



Financial prediction with constrained tail risk

A. Alexandre Trindade ^a, Stan Uryasev ^{b,*},
Alexander Shapiro ^c, Grigory Zrazhevsky ^b

^a *Department of Statistics, University of Florida, P.O. Box 118545, Gainesville, FL 32611, USA*

^b *Department of Industrial and Systems Engineering, University of Florida,
P.O. Box 116595, Gainesville, FL 32611, USA*

^c *School of Industrial and Systems Engineering, Georgia Institute of Technology,
765 Ferst Drive NW, Atlanta, GA 30332, USA*

Available online 29 April 2007

Abstract

A new class of asymmetric loss functions derived from the least absolute deviations or least squares loss with a constraint on the mean of one tail of the residual error distribution, is introduced for analyzing financial data. Motivated by risk management principles, the primary intent is to provide “cautious” forecasts under uncertainty. The net effect on fitted models is to shape the residuals so that on average only a prespecified proportion of predictions tend to fall above or below a desired threshold. The loss functions are reformulated as objective functions in the context of parameter estimation for linear regression models, and it is demonstrated how optimization can be implemented via linear programming. The method is a competitor of quantile regression, but is more flexible and broader in scope. An application is illustrated on prediction of NDX and SPX index returns data, while controlling the magnitude of a fraction of worst losses.

© 2007 Elsevier B.V. All rights reserved.

JEL classification: C22; C51; C53; C61

Keywords: Risk measure; Value-at-risk; Asymmetric loss; Quantile regression; Constrained regression

* Corresponding author. Tel.: +1 352 392 1464x2023; fax: +1 352 392 3537.
E-mail address: uryasev@ufl.edu (S. Uryasev).

1. Introduction

In today's increasingly volatile financial markets, the ability to make "safe" predictions regarding future values can be a highly-valued skill in the eyes of the cautious investor. For example, it may be advantageous to obtain good predictions in such a way so as to also control, on average, some fraction of the "worst cases". Traditional econometric forecasts with their focus on expected values and associated symmetric confidence bands, indicate little else beyond the fact that future outcomes are as likely to be above as below the predicted mean. In this context, one typically draws inferences by minimizing either the *absolute value* (ℓ_1 -norm) or *squared value* (ℓ_2 -norm) of some appropriate vector of residual errors. From a risk management point of view therefore, the heavy-tailed nature of financial data and the markedly different attitudes toward over and under-prediction prevalent with many investors, render the symmetric absolute error and squared error loss inadequate objective functions for econometric model-fitting.

The concepts of asymmetric loss and asymmetric risk, although not new, have been receiving renewed attention in recent years, particularly in the econometrics literature. Christoffersen and Diebold (1997) study the optimal prediction problem under general loss structures and characterize the optimal predictor. There has been a steady increase in research activity in *quantile regression* since its introduction by Koenker and Bassett (1978), culminating in the monograph by Koenker (2005). This criterion extends the idea of ℓ_1 -norm or median-based regression to an arbitrary quantile. Rockafellar et al. (2006a,b) use various asymmetric *deviation measures* as objective functions in the context of linear regression and portfolio optimization problems.

In some instances it may be desirable that the loss function be asymmetric in its treatment of the errors, by not only penalizing positive and negative residuals differently, but by also taking their magnitude into account. In this paper, we introduce an approach that allows for direct and asymmetric control over the distribution of residual errors, by constraining one of its tail means not to exceed some prespecified value. In a model-based setting, this approach can be viewed as a shaping of the residual distribution so that on average only a prespecified proportion of estimates or predictions tend to fall above or below (depending on the tail being constrained) a desired threshold. These loss functions, formulated as objective functions in an empirical setting, can be efficiently optimized via linear programming techniques.

In order to assess the tail risk in question, we propose to use Value-at-Risk (VaR), and especially its increasingly popular variant, Conditional Value-at-Risk (CVaR). For a given probability level α of cumulative losses, the former is the value beyond which higher losses only occur with probability $1 - \alpha$, while the latter is the average of this $1 - \alpha$ fraction of worst losses. The conceptual simplicity of VaR has resulted in its widespread incorporation into standard banking regulation, where it is imperative to assess downside risk. Various undesirable conceptual and mathematical properties inherent in VaR, led Artzner et al. (1997, 1999) to introduce a set of axioms for the proper construction of risk measures. One of these is the notion of a *coherent* risk measure, an example of which is CVaR (Rockafellar and Uryasev, 2000, 2002), constituting a computationally attractive alternative to VaR. A thorough axiomatic treatment of such *static* risk measures can also be found in Föllmer and Schied (2004, (Chapter 4)).

These risk measures can be used alone or in conjunction with objective functions that characterize the optimality of the desired solution. For example, one might wish to

allocate resources to a set of financial instruments in such a way so as to minimize an α -level CVaR risk, or maximize the expected return while constraining the CVaR risk to be no greater than some specified value. Palmquist et al. (2002), demonstrate applications of financial portfolio optimization with CVaR constraints on the potential losses. Rockafellar and Uryasev (2002) extend the definition of CVaR to encompass general loss distribution functions, and use CVaR constraints to *shape* the loss distribution in portfolio replication problems, thereby improving out-of-sample performance. Similarly, Krokmal et al. (2002) study the out-of-sample performance of portfolios of hedge funds by introducing constraints on a variety of risk measures, among them VaR and CVaR. Rockafellar and Uryasev (2000), provide a case study of portfolio optimization by minimizing CVaR, and introduce a different definition of VaR and CVaR that enables their implementation in computationally efficient ways. Other recent noteworthy works on portfolio optimization under downside risk constraints include Basak and Shapiro (2001), Gabih et al. (2005) and Gundel and Weber (2005).

This paper shows that similar procedures can be used for constructing numerical percentile bounds for the residual error distribution in linear regression, thus shaping the resulting fitted response surfaces to control for under or over-performance. In particular, we seek to make “cautious” predictions, in the sense that the magnitude of a specified fraction of worst predictions is bounded *a priori*. After introducing VaR and CVaR and discussing their asymptotic properties in Section 2, we show in Section 3 how this control can be accomplished via CVaR-constrained ℓ_1 and ℓ_2 loss functions in the context of a linear regression model. With VaR constraints, a competing approach to our method is quantile regression, which seeks to estimate a percentile of the response surface. Hence another way to view our CVaR-constrained loss functions is as an estimated CVaR response surface. This connection, as well as an empirical formulation of the loss functions in terms of objective functions in a linear regression setting, is made in Section 4. We conclude in Section 5 by applying the methodology to forecasting NDX and SPX index returns. As practitioners increasingly opt for the use of CVaR, we foresee an important role for our methodology in future risk management practice.

2. Risk measures: VaR and CVaR

Financial *risk measures* provide a useful class of objective functions with a wide spectrum of applications. Significant progress has recently been made in understanding their properties. Throughout our discussion we consider only *static* risk measures, that is they are not viewed as evolving over time even though the data itself may be of a temporal nature.

To introduce the concepts, let Z be a random variable of interest. For example, Z can represent the cost or loss associated with a class of decisions parameterized by a (finite) number of parameters. For the moment we suppress dependence of Z on the parameter vector, and denote by $F(z) = \text{Prob}(Z \leq z)$ the cumulative distribution function (cdf) of Z . For $\alpha \in (0, 1)$, the α -quantile of Z is defined to be the smallest value $F^{-1}(\alpha)$ such that with probability α the cost Z will not exceed $F^{-1}(\alpha)$, i.e.,

$$F^{-1}(\alpha) := \inf\{z : F(z) \geq \alpha\}. \quad (1)$$

In the context of monetary cost (loss), the quantile $F^{-1}(\alpha)$ is a measure of risk commonly used in the finance industry, known as *Value-at-Risk* (VaR) and denoted $\text{VaR}_\alpha(Z) :=$

$F^{-1}(\alpha)$. A risk measure then, evaluates the overall seriousness of possible losses associated with Z .

