**MATHEMATICS OF OPERATIONS RESEARCH**
Vol. 33, No. 3, August 2008, pp. 712–729
ISSN 0364-765X | EISSN 1526-5471 | 08 | 3303 | 0712

DOI 10.1287/moor.1080.0313


# Risk Tuning with Generalized Linear Regression

R. Tyrrell Rockafellar
Department of Mathematics, University of Washington, Seattle, Washington 98195,
rtr@math.washington.edu, http://www.math.washington.edu/~rtr/mypage.html

Stan Uryasev
ISE Department, University of Florida, Gainesville, Florida 32611,
uryasev@ufl.edu, http://www.ise.ufl.edu/uryasev

Michael Zabarankin
Department of Mathematical Sciences, Stevens Institute of Technology, Hoboken, New Jersey 07030,
mzabaran@stevens.edu, http://personal.stevens.edu/~mzabaran

A framework is set up in which linear regression, as a way of approximating a random variable by other random variables, can be carried out in a variety of ways, which, moreover, can be tuned to the needs of a particular model in finance, or operations research more broadly. Although the idea of adapting the form of regression to the circumstances at hand has already found advocates in promoting quantile regression as an alternative to classical least-squares approaches, it is carried here much farther than that. Axiomatic concepts of error measure, deviation measure, and risk measure are coordinated with certain "statistics" that likewise say something about a random variable. Problems of regression utilizing these concepts are analyzed and the character of their solutions is explored in a range of examples. Special attention is paid to parametric forms of regression which arise in connection with factor models. It is argued that when different aspects of risk enter an optimization problem, different forms of regression ought to be invoked for each of those aspects.

*Key words*: linear regression; error measures; deviation measures; risk measures; risk management; factor models; portfolio optimization; value-at-risk; conditional value-at-risk; quantile regression
*MSC2000 subject classification*: Primary: 90C25, 90C46, 91B30, 91B82; secondary: 90C90, 90B50
*OR/MS subject classification*: Primary: decision analysis/risk, forecasting/regression, mathematics/convexity; secondary: finance/portfolio
*History*: Received May 17, 2007; revised December 3, 2007.

**1. Introduction.** A common task in stochastic modeling is that of approximating a complex, or only partially understood, random variable $Y$ by a linear combination

$$\widehat{Y} = c_0 + c_1 X_1 + \cdots + c_n X_n$$

of random variables $X_j$ which are better known, or more accessible, than $Y$, or thought to express the major influences on $Y$. The coefficients $c_0, c_1, \ldots, c_n$ are chosen to control the approximation error $Y - \widehat{Y}$ as a random variable and optimize it from some standpoint. This is the essence of *linear regression*, but for the most part only a few cases have received much attention.

The classical approach is to use least squares, in which the expectation of $[Y - \widehat{Y}]^2$, or equivalently the square root of that expectation, is minimized; this amounts to minimizing the standard deviation of $Y - \widehat{Y}$ subject to the expectation of $Y - \widehat{Y}$ being 0. In quantile regression (Koenker [9], Koenker and Bassett [10]), a different measure of the error is minimized with the aim of providing direct estimates for quantiles of $Y$. Other approaches could also be taken, however. What might be said about them and why might one be preferred over another in some situations? How might they affect model building itself?

Our goal in this paper is to build a bridge between statistical ideas and their application to diverse areas of operations research, especially finance, which opens a broad way to posing and answering such questions. We develop an axiomatic framework, encompassing least squares and quantile regression along with much more, in which *measures of error* $\mathcal{E}$ on random variables $X$ (quantifying how far any $X$ is from being just the 0 random variable) are coordinated with generalized *measures of deviation* $\mathcal{D}$ beyond standard deviation (which quantify how far $X$ is from being constant, or "certain"). Such measures of deviation $\mathcal{D}$ are naturally paired moreover with *measures of risk* $\mathcal{R}$ (providing numerical surrogates for the overall "loss" associated with a prospect $X$).

The measures $\mathcal{E}$, $\mathcal{D}$, and $\mathcal{R}$ bring out features of interest in many situations—in particular for risk management in finance—and furthermore tie in with a "statistic" $\mathcal{S}$ which identifies, from the perspective of the chosen measure of error, the constant $C$ that is closest to a given random variable $X$. By putting all these notions together for the first time in a single package, we come to a view of generalized linear regression in which optimization is at the forefront, and techniques for estimating and approximating random variables can be tuned to the purposes of a particular model. Quantile regression already offers some precedents for taking this optimization perspective (as explained well in the book by Koenker [9]), but we go much further.

It is important then to address the issue that arises, in this new horizon of capabilities for generalized linear regression, of how the available choices can be used advantageously in particular applications. In finance, for example, where so much seems at present to hinge on variance and covariance, practitioners could well wonder why anything other than least squares approaches to estimation should ever be brought in. But regression based on least squares concentrates basically on estimating expectations, while in financial modeling tail behavior is often of greater concern. Quantile regression could, for instance, help with that. A recent paper (Trindade et al. [21]) goes in a direction related to quantile regression that and adds constraints.

A key reason why one mode of regression might be preferable over another in a practical setting is provided by factor models in finance. A brief explanation of this will provide advance motivation for the theory to be developed.

A portfolio is commonly described by a random variable $x_1r_1+\cdots+x_mr_m$ in which the $r_i$'s are the returns, or rates of return, of the instruments $i=1,\ldots,m$ incorporated in the portfolio, and the $x_i$'s are their weights. The weights are to be adjusted in accordance with some objective, and this involves paying attention to functions having the form

$$f(x_1,\ldots,x_m)=\mathcal{F}(x_1r_1+\cdots+x_mr_m) \tag{1.1}$$

for a functional $\mathcal{F}$ that acts on random variables. For instance, $\mathcal{F}$ might quantify some aspect of the risk in the portfolio. (Deviation measures $\mathcal{D}$ and risk measures $\mathcal{R}$ both do this, but we do not exclude still other possibilities when we speak of "aspects" of risk.) The random variables $r_i$ may be difficult or complicated to handle directly, however. This leads, in a factor model, to approximating each $r_i$ by a linear combination of a much smaller number of other random variables, called factors. In other words, $r_i$ is approximated by some $\hat{r}_i$ obtained through linear regression on those factors. This means that $f(x_1,\ldots,x_m)$ is replaced by

$$\hat{f}(x_1,\ldots,x_m)=\mathcal{F}(x_1\hat{r}_1+\cdots+x_m\hat{r}_m). \tag{1.2}$$

But what can be said then about the difference $f(x_1,\ldots,x_m)-\hat{f}(x_1,\ldots,x_m)$ and its behavior as a function of the weights $x_i$?

This difference, constituting an error of another sort, is affected by the approach taken to regression in the factor model, so the idea comes up of tuning that approach to achieve optimal behavior. We will be able to show that, for functionals $\mathcal{F}$ of type $\mathcal{F}=\mathcal{D}$ or $\mathcal{F}=\mathcal{R}$ associated with aspects of risk, such tuning can definitely be carried out. This will emerge from the way risk and error are coordinated in the framework that we set up.

We start §2 with a summary of some concepts and results in our recent paper (Rockafellar et al. [16]) on which we need to rely. We then develop the notion of "error projection" and how it leads from an error measure $\mathcal{E}$ not only to a deviation measure $\mathcal{D}$ but also to an associated "statistic" $\mathcal{S}$. Various examples are laid out. Among other new results is the identification of an error measure that projects to a given *mixture* of deviation measures, such as may be associated with the risk profile of an investor.

In §3, we go on to generalized linear regression and the basic results about it. In §4, we characterize regression coefficients from the angle of optimality conditions with respect to the minimization of error. These results are concerned with approximating a single given random variable $Y$ by other random variables in the setting of a particular linear space of random variables. They do not relate yet to what happens in a *parametric* situation as explained in connection with factor models, where the approximation is to be controlled somehow for a whole range of parameter vectors $(x_1,\ldots,x_n)$. That is the topic in §5.

In the long run, it will be important from the angle of statistical estimation to explain also how these ideas can be supported by asymptotic results about the use of empirical data and sampling, rather than just the ideal approximations obtained with "true" random variables. Such results, which are well known, of course, for least squares and quantile regression, have been further investigated for financial prediction with constrained tail risk in Trindade et al. [21]. However, for our general context this will entail a sizable development beyond what can fit in this paper, so we reserve it for later.

**2. Measures of error, deviation, and risk.** In the background for our work, there is a space $\Omega$ of future states $\omega$ which has been supplied with the structure of a probability space: a field $\mathcal{A}$ of measurable sets on which there is a probability measure $P$. Random variables are construed as (measurable) functions $X$ from $\Omega$ to $\mathbb{R}$ such that $\int_\Omega |X(\omega)|^2\,dP(\omega)$ is finite. In other words, we focus on random variables as elements of the linear space $\mathcal{L}^2(\Omega,\mathcal{A},P)$, which we denote simply by $\mathcal{L}^2(\Omega)$. The choice of this space simplifies the picture of duality which will appear later and coordinates with the role of $\mathcal{L}^2$-concepts in classical statistics and linear regression.

The integral of $X$ with respect to $P$ is the expectation $EX$ of $X$. The standard deviation of $X$ is $\sigma(X) = \|X - EX\|_2$ for the norm $\|X\|_2 = (E[X^2])^{1/2}$. The symbol $C$ will typically stand for a constant in $\mathbb{R}$, or for the corresponding constant random variable; we simply write $X = C$, meaning technically that $X(\omega) = C$ almost surely. Similarly, for the essential supremum and infimum of $X$ we just write $\sup X$ and $\inf X$. Thus, $\inf X$ is the supremum of the set of all $C \in \mathbb{R}$ such that $X(\omega) \geq C$ almost surely, and likewise for $\sup X$.

