^r\lambda_j\mathsf{V@R}_{\alpha_j}(Y)$ as a discretization of the integral $\frac{1}{1-\alpha}\int_\alpha^1\mathsf{V@R}_\tau(Y)d\tau$ if we set $\Delta:=(1-\alpha)/r$ and take

$$\lambda_j:=(1-\alpha)^{-1}\Delta,\ \alpha_j:=\alpha+(j-0.5)\Delta,\ j=1,...,r. \tag{17}$$

For this choice of λ_j, α_j, and by formula (3), we have that

$$\mathsf{AV@R}_\alpha(Y)=\frac{1}{1-\alpha}\int_\alpha^1\mathsf{V@R}_\tau(Y)d\tau\approx\sum_{j=1}^r\lambda_j\mathsf{V@R}_{\alpha_j}(Y)=S(Y). \tag{18}$$

Consider now the problem

$$\operatorname*{Min}_{\beta_0,\boldsymbol{\beta}}\mathcal{E}(Y-\beta_0-\boldsymbol{\beta}^\mathsf{T}\boldsymbol{X}). \tag{19}$$

By the definition (13) of $\mathcal{E}(\cdot)$, we can write this problem in the following equivalent form

$$\operatorname*{Min}_{\boldsymbol{\tau},\beta_0,\boldsymbol{\beta}} \mathbb{E}\left[\sum_{j=1}^r \psi_j(Y-\beta_0-\boldsymbol{\beta}^\top \boldsymbol{X}-\tau_i)\right] \text{ s.t. } \sum_{j=1}^r \lambda_j\tau_j=0. \tag{20}$$

The so-called Sample Average Approximation (SAA) of this problem is

$$\operatorname*{Min}_{\boldsymbol{\tau},\beta_0,\boldsymbol{\beta}} \frac{1}{N}\sum_{i=1}^N\sum_{j=1}^r \psi_j(Y_i-\beta_0-\boldsymbol{\beta}^\top \boldsymbol{X}_i-\tau_i) \text{ s.t. } \sum_{j=1}^r \lambda_j\tau_j=0. \tag{21}$$

The above problem (21) can be formulated as a linear programming problem. Following Rockafellar et al. (2008), we consider the following estimators.

Mixed quantile estimator for $\mathsf{AV@R}_\alpha(\mathbf{Y}|\boldsymbol{x})$

We refer to $\check{\beta}_0+\check{\boldsymbol{\beta}}^\top\boldsymbol{x}$ as the *mixed quantile estimator* of $\mathsf{AV@R}_\alpha(Y|\boldsymbol{x})$, where $(\check{\boldsymbol{\tau}},\check{\beta}_0,\check{\boldsymbol{\beta}})$ is an optimal solution of problem (21).

This idea can be extended to a larger class of law invariant risk measures. For example, consider a risk measure

$$\rho(Y):=c\mathbb{E}[Y]+(1-c)\mathsf{AV@R}_\alpha(Y) \tag{22}$$

for some constants $c\in[0,1]$ and $\alpha\in(0,1)$. Recall that the minimizer of $\mathbb{E}[(Y-t)^2]$ is $t^*=\mathbb{E}[Y]$. Therefore, by taking functions $\psi_0(t):=t^2$, functions $\psi_j(t)$ of the form (16), λ_j, and α_j given in (17), we can construct the corresponding error function

$$\mathcal{E}(Y):=\inf_{\boldsymbol{\tau}\in\mathbb{R}^{r+1}}\left\{\mathbb{E}\left[\psi_0(Y-\tau_0)+\sum_{j=1}^r\psi_j(Y-\tau_j)\right]: c\tau_0+\sum_{j=1}^r(1-c)\lambda_j\tau_j=0\right\}. \tag{23}$$

As another example, consider risk measures of the form

$$\rho(Y):=\int_0^1 \mathsf{AV@R}_\alpha(Y)d\mu(\alpha), \tag{24}$$

where μ is a probability measure on the interval $[0,1)$. By a result due to Kusuoka (2001), this measures form a class of the comonote law invariant coherent risk measures. By (3), we can write such risk measure as

$$\rho(Y)=\int_0^1\int_\alpha^1(1-\alpha)^{-1}\mathsf{V@R}_\tau(Y)d\tau d\mu(\alpha)=\int_0^1 w(\tau)\mathsf{V@R}_\tau(Y)d\tau, \tag{25}$$

where $w(\tau) := \int_0^\tau (1-\alpha)^{-1} d\mu(\alpha)$. Such risk measures are also called spectral risk measures (Acerbi 2002). By making a discretization of the above integral (25), we can proceed as above.

It could be remarked here that while the LSR approach is quite general, the approach based on mixing M-estimators is somewhat restrictive. Constructing an appropriate error function for a particular risk measure could be quite involved.

3. Large Sample Statistical Inference

In the previous section, we formulated two approaches, the LSR estimators and mixed M-estimators, to estimation of the true (population) values of parameters $\beta_0^*, \boldsymbol{\beta}^*$ of the linear model (5) such that $\rho(\varepsilon) = 0$. For the $\mathsf{V@R}_\alpha$ risk measure, the corresponding M-estimators $\hat{\beta}_0$ and $\hat{\boldsymbol{\beta}}$ are taken as solutions of the optimization problem (11), with the error function (12), and referred to as the quantile regression estimators. For the $\mathsf{AV@R}_\alpha$ risk measure and more generally comonotone risk measures of the form (25), we constructed the corresponding mixed quantile estimators $\check{\boldsymbol{\tau}}, \check{\beta}_0, \check{\boldsymbol{\beta}}$. In this section, we discuss statistical properties of these estimators. In particular, we address the question of which of these two estimation procedures is more efficient by computing corresponding asymptotic variances.

3.1. Statistical Inference of Least Squares Residual Estimators

The linear model (7) can be written as

$$\boldsymbol{Y} = \mathbb{X}[\beta_0; \boldsymbol{\beta}] + \boldsymbol{\epsilon}, \tag{26}$$

where $\boldsymbol{Y} = (Y_1, ..., Y_N)^\mathsf{T}$ is $N \times 1$ vector of responses, $\mathbb{X}$ is $N \times (k+1)$ data matrix of predictor variables with rows $(1, \boldsymbol{X}_i^\mathsf{T})$, $i = 1, ..., N$, (i.e., first column of $\mathbb{X}$ is column of ones), $\boldsymbol{\beta} = (\beta_1, ..., \beta_k)^\mathsf{T}$ vector of parameters and $\boldsymbol{\epsilon} = (\varepsilon_1, .., \varepsilon_N)^\mathsf{T}$ is $N \times 1$ vector of errors. By $[\beta_0; \boldsymbol{\beta}]$, we denote $(k+1) \times 1$ vector $(\beta_0, \boldsymbol{\beta}^\mathsf{T})^\mathsf{T}$. We assume that the conditions (i) and (ii), specified at the beginning of section 2, hold. It is also possible to view data points $\boldsymbol{X}_i$ as deterministic. In that case, we assume that $\mathbb{X}$ has full column rank $k+1$.

Let $\tilde{\beta}_0$ and $\tilde{\boldsymbol{\beta}}$ be the least squares estimators of the respective parameters of the linear model (7).

Recall that these estimators are given by $[\tilde\beta_0;\tilde{\boldsymbol\beta}]=(\mathbb{X}^\top\mathbb{X})^{-1}\mathbb{X}^\top\boldsymbol{Y}$, vector of residuals $\boldsymbol{e}:=\boldsymbol{Y}-\mathbb{X}[\tilde\beta_0,\tilde{\boldsymbol\beta}]$ is given by

$$\boldsymbol{e}=(\boldsymbol{I}_N-\boldsymbol{H})\boldsymbol{Y}=(\boldsymbol{I}_N-\boldsymbol{H})\boldsymbol{\epsilon},$$

where $\boldsymbol{I}_N$ is the $N\times N$ identity matrix, and $\boldsymbol{H}=\mathbb{X}(\mathbb{X}^\top\mathbb{X})^{-1}\mathbb{X}^\top$ is the so-called hat matrix. Note that trace$(\boldsymbol{H})=k+1$ and we have that

$$\varepsilon_i-e_i=[1;\boldsymbol{X}_i^\top](\mathbb{X}^\top\mathbb{X})^{-1}\mathbb{X}^\top\boldsymbol{\epsilon},\ \ i=1,...,N. \tag{27}$$

If we knew errors $\varepsilon_1,..,\varepsilon_N$, we could estimate $\rho(\varepsilon)$ by the corresponding sample estimate based on the empirical cdf

$$\hat F_{\varepsilon,N}(\cdot)=N^{-1}\sum_{i=1}^{N}\mathbb{I}_{[\varepsilon_i,\infty)}(\cdot), \tag{28}$$

where $\mathbb{I}_A(\cdot)$ denotes the indicator function of set A. However, the true values of the errors are unknown. Therefore, in the LSR approach we replace them by the residuals computed by the least squares method and hence estimate $\rho(\varepsilon)$ by employing the respective empirical cdf $\hat F_{e,N}(\cdot)$ instead of $\hat F_{\varepsilon,N}(\cdot)$.

The first natural question is whether the LSR estimators are consistent, i.e., converge w.p.1 to their true values as the sample size N tends to infinity. It is well known that, under the specified assumptions, the LS estimators $\tilde\beta_0$ and $\tilde{\boldsymbol\beta}$ are consistent, with $\tilde\beta_0$ being consistent under the condition $\mathbb{E}[\varepsilon]=0$. The question of consistency of empirical estimates of law invariant coherent risk measures was studied in Wozabal and Wozabal (2009). It was shown that, under mild regularity conditions, such estimators are consistent. In particular, the consistency holds for the comonotone risk measures of the form (25), i.e., $\rho(\hat F_{\varepsilon,N})$ converges w.p.1 to $\rho(F_\varepsilon)$ as $N\to\infty$. It is also possible to show that the difference $\rho(\hat F_{\varepsilon,N})-\rho(\hat F_{e,N})$ tends w.p.1 to zero and hence $\rho(\hat F_{e,N})$ converges w.p.1 to $\rho(F_\varepsilon)$ as well. A rigorous proof of this could be quite technical and will be beyond the scope of this paper.

We have that the LS estimator $\left[\tilde\beta_0;\tilde{\boldsymbol\beta}\right]$ asymptotically has normal distribution with the asymptotic covariance matrix $N^{-1}\sigma^2\boldsymbol{\Omega}^{-1}$, where $\boldsymbol\mu:=\mathbb{E}[\boldsymbol{X}]$, $\boldsymbol\Sigma:=\mathbb{E}\left[\boldsymbol{X}\boldsymbol{X}^\top\right]$ and $\boldsymbol\Omega:=\begin{bmatrix}1 & \boldsymbol\mu^\top\\ \boldsymbol\mu & \boldsymbol\Sigma\end{bmatrix}$. Consequently, for a given $\boldsymbol{x}$, the estimate $\tilde\beta_0+\boldsymbol{x}^\top\tilde{\boldsymbol\beta}$ asymptotically has normal distribution with the asymptotic variance $N^{-1}\sigma^2[1;\boldsymbol{x}^\top]\boldsymbol\Omega^{-1}[1;\boldsymbol{x}^\top]^\top$.

We also have that random vectors $(\tilde\beta_0, \tilde{\boldsymbol\beta})$ and $\boldsymbol{e}$ are uncorrelated. Therefore, if errors ε_i have normal distribution, then vectors $(\tilde\beta_0, \tilde{\boldsymbol\beta})$ and $\boldsymbol{e}$ jointly have a multivariate normal distribution and these vectors are independent. Consequently, $\tilde\beta_0 + \boldsymbol{x}^\mathsf{T}\tilde{\boldsymbol\beta}$ and $\rho(\hat F_{e,N})$ are independent. For nonnormal distribution, this independence holds asymptotically and thus asymptotically $\tilde\beta_0 + \boldsymbol{x}^\mathsf{T}\tilde{\boldsymbol\beta}$ and $\rho(\hat F_{e,N})$ are uncorrelated.