If Value-at-Risk stands for the maximum loss (damage, incurred cost, etc.) that can be exceeded only in $(1 - \alpha)100\%$ of cases, then *Conditional Value-at-Risk* (CVaR), denoted by $\text{CVaR}_\alpha(Z)$, can be thought of as the conditional expectation of losses that exceed the $\text{VaR}_\alpha(Z)$ level. In case Z has a continuous distribution, $\text{CVaR}_\alpha(Z)$ is given by the expectation of the right $(1 - \alpha)$ -tail of the cost distribution, i.e.,

$$\text{CVaR}_\alpha(Z) = \mathbb{E}[Z|Z \geq \text{VaR}_\alpha(Z)], \tag{2}$$

whereas for general distributions $\text{CVaR}_\alpha(Z)$ is defined as a convex combination of $\text{VaR}_\alpha(Z)$ and the conditional expectation of losses strictly exceeding the $\text{CVaR}_\alpha(Z)$ level. As noted in Föllmer and Schied (2004, (Chapter 4)), this can be succinctly expressed as

$$\text{CVaR}_\alpha(Z) = \frac{1}{1 - \alpha} \int_\alpha^1 \text{VaR}_\gamma(Z) d\gamma.$$

Also, and of more interest to our work, Pflug (2000) showed that $\text{CVaR}_\alpha(Z)$ can be viewed as the solution of an optimization problem,

$$\text{CVaR}_\alpha(Z) = \inf_{t \in \mathbb{R}} \left\{ t + \frac{1}{1 - \alpha} \mathbb{E}[Z - t]^+ \right\}, \tag{3}$$

where $[a]^+ := \max\{0, a\}$ denotes the positive part of $a \in \mathbb{R}$.

For continuous distributions CVaR is known variously in the financial literature as conditional tail expectation, tail VaR, expected shortfall, and mean excess loss. The term Conditional Value-at-Risk was introduced by Rockafellar and Uryasev (2000). Acerbi and Tasche (2002) established relationships of CVaR with the expected shortfall risk measures, and investigated various mathematical properties thereof. A rigorous axiomatic treatment of these static risk measures and associated variants can be found in Föllmer and Schied (2004, (Chapter 4)).

Suppose now that we have an independent identically distributed (iid) sample Z_1, \dots, Z_n of n realizations of the random variable Z . The corresponding empirical cdf is defined as

$$\hat{F}_n(z) := \frac{\#(Z_i \leq z)}{n}, \tag{4}$$

where $\#(Z_i \leq z)$ denotes the number of times that $Z_i = 1, \dots, n$, is less than or equal to z . The sample α -quantile is given by $\hat{q}_n := \hat{F}_n^{-1}(\alpha)$ and is viewed as an estimator of $q := \text{VaR}_\alpha(Z)$.

Remark 1. Let $r := \sup\{t : F(t) \leq \alpha\}$ be the respective right side quantile. It is well known that \hat{q}_n is a consistent estimator in the sense that \hat{q}_n converges with probability one (w.p.1) to the interval $[q, r]$ as n tends to infinity. If $r = q$ it is also known that \hat{q}_n is asymptotically normal provided that the distribution of Z has density $f(q) = [dF(z)/dz]_z = q$ at q . That is, $n^{1/2}(\hat{q}_n - q)$ converges in distribution to a normal $N(0, \alpha(1 - \alpha)/f(q)^2)$.

It is natural to estimate $\theta^* := \text{CVaR}_\alpha(Z)$ by replacing the expectation in the right-hand side of (3) with the corresponding sample average. Note that $\mathbb{E}_{\hat{F}_n} [Z - t]^+ = \sum_{i=1}^n [Z_i - t]^+ / n$, where $\mathbb{E}_{\hat{F}_n}$ means that the expectation is taken with respect to the empirical cdf \hat{F}_n . This suggests

$$\hat{\theta}_n := \inf_{t \in \mathbb{R}} \left\{ t + \frac{1}{n(1-\alpha)} \sum_{i=1}^n [Z_i - t]^+ \right\} \tag{5}$$

as an estimator of θ^* . In order to discuss the asymptotic behavior of $\hat{\theta}_n$, consider the function

$$h_\alpha(z, t) := t + \frac{1}{1-\alpha} [z - t]^+. \tag{6}$$

Note that we have, $\theta^* = \inf_{t \in \mathbb{R}} \mathbb{E}_F[h_\alpha(Z, t)]$, and $\hat{\theta}_n = \inf_{t \in \mathbb{R}} \mathbb{E}_{F_n}^\wedge[h_\alpha(Z, t)]$. The proof of the following theorem is deferred to [Appendix](#).

Theorem 2 (Asymptotics of CVaR). *Consider numbers $q = \text{VaR}_\alpha[Z]$ and $r = \sup\{t : F(t) \leq \alpha\}$. Suppose that $\mathbb{E}[Z^2]$ is finite. Then we have that,*

$$\hat{\theta}_n = \inf_{t \in [q, r]} \frac{1}{n} \sum_{i=1}^n h_\alpha(Z_i, t) + o_p(n^{-1/2}), \tag{7}$$

is a consistent estimator of θ^* with negative bias, i.e. $\mathbb{E}[\hat{\theta}_n] - \theta^* \leq 0$. There are two cases to consider:

(i) If $r = q$, then

$$\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^n h_\alpha(Z_i, q) + o_p(n^{-1/2}), \tag{8}$$

and $n^{1/2}(\hat{\theta}_n - \theta^*)$ converges in distribution to a normal $N(0, \sigma^2)$, where

$$\sigma^2 = \text{Var}[h_\alpha(Z, q)] = (1-\alpha)^{-2} \text{Var}([Z - q]^+). \tag{9}$$

(ii) If $r > q$, then the limiting distribution of $n^{1/2}(\hat{\theta}_n - \theta^*)$ is not normal.

Remark 3. The asymptotic result of [Theorem 2](#) was first established by [Nagaraja \(1982\)](#) in the context of the *selection differential*, a standardized version of CVaR of interest in population genetics. Taking as estimate of CVaR the sample mean of the order statistics $Z_{(\lfloor n\alpha \rfloor)} < \dots < Z_{(n)}$, which agrees with $\hat{\theta}_n$ w.p.1, where $\lfloor n\alpha \rfloor$ is the closest integer to $n\alpha$, he gives explicit expressions for the limiting variables in both the $r > q$ and $r = q$ cases in a slightly different form from ours. In the latter case he obtains the equivalent representation of [\(9\)](#) as, $\sigma^2 = [\tau^2 + \alpha(\theta^* - q)^2]/(1-\alpha)$, where $\tau^2 = \mathbb{E}[(Z - \theta^*)^2 | Z \geq q]$.

Note that a consistent estimate of σ^2 is given by the sample variance of the data $W_i := (1-\alpha)^{-1}[Z_i - \hat{q}_n]^+$, $i = 1, \dots, n$. (Recall that $\hat{q}_n = \hat{F}_n^{-1}(\alpha)$ is the sample estimate of $q = \text{VaR}_\alpha[Z]$, and that the sample mean and sample variance of a random sample W_1, \dots, W_n are, respectively $\bar{W} = n^{-1} \sum_{i=1}^n W_i$, and $S^2 = (n-1)^{-1} \sum_{i=1}^n (W_i - \bar{W})^2$.)