Of immediate interest to us is the case where a random variable $X$ may represent an approximation error, as in linear regression. If $X = 0$, the error vanishes, but if $X \neq 0$, it may be assessed in different ways. By a *measure of error* (error quantifier), we will have in mind a functional $\mathcal{E}\colon \mathcal{L}^2(\Omega) \to [0, \infty]$ satisfying the axioms

(E1) $\mathcal{E}(0) = 0$ but $\mathcal{E}(X) > 0$ when $X \neq 0$; also, $\mathcal{E}(C) < \infty$ for constants $C$.
(E2) $\mathcal{E}(\lambda X) = \lambda \mathcal{E}(X)$ when $\lambda > 0$ (positive homogeneity).
(E3) $\mathcal{E}(X + X') \leq \mathcal{E}(X) + \mathcal{E}(X')$ for all $X$, $X'$ (subadditivity).
(E4) $\{X \in \mathcal{L}^2(\Omega) \mid \mathcal{E}(X) \leq c\}$ is closed for all $c < \infty$ (lower semicontinuity).

A closely related notion is that of a *measure of deviation* (deviation quantifier), which, in contrast to a measure of error, assesses the degree of uncertainty (inconstancy) in a random variable $X$. In that case we will have in mind a functional $\mathcal{D}\colon \mathcal{L}^2(\Omega) \to [0, \infty]$ satisfying the axioms

(D1) $\mathcal{D}(X) = 0$ for constant $X$, but $\mathcal{D}(X) > 0$ otherwise.
(D2) $\mathcal{D}(\lambda X) = \lambda \mathcal{D}(X)$ when $\lambda > 0$ (positive homogeneity).
(D3) $\mathcal{D}(X + X') \leq \mathcal{D}(X) + \mathcal{D}(X')$ for all $X$, $X'$ (subadditivity).
(D4) $\{X \in \mathcal{L}^2(\Omega) \mid \mathcal{D}(X) \leq c\}$ is closed for all $c < \infty$ (lower semicontinuity).

Axioms (E2) and (E3) and (D2) and (D3), respectively, imply that $\mathcal{E}$ and $\mathcal{D}$ are convex functionals on the linear space $\mathcal{L}^2(\Omega)$. For some purposes these conditions might be replaced in each case by a single axiom merely requiring that convexity so as to enlarge the definitions beyond the property of positive homogeneity in (E2) and (D2). But that property will be important to us, and we therefore hold back here from such an extension. Axiom (D1) has the consequence, shown in Rockafellar et al. [16], that

$$\mathcal{D}(X - C) = \mathcal{D}(X) \quad \text{for all constants } C. \tag{2.1}$$

On the other hand, the second part of (E1) ensures through (E3) that

$$\text{if } \mathcal{E}(X - C) < \infty \quad \text{for some } C, \qquad \text{then } \mathcal{E}(X - C) < \infty \quad \text{for every } C. \tag{2.2}$$

Axioms (E4) and (D4) require $\mathcal{E}$ and $\mathcal{D}$ to be lower semicontinuous as functionals on $\mathcal{L}^2(\Omega)$, this being a natural property which serves technical needs when working in that chosen space. When $\Omega$ consists only of a finite set of future states $\omega$, and $\mathcal{E}$ and $\mathcal{D}$ do not take on $\infty$, this lower semicontinuity is automatic.

The definitions do not require $\mathcal{E}$ or $\mathcal{D}$ to be symmetric: perhaps $\mathcal{E}(-X) \neq \mathcal{E}(X)$ or $\mathcal{D}(-X) \neq \mathcal{D}(X)$. This is essential in our framework because we want to allow for applications to finance in which gains and losses might not be viewed symmetrically. Some examples of such error measures are

$$\mathcal{E}(X) = \|a \max\{0, X\} + b \max\{0, -X\}\|_p \quad \text{with } a > 0, \quad b > 0, \quad 1 \leq p \leq \infty. \tag{2.3}$$

Here $||X||_p$ is the usual $\mathcal{L}^p$ norm (which is well defined on $\mathcal{L}^2(\Omega)$ but able to take on $\infty$ when $p > 2$, unless $\Omega$ is a finite set so that $\mathcal{L}^2(\Omega)$ is finite dimensional).

Many people think of the risk inherent in a financial random variable as tied entirely to the uncertainty (inconstancy) in that variable. Measures of deviation were introduced in our working paper (Rockafellar et al. [15]) in order to quantify that aspect of risk and keep it distinct from another aspect which has come to be quantified instead by "risk measures," most notably the coherent measures of risk developed by Artzner et al. [3, 4]. Risk measures in the vein of Artzner et al. [3, 4], will be important here as well, but our project requires us to concentrate on an "averse" class which differs in some crucial assumptions from the "coherent" class in Artzner et al. [3, 4]. The overlap is substantial, however. A full discussion of the relationships is available in Rockafellar et al. [16].

The class we need to work with is defined as follows. By an *averse measure of risk* (risk quantifier), we will mean a functional $\mathcal{R}\colon \mathcal{L}^2(\Omega) \to (-\infty, \infty]$ satisfying the axioms

(R1) $\mathcal{R}(C) = -C$ for constants $C$, but $\mathcal{R}(X) > E[-X]$ for nonconstant $X$.
(R2) $\mathcal{R}(\lambda X) = \lambda \mathcal{R}(X)$ when $\lambda > 0$ (positive homogeneity).
(R3) $\mathcal{R}(X + X') \leq \mathcal{R}(X) + \mathcal{R}(X')$ for all $X$, $X'$ (subadditivity).
(R4) $\{X \in \mathcal{L}^2(\Omega) \mid \mathcal{R}(X) \leq c\}$ is closed for all $c < \infty$ (lower semicontinuity).

This concept depends on a particular "orientation" being implicit in how random variables $X$ are regarded: positive outcomes $X(\omega)$ are deemed good (associated with profits or gains, say), whereas negative outcomes $X(\omega)$ are deemed bad (through connection with costs or losses). The value $\mathcal{R}(X)$ assigned to $X$, which we call the $\mathcal{R}$*-risk* of $X$, is intended as a numerical surrogate for the overall "badness" of $X$. This being the perspective, the definition is adopted as in Artzner et al. [4] that,

$$\text{the } \mathcal{R}\text{-risk in } X \text{ is } \textit{acceptable} \quad \text{if and only if} \quad \mathcal{R}(X) \le 0. \tag{2.4}$$

Axioms (R2), (R3), and (R4) guarantee, as in the case of the preceding functionals $\mathcal{E}$ and $\mathcal{D}$, that $\mathcal{R}$ is a lower semicontinuous, convex functional on $\mathcal{L}^2(\Omega)$ with $\mathcal{R}(0) = 0$. But Axiom (R1) implies

$$\mathcal{R}(X - C) = \mathcal{R}(X) + C \quad \text{for all constants } C, \tag{2.5}$$

which contrasts sharply with the property of deviations measures $\mathcal{D}$ in (2.1).

It was shown in Rockafellar et al. [16, Theorem 2] (and earlier in the working paper Rockafellar et al. [15]) that *averse measures of risk* $\mathcal{R}$ *and measures of deviation* $\mathcal{D}$ *correspond one-to-one* through the relations

$$\mathcal{R}(X) = E[-X] + \mathcal{D}(X), \qquad \mathcal{D}(X) = \mathcal{R}(X - EX). \tag{2.6}$$

The axioms for $\mathcal{D}$ and $\mathcal{R}$, as laid out here, differ slightly in their statement from the ones in Rockafellar et al. [16] but are equivalent to them on the basis of results in that paper. Our terminology in Rockafellar et al. [16] of a "strictly expectation bounded" measure of risk has been shortened to that of an "averse" measure of risk, which more accurately reflects the concept.

Connections with coherency have been fully worked out in Rockafellar et al. [16] and more will be said about this later. The axioms for coherency can be presented in various ways, but they differ essentially in that the strict inequality in (R1) is dropped (the weak inequality already follows from the combination of (R1) and (R3)), and on the other hand the monotonicity rule is added that $\mathcal{R}(X) \le \mathcal{R}(Y)$ when $X \ge Y$. Not every averse measure of risk is coherent, nor is every coherent measure of risk necessarily averse. An example of an averse measure of risk which lacks coherency (because it fails to obey the monotonicity rule), but is popular in finance and other situations that require safeguarding, is

$$\mathcal{R}(X) = E[-X] + \lambda\sigma(X), \quad \lambda > 0. \tag{2.7}$$

The corresponding measure of deviation, through (2.6), is obviously

$$\mathcal{D}(X) = \lambda\sigma(X). \tag{2.8}$$

The acceptability of the $\mathcal{R}$-risk in $X$ then revolves around having $EX$ be at least $\lambda$ standard deviation units above 0. To put it another way, $X$ is acceptable provided that the "losses" associated with realizations $X(\omega) < 0$ only occur in the tail of the distribution of $X$ that lies more than $\lambda$ standard deviation units below the expectation. Although good reasons abound for emphasizing risk measures that are coherent (see the tutorial paper by Rockafellar [12]), we are obviously obliged in studying regression to include $\sigma$-based deviation measures like (2.8), and therefore, through the one-to-one correspondence in (2.6), must allow incoherent risk measures like (2.7) to enter the picture.

An elementary example of a coherent measure of risk that is not averse is $\mathcal{R}(X) = E[-X]$. Acceptability merely corresponds in that case to having $EX \ge 0$. In general, *coherent* measures of risk are not obliged to satisfy the strict inequality in (R1), but on the other hand must obey the rule that $\mathcal{R}(X) \le \mathcal{R}(Y)$ when $X \ge Y$. (That fails for the risk measure in (2.7), in particular.)

More will soon be offered in the way of examples of deviation measures and risk measures, but we must now get back to measures of error, since they will be primary in our study of linear regression. Other explorations in the direction of quantifying risk by a functional $\mathcal{R}$ can be found in many places, including Acerbi [1], Bassett et al. [5], Ben Tal and Teboulle [6], Föllmer and Schied [8], and Ruszczyński and Shapiro [19, 20].