Asymptotics of empirical estimators of law invariant coherent risk measures were studied in Pflug and Wozabal (2010) and Shapiro et al. (2009, section 6.5). Derivation of the asymptotic variance of $\rho(\hat F_{\varepsilon,N})$, for a general law invariant risk measure, could be quite involved. Let us consider two important cases of the $\mathsf{V@R}_\alpha$ and $\mathsf{AV@R}_\alpha$ risk measures. We give below a summary of basic results, for a more technical discussion we refer to the Appendix.

In case of $\rho := \mathsf{V@R}_\alpha$, the LSR estimate of $\mathsf{V@R}_\alpha(\varepsilon)$ becomes

$$\widehat{\mathsf{V@R}}_\alpha(e) := \hat F^{-1}_{e,N}(\alpha) = e_{(\lceil N\alpha\rceil)}, \tag{29}$$

where $e_{(1)} \le ... \le e_{(N)}$ are order statistics (i.e., numbers $e_1, ..., e_N$ arranged in the increasing order), and $\lceil a\rceil$ denotes the smallest integer $\ge a$. Suppose that the cdf $F_\varepsilon(\cdot)$ has nonzero density $f_\varepsilon(\cdot) = F'_\varepsilon(\cdot)$ at $F^{-1}_\varepsilon(\alpha)$ and let

$$\omega^2 := \frac{\alpha(1-\alpha)}{\left[f_\varepsilon\left(F^{-1}_\varepsilon(\alpha)\right)\right]^2}. \tag{30}$$

LSR estimator of $\mathsf{V@R}_\alpha(\mathbf{Y}|\boldsymbol{x})$

Consider the LSR estimator $\tilde\beta_0 + \boldsymbol{x}^\mathsf{T}\tilde{\boldsymbol\beta} + \widehat{\mathsf{V@R}}_\alpha(e)$ of $\mathsf{V@R}_\alpha(\boldsymbol{Y}|\boldsymbol{x})$. Suppose that the set of population α-quantiles is a singleton. Then the LSR estimator is a consistent estimator of $\mathsf{V@R}_\alpha(\boldsymbol{Y}|\boldsymbol{x})$, and the asymptotic variance of this estimator can be approximated by

$$N^{-1}\left(\omega^2 + \sigma^2[1;\boldsymbol{x}^\mathsf{T}]\boldsymbol{\Omega}^{-1}[1;\boldsymbol{x}^\mathsf{T}]^\mathsf{T}\right). \tag{31}$$

Detailed derivation of above asymptotics is discussed in Appendix A.

For the $\rho := \mathsf{AV@R}_\alpha$ risk measure, the LSR estimate of $\mathsf{AV@R}_\alpha(\varepsilon)$ is given by

$$\begin{aligned}\widehat{\mathsf{AV@R}}_\alpha(e) &= \inf_{t\in\mathbb{R}}\left\{t + \tfrac{1}{(1-\alpha)N}\textstyle\sum_{i=1}^N [e_i - t]_+\right\}\\ &= \widehat{\mathsf{V@R}}_\alpha(e) + \tfrac{1}{(1-\alpha)N}\textstyle\sum_{i=1}^N\left[e_i - \widehat{\mathsf{V@R}}_\alpha(e)\right]_+\\ &= e_{(\lceil N\alpha\rceil)} + \tfrac{1}{(1-\alpha)N}\textstyle\sum_{i=\lceil N\alpha\rceil+1}^N\left(e_{(i)} - e_{(\lceil N\alpha\rceil)}\right).\end{aligned} \tag{32}$$

LSR estimator of $\mathsf{AV@R}_\alpha(\boldsymbol{Y}|\boldsymbol{x})$

Consider the LSR estimator $\tilde{\beta}_0 + \boldsymbol{x}^\mathsf{T}\tilde{\boldsymbol{\beta}} + \widehat{\mathsf{AV@R}}_\alpha(e)$ of $\mathsf{AV@R}_\alpha(\boldsymbol{Y}|\boldsymbol{x})$. This estimator is consistent and its asymptotic variance is given by

$$N^{-1}\left(\gamma^2 + \sigma^2[1;\boldsymbol{x}^\mathsf{T}]\boldsymbol{\Omega}^{-1}[1;\boldsymbol{x}^\mathsf{T}]^\mathsf{T}\right), \tag{33}$$

where $\gamma^2 := (1-\alpha)^{-2}\mathrm{Var}\big([\varepsilon - \mathsf{V@R}_\alpha(\varepsilon)]_+\big)$, $\boldsymbol{\Omega} := \begin{bmatrix} 1 & \boldsymbol{\mu}^\mathsf{T} \\ \boldsymbol{\mu} & \boldsymbol{\Sigma} \end{bmatrix}$, $\boldsymbol{\mu} := \mathbb{E}[\boldsymbol{X}]$ and $\boldsymbol{\Sigma} := \mathbb{E}\left[\boldsymbol{X}\boldsymbol{X}^\mathsf{T}\right]$.

The above asymptotics are discussed in Appendix B.

Remark 1. It should be remembered that the above approximate variances are asymptotic results. Suppose for the moment that $N < (1-\alpha)^{-1}$. Then $\lceil N\alpha \rceil = N$ and hence $\widehat{\mathsf{V@R}}_\alpha(\varepsilon) = \max\{\varepsilon_1, ..., \varepsilon_N\}$. Consequently $\left[\varepsilon_i - \widehat{\mathsf{V@R}}_\alpha(\varepsilon)\right]_+ = 0$ for all $i = 1, ..., N$, and hence

$$\widehat{\mathsf{AV@R}}_\alpha(\varepsilon) = \widehat{\mathsf{V@R}}_\alpha(\varepsilon) = \max\{\varepsilon_1, ..., \varepsilon_N\}.$$

In that case the above asymptotics are inappropriate. In order for these asymptotics to be reasonable, N should be significantly bigger than $(1-\alpha)^{-1}$.

LSR approach can be easily applied to a considerably larger class of law invariant risk measures. For example, let us consider the entropic risk measure $\rho(Y) := \alpha^{-1}\log \mathbb{E}[e^{\alpha Y}]$, where $\alpha > 0$ is a positive constant. This risk measure satisfies axioms (A1)–(A3), but it is not positively homogeneous (see Giesecke and Weber 2008, for the general discussion of utility-based shortfall risk including entropic risk measure). The empirical estimate of $\rho(\varepsilon)$ is

$$\rho(\hat{F}_{\varepsilon,N}) = \alpha^{-1}\log\left(N^{-1}\textstyle\sum_{i=1}^N e^{\alpha\varepsilon_i}\right). \tag{34}$$

Of course, as it was discussed above, the errors ε_i should be replaced by the respective residuals e_i in the construction of the corresponding LSR estimators. By using linearizations $e^{\alpha\varepsilon} = 1 + \alpha\varepsilon + o(\alpha\varepsilon)$ and $\log(1+x) = x + o(x)$, we obtain that $N^{1/2}[\rho(\hat{F}_{\varepsilon,N}) - \rho(\varepsilon)]$ converges in distribution to normal with zero mean and variance σ^2 (by the Delta Theorem).

3.2. Statistical Inference of Quantile and Mixed Quantile Estimators

As it was discussed in section 2, the quantile regression is a particular case of the M-estimation method with the error function $\psi(\cdot)$ of the form (12). By the Law of Large Numbers (LLN), we have that N^{-1} times the objective function in (11) converges (pointwise) w.p.1 to the function $\Psi(\beta_0, \boldsymbol{\beta}) := \mathbb{E}\left[\psi(Y - \beta_0 - \boldsymbol{\beta}^\top \boldsymbol{X})\right]$. We also have

$$\begin{aligned} \Psi(\beta_0, \boldsymbol{\beta}) &= \mathbb{E}\left[\psi\left(\beta_0^* + \boldsymbol{\beta}^{*\top} \boldsymbol{X} + \varepsilon - \beta_0 - \boldsymbol{\beta}^\top \boldsymbol{X}\right)\right] \\ &= \mathbb{E}[\psi\left(\varepsilon - (\beta_0 - \beta_0^*) - (\boldsymbol{\beta} - \boldsymbol{\beta}^*)^\top \boldsymbol{X}\right)]. \end{aligned} \tag{35}$$

Under mild regularity conditions, derivatives of $\Psi(\beta_0, \boldsymbol{\beta})$ can be taken inside the integral (expectation) and hence

$$\begin{aligned} \nabla_{\beta_0} \Psi(\beta_0, \boldsymbol{\beta}) &= \mathbb{E}[\nabla_{\beta_0} \psi\left(\varepsilon - (\beta_0 - \beta_0^*) - (\boldsymbol{\beta} - \boldsymbol{\beta}^*)^\top \boldsymbol{X}\right)] \\ &= -\mathbb{E}[\psi'\left(\varepsilon - (\beta_0 - \beta_0^*) - (\boldsymbol{\beta} - \boldsymbol{\beta}^*)^\top \boldsymbol{X}\right)], \end{aligned} \tag{36}$$

$$\begin{aligned} \nabla_{\boldsymbol{\beta}} \Psi(\beta_0, \boldsymbol{\beta}) &= \mathbb{E}[\nabla_{\boldsymbol{\beta}} \psi\left(\varepsilon - (\beta_0 - \beta_0^*) - (\boldsymbol{\beta} - \boldsymbol{\beta}^*)^\top \boldsymbol{X}\right)] \\ &= -\mathbb{E}[\psi'\left(\varepsilon - (\beta_0 - \beta_0^*) - (\boldsymbol{\beta} - \boldsymbol{\beta}^*)^\top \boldsymbol{X}\right) \boldsymbol{X}]. \end{aligned} \tag{37}$$

Since ε and $\boldsymbol{X}$ are independent, we obtain that derivatives of $\Psi(\beta_0, \boldsymbol{\beta})$ are zeros at $(\beta_0^*, \boldsymbol{\beta}^*)$ if the following condition holds

$$\mathbb{E}[\psi'(\varepsilon)] = 0. \tag{38}$$

Since function $\Psi(\cdot, \cdot)$ is convex, it follows that if condition (38) holds, then $\Psi(\cdot, \cdot)$ attains its minimum at $(\beta_0^*, \boldsymbol{\beta}^*)$. If the minimizer $(\beta_0^*, \boldsymbol{\beta}^*)$ is unique, then the estimator $(\hat{\beta}_0, \hat{\boldsymbol{\beta}})$ converges w.p.1 to the population value $(\beta_0^*, \boldsymbol{\beta}^*)$ as $N \to \infty$, i.e., $(\hat{\beta}_0, \hat{\boldsymbol{\beta}})$ is a *consistent* estimator of $(\beta_0^*, \boldsymbol{\beta}^*)$ (cf. Huber 1981). That is, (38) is the basic condition for consistency of $(\hat{\beta}_0, \hat{\boldsymbol{\beta}})$.