3. Linear regression with CVaR constraints

Consider now the following linear regression model

$$Y_i = g(\mathbf{X}_i, \boldsymbol{\beta}^*) + \varepsilon_i, \quad i = 1, \dots, n, \tag{10}$$

where $\mathbf{X}_i = (X_{i1}, \dots, X_{ik}) \in \mathbb{R}^k$, $\boldsymbol{\beta} = (\beta_0, \dots, \beta_k) \in \mathbb{R}^p$, $p = k + 1$, and $g(\mathbf{x}, \boldsymbol{\beta}) := \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$. We estimate the true parameter vector $\boldsymbol{\beta}^*$ as a solution of the problem:

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \sum_{i=1}^n \rho(e_i) \quad \text{s.t.} \quad \inf_{t \in \mathbb{R}} \left\{ \frac{1}{n} \sum_{i=1}^n h_\alpha(e_i, t) \right\} \leq \eta, \tag{11}$$

where $e_i := Y_i - g(\mathbf{X}_i, \boldsymbol{\beta})$, $\eta \in \mathbb{R}$ is a given constant, the function $h_\alpha(\cdot, \cdot)$ is defined in (6), and $\rho : \mathbb{R} \rightarrow \mathbb{R}_+$ is a convex nonnegative valued function such that $\rho(t) = 0$ iff $t = 0$. Without the additional (CVaR) constraint, the optimization problem (11) corresponds to the standard robust approach to linear regression in the sense of Huber (1981). In particular, if $\rho(t) := t^2$, then $\sum_{i=1}^n \rho(e_i)$ is the squared ℓ_2 (Euclidean) norm of the residuals vector and (11) corresponds to the least squares method. If $\rho(t) := |t|$, then this corresponds to the least absolute deviations method (ℓ_1 norm). Note that (11) can also equivalently be written in the form

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^p, t \in \mathbb{R}} \sum_{i=1}^n \rho(e_i) \quad \text{such that} \quad n(1 - \alpha)t + \sum_{i=1}^n [e_i - t]_+ \leq n(1 - \alpha)\eta, \tag{12}$$

and is a convex problem. In particular, if $\rho(t) := |t|$ this becomes a linear programming problem, and if $\rho(t) := t^2$ it can be formulated as a quadratic programming problem.

The following theorem characterizes the solutions of the linear regression model with CVaR constraints, and establishes the consistency of associated estimators. A proof is provided in Appendix.

Theorem 4. *Assume that in model (10) the errors $\varepsilon_1, \dots, \varepsilon_n$ are iid, and that $\mathbf{X}_1, \dots, \mathbf{X}_n$ are iid random vectors independent of the errors. Let (Y, \mathbf{X}) be a random vector with the same probability distribution as (Y_i, \mathbf{X}_i) . Define $\psi(t) = d\rho(t)/dt$, and suppose further that $\mathbb{E}[\psi(\varepsilon)] = 0$. Then an optimal solution $\hat{\boldsymbol{\beta}}$ of problem (11) converges w.p.1 to the optimal solution of the problem*

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \mathbb{E}[\rho(Y - g(\mathbf{X}, \boldsymbol{\beta}))] \quad \text{s.t.} \quad \inf_{t \in \mathbb{R}} \{ \mathbb{E}[h_\alpha(Y - g(\mathbf{X}, \boldsymbol{\beta}), t)] \} \leq \eta, \tag{13}$$

provided such optimal solution, $\boldsymbol{\beta}^*$, is unique. This occurs if and only if

$$\text{CVaR}_\alpha(\varepsilon) \leq \eta. \tag{14}$$

Remark 5. A few comments about the assumption $\mathbb{E}[\psi(\varepsilon)] = 0$ are in order. If $\rho(t) := t^2$, then $\psi(t) = 2t$ and hence this assumption means that $\mathbb{E}[\varepsilon] = 0$. If $\rho(t) := |t|$, then $\psi(t) = \text{sign}(t)$ for $t \neq 0$. In that case this assumption means that the median of ε is zero, provided that $\text{Prob}(\varepsilon = 0) = 0$.

Now if the strict inequality in (14) holds, i.e., $\text{CVaR}_\alpha(\varepsilon) < \eta$, then the CVaR constraint is not active in the limit, and the asymptotics of $\hat{\boldsymbol{\beta}}$ is the same as in the unconstrained case. If $\text{CVaR}_\alpha(\varepsilon) = \eta$, then $\hat{\boldsymbol{\beta}}$ is still a consistent estimator of $\boldsymbol{\beta}^*$, but its asymptotics are different. In this case $\hat{\boldsymbol{\beta}}$ is not asymptotically normal. If $\text{CVaR}_\alpha(\varepsilon) > \eta$, then $\boldsymbol{\beta}^*$ is not a feasible point of (13), and $\hat{\boldsymbol{\beta}}$ is not a consistent estimator of the true parameter value $\boldsymbol{\beta}^*$.

The above discussion motivates the question of testing the hypothesis

$$H_0 : \text{CVaR}_\alpha(\varepsilon) \leq \eta \quad \text{versus} \quad H_1 : \text{CVaR}_\alpha(\varepsilon) > \eta. \tag{15}$$

In order proceed with statistical testing of (15) let us consider an unconstrained estimator $\tilde{\beta}$ of β^* given by

$$\tilde{\beta} = \arg \min_{\beta \in \mathbb{R}^p} \sum_{i=1}^n \rho(Y_i - g(\mathbf{X}_i, \beta)). \tag{16}$$

Consider the corresponding residuals $\tilde{e}_i = Y_i - g(\mathbf{X}_i, \tilde{\beta})$. We then have that

$$\tilde{\theta}_n := \inf_{t \in \mathbb{R}} \left\{ \frac{1}{n} \sum_{i=1}^n h_\alpha(\tilde{e}_i, t) \right\}, \tag{17}$$

is a consistent estimator of $\theta^* := \text{CVaR}_\alpha(\varepsilon)$. By Theorem 2 we can assume that $\tilde{\theta}_n$ has approximately a normal distribution with mean θ^* and standard error s_n/\sqrt{n} , where s_n^2 is the sample variance of the data $(1 - \alpha)^{-1}[\tilde{e}_i - \tilde{q}_n]^+$ and \tilde{q}_n the sample α -quantile of $\tilde{e}_1, \dots, \tilde{e}_n$. We would reject H_0 if $\tilde{\theta}_n$ is sufficiently bigger than η , that is, if the statistic $T := (\tilde{\theta}_n - \eta)/(s_n/\sqrt{n})$ is bigger than the critical value $\Phi^{-1}(1 - \gamma)$, where $\Phi(\cdot)$ is the cumulative distribution function of a standard normal distribution, and γ is a chosen significance level.

4. CVaR-constrained regression and quantile regression

The CVaR-constrained linear regression model (10), (11) of Section 3 enables us to exercise control over under-predictions (positive errors) by capping them at the value η . In many financial applications it is more important to control *over-predictions* (negative errors), which would involve replacing (11) with

$$\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^n \rho(e_i) \quad \text{s.t.} \quad \inf_{t \in \mathbb{R}} \left\{ \frac{1}{n} \sum_{i=1}^n h_\alpha(-e_i, t) \right\} \leq \omega. \tag{18}$$

However, since Pflug (2000) shows that for any random variable ε , $\mathbb{E}(\varepsilon) = (1 - \alpha) \text{CVaR}_\alpha(\varepsilon) - \alpha \text{CVaR}_{1-\alpha}(-\varepsilon)$, we have that

$$\text{CVaR}_\alpha(-\varepsilon) \leq \omega \iff \text{CVaR}_{1-\alpha}(\varepsilon) \leq \frac{(1 - \alpha)\omega + \mathbb{E}(\varepsilon)}{\alpha},$$

so that constraints on ε translate into equivalent constraints on $-\varepsilon$ by simply changing α and η in (11). Therefore, all the asymptotic results of Section 3 carry immediately over to the setting of (18).