There are two fundamental modes of relationship between measures of error and measures of deviation. The simplest is to start with any measure of error $\mathcal{E}$ and take $\mathcal{D}(X) = \mathcal{E}(X - EX)$. Then $\mathcal{D}$ is a measure of deviation, as can readily be seen from the axioms. Examples in that direction have been analyzed in detail in Rockafellar et al. [16]. In this paper, our attention will be centered instead on a different way of obtaining a deviation measure $\mathcal{D}$ from an error measure $\mathcal{E}$. For that, we will want from $\mathcal{E}$ the following property, which we call *nondegeneracy*:

$$\inf_{X:\, EX = C} \mathcal{E}(X) > 0 \quad \text{for constants } C \ne 0. \tag{2.9}$$

All the examples in (2.3) are nondegenerate measures of error, inasmuch as $||X||_p \geq ||X||_1 \geq |EX|$. When $\Omega$ is a finite set, so that $\mathcal{L}^2(\Omega)$ is finite dimensional, nondegeneracy of $\mathcal{E}$ is an automatic consequence of axioms (E1)–(E4). When $\mathcal{L}^2(\Omega)$ is infinite dimensional, however, this property is not automatic, and there are examples for which it fails. A condition obviously sufficient for (2.9) is:

$$\text{there exists } \delta > 0 \quad \text{such that} \quad \mathcal{E}(X) \geq \delta |EX| \quad \text{for all } X. \tag{2.10}$$

The examples in (2.3) satisfy (2.10) with $\delta = \min\{a, b\}$, for instance. In fact, (2.10) is also necessary for nondegeneracy:

Conditions (2.9) and (2.10) are equivalent expressions of nondegeneracy.

(Let $\phi(C)$ denote the infimum in (2.9). Then from (E2) one has $\phi(C) = C\phi(1)$ for $C > 0$ and $\phi(C) = |C|\phi(-1)$ for $C < 0$, where $\phi(1) > 0$ and $\phi(-1) > 0$; therefore $\phi(C) \geq \delta|C|$ for some $\delta > 0$.)

Through the notion of nondegeneracy, we are able to extend, with a similar proof, a previous result in Rockafellar et al. [16, Theorem 4] which was articulated only for error measures $\mathcal{E}$ belonging to the category in (2.3). In this version we also say more about $\mathcal{S}(X)$ and provide a bound for it.

THEOREM 2.1 (ERROR PROJECTION). *For any nondegenerate error measure $\mathcal{E}$, let $\mathcal{D}$ be defined by*

$$\mathcal{D}(X) = \inf_C \mathcal{E}(X - C), \tag{2.11}$$

*where the minimization takes place over all $C \in \mathbb{R}$. Let $\mathcal{S}(X)$ be the set of $C$ for which the minimum is attained and finite*:

$$\mathcal{S}(X) = \arg\min_C \mathcal{E}(X - C). \tag{2.12}$$

*Then $\mathcal{D}$ is a deviation measure having $\{X \mid \mathcal{D}(X) < \infty\} = \{X \mid \mathcal{E}(X) < \infty\}$, and for each $X$ in that set, $\mathcal{S}(X)$ is a nonempty closed interval in $\mathbb{R}$ (likely reducing to just a single number). Moreover for $\delta > 0$ expressing nondegeneracy as in (2.10), there is the bound*:

$$C \in \mathcal{S}(X) \quad \textit{implies} \quad EX - \delta^{-1}\mathcal{E}(X) \leq C \leq EX + \delta^{-1}\mathcal{E}(X). \tag{2.13}$$

PROOF. First, fix $X$ and consider $\psi(C) = \mathcal{E}(X - C)$ as a function of $C \in \mathbb{R}$, noting that it is nonnegative by (E1), convex by (E2) and (E3), and lower semicontinuous by (E4). Clearly $\mathcal{D}(X) < \infty$ if and only if $\psi(C) < \infty$ for some $C$. Also $\psi(C) < \infty$ for every other constant $C$ as well; cf. (2.2). In particular, $\mathcal{D}(X) < \infty$ if and only if $\psi(0) < \infty$, which is the same as $\mathcal{E}(X) < \infty$. Thus, under the assumption henceforth on $X$ that $\mathcal{E}(X)$ is finite, we have $\psi$ finite on $\mathbb{R}$, and because $\psi$ inherits the convexity of $\mathcal{E}$, also continuous on $\mathbb{R}$. The set of $C$ where $\psi$ attains its minimum over $\mathbb{R}$, which is $\mathcal{S}(X)$, is therefore closed and convex, hence an interval if nonempty. It is the same as the set of $C$ where $\psi$ attains its minimum subject to the constraint $\psi(C) \leq \psi(0)$, so to be sure of nonemptiness we only need to know that the interval $\{C \mid \psi(C) \leq \psi(0)\} = \{C \mid \mathcal{E}(X - C) \leq \mathcal{E}(X)\}$ is bounded. That follows from the nondegeneracy of $\mathcal{E}$ as expressed in (2.10), which gives us $\mathcal{E}(X - C) \geq \delta|EX - C|$ and implies from $\mathcal{E}(X - C) \leq \mathcal{E}(X)$ that $|EX - C| \leq \delta^{-1}\mathcal{E}(X)$. That is the source also of (2.13).

We shift our attention now to properties of $\mathcal{D}$. It is elementary from (E2) and (E3) that $\mathcal{D}$ satisfies (D2) and (D3). Also, we have from (E1) and the formula (2.11) that $\mathcal{D}(X) = 0$ for constant $X$ and $\mathcal{D}(X) \geq 0$ otherwise; but we have to verify that this inequality is strict for nonconstant $X$. We have already seen the existence of $C$ such that $\mathcal{D}(X) = \mathcal{E}(X - C)$, namely any $C \in \mathcal{S}(X)$. But $\mathcal{E}(X - C) > 0$ by (E1) unless $X$ is itself the constant $C$. Hence indeed, $\mathcal{D}(X) > 0$ unless $X$ is constant.

The last task is confirming (D4). Due to (D2), we can concentrate on $c = 1$. Suppose we have a sequence of random variables $X_i$ satisfying $\mathcal{D}(X_i) \leq 1$ and converging in $\mathcal{L}^2(\Omega)$ to $\bar{X}$. We have to show that $\mathcal{D}(\bar{X}) \leq 1$ as well. For each $i$ there exists $C_i \in \mathcal{S}(X_i)$. Then $\mathcal{E}(X_i - C_i) = \mathcal{D}(X_i) \leq 1$; hence $\delta|EX_i - C_i| \leq 1$ by the nondegeneracy in (2.10). Since $EX_i$ converges to $E\bar{X}$, this implies that the sequence of constants $C_i$ is bounded. By passing to subsequences if necessary, we can suppose that $C_i$ converges to a constant $\bar{C}$ while $X_i$ converges to $\bar{X}$, in which case $X_i - C_i$ converges to $\bar{X} - \bar{C}$. Because $\mathcal{E}(X_i - C_i) \leq 1$ we also have $\mathcal{E}(\bar{X} - \bar{C}) \leq 1$ by (E4). On the other hand, $\mathcal{D}(\bar{X}) \leq \mathcal{E}(\bar{X} - \bar{C})$ by (2.11). We conclude that $\mathcal{D}(\bar{X}) \leq 1$, as required. □

In the circumstances of Theorem 2.1, we will call $\mathcal{D}(X)$ the *deviation of $X$ projected from $\mathcal{E}$*, and $\mathcal{S}(X)$ the *statistic of $X$ associated with $\mathcal{E}$*. We get from $\mathcal{S}(X)$ a best approximation of $X$ by a constant $C$, with respect to the criterion set by $\mathcal{E}$, and then from $\mathcal{D}(X)$ the magnitude of the residual uncertainty in that approximation.

Although $\mathcal{S}(X)$ may contain more than one $C$, we have by the definitions of $\mathcal{D}(X)$ and $\mathcal{E}(X)$ in (2.12) and (2.13) that

$$\mathcal{D}(X) = \mathcal{E}(X - C) \quad \text{for any } C \in \mathcal{S}(X), \qquad \text{but } \mathcal{D}(X) < \mathcal{E}(X - C) \quad \text{when } C \notin \mathcal{S}(X).$$

It will be valuable later, in appreciating how generalized linear regression effectively operates, to note that this implies

$$\mathcal{D}(X) = \mathcal{E}(X) \quad \text{when } 0 \in \mathcal{S}(X), \qquad \text{but } \mathcal{D}(X) < \mathcal{E}(X) \quad \text{when } 0 \notin \mathcal{S}(X). \tag{2.14}$$

Moreover the rules hold that

$$\begin{aligned} \mathcal{S}(X - C) &= \mathcal{S}(X) - C \quad \text{for any constant } C, \\ \mathcal{S}(\lambda X) &= \lambda \mathcal{S}(X) \quad \text{for any } \lambda > 0. \end{aligned} \tag{2.15}$$

Because $\mathcal{S}(X)$ is a nonempty closed interval in $\mathbb{R}$, we can give it, when convenient, the expression

$$\mathcal{S}(X) = [\mathcal{S}^-(X), \mathcal{S}^+(X)] \quad \text{for} \begin{cases} \mathcal{S}^-(X) = \min\{C \mid C \in \mathcal{S}(X)\}, \\ \mathcal{S}^+(X) = \max\{C \mid C \in \mathcal{S}(X)\}. \end{cases} \tag{2.16}$$

Much of the time this interval reduces to a single number; however, we have $\mathcal{S}^-(X) = \mathcal{S}^+(X)$, and in that case are able to use the same notation $\mathcal{S}(X)$ to stand for that number.

Example 2.1 (Standard Deviation). Let $\mathcal{E}(X) = \lambda \|X\|_2$ for an arbitrary $\lambda > 0$. Then $\mathcal{E}$ is a nondegenerate measure of error for which the projected measure of deviation $\mathcal{D}$ is given by (2.8) and its corresponding measure of risk by (2.7). The associated statistic is $\mathcal{S}(X) = EX$.