For the error function (12) of the quantile regression, we have

$$\psi'(t) = \begin{cases} \alpha - 1 & \text{if } t < 0, \\ \alpha & \text{if } t > 0. \end{cases} \tag{39}$$

(Note that here the error function $\psi(t)$ is not differentiable at $t = 0$ and its derivative $\psi'(t)$ is discontinuous at $t = 0$. Nevertheless, all arguments can go through provided that the error term has a continuous distribution.) Consequently,

$$\mathbb{E}[\psi'(\varepsilon)] = (\alpha - 1) F_\varepsilon(0) + \alpha(1 - F_\varepsilon(0)) = \alpha - F_\varepsilon(0), \tag{40}$$

and hence condition (38) holds iff $F_\varepsilon(0)=\alpha$, or equivalently $F_\varepsilon^{-1}(\alpha)=0$ provided this quantile is unique. In that case, the estimator $(\hat\beta_0,\hat{\boldsymbol\beta})$ is consistent if the population value β_0^* is normalized such that $\mathsf{V@R}_\alpha(\varepsilon)=0$. That is, for this error function, $\hat\beta_0+\hat{\boldsymbol\beta}^\top\boldsymbol{x}$ is a consistent estimator of the conditional Value-at-Risk $\mathsf{V@R}_\alpha(Y|\boldsymbol{x})$ of Y given $\boldsymbol{X}=\boldsymbol{x}$.

It is also possible to derive asymptotics of the estimator $(\hat\beta_0,\hat{\boldsymbol\beta})$. That is, suppose that the cdf $F_\varepsilon(\cdot)$ has nonzero density $f_\varepsilon(\cdot)=F_\varepsilon'(\cdot)$ at $F_\varepsilon^{-1}(\alpha)$ and consider ω^2 defined in (30). Then $N^{1/2}\left[\hat\beta_0-\beta_0^*;\hat{\boldsymbol\beta}-\boldsymbol\beta^*\right]$ converges in distribution to normal with zero mean vector and covariance matrix (cf., Koenker 2005)

$$\omega^2[1;\boldsymbol{x}^\top]\boldsymbol{\Omega}^{-1}[1;\boldsymbol{x}^\top]^\top, \tag{41}$$

i.e., N^{-1} times the matrix given in (41) is the asymptotic covariance matrix of $\left[\hat\beta_0;\hat{\boldsymbol\beta}\right]$.

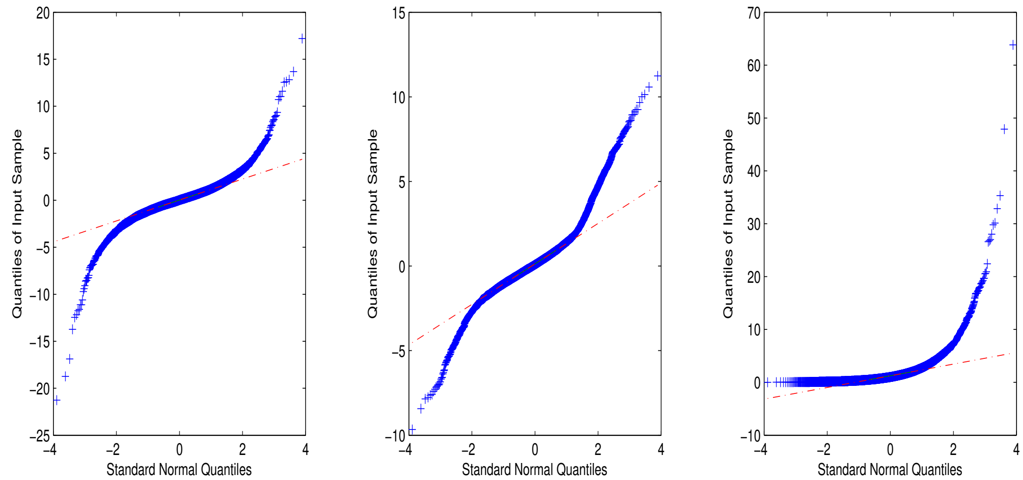
Remark 2. Note that by LLN, we have that $N^{-1}\sum_{i=1}^N\boldsymbol{X}_i$ and $N^{-1}\sum_{i=1}^N\boldsymbol{X}_i\boldsymbol{X}_i^\top$ converge w.p.1 as $N\to\infty$ to the vector $\boldsymbol\mu$ and matrix $\boldsymbol\Sigma$ respectively and that $\boldsymbol\Sigma-\boldsymbol\mu\boldsymbol\mu^\top$ is the covariance matrix of $\boldsymbol{X}$. In case of deterministic $\boldsymbol{X}_i$, we simply define vector $\boldsymbol\mu$ and matrix $\boldsymbol\Sigma$ as the respective limits of $N^{-1}\sum_{i=1}^N\boldsymbol{X}_i$ and $N^{-1}\sum_{i=1}^N\boldsymbol{X}_i\boldsymbol{X}_i^\top$, assuming that such limits exist. It follows then that $N^{-1}\mathbb{X}^\top\mathbb{X}\to\boldsymbol\Omega$ as $N\to\infty$.

The mixed quantile estimator $\check\beta_0+\check{\boldsymbol\beta}^\top\boldsymbol{x}$ can be justified by the following arguments. We have that an optimal solution $(\check{\boldsymbol\tau},\check\beta_0,\check{\boldsymbol\beta})$ of problem (21) converges w.p.1 as $N\to\infty$ to the optimal solution $(\boldsymbol\tau^\star,\beta_0^\star,\boldsymbol\beta^\star)$ of problem (20), provided (20) has unique optimal solution. Because of the linear model (5), we can write problem (20) as

$$\operatorname*{Min}_{\boldsymbol\tau,\beta_0,\boldsymbol\beta}\mathbb{E}\left[\sum_{j=1}^r\psi_{\alpha_j}(\varepsilon+\beta_0^*-\beta_0+(\boldsymbol\beta^*-\boldsymbol\beta)^\top\boldsymbol{X}-\tau_j)\right]\ \text{s.t.}\ \sum_{j=1}^r\lambda_j\tau_j=0, \tag{42}$$

where β_0^* and $\boldsymbol\beta^*$ are population values of the parameters. Similar to the proof of Proposition 1, by making change of variables $\eta_j=\beta_0+\tau_j$, $j=1,\dots,r$, we can write problem (42) in the following equivalent form

$$\operatorname*{Min}_{\boldsymbol\eta,\beta_0,\boldsymbol\beta}\mathbb{E}\left[\sum_{j=1}^r\psi_{\alpha_j}(\varepsilon+\beta_0^*-\eta_j+(\boldsymbol\beta^*-\boldsymbol\beta)^\top\boldsymbol{X})\right]\ \text{s.t.}\ \sum_{j=1}^r\lambda_j\eta_j=\beta_0. \tag{43}$$



(a) N(0,1) vs. t(3) (b) N(0,1) vs. CN(1,9) (c) N(0,1) vs. LN(0,1)

Figure 1 Normal Q-Q plot for different error distributions

It follows that if

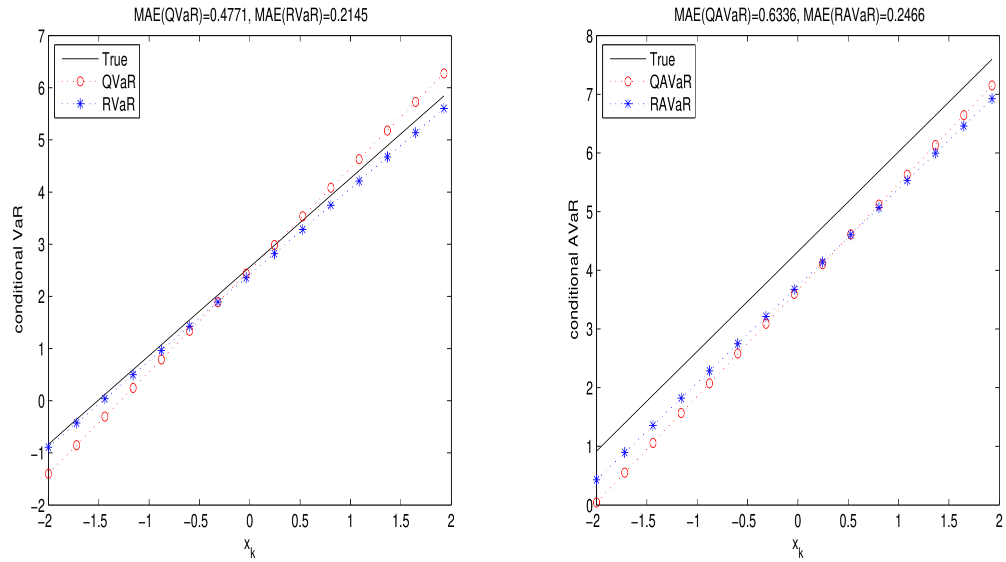
$$\sum_{j=1}^{r} \lambda_j \mathsf{V@R}_{\alpha_j}(\varepsilon) = 0, \tag{44}$$

then $(\beta_0^\star, \boldsymbol{\beta}^\star) = (\beta_0^*, \boldsymbol{\beta}^*)$. That is, $\check{\beta}_0 + \check{\boldsymbol{\beta}}^\mathsf{T}\boldsymbol{x}$ is a consistent estimator of $\sum_{j=1}^{r} \lambda_j \mathsf{V@R}_{\alpha_j}(Y|\boldsymbol{x})$. Consequently for λ_j and α_j given in (17), we can use $\check{\beta}_0 + \check{\boldsymbol{\beta}}^\mathsf{T}\boldsymbol{x}$ as an approximation of $\mathsf{AV@R}_\alpha(Y|\boldsymbol{x})$.

Asymptotics of the mixed quantile estimators are more involved. These asymptotics are discussed in Appendix C.

4. Simulation Study

To illustrate the performance of the considered estimators, we perform the Monte Carlo simulations where errors (innovations) in linear model (7) are generated from following different distributions; (1) Standard Normal (denoted as $N(0,1)$), (2) Student's t distribution with 3 degrees of freedom (denoted as $t(3)$), (3) Skewed Contaminated Normal where standard normal is contaminated with 20% $N(1,9)$ errors (denoted as $CN(1,9)$), (4) Log-Normal with parameter 0 and 1 (denoted as $LN(0,1)$). Note that error distributions (2)-(4) are heavy-tailed in contrast to the normal errors as shown in Figure 4. In fact, financial innovations often follow heavy-tailed distributions. We consider $\alpha = 0.9, 0.95, 0.99$, sample size $N = 500, 1000, 2000$ and $R = 500$ replications for each sample size. Conditional Value-at-Risk (VaR) and Average Value-at-Risk (AVaR) are estimated and compared



(a) Conditional VaR (b) Conditional AVaR

Figure 2 Conditional VaR and AVaR: True vs. Estimated (Errors$\sim CN(1,9)$, $\alpha = 0.95$, $N = 1000$)

with true (theoretical) values at given 500 test points x_k ($k = 1, 2, \ldots, 500$), which are equally spaced between -2 and 2 for each replication. Estimators obtained from different methods are computed; quantile based estimator (referred to as "QVaR") and LSR estimator (referred to as "RVaR") for the conditional VaR, mixed quantile estimator (referred to as "QAVaR") and LSR estimator (referred to as "RAVaR") for the conditional AVaR (as described in Section 2).

Figure 4 displays an example of estimation results where solid line is true (theoretical) VaR (AVaR), dash-circle line is QVaR (QAVaR), and dash-cross line is RVaR (RAVaR) given test points x_k. In this example, errors follow $CN(1,9)$, $\alpha = 0.95$ and $N = 1000$. In Figure 4-(a), RVaR estimates are closer to true VaR values as Mean Absolute Error (MAE) confirms (MAE(QVaR)=0.4771 vs. MAE(RVaR)=0.2145). Performance of both estimators are worse for AVaR, yet RAVaR estimates are still closer to true AVaR values than QAVaR (MAE(QAVaR)=0.6336 vs. MAE(RAVaR)=0.2466) as shown in Figure 4-(b).