Incorporating these ideas into new optimality criteria, we propose the following empirical CVaR-constrained regression estimators:

$$\hat{\beta}_{\ell_1} = \text{minimizer of (18) with the } \ell_1 \text{ norm on } \rho(\cdot), \tag{19}$$

$$\hat{\beta}_{\ell_2} = \text{minimizer of (18) with the } \ell_2 \text{ norm on } \rho(\cdot). \tag{20}$$

These criteria are clearly asymmetric in that they pay different degrees of attention to under and over-prediction. Increasing ω has the effect of loosening the constraints on the risk of over-predicting. In fact, setting $\alpha = 0$ and $\omega = \infty$ in both (19) and (20), we recover precisely the usual unconstrained least absolute deviations and least squares regression, respectively. For continuous distributions, the constraint $\text{CVaR}_\alpha(-\varepsilon) \leq \omega$, caps the average of the “worst” $(1 - \alpha)100\%$ of over-predictions at the value ω . The imposition of these

(possibly multiple) constraints allows one to control the shape of the resulting residual error distribution.

This CVaR-constrained regression idea is similar to the concept of *quantile regression* (Koenker, 2005). To illustrate the connection, consider a random sample of observations from linear regression model (10), which can be concisely written as

$$Y = X\boldsymbol{\beta} + \varepsilon, \tag{21}$$

where X denotes the n by p design matrix whose i th row is $[1, X_{i1}, \dots, X_{ik}] := \mathbf{X}'_i$. For design point \mathbf{x} consisting of a realization of the covariate vector $\{1, X_1, \dots, X_k\}$, interest typically centers around optimal estimation of some measure of location of $Y|\mathbf{x} = \mathbf{x}'\boldsymbol{\beta}$ such as the median (ℓ_1 regression) or mean (ℓ_2 regression). If we assume the $\{\varepsilon_i\}$ to be iid,

$$F_{Y|\mathbf{x}}(y) = P(Y \leq y|\mathbf{x}) = P(\varepsilon \leq y - \mathbf{x}'\boldsymbol{\beta}|\mathbf{x}) = F_{\varepsilon|\mathbf{x}}(y - \mathbf{x}'\boldsymbol{\beta}),$$

from which it follows immediately that the conditional quantile functions of Y and ε given \mathbf{x} , $F_{Y|\mathbf{x}}^{-1}(\cdot)$ and $F_{\varepsilon|\mathbf{x}}^{-1}(\cdot)$, are related according to

$$F_{Y|\mathbf{x}}^{-1}(\alpha) = \mathbf{x}'\boldsymbol{\beta} + F_{\varepsilon|\mathbf{x}}^{-1}(\alpha) := \mathbf{x}'\boldsymbol{\beta}(\alpha),$$

where $\beta_0(\alpha) = \beta_0 + F_{\varepsilon|\mathbf{x}}^{-1}(\alpha)$, and $\beta_j(\alpha) = \beta_j$, for $j = 1, \dots, k$. Viewed as a function of \mathbf{x} , $F_{Y|\mathbf{x}}^{-1}(\alpha)$ describes how the α -quantile surface of Y varies as a function of the covariates. With $[t]^- = \min\{0, t\}$ denoting the negative part of t , Koenker and Bassett (1978) propose using the function

$$\rho(t) = \alpha[t]^+ - (1 - \alpha)[t]^-,$$

in (18) (with no CVaR constraint), for consistent estimation of $F_{Y|\mathbf{x}}^{-1}(\alpha)$. This is known as quantile regression, the $\hat{\boldsymbol{\beta}}(\alpha)$ being the *regression quantiles*. The resulting estimated α -quantile function of $Y|\mathbf{x}$ is then $\hat{F}_{Y|\mathbf{x}}^{-1}(\alpha) = \mathbf{x}'\hat{\boldsymbol{\beta}}(\alpha)$.

In a similar vein, direct control over potential *overestimates* of Y can be achieved by using the estimators $\hat{\boldsymbol{\beta}}_{\ell_1}$ and $\hat{\boldsymbol{\beta}}_{\ell_2}$, thereby producing “good” predictors with residuals whose upper $(1 - \alpha)$ tail mean does not exceed ω . We will refer to the optimality criteria that give rise to these estimators as $CL_1(\hat{\boldsymbol{\beta}}_{\ell_1})$ and $CL_2(\hat{\boldsymbol{\beta}}_{\ell_2})$ norm.

In a way this approach is philosophically different from quantile regression, but the intent is similar, to estimate some measure of location of $Y|\mathbf{x}$. Using a VaR constraint in (18) makes it a direct competitor of quantile regression; with a CVaR constraint the approach can be regarded as a sort of *CVaR regression*. We can also view the effect of (18) as a *haping* of the residual error distribution ε , so that $Y|\mathbf{x}$ will have certain desirable properties. For a particular value of the pair $\{\alpha, \omega\}$, the above fitted response surface can be heuristically interpreted as follows:

If the fitted value \hat{Y}_i is used as a predictor/estimator of Y_i , the average of the resulting $(1 - \alpha)100\%$ over-predictions/estimates $\hat{Y}_i - Y_i$, will not exceed ω .

Note that for fixed α , decreasing the value of ω in (18) leads to a fitted response surface that is lower (on average) than the one for a higher value of ω .

The following easily verifiable proposition relates the α -level CVaRs of $Y|\mathbf{x} = \mathbf{x}'\boldsymbol{\beta} + \varepsilon$, and ε in a linear regression model.

Proposition 6 (CVaR relations in a regression model). *For a realization $\mathbf{x}' = [1, x_1, \dots, x_k]'$ from the covariate vector in model (10) with $\{\varepsilon_i\}$ iid, we have the CVaR relations:*

$$\begin{aligned}
 \text{(i)} \quad & \text{CVaR}_\alpha(Y|\mathbf{x}) = \mathbf{x}'\boldsymbol{\beta} + \text{CVaR}_\alpha(\varepsilon), \text{ and } \text{CVaR}_\alpha(-Y|\mathbf{x}) = -\mathbf{x}'\boldsymbol{\beta} + \text{CVaR}_\alpha(-\varepsilon). \\
 \text{(ii)} \quad & \text{CVaR}_{1-\alpha}(Y|\mathbf{x}) = \mathbf{x}'\boldsymbol{\beta} + \frac{1}{\alpha}\mathbb{E}(\varepsilon) + \frac{1-\alpha}{\alpha}\text{CVaR}_\alpha(-\varepsilon). \tag{22}
 \end{aligned}$$

From (i) we have the equivalence

$$\text{CVaR}_\alpha(-\varepsilon) \leq \omega \iff \text{CVaR}_\alpha(-Y_x) \leq \omega - \mathbf{x}'\boldsymbol{\beta},$$

so that the constraint on the residuals is equivalent to a constraint on the response.

5. Case study: NDX and SPX indices

In this section, we illustrate how the ideas proposed in the previous sections can be applied in the context of financial data. Specifically, consider historical rates of semi-daily returns for the NDX and SPX indices from 07/14/2000 to 06/27/2002. For any given day, t , in this time frame, $t = 1, \dots, 480$, let the variables defined in Table 1 denote the semi-daily rates of return for the two indices. Selection of these half-day returns was motivated by the trading schedule of the index funds (Velocity 100, RYVYX) and (Titan 500, RYTNX) from the RYDEX family of funds. These funds replicate the NDX and SPX indices, and can be traded twice daily at the listed times without any transaction costs.

Time series plots of the semi-daily returns for the two indices over the 480 day period are shown in Fig. 1. As is characteristic of returns data, the series are serially uncorrelated (and uncorrelated with each other), but sample autocorrelation plots of the squares and absolute values suggest some dependence. Note that the SPX series is substantially less variable than the NDX series.