Example 2.2 (Range Deviation). Let $\mathcal{E}(X) = \|X\|_\infty$. Then $\mathcal{E}$ is a nondegenerate measure of error, and $\mathcal{E}(X) < \infty$ if and only if $X$ satisfies both $\sup X < \infty$ and $\inf X > -\infty$. The measure of deviation $\mathcal{D}$ projected from $\mathcal{E}$, and the associated statistic $\mathcal{S}$ for bounded $X$, are the *radius* of the range of $X$, and its *center*, given by

$$\mathcal{D}(X) = \tfrac{1}{2}[\sup X - \inf X], \qquad \mathcal{S}(X) = \tfrac{1}{2}[\sup X + \inf X].$$

Example 2.3 (Median Absolute Deviation). Let $\mathcal{E}(X) = \|X\|_1 = E|X|$. Then $\mathcal{E}$ is a nondegenerate measure of error, and $\mathcal{E}(X) < \infty$ for all $X \in \mathcal{L}^2(\Omega)$. The statistic $\mathcal{S}(X)$ is $\mathrm{med}(X)$, the *median* of $X$, and the projected measure of deviation is $\mathcal{D}(X) = E|X - \mathrm{med}(X)|$ under the interpretation that this expression stands for $E|X - C|$ for an arbitrary choice of $C \in \mathrm{med}(X)$ (the particular choice makes no difference).

The notation in this example requires some explaining, as it will also come up in other situations later. We are coping with the fact that $\mathrm{med}(X)$ is in principle an interval which, in the manner of (2.16), could be written as $[\mathrm{med}^-(X), \mathrm{med}^+(X)]$. It is conventional to select just the value $\mathrm{med}^-(X)$ from this interval and dub it as "the median," but we prefer, at least in this paper, to think of the whole interval as constituting "the median," because that is what fits the framework of deriving medians from error minimization.

The next example brings in the *value-at-risk* $\mathrm{VaR}_\alpha(X)$ and *conditional value-at-risk* $\mathrm{CVaR}_\alpha(X)$ of a random variable $X$ at a probability level $\alpha$. We refer to Rockafellar and Uryasev [14] and Föllmer and Schied [8] for more on these concepts.[1] These notions are tied to quantiles and, as we shall see later, to quantile regression. In order to define them, we consider for any random variable $X$ its cumulative distribution function $F_X$ and, for any $\alpha \in (0, 1)$, the interval

$$q_\alpha(X) = [q_\alpha^-(X), q_\alpha^+(X)] \quad \text{for} \begin{cases} q_\alpha^-(X) = \inf\{x \mid F_X(x) \geq \alpha\}, \\ q_\alpha^+(X) = \sup\{x \mid F_X(x) \leq \alpha\}. \end{cases}$$

We call this the *$\alpha$-quantile (interval)* for $X$, although the convention in statistics is to fix on just $q_\alpha^-(X)$ as the "quantile" in question. Note that $q_\alpha(X)$ becomes $\mathrm{med}(X)$ when $\alpha = 1/2$.

With the orientation we have imposed on random variables in connection with evaluating "risk" (positive values good, negative values bad), the value-at-risk of $X$ at level $\alpha$ is

$$\mathrm{VaR}_\alpha(X) = -q_\alpha^+(X) = q_{1-\alpha}^-(-X),$$

[1] Conditional value-at-risk and the CVaR notation originated in Rockafellar and Uryasev [13], but other terms that have come into play for the same concept, under varying degrees of generality, are expected shortfall (Acerbi and Tasche [2]) and average value-at-risk (Föllmer and Schied [8]).

while $\mathrm{CVaR}_\alpha(X)$ is the conditional expectation of the "loss" $-X$ in the lower $\alpha$-tail of its distribution (having probability $\alpha$). But the definition of that tail requires careful description when a probability atom sits at its endpoint or, in other words, when $F_X$ has a jump at $-\mathrm{VaR}_\alpha(X)$; see Rockafellar and Uryasev [14]. Through a change of variables, this develops into the formula that

$$\mathrm{CVaR}_\alpha(X) = \frac{1}{\alpha}\int_0^\alpha \mathrm{VaR}_\beta(X)\,d\beta.$$

Anyway, the precise relationships can be expressed most interestingly for our purposes through error projection, as follows.

Example 2.4 (CVaR Deviation and Quantiles). For any choice of $\alpha \in (0, 1)$, the formula

$$\mathscr{E}(X) = E[\max\{0, X\} + (\alpha^{-1} - 1)\max\{0, -X\}] \tag{2.17}$$

defines a nondegenerate measure of error for which the projected measure of deviation is

$$\mathscr{D}(X) = \mathrm{CVaR}_\alpha(X - EX),$$

corresponding to the averse measure of risk

$$\mathscr{R}(X) = \mathrm{CVaR}_\alpha(X),$$

and the associated statistic is

$$\mathscr{S}(X) = q_\alpha(X).$$

Observe that the expression for $\mathscr{E}$ in (2.17) is the case of (2.3) for $a = 1$, $b = \alpha^{-1} - 1$, and $p = 1$. The connection in Example 2.4 between $\mathscr{E}$ and $\mathscr{S}$ goes to the origins of quantile regression (cf. Koenker and Bassett [10], Koenker [9]), but the CVaR tie-in with $\mathscr{D}$ and $\mathscr{E}$ was first brought out in Rockafellar and Uryasev [13, 14].

There can well be more than one error measure $\mathscr{E}$ projecting onto a given deviation measure $\mathscr{D}$. Uniqueness cannot be expected, and is not necessarily even desirable. This is underscored by the next example.

Example 2.5 (Inverse Projection). For any $\mathscr{D}$ deviation measure, let $\mathscr{E}(X) = \mathscr{D}(X) + |EX|$. Then $\mathscr{E}$ is a nondegenerate measure of error projecting to $\mathscr{D}$, and the associated statistic is $\mathscr{S}(X) = EX$.

Example 2.5 settles the question of whether every deviation measure can be viewed as coming from *some* nondegenerate error measure in the manner of (2.11), but the error measure it furnishes might not always be the most interesting or convenient. The deviation measures $\mathscr{D}$ in Examples 2.3, 2.4, and 2.5, for instance, arise also from other error measures $\mathscr{E}$ of deeper significance. This raises a challenge. Ways are known for combining known measures of deviation or risk to get new ones, but how should these operations be coordinated with corresponding "natural" measures of error?

For example, given probability levels $\alpha_k \in (0, 1)$ for $k = 1, \ldots, r$ and weights $\lambda_k > 0$ adding to 1, the *mixed* CVaR formula

$$\mathscr{R}(X) = \lambda_1 \mathrm{CVaR}_{\alpha_1}(X) + \cdots + \lambda_r \mathrm{CVaR}_{\alpha_r}(X) \tag{2.18}$$

defines an averse measure of risk which is partnered with the mixed CVaR deviation measure

$$\mathscr{D}(X) = \lambda_1 \mathrm{CVaR}_{\alpha_1}(X - EX) + \cdots + \lambda_r \mathrm{CVaR}_{\alpha_r}(X - EX). \tag{2.19}$$

But is there an error measure $\mathscr{E}$ which projects to $\mathscr{D}$ in a manner akin to that in Example 2.4? The answer is yes. It comes out of the new and much broader result about error projection which we establish next.

Theorem 2.2 (Mixed Deviations and Weighted Statistic). *For $k = 1, \ldots, r$ let $\mathscr{D}_k$ be a measure of deviation, with $\mathscr{R}_k$ as the corresponding averse measure of risk, and let $\mathscr{E}_k$ be a nondegenerate measure of error that projects to $\mathscr{D}_k$. Then, for any weights $\lambda_k > 0$ adding to 1,*

$$\mathscr{D}(X) = \lambda_1 \mathscr{D}_1(X) + \cdots + \lambda_r \mathscr{D}_r(X)$$

*is a measure of deviation for which the corresponding averse measure of risk is*

$$\mathscr{R}(X) = \lambda_1 \mathscr{R}_1(X) + \cdots + \lambda_r \mathscr{R}_r(X),$$

*and the formula*

$$\mathscr{E}(X) = \inf_{\substack{C_1, \ldots, C_r \\ \lambda_1 C_1 + \cdots + \lambda_r C_r = 0}} \{\lambda_1 \mathscr{E}_1(X - C_1) + \cdots + \lambda_r \mathscr{E}_r(X - C_r)\} \tag{2.20}$$

*defines a nondegenerate measure of error which projects to $\mathscr{D}$ and furnishes, as the associated statistic,*

$$\mathscr{S}(X) = \lambda_1 \mathscr{S}_1(X) + \cdots + \lambda_r \mathscr{S}_r(X). \tag{2.21}$$

PROOF. The initial assertions about $\mathcal{D}$ and $\mathcal{R}$ are evident from (2.6) and the axioms (and were noted previously in Rockafellar et al. [16]), but the existence of a corresponding error measure $\mathcal{E}$ with the specified formula is new. The fact that $\mathcal{E}$, defined by (2.20), is nonnegative with $\mathcal{E}(0) = 0$ and $\mathcal{E}(C) < \infty$, and satisfies (E2) and (E3), is elementary from those properties holding for each $\mathcal{E}_k$. The calculation that $\mathcal{E}$ projects to $\mathcal{D}$ is also straightforward:

$$\begin{aligned}
\inf_C \mathcal{E}(X - C) &= \inf_{\substack{C, C_1, \ldots, C_r \\ \lambda_1 C_1 + \cdots + \lambda_r C_r = 0}} \{\lambda_1 \mathcal{E}_1(X - C - C_1) + \cdots + \lambda_r \mathcal{E}_r(X - C - C_r)\} \\
&= \inf_{\substack{C, C_1', \ldots, C_r' \\ \lambda_1 [C_1' - C] + \cdots + \lambda_r [C_r' - C] = 0}} \{\lambda_1 \mathcal{E}_1(X - C_1') + \cdots + \lambda_r \mathcal{E}_r(X - C_r')\} \\
&= \inf_{\substack{C, C_1', \ldots, C_r' \\ \lambda_1 C_1' + \cdots + \lambda_r C_r' = C}} \{\lambda_1 \mathcal{E}_1(X - C_1') + \cdots + \lambda_r \mathcal{E}_r(X - C_r')\} \\
&= \inf_{C_1', \ldots, C_r'} \{\lambda_1 \mathcal{E}_1(X - C_1') + \cdots + \lambda_r \mathcal{E}_r(X - C_r')\} \\
&= \lambda_1 \inf_{C_1'} \mathcal{E}_1(X - C_1') + \cdots + \lambda_r \inf_{C_r'} \mathcal{E}_r(X - C_1') \\
&= \lambda_1 \mathcal{D}_1(X) + \cdots + \lambda_r \mathcal{D}_r(X) = \mathcal{D}(X).
\end{aligned}$$

Moreover, $\mathcal{S}(X)$ is revealed in this way as consisting of the values of $\lambda_1 C_1' + \cdots + \lambda_r C_r'$ generated by taking $C_k' \in \mathcal{S}_k(X)$, thus confirming (2.21). Also, the infimum in (2.20) is attained when finite.