To compare estimators under different error distributions, MAE (averaged over all test points) and variance of MAE (in parenthesis) across 500 replications are obtained as shown in Table 1. Regardless of the error distributions, RVaR (RAVaR) works better than QVaR (QAVaR); MAE and the variance of MAE are smaller. As we can expect, both estimators perform better for the

Table 1 MAE for different error distributions $\alpha = 0.95, N = 1000$ (averaged over all test points)

Error	QVaR	RVaR	QAVaR	RAVaR
$N(0,1)$	0.0762	0.0575	0.0990	0.0674
	(0.0037)	(0.0020)	(0.0058)	(0.0026)
$t(3)$	0.1758	0.1290	0.4255	0.3232
	(0.0188)	(0.0095)	(0.0808)	(0.0623)
$CN(1,9)$	0.3006	0.1955	0.3844	0.2311
	(0.0563)	(0.0225)	(0.0882)	(0.0316)
$LN(0,1)$	0.3905	0.2670	0.8957	0.6432
	(0.0959)	(0.0430)	(0.3896)	(0.2481)

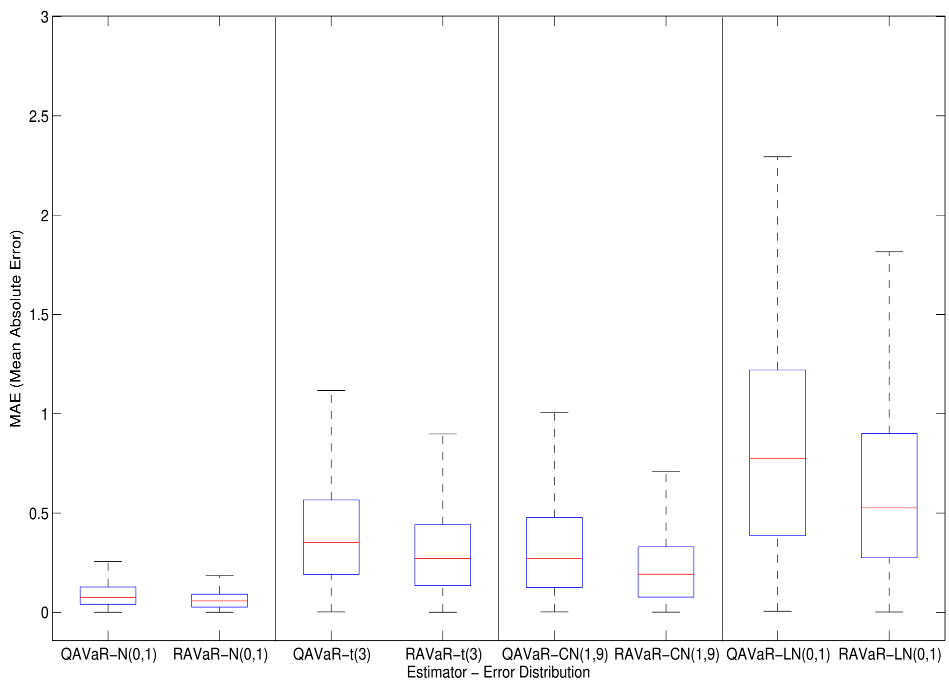


Figure 3 MAE for conditional AVaR given $x = 1.006$ under different error distributions ($\alpha = 0.95,\ N = 1000$)

conditional VaR than AVaR.

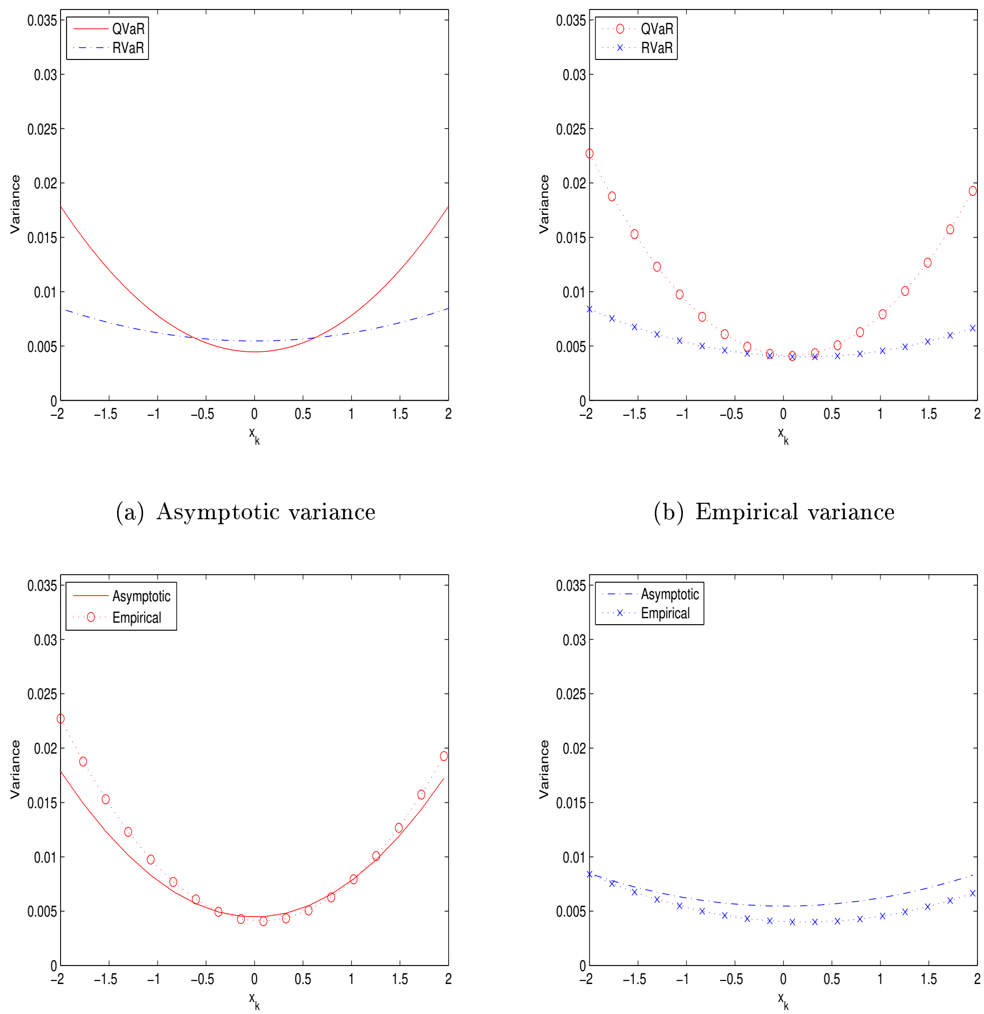
Figure 3 presents box-plots for both estimators (QAVaR and RAVaR) given $x = 1.006$ across 500 replications. Findings are similar to the one from Table 1; there are some evidence to suggest that RAVaR has smaller MAE than QAVaR. Also, RAVaR is more stable than QAVaR (MAE of QAVaR is more spread). Note that both estimators work better for normal distributions than other heavy-tailed distributions. We could observe the similar pattern for conditional VaR.

Table 2 MAE for different sample size N with $\alpha = 0.95$ (averaged over all test points)

Error	Estimator	$N=500$	$N=1000$	$N=2000$
$N(0,1)$	QVaR	0.1129	0.0762	0.0569
	RVaR	0.0849	0.0575	0.0418
	QAVaR	0.1390	0.0990	0.0737
	RAVaR	0.0992	0.0674	0.0498
$t(3)$	QVaR	0.2420	0.1758	0.1277
	RVaR	0.1785	0.1290	0.0942
	QAVaR	0.5385	0.4255	0.3207
	RAVaR	0.4517	0.3232	0.2085
$CN(1,9)$	QVaR	0.4322	0.3006	0.2180
	RVaR	0.2928	0.1955	0.1447
	QAVaR	0.5471	0.3844	0.2658
	RAVaR	0.3373	0.2311	0.1636
$LN(0,1)$	QVaR	0.5814	0.3905	0.2959
	RVaR	0.4095	0.2670	0.1975
	QAVaR	1.1986	0.8957	0.7275
	RAVaR	0.9503	0.6432	0.4754

Table 2 illustrates sample size effect on MAE of estimators. As expected, both estimators perform better as sample size increases. MAE of RVaR (RAVaR) is still smaller than that of QVaR (QAVaR) across all sample sizes.

Next, we obtain asymptotic variances (derived in Section 3) and compare that with empirical (finite sample) variances of both estimators. Figure 4 reports asymptotic and finite sample efficiencies of both estimators for the conditional VaR where $R = 500$, and error follows $N(0,1)$ (results are similar for other error distributions). In Figure 4-(a), we see that asymptotic variance of RVaR (dash-dot line) is smaller than that of QVaR (solid line) except at x_k near 0. In fact, asymptotic variance is affected by how far x_k is away from 0 (which is the mean of explanatory variable in the simulation); when x_k is further from the mean, the difference between asymptotic variances of both estimators is bigger. Figure 4-(b) provides empirical variance of both estimators across 500 replications. Empirical variance of RVaR is (equal or) smaller than that of QVaR at all x_k. Figure 4-(c) and Figure 4-(d) compare asymptotic variances to empirical variances of both estimators. It is clear that asymptotic variances are to provide a good approximation to the empirical ones for both estimators.



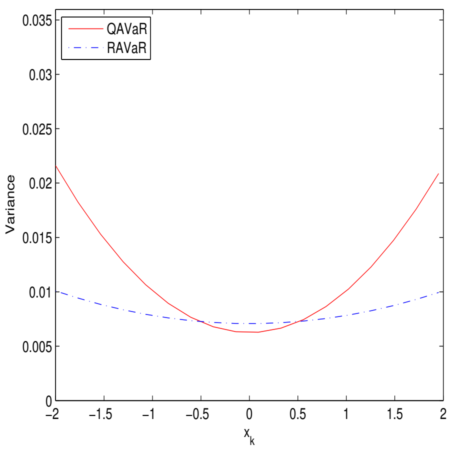
(a) Asymptotic variance (b) Empirical variance

(c) QVaR: Asymptotic vs. Empirical (d) RVaR: Asymptotic vs. Empirical

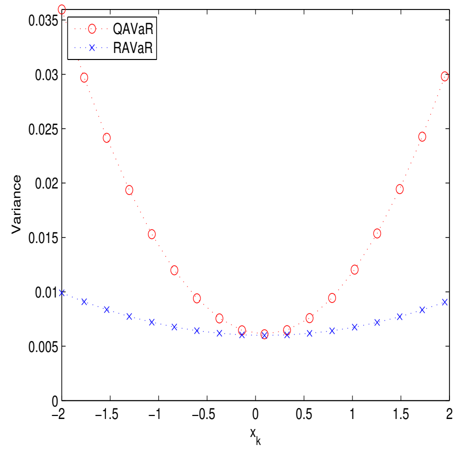
Figure 4 Conditional VaR: asymptotic and empirical variance (Error$\sim N(0,1)$, $\alpha = 0.95$, $N = 1000$, $R = 500$)

Figure 4 illustrates asymptotic and empirical variances of both estimators for AVaR. Insights obtained from the results are similar to the VaR case. However, Figure 4-(c) indicates that empirical variances of QAVaR are larger than asymptotic variances, especially when x_k is far from the mean. For this case, asymptotic efficiency of QAVaR may not very informative on its behavior in finite sample. Results are similar for other error distributions except $t(3)$. When the error follows $t(3)$, asymptotic (empirical) variances of QAVaR are smaller than that of RAVaR except when x_k is close to the boundary (as shown in Figure 4).