Now consider triplets $\{Y_N(t), X_N(t), X_S(t)\}$ and $\{Y_S(t), X_N(t), X_S(t)\}$, in each case treating $Y_N(t)$ and $Y_S(t)$ as the response, and the pair $\{X_N(t), X_S(t)\}$ as associated covariates. Allowing powers of the covariates and interactions up to degree p , we proceed by fitting multiple polynomial regressions of the form

$$Y_N(t) = \sum_{l=0}^p \sum_{j=0}^l \beta_{jl}^{(N)} X_N^j(t) X_S^{l-j}(t) + \varepsilon_t, \tag{23}$$

$$Y_S(t) = \sum_{l=0}^p \sum_{j=0}^l \beta_{jl}^{(S)} X_N^j(t) X_S^{l-j}(t) + \varepsilon_t. \tag{24}$$

Table 1
 Definitions of semi-daily NDX & SPX rate of return variables for day t , $t = 1, \dots, 480$

Variable name	Index name	Rate of return	
		From	To
$X_N(t)$	NDX	4:00 PM day $t - 1$	10:30 AM day t (overnight)
$Y_N(t)$	NDX	10:30 AM day t	4:00 PM day t (over-day)
$X_S(t)$	SPX	4:00 PM day $t - 1$	10:30 AM day t (overnight)
$Y_S(t)$	SPX	10:30 AM day t	4:00 PM day t (over-day)

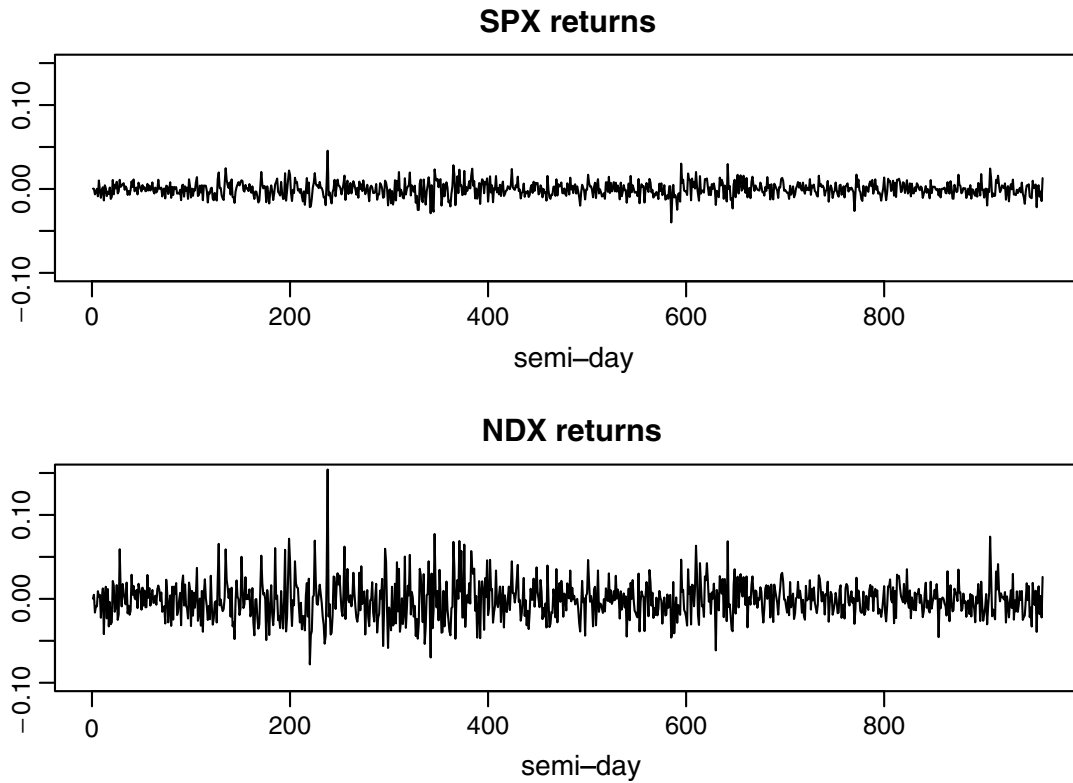


Fig. 1. Time series of SPX and NDX returns for the period 07/14/2000 to 06/27/2002.

The proposed CL_1 and CL_2 loss functions were used to estimate each of the $(p + 1)(p + 2)/2$ sets of coefficients. Choosing $p = 4$ provided flexible enough response surfaces to capture the essential features of the data and accommodate high-order nonlinearities. This meant estimating 15 coefficients in each of the models (23) and (24).

For the purpose of cross-validation, the data was divided up into 10 groups, 48 observations (days) per group, every tenth day to a group. For example, group k (henceforth g_k) consisted of $\{Y_N(t), X_N(t), Y_S(t), X_S(t)\}$, for $i = k, 10 + k, 20 + k, \dots, 470 + k$. Each of the 10 groups g_1, \dots, g_{10} was in turn selected and omitted from the analysis to form the *test set* (48 observations). The remaining 9 groups comprising the *training set* (432 observations), were used to fit the model and compute the in-sample CVaR. This fitted model was then used to calculate the CVaR of the test set (out-of-sample attained CVaR). Let $\hat{\theta}_{\alpha,\omega}(g_k)$ denote this out-of-sample α -level CVaR when g_k is used as the test set, and the constraint is set to ω (in-sample CVaR constrained not to exceed ω).

Fig. 2 compares the out-of-sample and in-sample attained CVaR values, across various $\{\alpha, \omega\}$ pairs used to fit model (23) under CL_1 loss. Specifically, and to make more meaningful comparisons, the boxplots shown are of percent differences between out and in-sample attained CVaR values, relative to the in-sample CVaR, ω . That is, if we define

$$z_{\alpha,\omega}(g_k) = 100 \frac{\hat{\theta}_{\alpha,\omega}(g_k) - \omega}{1 + \omega}, \tag{25}$$

each boxplot is based on the 10 values $z_{\alpha,\omega}(g_1), \dots, z_{\alpha,\omega}(g_{10})$. (The addition of 1 to the denominator corrects for CVaR values near zero.) For example, the top leftmost boxplot indicates (with horizontal lines) the location of the 0th, 25th, 50th, 75th, and 100th percentiles, for the 10 values $z_{0.5,0}(g_1), \dots, z_{0.5,0}(g_{10})$. Outliers appear as circles. All graphics in this

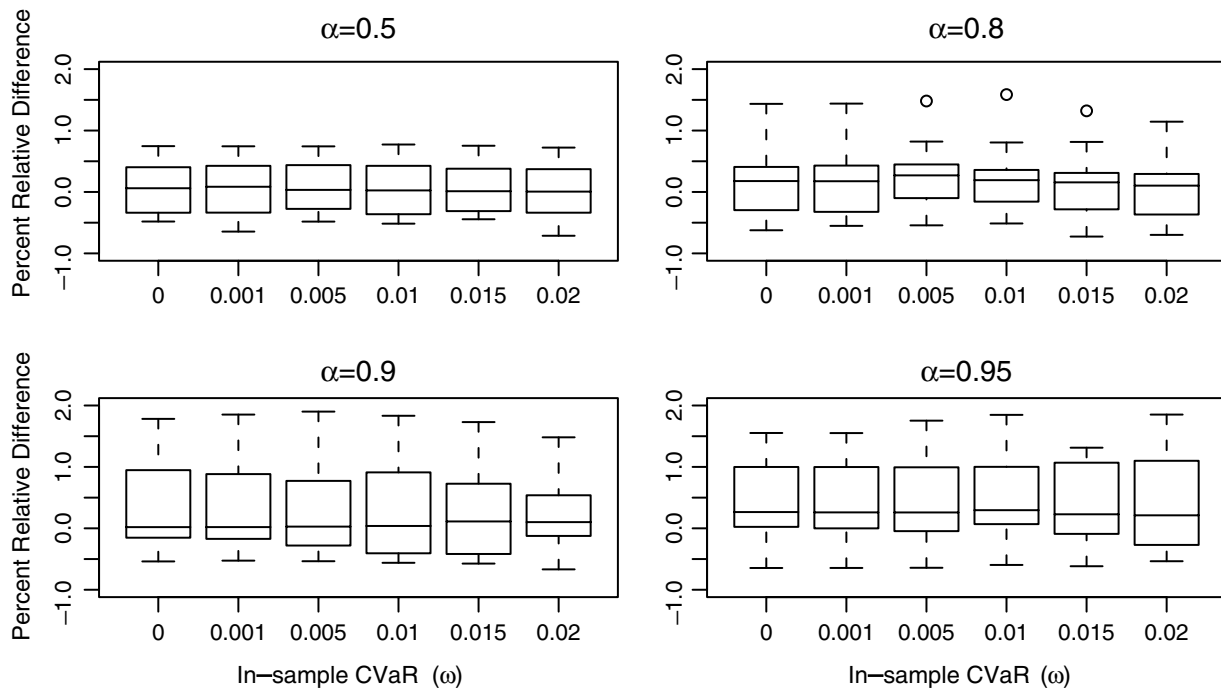


Fig. 2. Boxplots of percent relative differences between out-of-sample and corresponding in-sample CVaR values, for all training/test set groupings of the triplet $\{Y_N(t), X_N(t), X_S(t)\}$ fitted via model (23) under CL_1 loss, across various values of α and ω . The in-sample CVaR is based on the training set, and is constrained to equal ω . Each boxplot is based on 10 CVaR values, one from each test set. For example, the first boxplot on top left shows that out-of-sample CVaR values for the 50% upper residual tail fall within -0.5% to 1% of corresponding in-sample values, the latter being zero.