We show next that $\mathcal{E}$ is nondegenerate. The nondegeneracy of each $\mathcal{E}_k$ provides through (2.10) a $\delta_k > 0$ such that $\mathcal{E}_k(X) \geq \delta_k |EX|$. Choose $\delta = \min\{\delta_1, \ldots, \delta_r\}$, so that $\mathcal{E}_k(X) \geq \delta |EX|$ for every $k$. Then, in recalling that the $\lambda_k$'s in (2.20) are positive weights adding 1, we get (through the convexity of the absolute value function on $\mathbb{R}$, and the constraint $\lambda_1 C_1 + \cdots + \lambda_r C_r = 0$) that

$$\begin{aligned}
\lambda_1 \mathcal{E}_1(X - C_1) + \cdots + \lambda_r \mathcal{E}_r(X - C_r) &\geq \lambda_1 \delta |EX - C_1| + \cdots + \lambda_r \delta |EX - C_r| \\
&\geq \delta |\lambda_1 (EX - C_1) + \cdots + \lambda_r (EX - C_r)| = \delta |EX - [\lambda_1 C_1 + \cdots + \lambda_r C_r]| = \delta |EX|.
\end{aligned}$$

In the minimum, therefore, we have $\mathcal{E}(X) \geq \delta |EX|$ as in (2.10), so $\mathcal{E}$ is nondegenerate.

We still have to confirm that $\mathcal{E}$ obeys (E4). In view of (E2), that amounts to considering a sequence of random variables $X_i$ satisfying $\mathcal{E}(X_i) \leq 1$ which converges to a random variable $\overline{X}$, and then demonstrating that $\mathcal{E}(\overline{X}) \leq 1$. We already know that the infimum in the formula defining $\mathcal{E}(X_i)$ is attained. Thus, there are constants $C_{ik}$ such that $\lambda_1 C_{i1} + \cdots + \lambda_r C_{ir} = 0$ and

$$\lambda_1 \mathcal{E}_1(X_i - C_{i1}) + \cdots + \lambda_r \mathcal{E}_r(X_i - C_{ir}) = \mathcal{E}(X_i) \leq 1. \tag{2.22}$$

We also have, from our dealings with nondegeneracy, a $\delta > 0$ such that $\mathcal{E}_k(X_i - C_{ik}) \geq \delta |EX - C_{ik}|$ for every $k$. Because $|EX_i - C_{ik}| \geq |C_{ik}| - E|X_i|$, this entails through (2.22) that

$$1 \geq \lambda_1 \delta |E(X_i - C_{i1})| + \cdots + \lambda_r \delta |E(X_i - C_{ir})| \geq \delta(\lambda_1 |C_{i1}| + \cdots + \lambda_r |C_{ir}| - E|X_i|),$$

or, in other words, $\lambda_1 |C_{i1}| + \cdots + \lambda_r |C_{ir}| \leq 1 + E|X_i|$. Since $X_i$ converges to $\overline{X}$, we have $E|X_i|$ converging to $E|\overline{X}|$, and the sequence of vectors $(C_{i1}, \ldots, C_{ir}) \in \mathbb{R}^r$ must therefore be bounded. We can suppose, by passing to a subsequence if necessary, that it converges to some $(\overline{C}_1, \ldots, \overline{C}_r) \in \mathbb{R}^r$. Then from (2.22), the lower semicontinuity of each $\mathcal{E}_k$, and the formula for $\mathcal{E}(\overline{X})$, we have

$$1 \geq \lambda_1 \mathcal{E}_1(\overline{X} - \overline{C}_1) + \cdots + \lambda_r \mathcal{E}_r(\overline{X} - \overline{C}_r) \geq \mathcal{E}(\overline{X}),$$

which furnishes the inequality we needed to establish. □

The prescription in Theorem 2.2 can be applied in the case where $\mathcal{E}_k$ is the quantile error measure for $\alpha_k$ in (2.17), which, because $\max\{0, X\} - \max\{0, -X\} = X$, can be expressed equivalently in the form

$$\mathcal{E}_k(X) = EX + \alpha_k^{-1} E[\max\{0, -X\}].$$

The formula in (2.20) then yields to simplifications, and the following result is obtained.

EXAMPLE 2.6 (MIXED QUANTILES AND MIXED CVaR DEVIATION). For any choice of probability thresholds $\alpha_k \in (0,1)$ and weights $\lambda_k > 0$ adding to 1, the formula

$$\mathcal{E}(X) = EX + \inf_{\substack{C_1,\dots,C_r \\ \lambda_1 C_1 + \cdots + \lambda_r C_r = 0}} \left\{ \frac{\lambda_1}{\alpha_1} E[\max\{0, C_1 - X\}] + \cdots + \frac{\lambda_r}{\alpha_r} E[\max\{0, C_r - X\}] \right\} \tag{2.23}$$

defines a nondegenerate error measure which projects to the mixed CVaR deviation measure $\mathcal{D}$ in (2.19). The associated statistic is

$$\mathcal{S}(X) = \lambda_1 q_{\alpha_1}(X) + \cdots + \lambda_r q_{\alpha_r}(X). \tag{2.24}$$

Weighted combinations of quantiles have a considerable history in statistics, as recounted by Koenker in [9, Chapter 5]. However, their usage there has revolved mainly around certain "smoothing effects" in estimation, without the naturally associated error measure in (2.23) ever coming up. In fact there is much more to the topic than this, because the mixed-CVaR risk measures and deviation measures in (2.18) and (2.19) are deeply connected with representations of subjective risk aversion; see Acerbi [1] and Rockafellar et al. [16, Proposition 5]. The error measure in (2.23) will be especially important for us later when we deal with regression in "factor models" of the sort mentioned in our introduction.

We finish this section with an application of the same idea to the symmetrization of a deviation measure.

EXAMPLE 2.7 (SYMMETRIZED DEVIATION). Let $\mathcal{D}$ be a deviation measure projected from a nondegenerate error measure $\mathcal{E}$, and let $\widehat{\mathcal{D}}$ be the symmetrization of $\mathcal{D}$:

$$\widehat{\mathcal{D}}(X) = \tfrac{1}{2}[\mathcal{D}(X) + \mathcal{D}(-X)].$$

Then the formula

$$\widehat{\mathcal{E}}(X) = \tfrac{1}{2} \inf_C \{\mathcal{E}(X - C) + \mathcal{E}(C - X)\}$$

furnishes a nondegenerate error measure which projects to $\widehat{\mathcal{D}}$.

This is the case in Theorem 2.2, where $\mathcal{D}_1(X) = \mathcal{D}(X)$, $\mathcal{D}_2(-X) = \mathcal{D}(-X)$, and $\lambda_1 = \lambda_2 = 1/2$. Incidentally, the averse measure of risk corresponding to the symmetrized deviation measure $\widehat{\mathcal{D}}$ is $\hat{R}(X) = (1/2)[\mathcal{R}(X) + \mathcal{R}(-X)] - EX$.

**3. Linear regression.** With this background at our disposal, we are ready to look at how regression might be carried out with a general measure of error and what implications that might have in terms of risk and deviation.

We take $\mathcal{E}$ now to be any measure of error and consider a random variable $Y$ which is to be approximated by a linear combination of given random variables $X_j$, $j = 1, \dots, n$, and a constant term. Specifically, we wish to solve the following *linear regression* problem:

$$\mathcal{LR}(Y) \qquad \begin{array}{l} \text{choose } c_0, c_1, \dots, c_n \text{ to minimize } \mathcal{E}(Z(c_0, c_1, \dots, c_n)), \\ \text{where } Z(c_0, c_1, \dots, c_n) = Y - [c_0 + c_1 X_1 + \cdots + c_n X_n]. \end{array}$$

Here $Z(c_0, c_1, \dots, c_n)$ is the error in approximating $Y$ by $c_0 + c_1 X_1 + \cdots + c_n X_n$, seen as a *random variable*, whereas $\mathcal{E}(Z(c_0, c_1, \dots, c_n))$ is the *magnitude* of this random variable, as assessed by $\mathcal{E}$.

Note that because $\mathcal{E}$ might not be symmetric, minimizing the magnitude of $Z(c_0, c_1, \dots, c_n)$ need not turn out to be the same as minimizing the magnitude of the random variable

$$-Z(c_0, c_1, \dots, c_n) = [c_0 + c_1 X_1 + \cdots + c_n X_n] - Y.$$

Minimizing $\mathcal{E}(-Z(c_0, c_1, \dots, c_n))$ would amount to the regression problem for $-Y$ with respect to $-X_1, \dots, -X_n$, which could be different in this general setting. The possible asymmetry might be helpful in situations where there is more concern about overestimating $Y$ than underestimating it. Positive realizations of $Z(c_0, c_1, \dots, c_n)$ could be penalized less severely than negative realizations.

Due to the convexity of $\mathcal{E}$ as a functional on $\mathcal{L}^2(\Omega)$, coming from axioms (E2) and (E3), $\mathcal{LR}(Y)$ is an optimization problem of convex type in $\mathbb{R}^{n+1}$. This is valuable in drawing a number of conclusions.