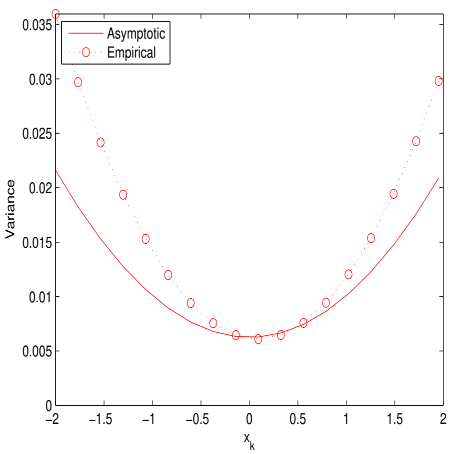
To further investigate the finite sample efficiencies and robustness of both estimators compared to the asymptotic ones, we provide empirical coverage probabilities (CP) of a two-sided 95% (nominal)



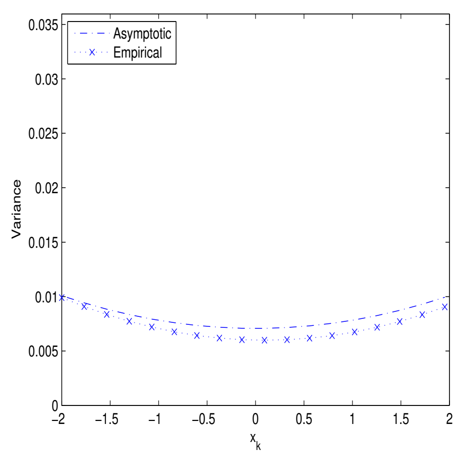
(a) Asymptotic variance



(b) Empirical variance



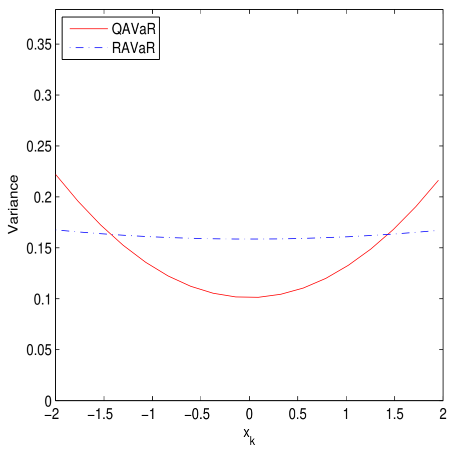
(c) QAVaR: Asymptotic vs. Empirical



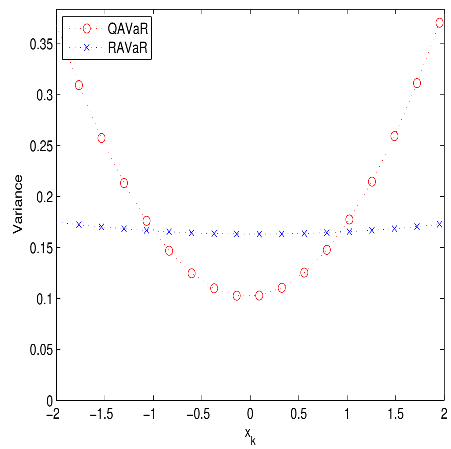
(d) RAVaR: Asymptotic vs. Empirical

Figure 5 Conditional AVaR: asymptotic and empirical variance (Error$\sim N(0,1)$, $\alpha = 0.95$, $N = 1000$, $R = 500$)

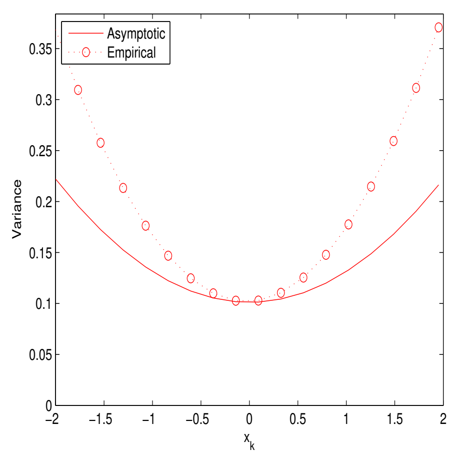
confidence interval (CI) in Table 3 (difference between CP and 0.95 is given in parentheses). For each replication, the empirical confidence interval is calculated from the sample version of asymptotic variance (when applied to the values of an observed sample of a given size). Then, for given x_k, the proportion of the 500 replications where the obtained confidence interval contains the true (theoretical) value is calculated, and these proportions are averaged across all test points. For $N(0,1)$ and $CN(1,9)$ error distributions, the resulting CP of RVaR (RAVaR) is very close to 0.95 while empirical CI for QVaR (QAVaR) under-covers (resulting CP is smaller than 0.95). For $t(3)$ and $LN(0,1)$ error distributions, CP of RVaR (RAVaR) drops, yet maintains somewhat adequate CP which is a lot better than CP of QVaR (QAVaR). CI of QAVaR under-covers seriously



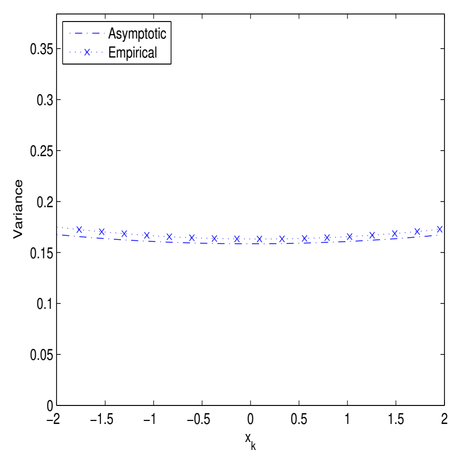
(a) Asymptotic variance



(b) Empirical variance



(c) QAVaR: Asymptotic vs. Empirical



(d) RAVaR: Asymptotic vs. Empirical

Figure 6 Conditional AVaR: asymptotic and empirical variance (Error$\sim t(3)$, $\alpha = 0.95$, $N = 1000$, $R = 500$)

(resulting CP is about 0.7) and this indicates QAVaR procedure may be very unstable and needs rather wider CI than other estimators to overcome its sensitivity. Note that RVaR (RAVaR) is more conservative than QVaR (QAVaR) regardless of the error distributions.

We could draw similar conclusions for other sample sizes and α values. That is, RVaR (RAVaR) performs better and provides stable results than QVaR (QAVaR) under different error distributions.

5. Illustrative Empirical Examples

In this section, we demonstrate considered methods to estimate conditional VaR and AVaR with real data; different financial asset classes. Let us first present an example of Credit Default Swap

Table 3 Coverage probability with $\alpha = 0.95, N = 1000$ (averaged over all test points)

Error	QVaR	RVaR	QAVaR	RAVaR
$N(0,1)$	0.9167	0.9551	0.8442	0.9552
	(0.0333)	(-0.0051)	(0.1058)	(-0.0052)
$t(3)$	0.9044	0.9269	0.7088	0.9080
	(0.0456)	(0.0231)	(0.2412)	(0.0420)
$CN(1,9)$	0.9262	0.9428	0.8824	0.9548
	(0.0238)	(0.0072)	(0.0676)	(-0.0048)
$LN(0,1)$	0.9185	0.9276	0.6930	0.9185
	(0.0315)	(0.0224)	(0.2570)	(0.0315)

(CDS). CDS is the most popular credit derivative in the rapidly growing credit markets (see FitchRatings 2006, for a detailed survey of the credit derivatives market). CDS contract provides insurance against a default by a particular company, a pool of companies, or sovereign entity. The rate of payments made per year by the buyer is known as the CDS spread (in basis points). We focus on the risk of CDS trading (long or short position) rather than on the use of a CDS to hedge credit risk. The CDS dataset obtained from Bloomberg consists of 1006 daily observations from January 2007 to January 2011. Let the dependent variable Y be daily percent change, $(Y(t+1) - Y(t))/Y(t) * 100$, of Bank of America Corp (NYSE:BAC) 5-year CDS spread, explanatory variables X_1 be daily return of BAC stock price, and X_2 be daily percent change of generic 5-year investment grade CDX spread (CDX.IG). We use the term "percent change" rather than return because the return of CDS contract is not same as the return of CDS spread (e.g., see O'Kane and Turnbull 2003, for an overview of CDS valuation models). Residuals obtained from this dataset are heavy-tailed distributed (similar to Figure 4-(b)).

Figure 7 shows estimated conditional VaR (RVaR) of BAC CDS spread percent change (result of QVaR is similar). Since one can take either short or long position, we present both tail risk with all values of α which ranges from 0.01 to 0.99; $\alpha < 0.5$ corresponds to the left tail (short position) and right tail (long position), otherwise. It is clear that RVaR of certain dates are much higher (lower) than normal level due to the different daily economic conditions reflected by BAC stock price and CDX spread. This indicates the specific (daily) economic conditions should be taken account for the accurate estimation of risk, and therefore emphasize the importance of conditional

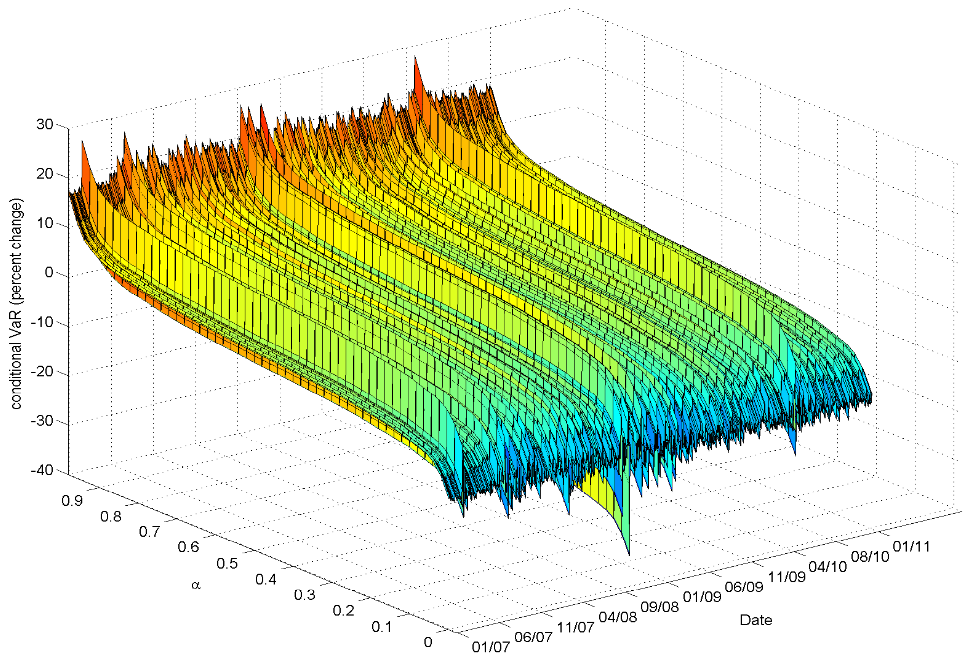


Figure 7 Estimated conditional VaR (RVaR) for BAC CDS spread percent change for $\alpha = 0.01, \ldots, 0.99$

risk measures. Note that given a specific date, estimated RVaR curve along the different α values is asymmetric since the distribution of CDS spread percent change is not symmetric.