paper were produced with *R*. Complete details on the *R* definition of a boxplot and other statistical tools can be found in Venables and Ripley (2002).

On the whole we see close agreement between in and out-of-sample CVaR values, with boxplots for higher values of α showing slightly larger variability and greater proportion of out-of-sample CVaRs exceeding the in-sample ones. (Recall that in-sample CVaRs are constrained to equal ω .) No out-of-sample CVaR values exceeded the corresponding in-sample attained CVaR by more than 2%, and no out-of-sample CVaR value below 1% of the in-sample CVaR was recorded.

Fig. 3 shows a contour plot of the over-day NDX return, $Y_N(t)$, surface in model (23), fitted under CL_1 loss with $\alpha = 0.95$ and $\omega = 0.01$. The data used is that corresponding to the training set comprised of groups $\{g_2, \dots, g_{10}\}$. For given values of the SPX ($X_S(t)$) and NDX ($X_N(t)$) overnight returns, regions where the fitted surface is positive flag when favorable NDX returns might be expected from morning to evening of the same day (over-day). If investments are made whenever these positive NDX returns are predicted, the CVaR constraint tries to ensure that any resulting losses are not “too bad”. Note that “losses” here are not the same as “differences between observed and predicted”, which is what the CVaR constraint applies to. Although the relationship between the two is difficult to quantify exactly, it is clear that stricter CVaR constraints in the sense of $\alpha \rightarrow 1$ and $\omega \rightarrow 0$, will lead to less severe losses.

Fig. 4 illustrates what the CVaR constraint tries to accomplish. The solid line is the training set of observed over-day NDX returns considered in Fig. 3, while the dashed line are the CL_1 loss (with $\alpha = 0.95$ and $\omega = 0.01$) predicted values. The regular least absolute

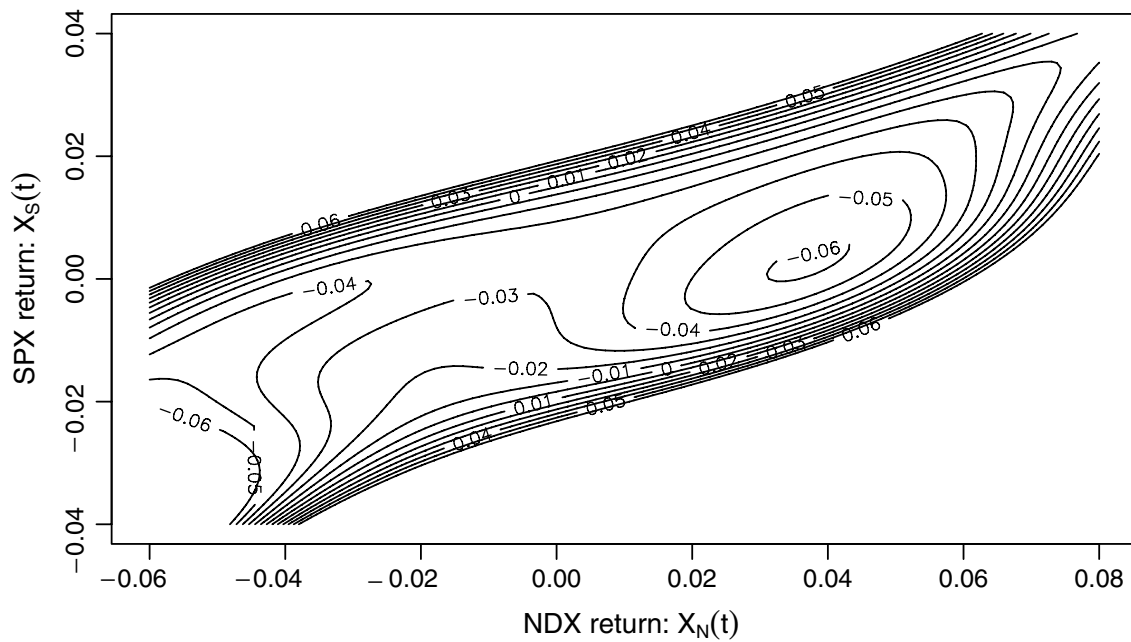


Fig. 3. Contour plot of the over-day NDX return fitted response surface ($\hat{Y}_N(t)$), with the SPX ($X_S(t)$) and NDX ($X_N(t)$) overnight returns as covariates, under polynomial regression model (23). The model is fitted to the data of groups $\{g_2, \dots, g_{10}\}$ under CL_1 loss, with $\alpha = 0.95$ and $\omega = 0.01$.

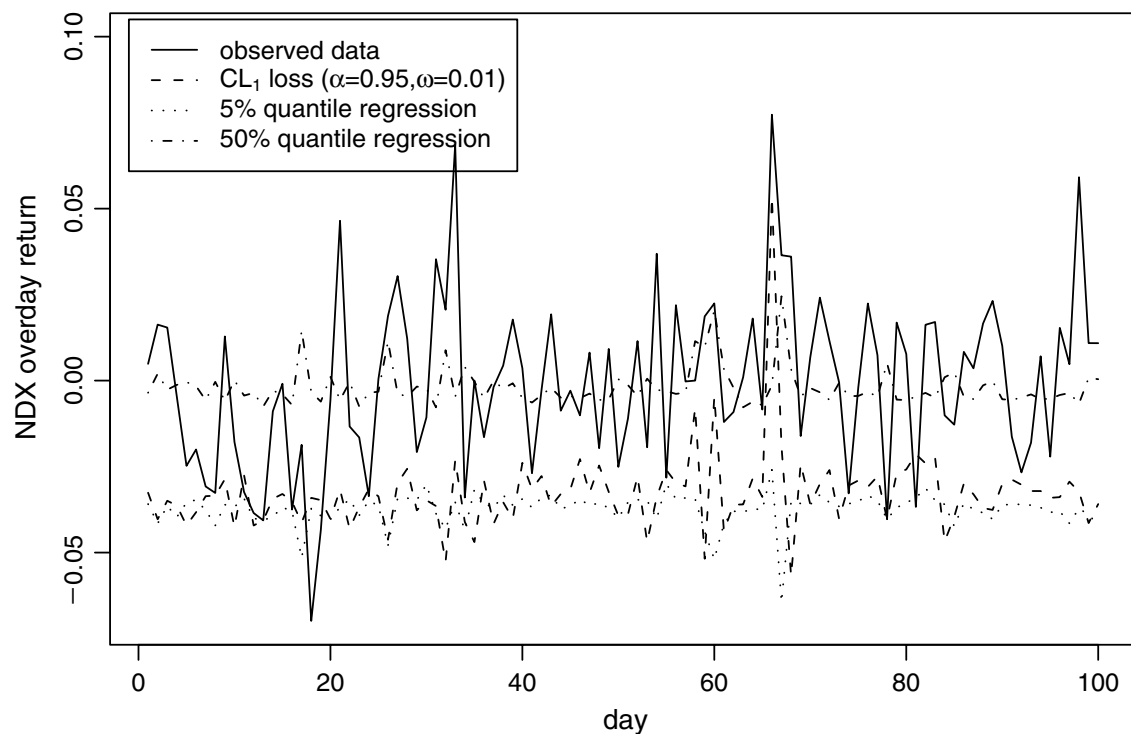


Fig. 4. Time series plot of the first 100 observed over-day NDX returns for training set $\{g_2, \dots, g_{10}\}$ (solid), along with predicted values obtained from CL_1 and 50% and 5% quantile regression fits. The time scale has been compressed to show 100 consecutive days.