THEOREM 3.1 (EXISTENCE AND POTENTIAL UNIQUENESS OF REGRESSION COEFFICIENTS). *As long as* $\mathscr{E}(Y) < \infty$, *the solution set*

$$\mathscr{C}(Y) = \{(\bar{c}_0, \bar{c}_1, \ldots, \bar{c}_n) \text{ furnishing the minimum in problem } \mathscr{LR}(Y)\}$$

*is a nonempty, closed, convex subset of* $\mathbb{R}^{n+1}$. *It is bounded if and only if the random variables* $X_j$ *satisfy the linear independence condition that*

$$c_1X_1 + \cdots + c_nX_n \quad \text{is not constant unless } c_1 = \cdots = c_n = 0. \tag{3.1}$$

*In those circumstances,* $\mathscr{C}(Y)$ *still might not reduce to a single vector, but this does hold in the special case where* $\mathscr{E}$ *has, beyond* (E3), *the property that*

$$\mathscr{E}(X + X') < \mathscr{E}(X) + \mathscr{E}(X') \quad \text{for nonzero } X \text{ and } X'$$
$$\text{as long as } X' \text{ is not just a positive multiple of } X. \tag{3.2}$$

PROOF. Let $f(d, c_0, \ldots, c_n) = \mathscr{E}(dY - [c_0 + c_1X_1 + \cdots + c_nX_n])$. The axioms (E1)–(E4) on $\mathscr{E}$ imply that $f$ is convex, positively homogeneous, and lower semicontinuous on $\mathbb{R}^{n+2}$ with $f(1, 0, \ldots, 0) < \infty$ and $f(0, 0, \ldots, 0) < \infty$. We are concerned with minimizing $f$ subject to the constraint that $d = 1$.

The solution set is guaranteed to be convex and closed, but is it nonempty? A sufficient condition for that is available from convex analysis in terms of the so-called recession vectors $(b, a_0, \ldots, a_n)$ for $f$, which are characterized equivalently at any point $(d, c_0, \ldots, c_n)$ where $f$ is finite, by the condition that $f(d + tb, c_0 + ta_0, \ldots, c_n + ta_n)$ is nonincreasing as a function of $t > 0$. Invoking that characterization at $(d, c_0, \ldots, c_n) = (0, 0, \ldots, 0)$ and recalling the formula for $f$, we see that $(b, a_0, \ldots, a_n)$ is a recession vector if and only if $\mathscr{E}(tbY - [ta_0 + ta_1X_1 + \cdots + ta_nX_n])$ is nondecreasing as a function of $t > 0$, which in view of (E2) is the same as $\mathscr{E}(bY - [a_0 + a_1X_1 + \cdots + a_nX_n]) = 0$ and, through (E1), $bY - [a_0 + a_1X_1 + \cdots + a_nX_n] = 0$.

The criterion we wish to invoke from convex analysis is the following (Rockafellar [11, Corollary 27.3.1]) $f$ attains its minimum relative to the constraint $d = 1$ if every recession vector of the form $(0, a_0, \ldots, a_n)$ also has $-(0, a_0, \ldots, a_n)$ as a recession vector. This results in saying that whenever $a_0 + a_1X_1 + \cdots + a_nX_n = 0$, then $(-a_0) + (-a_1)X_1 + \cdots + (-a_n)X_n = 0$, which is trivial. Hence $\mathscr{C}(Y) \neq \varnothing$.

The boundedness of $\mathscr{C}(Y)$ corresponds to the boundedness of the set where $f$ attains its minimum relative to the constraint $d = 1$, and that, too, has a characterization in terms of recession vectors (Rockafellar [11, Theorem 27.1(d)]). It holds if and only if there is no nonzero recession vector. That reduces to the condition in (3.1).

The question now is whether $\mathscr{C}(Y)$, although bounded, might contain different solutions $(\bar{c}_0, \ldots, \bar{c}_n)$ and $(\tilde{c}_0, \ldots, \tilde{c}_n)$. Then $Z(\bar{c}_0, \ldots, \bar{c}_n)$ and $Z(\tilde{c}_0, \ldots, \tilde{c}_n)$ have the same $\mathscr{E}$-magnitude, say $\alpha$, and by (E2) cannot just be positive multiples of each other unless $\alpha = 0$, which is impossible. (That would make $Z(\bar{c}_0, \ldots, \bar{c}_n) = 0$ and $Z(\tilde{c}_0, \ldots, \tilde{c}_n) = 0$ by (E1), implying that $\bar{c}_0 + \bar{c}_1X_1 + \cdots + \bar{c}_nX_n = \tilde{c}_0 + \tilde{c}_1X_1 + \cdots + \tilde{c}_nX_n$ and furnishing a contradiction to (3.1).) Consider

$$(c_0, \ldots, c_n) = \tfrac{1}{2}(\bar{c}_0, \ldots, \bar{c}_n) + \tfrac{1}{2}(\tilde{c}_0, \ldots, \tilde{c}_n),$$

noting that $2Z(c_0, \ldots, c_n) = Z(\bar{c}_0, \ldots, \bar{c}_n) + Z(\tilde{c}_0, \ldots, \tilde{c}_n)$. We then have, by (E2) and (3.2),

$$2\alpha \leq 2\mathscr{E}(Z(c_0, \ldots, c_n)) < \mathscr{E}(Z(\bar{c}_0, \ldots, \bar{c}_n)) + \mathscr{E}(Z(\tilde{c}_0, \ldots, \tilde{c}_n)) = \alpha + \alpha,$$

which is an inconsistent inequality revealing that our assumption of the existence of two different solutions is untenable. □

Although the additional condition in (3.2) holds for some measures of error, such as in the classical case of Example 2.1 where $\mathscr{E}$ is the $\mathscr{L}^2$ norm, it fails in many other cases of significance, such as the quantile error measure in Example 2.2. Therefore, the possibility of nonuniqueness cannot be swept aside.

A noteworthy consequence of the positive homogeneity property (E2) of $\mathscr{E}$ in problem $\mathscr{LR}(Y)$ is the obvious scaling effect it has on solutions:

$$\mathscr{C}(\lambda Y) = \lambda\mathscr{C}(Y) \quad \text{for any } \lambda > 0.$$

In other words, if the random variable $Y$ being approximated is rescaled by $\lambda$, the regression coefficients are simply rescaled accordingly, and the *ratios* among them are unchanged. Shifting $Y$ to $Y - C$ likewise leaves these ratios unchanged for $k = 1, \ldots, n$ but moves $\bar{c}_0$ to $\bar{c}_0 - C$.

Thus, the constant $\bar{c}_0$ determined in the regression has a distinct role of its own, which we discuss next. We let

$$Z_0(c_1, \ldots, c_n) = Y - [c_1X_1 + \cdots + c_nX_n], \quad \text{so that} \quad Z(c_0, c_1, \ldots, c_n) = Z_0(c_1, \ldots, c_n) - c_0. \tag{3.3}$$

THEOREM 3.2 (ERROR-SHAPING DECOMPOSITION OF REGRESSION). *Let $\mathcal{E}$ be a nondegenerate measure of error, let $\mathcal{D}$ be the projected measure of deviation, and let $\mathcal{S}$ be the associated statistic. Then, in the notation (3.3), the solutions to problem $\mathcal{LR}(Y)$ are characterized by the prescription that*

$$(\bar{c}_0, \bar{c}_1, \ldots, \bar{c}_n) \in \mathcal{C}(Y) \quad \text{if and only if} \quad \begin{cases} (\bar{c}_1, \ldots, \bar{c}_n) \in \arg\min\limits_{c_1,\ldots,c_n} \mathcal{D}(Z_0(c_1, \ldots, c_n)), \\ \bar{c}_0 \in \mathcal{S}(Z_0(\bar{c}_1, \ldots, \bar{c}_n)). \end{cases}$$

*In particular, the regression coefficients $\bar{c}_1, \ldots, \bar{c}_n$ depend only on $\mathcal{D}$. Indeed, problem $\mathcal{LR}(Y)$ is equivalent to the following*:

$$\begin{aligned} &\text{choose } c_0, c_1, \ldots, c_n \text{ to minimize } \mathcal{D}(Z(c_0, c_1, \ldots, c_n)) \\ &\text{subject to the constraint that } 0 \in \mathcal{S}(Z(c_0, c_1, \ldots, c_n)). \end{aligned} \tag{3.4}$$

PROOF. The argument is based on a standard principle. Suppose we have a function $g(c_0, c_1, \ldots, c_n)$ on $\mathbb{R}^{n+1}$. Let $h(c_1, \ldots, c_n) = \inf_{c_0} g(c_0, c_1, \ldots, c_n)$ be its inf-projection on $\mathbb{R}^n$. Then, in order to have $(\bar{c}_0, \bar{c}_1, \ldots, \bar{c}_n)$ minimize $g$, it is necessary and sufficient to have $(\bar{c}_1, \ldots, \bar{c}_n)$ minimize $h$ and then take $\bar{c}_0$ to minimize $g(c_0, \bar{c}_1, \ldots, \bar{c}_n)$. Here we are merely applying this principle to $g(c_0, c_1, \ldots, c_n) = \mathcal{E}(Z(c_0, c_1, \ldots, c_n))$ and invoking the formulas for the corresponding $\mathcal{D}$ and $\mathcal{S}$ in Theorem 2.1, with $c_0$ replacing $C$. Through (2.15) we have $\bar{c}_0 \in \mathcal{S}(Z_0(\bar{c}_1, \ldots, \bar{c}_n))$ if and only if $0 \in \mathcal{S}(Z_0(\bar{c}_1, \ldots, \bar{c}_n) - \bar{c}_0) = \mathcal{S}(Z(\bar{c}_0, \bar{c}_1, \ldots, \bar{c}_n))$. □

This basic result provides important insights into the capabilities of generalized linear regression as a method for "shaping" the uncertain error $Z(c_0, c_1, \ldots, c_n) = Y - [c_0 + c_1X_1 + \cdots + c_nX_n]$ in contrast to "keeping it small." Further insights into the formulation in (3.4) are gained through the rule in (2.14), according to which

$$\mathcal{D}(Z(c_0, c_1, \ldots, c_n)) = \mathcal{E}(Z(c_0, c_1, \ldots, c_n)) \quad \text{when } 0 \in \mathcal{S}(Z(c_0, c_1, \ldots, c_n)).$$

Once the "statistic" of the error is fixed at 0 by the constraint in (3.4), the rest of the approximation revolves around adjusting the error so that it differs from that statistic as little as possible with respect to the penalties imposed by $\mathcal{E}$ for negative outcomes of $Z(c_0, c_1, \ldots, c_n)(\omega)$, which correspond to overapproximations of $Y(\omega)$ versus positive outcomes $Z(c_0, c_1, \ldots, c_n)(\omega)$, which correspond to underapproximations of $Y(\omega)$.