To compare the prediction performance of both estimators, we forecast 603 one-day-ahead (tomorrow's) VaR (AVaR) given the current (today's) value of explanatory variables using a rolling window of the previous 403 days. Figure 8 presents forecasting results of QVaR and RVaR with $\alpha = 0.05$ on 603 out-of-sample. Both estimators show similar behaviors, but RVaR seems little more stable. Following ideas in McNeil and Frey (2000) and Leorato et al. (2010), "violation event" is said to occur whenever observed CDS spread percent change falls below the predicted VaR (we can find a few violation events from Figure 8). Also, the forecast error of AVaR is defined as the difference between the observed CDS spread percent change and the predicted AVaR under the violation event. By definition, the violation event probability should be close to α and the forecast error should be close to zero. Table 4 presents the prediction performance (violation event probability for VaR, mean and MAE of forecast error for AVaR in parenthesis) of both estimators for $\alpha = 0.01$ and 0.05. In-sample statistics show that both estimators fit the data well; the violation

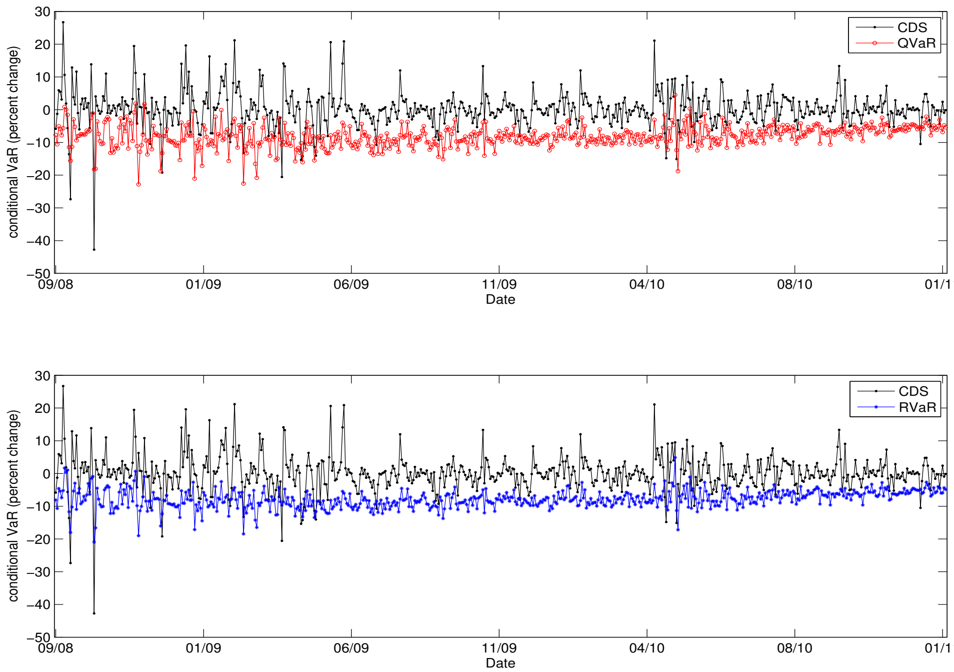


Figure 8 Risk prediction of BAC CDS: QVaR and RVaR (α=0.05)

Table 4 Risk prediction performance of BAC CDS

In-sample	α	Event(%)	Mean	MAE
QVaR(QAVaR)	0.01	0.9950	(0.1965)	(1.3118)
RVaR(RAVaR)	0.01	0.9950	(-0.8630)	(2.8183)
QVaR(QAVaR)	0.05	4.9751	(0.2287)	(2.5016)
RVaR(RAVaR)	0.05	4.9751	(-0.0269)	(2.8090)
Out-of-sample	α	Event(%)	Mean	MAE
QVaR(QAVaR)	0.01	0.8292	(1.4546)	(2.4421)
RVaR(RAVaR)	0.01	0.8292	(1.1052)	(4.0615)
QVaR(QAVaR)	0.05	3.6484	(1.3740)	(3.1099)
RVaR(RAVaR)	0.05	4.4776	(-0.3722)	(3.3681)

event probabilities are very close to α and forecast errors are very small. Out-of-sample performances of both estimators are very similar for $\alpha = 0.01$, even though the forecast errors increase a little compared to in-sample cases. For $\alpha = 0.05$, RVaR (RAVaR) seems perform better; event probabilities are closer to 0.05 and forecast errors are smaller.

Next, we apply considered methods to one of the US equities; International Business Machines Corp (NYSE). The dataset contains 1722 daily observation from December 2005 to December

Table 5 Risk prediction performance of IBM stock

In-sample	α	Event(%)	Mean	MAE
QVaR(QAVaR)	0.01	1.0180	(-0.1305)	(0.5727)
RVaR(RAVaR)	0.01	0.9397	(-0.3481)	(0.8926)
QVaR(QAVaR)	0.05	5.0117	(0.0468)	(1.0204)
RVaR(RAVaR)	0.05	4.9334	(-0.0225)	(1.1579)
Out-of-sample	α	Event(%)	Mean	MAE
QVaR(QAVaR)	0.01	2.3511	(0.6171)	(1.1028)
RVaR(RAVaR)	0.01	1.8809	(0.5023)	(0.6827)
QVaR(QAVaR)	0.05	6.7398	(0.4787)	(1.3086)
RVaR(RAVaR)	0.05	6.1129	(0.4778)	(1.2387)

2010. Let the dependent variable Y be the daily log return, 100*log(Y(t+1)/Y(t)), of IBM stock price, explanatory variables X_1 be the log return of S&P 500 index, and X_2 be the lagged log return. Similar to CDS example, we forecast 638 one-day-ahead (tomorrow's) VaR (AVaR) given the current (today's) value of explanatory variables using a rolling window of the previous 639 days. Residuals obtained from this dataset are heavy-tailed distributed. Table 4 compares the risk prediction performance of IBM stock return. Both estimators perform well for in-sample prediction. For out-of-sample prediction, both estimators behave similarly for $\alpha = 0.05$, but violation event probability is larger than 0.05. For $\alpha = 0.01$, RVaR (RAVaR) seems a bit better, but event probability exceeds 0.01.

Finally, we illustrate how crude oil price had impacted the US airlines' risk as we mentioned in Section 1. Crude oil prices had continued to rise since May 2007 and peaked all time high in July 2008, right before the brink of the US financial system collapse. We compare the movement of estimated VaR for three airline stocks given crude oil price change; Delta Airlines, Inc (NYSE:DAL), American Airlines, Inc (NYSE:AMR), and Southwest Airlines Co (NYSE:LUV). Figure 9 depicts RVaR movement with $\alpha = 0.05$ from May 2007 to July 2008 (QVaR shows similar patterns). For easy comparison, we standardize all units relative to the starting date. As we can see, crude oil price had jumped 150% during this time span. On the other hand, RVaR of LUV increased about 15% while that of AMR increased 120% and that of DAL increased 90% (in magnitude). In fact, different airlines have different strategies to hedge the risk on oil price fluctuations and this in turn affects the risk of airlines' stock movement. For example, Southwest Airlines is well known for

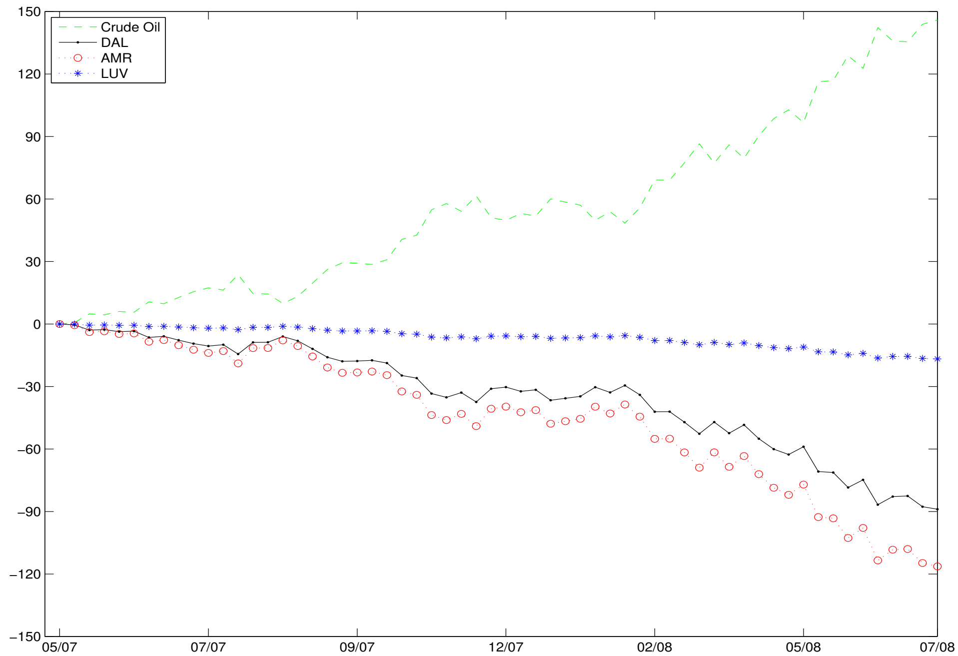


Figure 9 Airline equities: RVaR conditional on crude oil price (α=0.05)

hedging crude oil prices aggressively. On the other hand, Delta Airlines does little hedge against crude oil price, but operates a lot of international flights. American Airlines does not have strong hedging against crude oil price either, and operates less international flights than Delta Airlines. Our estimation results confirm the firm specific risk regarding crude oil price fluctuations.

6. Conclusions

Value-at-Risk and Average Value-at-Risk are widely used measures of financial risk. In order to accurately estimate risk measures, taking into account the specific economic conditions, we considered two estimation procedures for conditional risk measures; one is based on residual analysis of the standard least squares method (LSR estimator) and the other is based on mixed M-estimators (mixed quantile estimator). Large sample statistical inferences of both estimators are derived and compared. In addition, finite sample properties of both estimators are investigated and compared as well. Monte Carlo simulation results, under different error distributions, indicate that the LSR estimators perform better than their (mixed) quantiles counterparts. In general, MAE and asymptotic/empirical variance of the LSR estimators are smaller than that of quantile based estimators.

We also observe that asymptotic variance of estimators approximates the finite sample efficiencies well for reasonable sample sizes used in practice. However, we may need more samples to guarantee an acceptable efficiency of the quantile based estimator for Average Value-at-risk compared to other estimators. Prediction performances on the real data example suggest similar conclusions. In fact, residual based estimators can be calculated easily and therefore the LSR method could be implemented efficiently in practice. Moreover, LSR method can be easily applied to the general class of law invariant risk measures. In this study, we assume a static model with independent error distributions. Extension of considered estimation procedures incorporating different aspects of (dynamic) time series models could be an interesting topic for the further study.

Appendix A: Asymptotics for LSR Estimator of $\mathsf{V@R}_\alpha(Y|\boldsymbol{x})$

Suppose, for the sake of simplicity, that support of the distribution of $\boldsymbol{X}_i$ is bounded, i.e., $\boldsymbol{X}_i$ is bounded w.p.1. Since $N^{-1}\mathbb{X}^\mathsf{T}\mathbb{X}$ converges w.p.1 to $\boldsymbol{\Omega}$ and by (27), we have that

$$|\varepsilon_i - e_i| \le O_p(N^{-1}) \sum_{j=1}^{N} \varepsilon_j.$$

We can assume here that $\mathbb{E}[\varepsilon_i] = 0$, and hence $\sum_{j=1}^{N} \varepsilon_j = O_p(N^{1/2})$. It follows that

$$\left|\varepsilon_{(\lceil N\alpha \rceil)} - e_{(\lceil N\alpha \rceil)}\right| = O_p(N^{-1/2}). \tag{45}$$

Suppose now that the set of population α-quantiles is a singleton. Then $\hat{F}_\varepsilon^{-1}(\alpha)$ converges w.p.1 to the population quantile $F_\varepsilon^{-1}(\alpha) = \mathsf{V@R}_\alpha(\varepsilon)$, and hence by (45), we have that $e_{(\lceil N\alpha \rceil)}$ converges in probability to $F_\varepsilon^{-1}(\alpha)$. That is, $\widehat{\mathsf{V@R}}_\alpha(e)$ is a consistent estimator of $\mathsf{V@R}_\alpha(\varepsilon)$, and hence the estimator $\tilde{\beta}_0 + \boldsymbol{x}^\mathsf{T}\tilde{\boldsymbol{\beta}} + \widehat{\mathsf{V@R}}_\alpha(e)$ is a *consistent estimator* of $\mathsf{V@R}_\alpha(\boldsymbol{Y}|\boldsymbol{x})$.