deviations predictions, which are exactly the 50th quantile regression predictions, are shown for comparison as the dashed-dotted line. The fact that the CL_1 predictions are consistently below the observed values demonstrates how they try to curtail losses by being

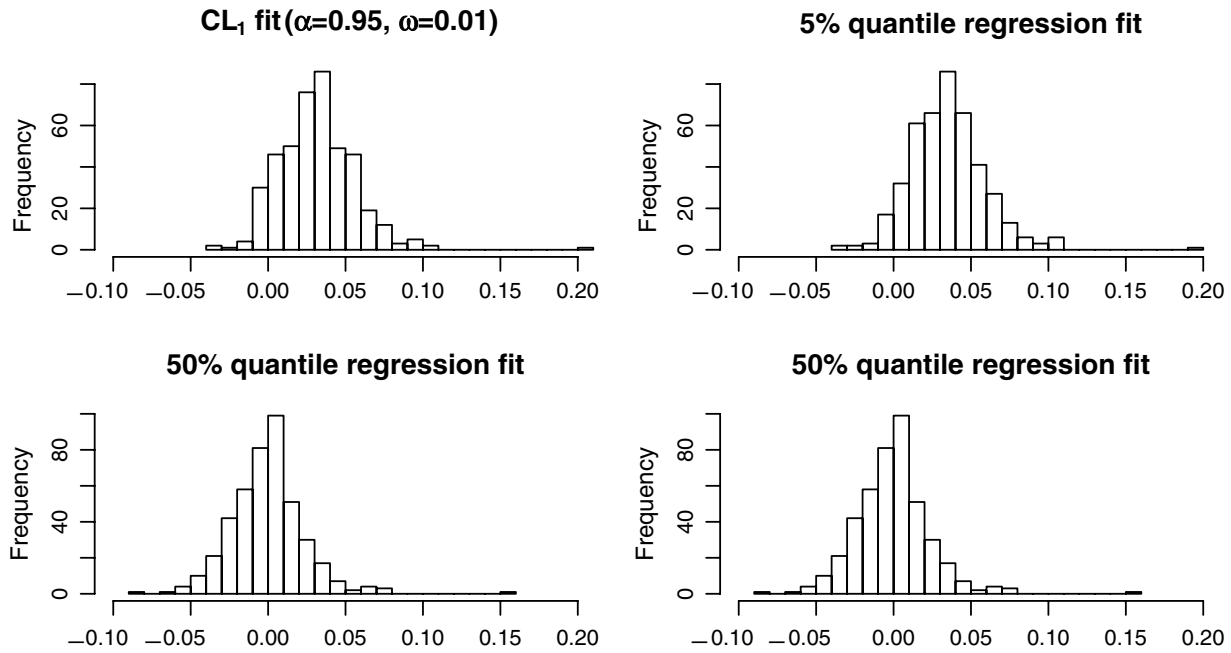


Fig. 5. Histograms of residuals corresponding to CL_1 and 5% quantile regression fits for the first 100 observed over-day NDX returns in training set $\{g_2, \dots, g_{10}\}$. The identical histograms at the bottom are the residuals from a 50% quantile regression fit, and are shown for comparison.

“safe”. Note that although the CL_1 predictions are very similar to those derived from a 5% quantile regression fit (dotted line), they are slightly better at capturing the magnitude of the fluctuations in the observed values. In order to discern patterns more clearly, only the first 100 consecutive days are shown; the remainder being similar. The cumulative returns for the entire observed period is -1.04 . The cumulative returns for the periods where positive returns are predicted by each of the CL_1 , 50%, and 5% quantile regression methods are, respectively, 0.16, 0.11, and 0 (all predictions were negative). In this context CL_1 therefore emerges as the winner.

Fig. 5 demonstrates the effect of the above CL_1 and 5% quantile regression model fits on the prediction error residuals (observed minus predicted). Compare each of these to the residuals from a 50% quantile regression fit. Note how the imposition of the CVaR constraint shapes the residual error distribution, re-centering and skewing it to the right. A similar effect is noted for the 5% quantile regression residuals.

Appendix A. Proof of Theorem 2

Let us observe that the set of minimizers in the right hand side of (3) is given by the (nonempty and bounded) interval $[q, r]$, where $q := \text{VaR}_\alpha(Z)$ and $r := \sup\{t: F(t) \leq \alpha\}$. Similarly, the set of minimizers in the right hand side of (5) is given by the interval $[\hat{q}_n, \hat{r}_n]$, where \hat{q}_n and \hat{r}_n are respective sample estimates obtained by replacing F with \hat{F}_n . As stated in Remark 1, the sample quantile \hat{q}_n converges w.p.1 to the interval $[q, r]$ as $n \rightarrow \infty$. Using this fact it is not difficult to show that $\hat{\theta}_n$ converges w.p.1 to θ^* , i.e., $\hat{\theta}_n$ is a consistent estimator of θ^* . It is also possible to deduce consistency of $\hat{\theta}_n$ directly from its definition as the optimal value of the right-hand side of (3).

Now, for any z the function $h_\alpha(z, \cdot)$ is Lipschitz continuous on \mathbb{R} with Lipschitz constant $1 + (1 - \alpha)^{-1}$ independent of z . Since $\mathbb{E}[Z^2]$ is finite, $\mathbb{E}[h_\alpha(Z, t)^2]$ is finite for every $t \in \mathbb{R}$. It

then follows by the Central Limit Theorem that $n^{-1/2} \sum_{i=1}^n (h_\alpha(Z_i, t) - \mathbb{E}[h_\alpha(Z, t)])$ converges in distribution to a normal with zero mean and variance $\sigma^2(t) = \text{Var}[h_\alpha(Z, t)]$. By using representation (3), it is then possible to apply a general result of Shapiro (1991) (see also Theorems 6.4.2 and 6.4.3 in Rubinstein and Shapiro, 1993) to obtain (i). Note also that

$$\theta^* = \inf_{t \in \mathbb{R}} \mathbb{E} \left[\frac{1}{n} \sum_{i=1}^n h_\alpha(Z_i, t) \right] \geq \mathbb{E} \left[\inf_{t \in \mathbb{R}} \frac{1}{n} \sum_{i=1}^n h_\alpha(Z_i, t) \right] = \mathbb{E}[\hat{\theta}_n],$$

that is, the estimator $\hat{\theta}_n$ has negative bias. (Finite sample evidence for this bias in the special case of random sampling from a Laplace distribution is provided by Trindade and Zhu, 2007).