In the familiar case of standard linear regression, corresponding to $\mathcal{E}$, $\mathcal{D}$, and $\mathcal{S}$ as in Example 2.1, the regression coefficients are chosen to *minimize the standard deviation (or variance) of the error subject to the expected value of the error being* 0. For the range deviation in Example 2.2, regression seeks to choose $(c_0, c_1, \ldots, c_n)$ to *shrink the range of the error as much as possible while keeping it centered at* 0. For the median absolute deviation in Example 2.3, in the case where the median reduces to a single value rather than an interval (for simplicity of interpretation), the task instead is to *minimize the $\mathcal{L}^1$-magnitude of the error subject to its median being* 0.

This generalizes to quantile regression beyond the median, with $\mathcal{E}$, $\mathcal{D}$ and $\mathcal{S}$ as in Example 2.4 for some $\alpha \in (0, 1)$. In fact this is the idea of quantile regression, going back to Koenker and Bassett [10] and explained at length in Koenker [9]. Again supposing for simplicity that the quantiles reduce from intervals to single values, the regression coefficients are chosen in this case to *minimize the* $\text{CVaR}_\alpha$ *deviation of the error subject to the $\alpha$-quantile of the probability distribution of the error being* 0, or in other words, subject to the probability of overestimation being just $\alpha$.

This feature of quantile regression extends in a similar manner to regression with mixed CVaR deviation as in (2.19), as we now demonstrate.

EXAMPLE 3.1 (MIXED QUANTILE REGRESSION). For $k = 1, \ldots, r$ consider probability thresholds $\alpha_k \in (0, 1)$ and weights $\lambda_k > 0$ adding to 1, and let $\mathcal{D}(X)$ be the corresponding mixed CVaR deviation in (2.19). Let $\mathcal{S}(X)$ be the mixed quantile statistic in (2.24). Then the error-shaping problem in (3.3) is the linear regression problem with respect to the error measure $\mathcal{E}$ in (2.23). Furthermore, this problem can be solved as follows. First,

$$\begin{aligned} &\text{minimize } E\left[Y - \sum_{j=1}^n c_jX_j\right] + \sum_{k=1}^r \lambda_k \left(\alpha_k^{-1} E\left[\max\left\{0, C_k - \sum_{j=1}^n c_jX_j\right\}\right] - C_k\right) \\ &\text{over all choices of } c_1, \ldots, c_n, C_1, \ldots, C_r, \text{ obtaining } \bar{c}_1, \ldots, \bar{c}_n, \bar{C}_1, \ldots, \bar{C}_r. \end{aligned} \tag{3.5}$$

Then set

$$\bar{c}_0 = \lambda_1\bar{C}_1 + \cdots + \lambda_r\bar{C}_r. \tag{3.6}$$

DETAIL. This utilizes the result in Example 2.6. Changing the notation in (2.23) from parameters $C_k$ to parameters $D_k$, we get

$$\mathcal{E}[Z(c_0,c_1,\ldots,c_n)] = E[Z_0(c_1,\ldots,c_n)]-c_0+\inf_{\substack{D_1,\ldots,D_r\\ \lambda_1D_1+\cdots+\lambda_rD_r=0}}\left\{\frac{\lambda_1}{\alpha_1}E[\max\{0,D_1-c_0-Z_0(c_1,\ldots,c_n)\}]\right.$$
$$\left.+\cdots+\frac{\lambda_r}{\alpha_r}E[\max\{0,D_r-c_0-Z_0(c_1,\ldots,c_n)\}]\right\}. \quad (3.7)$$

In terms of $C_k = D_k - c_0$, the condition that $\lambda_1D_1+\cdots+\lambda_rD_r=0$ translates to $\lambda_1C_1+\cdots+\lambda_rC_r=c_0$. Under this change of variables, the minimization in (3.5) reduces to that in (3.7), with the optimal value of $c_0$ given by (3.6). □

When the space $\Omega$ of future states is a finite set, the minimization problem in (3.5) can be converted to linear programming and solved efficiently by available software. The technique is explained in a similar situation in Rockafellar and Uryasev [14].

It has already been mentioned that classical regression with the error measure in Example 2.1 corresponds, in the pattern of Theorem 3.2, to minimizing $\sigma(Z(c_0,c_1,\ldots,c_n))$ subject to insisting that $E[Z(c_0,c_1,\ldots,c_n)]=0$. In general, however, the side condition in the risk-shaping problem in (3.4) focuses on the particular statistic associated with the error measure $\mathcal{E}$ being used. But as a matter of fact *it is always possible to revert to expectation as the statistic*, in the following manner.

EXAMPLE 3.2 (UNBIASED LINEAR REGRESSION WITH GENERAL DEVIATION MEASURES). For any deviation measure $\mathcal{D}$ projected from an error measure $\mathcal{E}$, the problem of choosing coefficients $c_0,c_1,\ldots,c_n$ so as to minimize $\mathcal{D}(Z(c_0,c_1,\ldots,c_n))$ subject to the constraint that

$$E[Z(c_0,c_1,\ldots,c_n)]=0$$

can be solved as follows. First, obtain $\bar{c}_0,\bar{c}_1,\ldots,\bar{c}_n$ by solving problem $\mathcal{LR}(Y)$ with respect to $\mathcal{E}$ as before. Then replace $\bar{c}_0$ by

$$\bar{c}_0' = E[Z_0(c_1,\ldots,c_n)] = EY-\bar{c}_1EX_1-\cdots-\bar{c}_nEX_n.$$

This is equivalent to solving the linear regression problem $\mathcal{LR}(Y)$ with respect to the error measure $\mathcal{E}'(X)=\mathcal{D}(X)+|EX|$ instead of $\mathcal{E}$. On the other hand, in terms of the risk measure $\mathcal{R}$ associated with $\mathcal{D}$, this is equivalent also to the problem

$$\text{minimize } \mathcal{R}(Z(c_0,c_1,\ldots,c_n)) \quad \text{subject to} \quad E[Z(c_0,c_1,\ldots,c_n)]=0.$$

This modification merely shifts the approximation of $Y$ up or down until the error random variable giving the difference has expectation equal to 0, disregarding the stipulation in (3.4) relative to the statistic associated with $\mathcal{E}$. The first equivalent interpretation is justified by the observation in Example 2.5 that $\mathcal{E}'$ projects to the same deviation measure $\mathcal{D}$, along with the fact that in Theorem 3.2, $\bar{c}_1,\ldots,\bar{c}_n$ depend only on $\mathcal{D}$. The second equivalence is immediate from the relationship between $\mathcal{D}$ and $\mathcal{R}$ in (2.6): one always has

$$\mathcal{D}(Z(c_0,c_1,\ldots,c_n)) = \mathcal{R}(Z(c_0,c_1,\ldots,c_n)) \quad \text{when } E[Z(c_0,c_1,\ldots,c_n)]=0.$$

Another such adjustment of basic linear regression, again involving the risk measure $\mathcal{R}$, could also be interesting in some situations.

EXAMPLE 3.3 (RISK-ACCEPTABLE REGRESSION). For any deviation measure $\mathcal{D}$, projected from a nondegenerate error measure $\mathcal{E}$, let $\mathcal{R}$ be the averse risk measure corresponding to $\mathcal{D}$. Then the problem

$$\text{choose } c_0,c_1,\ldots,c_n \text{ to minimize } \mathcal{D}(Z(c_0,c_1,\ldots,c_n))$$
$$\text{subject to the constraint that } \mathcal{R}(Z(c_0,c_1,\ldots,c_n))=0 \quad (3.8)$$

can be solved as follows. First, obtain $\bar{c}_0,\bar{c}_1,\ldots,\bar{c}_n$ by solving problem $\mathcal{LR}(Y)$ with respect to $\mathcal{E}$ as before, but then replace $\bar{c}_0$ by

$$\bar{c}_0^* = E[Z_0(\bar{c}_1,\ldots,\bar{c}_n)]-\mathcal{D}[Z_0(\bar{c}_1,\ldots,\bar{c}_n)].$$

Then too, the resulting error $Z(\bar{c}_0^*,\bar{c}_1,\ldots,\bar{c}_n)$ has

$$E[Z(\bar{c}_0^*,\bar{c}_1,\ldots,\bar{c}_n)] = \mathcal{D}[Z(\bar{c}_0^*,\bar{c}_1,\ldots,\bar{c}_n)] \ge 0. \quad (3.9)$$

This conclusion is evident from the observation that

$$\begin{aligned}\mathcal{R}(Z(c_0, c_1, \ldots, c_n)) &= \mathcal{D}(Z(c_0, c_1, \ldots, c_n)) - E[Z(\bar{c}_0, \bar{c}_1, \ldots, \bar{c}_n)] \\ &= \mathcal{D}(Z_0(c_1, \ldots, c_n)) - E[Z_0(\bar{c}_1, \ldots, \bar{c}_n)] + c_0.\end{aligned}$$

The rationale behind regression in the mode of (3.8) hinges on the concept of a random variable $X$ having an acceptable level of risk with respect to $\mathcal{R}$ when $\mathcal{R}(X) \le 0$. Here we apply that to $X = Z(c_0, c_1, \ldots, c_n)$. The risk perspective is to control the degree to which $Y$ might be overestimated by the regression approximation: overestimates correspond to negative outcomes of the random variable $Z(c_0, c_1, \ldots, c_n)$. Observe that the risk-acceptability constraint in (3.8) is equivalent through (3.9) to requiring overestimates to comprise the tail of the error distribution lying one full deviation unit (with respect to $\mathcal{D}$) below the expectation of the error.