Let us consider the asymptotic efficiency of the residual based $\mathsf{V@R}_\alpha$ estimator. It is known that $\tilde{\beta}_0 + \boldsymbol{x}^\mathsf{T}\tilde{\boldsymbol{\beta}}$ is an unbiased estimator of the true expected value $\beta_0 + \boldsymbol{x}^\mathsf{T}\boldsymbol{\beta}$ and $N^{1/2}\left[\tilde{\beta}_0 - \beta_0^* + \boldsymbol{x}^\mathsf{T}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)\right]$ converges in distribution to normal with zero mean and variance

$$\sigma^2[1; \boldsymbol{x}^\mathsf{T}]\boldsymbol{\Omega}^{-1}[1; \boldsymbol{x}^\mathsf{T}]^\mathsf{T}. \tag{46}$$

Also, $N^{1/2}\left(\varepsilon_{(\lceil N\alpha \rceil)} - \mathsf{V@R}_\alpha(\varepsilon)\right)$ converges in distribution to normal with zero mean and variance

$$\omega^2 := \frac{\alpha(1-\alpha)}{\left[f_\varepsilon\left(F_\varepsilon^{-1}(\alpha)\right)\right]^2}, \tag{47}$$

provided that distribution of ε has nonzero density $f_\varepsilon(\cdot)$ at the quantile $F_\varepsilon^{-1}(\alpha)$.

Let us also estimate the asymptotic variance of the right hand side of (27). We have that N times variance of the second term in the right hand side of (27) can be approximated by

$$\sigma^2 \mathbb{E}\big\{[1;\boldsymbol{X}_i^\mathsf{T}]\boldsymbol{\Omega}^{-1}[1;\boldsymbol{X}_i^\mathsf{T}]^\mathsf{T}\big\} = \sigma^2(k+1).$$

We also have that random vectors $(\tilde{\beta}_0, \tilde{\boldsymbol{\beta}})$ and $\boldsymbol{e}$ are uncorrelated. Therefore, if errors ε_i have normal distribution, then vectors $(\tilde{\beta}_0, \tilde{\boldsymbol{\beta}})$ and $\boldsymbol{e}$ have jointly a multivariate normal distribution and these vectors are independent. Consequently, $\tilde{\beta}_0 + \boldsymbol{x}^\mathsf{T}\tilde{\boldsymbol{\beta}}$ and $\widehat{\mathsf{V@R}}_\alpha(e)$ are independent. For not necessarily normal distribution, this independence holds asymptotically and thus asymptotically $\tilde{\beta}_0 + \boldsymbol{x}^\mathsf{T}\tilde{\boldsymbol{\beta}}$ and $\widehat{\mathsf{V@R}}_\alpha(e)$ are uncorrelated.

Now, we can calculate the asymptotic covariance of the corresponding terms $\big(\varepsilon_{(\lceil N\alpha\rceil)} - \mathsf{V@R}_\alpha(\varepsilon)\big)$ and $\big(\varepsilon_{(\lceil N\alpha\rceil)} - e_{(\lceil N\alpha\rceil)}\big)$ as $\frac{-\sigma^2\left(k+1\right)}{2}$. Thus, asymptotic variance of the residual based $\mathsf{V@R}_\alpha$ estimator can be approximated as

$$N^{-1}\left(\omega^2 + \sigma^2[1;\boldsymbol{x}^\mathsf{T}]\boldsymbol{\Omega}^{-1}[1;\boldsymbol{x}^\mathsf{T}]^\mathsf{T}\right). \tag{48}$$

Appendix B: Asymptotics for LSR Estimator of $\mathsf{AV@R}_\alpha(Y|\boldsymbol{x})$

The estimator $\widehat{\mathsf{AV@R}}_\alpha(e)$ can be compared with the corresponding random variable which is based on the errors instead of residuals

$$\begin{aligned} \widehat{\mathsf{AV@R}}_\alpha(\varepsilon) &:= \inf_{t\in\mathbb{R}}\left\{t + \tfrac{1}{(1-\alpha)N}\textstyle\sum_{i=1}^N[\varepsilon_i - t]_+\right\} \\ &= \widehat{\mathsf{V@R}}_\alpha(\varepsilon) + \tfrac{1}{(1-\alpha)N}\textstyle\sum_{i=1}^N\left[\varepsilon_i - \widehat{\mathsf{V@R}}_\alpha(\varepsilon)\right]_+ \\ &= \varepsilon_{(\lceil N\alpha\rceil)} + \tfrac{1}{(1-\alpha)N}\textstyle\sum_{i=\lceil N\alpha\rceil+1}^N\left[\varepsilon_{(i)} - \varepsilon_{(\lceil N\alpha\rceil)}\right]. \end{aligned} \tag{49}$$

Note that $\widehat{\mathsf{AV@R}}_\alpha(\varepsilon)$ is not an estimator since errors ε_i are unobservable.

By (45), we have that

$$\left|\widehat{\mathsf{V@R}}_\alpha(\varepsilon) - \widehat{\mathsf{V@R}}_\alpha(e)\right| = O_p(N^{-1/2}) \tag{50}$$

and it is known that $\widehat{\mathsf{AV@R}}_\alpha(\varepsilon)$ converges w.p.1 to $\mathsf{AV@R}_\alpha(\varepsilon)$ as $N\to\infty$, provided that ε has a finite first order moment. It follows that $\widehat{\mathsf{AV@R}}_\alpha(e)$ converges in probability to $\mathsf{AV@R}_\alpha(\varepsilon)$, and hence $\tilde{\beta}_0 + \boldsymbol{x}^\mathsf{T}\tilde{\boldsymbol{\beta}} + \widehat{\mathsf{AV@R}}_\alpha(e)$ is a consistent estimator of $\mathsf{AV@R}_\alpha(\boldsymbol{Y}|\boldsymbol{x})$.

Lets discuss asymptotic properties of the residual based $\mathsf{AV@R}_\alpha$ estimator. As it was pointed out in Appendix A, random vectors $(\tilde\beta_0, \tilde{\boldsymbol\beta})$ and $\boldsymbol{e}$ are uncorrelated, and hence asymptotically $\tilde\beta_0 + \boldsymbol{x}^\mathsf{T}\tilde{\boldsymbol\beta}$ and $\widehat{\mathsf{AV@R}}_\alpha(e)$ are independent and hence uncorrelated. Assuming that α-quantile of $F_\varepsilon(\cdot)$ is unique, we have by Delta theorem

$$\widehat{\mathsf{AV@R}}_\alpha(e) = \mathsf{V@R}_\alpha(\varepsilon) + (1-\alpha)^{-1}N^{-1}\sum_{i=1}^{N}[e_i - \mathsf{V@R}_\alpha(\varepsilon)]_+ + o_p(N^{-1/2}) \tag{51}$$

and

$$\widehat{\mathsf{AV@R}}_\alpha(\varepsilon) = \mathsf{V@R}_\alpha(\varepsilon) + (1-\alpha)^{-1}N^{-1}\sum_{i=1}^{N}[\varepsilon_i - \mathsf{V@R}_\alpha(\varepsilon)]_+ + o_p(N^{-1/2}). \tag{52}$$

Equation (52) leads to the following asymptotic result (cf. Trindade et al. 2007, Shapiro et al. 2009, section 6.5.1)

$$N^{1/2}\left[\widehat{\mathsf{AV@R}}_\alpha(\varepsilon) - \mathsf{AV@R}_\alpha(\varepsilon)\right] \xrightarrow{\mathcal{D}} \mathcal{N}(0,\gamma^2), \tag{53}$$

where $\gamma^2 = (1-\alpha)^{-2}\mathrm{Var}\big([\varepsilon - \mathsf{V@R}_\alpha(\varepsilon)]_+\big)$. Moreover, if distribution of ε has nonzero density $f_\varepsilon(\cdot)$ at $\mathsf{V@R}_\alpha(\varepsilon)$, then

$$\mathbb{E}\big[\widehat{\mathsf{AV@R}}_\alpha(\varepsilon)\big] - \mathsf{AV@R}_\alpha(\varepsilon) = -\frac{1-\alpha}{2Nf_\varepsilon(\mathsf{V@R}_\alpha(\varepsilon))} + o(N^{-1}). \tag{54}$$

From the equation (51) and (52), the asymptotic variance of $\big(\widehat{\mathsf{AV@R}}_\alpha(\varepsilon) - \widehat{\mathsf{AV@R}}_\alpha(e)\big)$ can be bounded by $(1-\alpha)^{-1}N^{-2}\sigma^2\big(k+1\big)$ and we can approximate the asymptotic covariance of the corresponding terms, $\Big(\widehat{\mathsf{AV@R}}_\alpha(\varepsilon) - \mathsf{AV@R}_\alpha(\varepsilon)\Big)$ and $\Big(\widehat{\mathsf{AV@R}}_\alpha(\varepsilon) - \widehat{\mathsf{AV@R}}_\alpha(e)\Big)$ as $\frac{-(1-\alpha)^{-1}N^{-2}\sigma^2\big(k+1\big)}{2}$. Thus, asymptotic variance of the residual based $\mathsf{AV@R}_\alpha$ estimator can be approximated as

$$N^{-1}\left(\gamma^2 + \sigma^2[1;\boldsymbol{x}^\mathsf{T}]\boldsymbol{\Omega}^{-1}[1;\boldsymbol{x}^\mathsf{T}]^\mathsf{T}\right). \tag{55}$$

Appendix C: Asymptotics for the Mixed Quantile Estimator

It is possible to derive asymptotics of the mixed quantile estimator. For the sake of simplicity, let us start with a sample estimate of $S(X)$, with λ_j and α_j, $j = 1,\dots,r$, given in (17). That is, let $X_1,\dots,X_N$ be an iid sample (data) of the random variable X, and $X_{(1)} \le \dots \le X_{(N)}$ be the corresponding order statistics. Then the corresponding sample estimate is obtained by replacing the true distribution F of X by its empirical estimate $\hat{F}$. Consequently, $(1-\alpha)^{-1}S(X)$ is estimated by

$$(1-\alpha)^{-1}\sum_{j=1}^{r}\lambda_j \hat{F}^{-1}(\alpha_j) = \frac{1}{r}\sum_{j=1}^{r} X_{(\lceil N\alpha_j \rceil)}. \tag{56}$$

This can be compared with the following estimator of $\mathsf{AV@R}_\alpha(X)$ based on sample version of (2):

$$\begin{array}{c} X_{(\lceil N\alpha\rceil)} + \frac{1}{(1-\alpha)N}\sum_{i=\lceil N\alpha\rceil+1}^{N}\left[X_{(i)} - X_{(\lceil N\alpha\rceil)}\right] = \\ \left(1 - \frac{N-\lceil N\alpha\rceil}{(1-\alpha)N}\right) X_{(\lceil N\alpha\rceil)} + \frac{1}{(1-\alpha)N}\sum_{i=\lceil N\alpha\rceil+1}^{N} X_{(i)}. \end{array} \tag{57}$$

Assuming that $N\alpha$ is an integer and taking $r := (1-\alpha)N$, we obtain that the right hand sides of (56) and (57) are the same.