To see (ii), if $r > q$ it follows from (7) that the limiting distribution of $n^{1/2}(\hat{\theta}_n - \theta^*)$ is given by the minimum of a set of correlated normally distributed random variables. In that case the limiting distribution of $n^{1/2}(\hat{\theta}_n - \theta^*)$ is not normal and has a negative mean. Consequently, if $r > q$, then one may expect that the (negative) bias of $\hat{\theta}_n$ is of order $O(n^{-1/2})$. On the other hand, suppose that $q = r$. Then it is possible to calculate a second order term in an asymptotic expansion of $\hat{\theta}_n$ as follows. Note that

$$\frac{\partial \mathbb{E}[h_\alpha(Z, t)]}{\partial t} = \mathbb{E} \left[\frac{\partial h_\alpha(Z, t)}{\partial t} \right] = 1 - \frac{1}{1 - \alpha} \text{Prob}(Z > t) = \frac{F(t) - \alpha}{1 - \alpha},$$

provided that $\text{Prob}(Z = t) = 0$. In particular, $\partial \mathbb{E}[h_\alpha(Z, q)] / \partial t = 0$. Also $\partial^2 \mathbb{E}[h_\alpha(Z, q)] / \partial t^2 = (1 - \alpha)^{-1} f(q)$, provided that the distribution of Z has density $f(q)$ at q . Consider now

$$\zeta_n := n^{-1} \sum_{i=1}^n \frac{\partial h_\alpha(Z_i, q)}{\partial t} = \frac{\hat{F}_n(q) - \alpha}{1 - \alpha},$$

and

$$\varphi(\zeta_n) := \inf_{\delta \in \mathbb{R}} \{ 2\delta\zeta_n + [\partial^2 \mathbb{E}[h_\alpha(Z, q)] / \partial t^2] \delta^2 \} = -(1 - \alpha)^{-1} f(q)^{-1} [\hat{F}_n(q) - \alpha]^2,$$

provided that $f(q) \neq 0$. Then we have that

$$\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^n h_\alpha(Z_i, q) + \frac{1}{2} \varphi(\zeta_n) + o_p(n^{-1}). \tag{A.1}$$

(See Theorem 4.4 in Shapiro, 2000.) Let us note that, under the above assumptions, $n^{1/2}[\hat{F}_n(q) - \alpha]$ converges in distribution to a normal $N(0, \alpha(1 - \alpha))$, and hence $n\varphi(\zeta_n)$ converges in distribution to $-\alpha\chi_1^2/f(q)$. That is, under the above assumptions, the asymptotic bias of $\hat{\theta}_n$ is of order $O(n^{-1})$, and is given by $-\alpha/[2f(q)n] + o(n^{-1})$.

Appendix B. Proof of Theorem 4

By standard arguments (e.g. Shapiro, 2003, Sec. 2) we have that an optimal solution $\hat{\beta}$ of problem (11) converges w.p.1 to the optimal solution of problem (13), provided such optimal solution is unique. Consider now the function $v(\beta) := \mathbb{E}[\rho(Y - g(\mathbf{X}, \beta))]$. We have that

$$\nabla v(\beta) = \mathbb{E}[\nabla_{\beta} \rho(Y - g(\mathbf{X}, \beta))] = -\mathbb{E}[\psi(Y - g(\mathbf{X}, \beta)) \nabla_{\beta} g(\mathbf{X}, \beta)],$$

where $\nabla v(\boldsymbol{\beta}) = [\partial v(\boldsymbol{\beta})/\partial \beta_0, \dots, \partial v(\boldsymbol{\beta})/\partial \beta_k]'$. In particular, for $\boldsymbol{\beta} = \boldsymbol{\beta}^*$ we obtain

$$\nabla v(\boldsymbol{\beta}^*) = -\mathbb{E}[\psi(\varepsilon)\nabla_{\boldsymbol{\beta}}g(\mathbf{X}, \boldsymbol{\beta}^*)] = -\mathbb{E}[\psi(\varepsilon)]\mathbb{E}[\nabla_{\boldsymbol{\beta}}g(\mathbf{X}, \boldsymbol{\beta}^*)], \quad (\text{B.1})$$

where the last equality follows by the independence of ε and \mathbf{X} . Since $\mathbb{E}[\psi(\varepsilon)] = 0$, we have by (B.1) that $\nabla v(\boldsymbol{\beta}^*) = 0$. Since the function $v(\boldsymbol{\beta})$ is convex, it follows that $\boldsymbol{\beta}^*$ is an (unconstrained) minimizer of $v(\boldsymbol{\beta})$. Consequently, $\boldsymbol{\beta}^*$ is an optimal solution of (13) iff condition (14) holds.

References

- Acerbi, C., Tasche, D., 2002. On the coherence of expected shortfall. *Journal of Banking and Finance* 26, 1487–1503.
- Artzner, P., Delbaen, F., Eber, J., Heath, D., 1997. Thinking coherently. *Risk* 10, 68–71.
- Artzner, P., Delbaen, F., Eber, J., Heath, D., 1999. Coherent measures of risk. *Mathematical Finance* 9, 203–228.
- Basak, S., Shapiro, A., 2001. Value-at-risk-based risk management: Optimal policies and asset prices. *The Review of Financial Studies* 14, 371–405.
- Christoffersen, P., Diebold, F., 1997. Optimal prediction under asymmetric loss. *Econometric Theory* 13, 808–817.
- Föllmer, H., Schied, A., 2004. *Stochastic Finance – An Introduction in Discrete Time*, second ed. Walter de Gruyter, Berlin.
- Gabih, A., Grecksch, W., Wunderlich, R., 2005. Dynamic portfolio optimization with bounded shortfall risks. *Stochastic Analysis and Applications* 3, 579–594.
- Gundel, A., Weber, S., 2005. Utility maximization under a shortfall risk constraint. Preprint, Cornell University.
- Huber, P., 1981. *Robust Statistics*. Wiley, New York.
- Koenker, R., 2005. *Quantile Regression*. Cambridge University Press, Econometric Society Monograph Series.
- Koenker, R., Bassett, G., 1978. Regression quantiles. *Econometrica* 46, 33–50.
- Krokhmal, P., Uryasev, S., Zrazhevsky, G., 2002. Risk management for hedge fund portfolios: A comparative analysis of linear portfolio rebalancing strategies. *Journal of Alternative Investments* 5, 10–29.
- Nagaraja, H., 1982. Some nondegenerate limit laws for the selection differential. *The Annals of Statistics* 10, 1306–1310.
- Palmquist, J., Uryasev, S., Krokhmal, P., 2002. Portfolio optimization with conditional value-at-risk objective and constraints. *The Journal of Risk* 4, 11–27.
- Pflug, G., 2000. Some remarks on the value-at-risk and the conditional value-at-risk. In: Uryasev, S. (Ed.), *Probabilistic Constrained Optimization: Methodology and Applications*. Kluwer, Dordrecht.
- Rockafellar, R.T., Uryasev, S., 2000. Optimization of conditional value-at-risk. *Journal of Risk* 2, 21–41.
- Rockafellar, R.T., Uryasev, S., 2002. Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance* 26, 1443–1471.
- Rockafellar, R.T., Uryasev, S., Zabarankin, M., 2006a. Master funds in portfolio analysis with general deviation measures. *Journal of Banking and Finance* 30, 743–778.
- Rockafellar, R.T., Uryasev, S., Zabarankin, M., 2006b. Generalized deviations in risk analysis. *Finance and Stochastics* 10, 51–74.
- Rubinstein, R.Y., Shapiro, A., 1993. *Discrete Event Systems: Sensitivity Analysis and Stochastic Optimization by the Score Function Method*. John Wiley & Sons, Chichester.
- Shapiro, A., 1991. Asymptotic analysis of stochastic programs. *Annals of Operations Research* 30, 169–186.
- Shapiro, A., 2000. Statistical inference of stochastic optimization problems. In: Uryasev, S. (Ed.), *Probabilistic Constrained Optimization: Methodology and Applications*. Kluwer, Dordrecht, pp. 282–304.
- Shapiro, A., 2003. Monte Carlo sampling methods. In: Ruczyński, A., Shapiro, A. (Eds.), *Stochastic programming*. In: *Handbooks in OR & MS*, vol. 10. North-Holland, Amsterdam.
- Trindade, A.A., Zhu, Y., 2007. Approximating the distributions of estimators of financial risk under an asymmetric Laplace law. *Computational Statistics and Data Analysis* 51, 3433–3447.
- Venables, W., Ripley, B., 2002. *Modern Applied Statistics with S*, fourth ed. Springer, New York.