Asymptotic variance of the mixed quantile estimator can be calculated as follows. Consider problem (43). The optimal solution of that problem is $\boldsymbol{\beta}^\star = \boldsymbol{\beta}^*$,

$$\eta_j^\star = \beta_0^* + \mathsf{V@R}_{\alpha_j}(\varepsilon) = \beta_0^* + F_\varepsilon^{-1}(\alpha_j),\ j = 1, ..., r,$$

and $\beta_0^\star = \sum_{j=1}^{r}\lambda_j \eta_j^\star = \beta_0^*$. Assume that ε has continuous distribution with cdf $F_\varepsilon(\cdot)$ and density function $f_\varepsilon(\cdot)$. Then conditional on $\boldsymbol{X}$, the asymptotic covariance matrix of the corresponding sample estimator $(\check{\boldsymbol{\beta}}, \check{\boldsymbol{\eta}})$ of $(\boldsymbol{\beta}^\star, \boldsymbol{\eta}^\star)$ is $N^{-1}\boldsymbol{H}^{-1}\boldsymbol{\Sigma}\boldsymbol{H}^{-1}$, where $\boldsymbol{H}$ is the Hessian matrix of second order partial derivatives of $\mathbb{E}\left[\sum_{j=1}^{r}\psi_{\alpha_j}(\varepsilon + \beta_0^* - \eta_j + (\boldsymbol{\beta}^* - \boldsymbol{\beta})^\mathsf{T}\boldsymbol{X})\right]$ at the point $(\boldsymbol{\beta}^\star, \boldsymbol{\eta}^\star)$, and $\boldsymbol{\Sigma}$ is the covariance matrix of the random vector

$$\boldsymbol{Z} := \sum_{j=1}^{r}\nabla\psi_{\alpha_j}\left(\varepsilon + \beta_0^* - \eta_j + (\boldsymbol{\beta}^* - \boldsymbol{\beta})^\mathsf{T}\boldsymbol{X}\right),$$

where the gradients are taken with respect to $(\boldsymbol{\beta}, \boldsymbol{\eta})$ at $(\boldsymbol{\beta}, \boldsymbol{\eta}) = (\boldsymbol{\beta}^\star, \boldsymbol{\eta}^\star)$ (e.g., Shapiro 1989). We have

$$\sum_{j=1}^{r}\nabla_{\boldsymbol{\beta}}\psi_{\alpha_j}\left(\varepsilon + \beta_0^* - \eta_j + (\boldsymbol{\beta}^* - \boldsymbol{\beta})^\mathsf{T}\boldsymbol{X}\right) = -\left(\sum_{j=1}^{r}\psi'_{\alpha_j}\left(\varepsilon + \beta_0^* - \eta_j + (\boldsymbol{\beta}^* - \boldsymbol{\beta})^\mathsf{T}\boldsymbol{X}\right)\right)\boldsymbol{X},$$

$$\nabla_{\eta_j}\psi_{\alpha_j}\left(\varepsilon + \beta_0^* - \eta_j + (\boldsymbol{\beta}^* - \boldsymbol{\beta})^\mathsf{T}\boldsymbol{X}\right) = -\psi'_{\alpha_j}\left(\varepsilon + \beta_0^* - \eta_j + (\boldsymbol{\beta}^* - \boldsymbol{\beta})^\mathsf{T}\boldsymbol{X}\right),$$

with $\psi'_{\alpha_j}(\cdot)$ is given in (39).

Note that $\mathbb{E}[\psi'_{\alpha_j}(\varepsilon - F_\varepsilon^{-1}(\alpha_j)] = 0$, $j = 1, ..., r$, (see (40)), and hence $\mathbb{E}[\boldsymbol{Z}] = 0$. Then $\boldsymbol{\Sigma} =$

$\mathbb{E}\left[\boldsymbol{Z}\boldsymbol{Z}^{\top}\right]$ and we can compute $\boldsymbol{\Sigma} = \begin{bmatrix} \kappa\mathbb{E}\left[\boldsymbol{X}\boldsymbol{X}^{\top}\right] & \boldsymbol{\Psi} \\ \boldsymbol{\Psi}^{\top} & \boldsymbol{\Delta} \end{bmatrix}$, where $\kappa = \mathbb{E}\left\{\left[\sum_{j=1}^{r} \psi'_{\alpha_j}\left(\varepsilon - F_{\varepsilon}^{-1}(\alpha_j)\right)\right]^2\right\}$, $\boldsymbol{\Psi} = [\boldsymbol{\Psi}_1, ..., \boldsymbol{\Psi}_r]$ with

$$\boldsymbol{\Psi}_j = \mathbb{E}\left[\left(\sum_{i=1}^{r} \psi'_{\alpha_i}\left(\varepsilon - F_{\varepsilon}^{-1}(\alpha_i)\right)\right)\psi'_{\alpha_j}\left(\varepsilon - F_{\varepsilon}^{-1}(\alpha_j)\right)\boldsymbol{X}\right], \, j = 1, ..., r,$$

and $\Delta_{ij} = \mathbb{E}\left[\psi'_{\alpha_i}\left(\varepsilon - F_{\varepsilon}^{-1}(\alpha_i)\right)\psi'_{\alpha_j}\left(\varepsilon - F_{\varepsilon}^{-1}(\alpha_j)\right)\right]$, $i, j = 1, ..., r$.

The Hessian matrix $\boldsymbol{H}$ can be computed as $\boldsymbol{H} = \begin{bmatrix} \gamma\mathbb{E}\left[\boldsymbol{X}\boldsymbol{X}^{\top}\right] & \boldsymbol{F} \\ \boldsymbol{F}^{\top} & \boldsymbol{D} \end{bmatrix}$, where $\gamma = \sum_{j=1}^{r} \gamma_j$ with

$$\begin{aligned} \gamma_j &= \frac{\partial\mathbb{E}\left[\psi'_{\alpha_j}(\varepsilon + \beta_0^* - \eta_j^* + t)\right]}{\partial t}\Big|_{t=0} \\ &= \frac{\partial\left[\alpha_j(1 - F_{\varepsilon}(F_{\varepsilon}^{-1}(\alpha_j) - t)) + (\alpha_j - 1)F_{\varepsilon}(F_{\varepsilon}^{-1}(\alpha_j) - t)\right]}{\partial t}\Big|_{t=0} \\ &= \alpha_j f_{\varepsilon}(F_{\varepsilon}^{-1}(\alpha_j)) - (1 - \alpha_j) f_{\varepsilon}(F_{\varepsilon}^{-1}(\alpha_j)) = f_{\varepsilon}(F_{\varepsilon}^{-1}(\alpha_j)), \, j = 1, ..., r, \end{aligned}$$

$\boldsymbol{F} = [\boldsymbol{F}_1, ..., \boldsymbol{F}_r]$ with $\boldsymbol{F}_j = \gamma_j\mathbb{E}[\boldsymbol{X}]$, $j = 1, ..., r$, and $\boldsymbol{D} = \mathrm{diag}(\gamma_1, ..., \gamma_r)$.

Since $\check{\beta}_0 = \boldsymbol{\lambda}^{\top}\check{\boldsymbol{\eta}}$, we have that $\check{\beta}_0 + \check{\boldsymbol{\beta}}^{\top}\boldsymbol{x} = [\boldsymbol{x}^{\top}; \boldsymbol{\lambda}^{\top}][\check{\boldsymbol{\beta}}; \check{\boldsymbol{\eta}}]$, and hence the asymptotic variance of $\check{\beta}_0 + \check{\boldsymbol{\beta}}_0^{\top}\boldsymbol{x}$ is given by $N^{-1}[\boldsymbol{x}^{\top}; \boldsymbol{\lambda}^{\top}]\boldsymbol{H}^{-1}\boldsymbol{\Sigma}\boldsymbol{H}^{-1}[\boldsymbol{x}; \boldsymbol{\lambda}]$.

References

Acerbi, C. 2002. Spectral measures of risk: a coherent representation of subjective risk aversion. *Journal of Banking & Finance* **26** 1505–1518.

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

Berkowitz, J, M Pritsker, M Gibson, H Zhou. 2002. How accurate are value-at-risk models at commercial banks. *Journal of Finance* **57** 1093–1111.

Bluhm, Christian, Ludger Overbeck, Christoph Wagner. 2002. *An Introduction to Credit Risk Modeling (Chapman & Hall/Crc Financial Mathematics Series)*. 1st ed. Chapman and Hall/CRC.

Cai, Z., X. Wang. 2008. Nonparametric estimation of conditional var and expected shortfall. *Journal of Econometrics* **147**(1) 120–130.

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

Chernozhukov, V., L. Umantsev. 2001. Conditional value-at-risk: Aspects of modeling and estimation. *Empirical Economics* **26**(1) 271–292.

Detlefsen, K., G. Scandolo. 2005. Conditional and dynamic convex risk measures. *Finance and Stochastics* .

Duffie, D., K. J. Singleton. 2003. *Credit Risk: Pricing, Measurement, and Management*. Princeton, Princeton University Press.

Engle, R. F., S Manganelli. 2004. Caviar: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics* **22** 367–381.

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

FitchRatings. 2006. Global credit derivatives survey.

Föllmer, H., A. Schied. 2011. *Stochastic finance: An introduction in discrete time*. Walter de Gruyter & Co., Berlin.

Frey, R., A. J. McNeil. 2002. Var and expected shortfall in portfolios of dependent credit risks: Conceptual and practical insights. *Journal of Banking & Finance* **26**(7) 1317–1334.

Gaglianone, W., L. Lima, O. Linton, Smith D. 2011. Var and expected shortfall in portfolios of dependent credit risks: Conceptual and practical insights. *Journal of Business & Economic Statistics* **29** 150–160.

Giesecke, K., S. Weber. 2008. Measuring the risk of large losses. *Journal of Investment Management* **6**(4) 1–15.

Gneiting, T. 2009. Evaluating point forecasts. Institut für Angewandte Mathematik, Universitpreprint, preprint.

Huber, P. J. 1981. *Robust Statistics*. Wiley, New York.

Jackson, Patricia, William Perraudin. 2000. Regulatory implications of credit risk modelling. *Journal of Banking & Finance* **24**(1-2) 1–14.

Jorion, P. 2003. *Financial Risk Manager Handbook*. 2nd ed. Wiley, New York.

Koenker, R. 2005. *Quantile Regression*. Cambridge University Press, Cambridge, UK.

Leorato, S., F. Peracchi, A. V. Tanase. 2010. Asymptotically effcient estimation of the conditional expected shortfall. EIEF Working Papers Series 1013, Einaudi Institute for Economic and Finance (EIEF).

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

O'Kane, D., S. Turnbull. 2003. Valuation of credit default swaps. Quantitative credit research quarterly, Lehman Brothers.

Peracchi, F., A. V. Tanase. 2008. On estimating the conditional expected shortfall. *Applied Stochastic Models in Business and Industry* **24** 471493.

Pflug, G., W. Römisch. 2007. *Modeling, Measuring and Managing Risk*. World Scientific Publishing Co., London.

Pflug, G., N. Wozabal. 2010. Asymptotic distribution of law-invariant risk functionals. *Finance and Stochastics* **14** 397–418.

Rockafellar, R. T., S. Uryasev. 2002. Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance* **26**(7) 1443–1471.

Rockafellar, R. T., S. Uryasev, M. Zabarankin. 2008. Risk tuning with generalized linear regression. *Mathematics of Operations Research* **33** 712–729.

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

Scaillet, O. 2005. Nonparametric estimation of conditional expected shortfall. *Insurance and Risk Management Journal* **72** 639–660.

Shapiro, A. 1989. Asymptotic properties of statistical estimators in stochastic programming. *Annals of Statistics* **17** 841–858.

Shapiro, A., D. Dentcheva, A. Ruszczynski. 2009. *Lectures on Stochastic Programming: Modeling and Theory*. SIAM, Philadelphia.

Trindade, A., S. Uryasev, A. Shapiro, G. Zrazhevsky. 2007. Financial prediction with constrained tail risk. *Journal of Banking & Finance* **31** 3524–3538.

Weber, S. 2006. Distribution-invariant risk measures, information, and dynamic consistency. *Mathematical Finance* **16**(2) 419–441.

Wozabal, D., N. Wozabal. 2009. Asymptotic consistency of risk functionals. *Journal of Nonparametric Statistics* **21**(8) 977–990.

Zhu, S., M. Fukushima. 2009. Worst-case conditional value-at-risk with application to robust portfolio management. *Operations Research* **57**(5) 1155–1168.