In contrast to the modified regression in Example 3.2, which could be construed as linear regression in the sense of problem $\mathcal{LR}(Y)$ for a modification $\mathcal{E}'$ of $\mathcal{E}$, it is unclear whether the modified regression in Example 3.3 can be given a parallel interpretation with $\mathcal{E}$ replaced by some $\mathcal{E}^*$.

**4. Characterization of regression coefficients.** With the generalized regression problem $\mathcal{LR}(Y)$ viewed as an optimization problem with respect to the variables $c_0, c_1, \ldots, c_n$, the next question is how to characterize $(\bar{c}_0, \bar{c}_1, \ldots, \bar{c}_n)$ as a solution. A basic step has already been taken toward answering that. We know from Theorem 3.2 that the value of $\bar{c}_0$ is immediately determined as soon as the optimal coefficient vector $(\bar{c}_1, \ldots, \bar{c}_n)$ has been determined. Moreover the vector $(\bar{c}_1, \ldots, \bar{c}_n)$ solves the problem:

$$\text{minimize } f(c_1, \ldots, c_n) \text{ over all } (c_1, \ldots, c_n) \in \mathbb{R}^n, \quad \text{where}$$
$$f(c_1, \ldots, c_n) = \mathcal{D}(Y - [c_1 X_1 + \cdots + c_n X_n]).$$

This is a finite-dimensional problem in which $f$ is a convex function by virtue of the convexity of $\mathcal{D}$ (coming from (D2) and (D3)). The problem appears to be unconstrained, but here a subtle issue arises. It is truly unconstrained if $f(c_1, \ldots, c_n) < \infty$ for all $(c_1, \ldots, c_n) \in \mathbb{R}^n$, but otherwise there is the implicit constraint of keeping to the subset of $\mathbb{R}^n$ where $f$ is finite (the so-called effective domain of $f$). That set might not be all of $\mathbb{R}^n$ if $\mathcal{D}$ is able to take on infinite values, which in our general framework would happen if the error measure $\mathcal{E}$ from which $\mathcal{D}$ is projected can take on infinite values. An illustration of this possibility is furnished by the range deviation measure in Example 2.2.

Here we wish to avoid the complications of such cases, and furthermore to make use of subgradient chain rules for composite functions like $f$ which we developed in Rockafellar et al. [17] for the case where not only are the deviation values in (5.1) finite, but those of other random variables as well. We therefore limit our analysis of optimality in this section to deviation measures $\mathcal{D}$ that are finite on $\mathcal{L}^2(\Omega)$ (with the same holding for $\mathcal{E}$ and the associated risk measure $\mathcal{R}$).

Note that $\mathcal{D}$ satisfies this finiteness condition on $\mathcal{L}^2(\Omega)$ in the CVaR and mixed-CVaR cases of Examples 2.5 and 2.8 (covering the median deviation of Example 2.3 in particular), and also, of course, when $\mathcal{D}$ is a multiple of standard deviation as in Example 2.1. It also satisfies it even for the range deviation of Example 2.2 when $\Omega$ is a finite set, standing for just finitely many "future scenarios." Thus, for practical purposes the restriction to finite $\mathcal{D}$ causes no inconvenience. (Incidentally, our need to encompass standard deviation along with the CVaR and mixed CVaR deviations in this manner has dictated our choice of $\mathcal{L}^2(\Omega)$ instead of $\mathcal{L}^1(\Omega)$ as the space for our random variables $X$.)

Optimality in this setting is addressed through the concept of the *risk envelope* $\mathcal{Q}$ associated with $\mathcal{D}$. As explained in Rockafellar et al. [16], convex analysis *uniquely* associates with any deviation measure $\mathcal{D}$ satisfying (D1)–(D4) a set $\mathcal{Q} \subset \mathcal{L}^2(\Omega)$ (the risk envelope in question) which satisfies the axioms

(Q1) $\mathcal{Q}$ is nonempty, closed and convex.

(Q2) $EQ = 1$ for all $Q \in \mathcal{Q}$.

(Q3) For each $X \neq 0$ there exists $Q \in \mathcal{Q}$ with $E[QX] < EX$.

This set $\mathcal{Q}$ expresses $\mathcal{D}$ through the formula that

$$\mathcal{D}(X) = EX - \inf_{Q \in \mathcal{Q}} E[QX] \quad \text{for all } X \in \mathcal{L}^2(\Omega). \tag{4.1}$$

In that association, our assumption that $\mathcal{D}$ is finite on $\mathcal{L}^2(\Omega)$ corresponds to $\mathcal{Q}$ being bounded, in which event $\mathcal{Q}$ is weakly compact in $\mathcal{L}^2(\Omega)$ and the infimum in (4.1) is sure to be attained for every random variable $X$. The nonempty set

$$\mathcal{Q}(X) = \arg\min_{Q \in \mathcal{Q}} E[QX]$$

(again convex and weakly compact) is the *risk identifier* for $X$ in the terminology of Rockafellar et al. [16] (going back to Rockafellar et al. [15]). It depends only on the nonconstant aspects of $X$: one has

$$\mathcal{Q}(X - C) = \mathcal{Q}(X) \quad \text{for all } X \text{ and constants } C. \tag{4.2}$$

The case where the risk measure $\mathcal{R}$ paired with $\mathcal{D}$ is *coherent* (in the sense pioneered in Artzner et al. [3, 4]) can also be identified as the case where $Q \geq 0$ for every $Q \in \mathcal{Q}$ (cf. Rockafellar et al. [16]). It allows the elements $Q$ of the risk envelope $\mathcal{Q}$ to be seen as probability densities with respect to the underlying probability measure $P$. Then the elements $Q \in \mathcal{Q}(X)$ maximize the corresponding expected loss $-E[QX]$ that could come from $X$. However, we hold back from treating only that case because it would exclude standard deviation, $\mathcal{D} = \sigma$. There is no sense in building a theory of generalized linear regression which excludes classical linear regression. Anyway, it deserves to be noted that all CVaR and mixed-CVaR risk measures are coherent, in particular.

Theorem 4.1 (Risk Identifier Characterization of Regression Coefficients). *For a nondegenerate error measure $\mathcal{E}$ that is finite everywhere on $\mathcal{L}^2(\Omega)$, let $\mathcal{D}$ be the projected deviation measure and let $\mathcal{Q}$ be its associated risk envelope. Then in the linear regression problem $\mathcal{LR}(Y)$, a vector $(\bar{c}_1, \ldots, \bar{c}_n)$ furnishes the regression coefficients for the random variables $X_1, \ldots, X_n$ if and only if, in terms of $\bar{Z}_0 = Y - [\bar{c}_1 X_1 + \cdots + \bar{c}_n X_n]$,*

$$\text{there exists } \bar{Q} \in \mathcal{Q}(\bar{Z}_0) \quad \text{such that} \quad E[\bar{Q} X_j] = EX_j \quad \text{for } j = 1, \ldots, n. \tag{4.3}$$

Proof. We apply our general result in Rockafellar et al. [17, Theorem 2] concerning the minimization of the deviation of a sum of random variables subject to linear constraints on the coefficients. Problem $\mathcal{LR}(Y)$ fits that framework as the very special case where we wish to

$$\begin{aligned} &\text{minimize } \mathcal{D}(c_1[-X_1] + \cdots + c_n[-X_n] + c_{n+1}[Y]) \text{ over all} \\ &(c_1, \ldots, c_n, c_{n+1}) \text{ satisfying } 0c_1 + \cdots + 0c_n + 1c_{n+1} = 1. \end{aligned}$$

(Here, and in the next display, implicit coefficients have been made explicit in order to connect clearly with the coefficient format in the cited result (Rockafellar et al. [17, Theorem 2]).) The optimality of $(\bar{c}_1, \ldots, \bar{c}_n, \bar{c}_{n+1})$, necessitating $\bar{c}_{n+1} = 1$ is characterized by the cited result as corresponding, for the random variable $\bar{c}_1[-X_1] + \cdots + \bar{c}_n[-X_n] + \bar{c}_{n+1}[Y]$, which here is $\bar{Z}_0$, to the existence of $\bar{Q} \in \mathcal{Q}(\bar{Z}_0)$ and a multiplier $\lambda \in \mathbb{R}$ such that

$$E[\bar{Q}(X_j - EX_j)] = \lambda 0 \quad \text{for } j = 1, \ldots, n \qquad \text{and} \qquad E[\bar{Q}(EY - Y)] = \lambda 1.$$

This characterization says the same thing about $(\bar{c}_1, \ldots, \bar{c}_n)$ as the assertion in the theorem. □

Note that having $\bar{Q} \in \mathcal{Q}(\bar{Z}_0)$ is the same as having $\bar{Q} \in \mathcal{Q}(\bar{Z})$ for $\bar{Z} = Y - [\bar{c}_0 + \bar{c}_1 X_1 + \cdots + \bar{c}_n X_n]$, due to the insensitivity to constants in (4.2). The first condition on $\bar{Q}$ in (4.3) is equivalent through (4.1) to having

$$\mathcal{D}(\bar{Z}_0) = E[\bar{Z}_0] - E[\bar{Q}\bar{Z}_0] = E[\bar{Z}] - E[\bar{Q}\bar{Z}],$$

but through axiom (Q2) and the expectation relations in (4.3), that comes out as

$$\mathcal{D}(\bar{Z}_0) = EY - E[\bar{Q}Y] = -\operatorname{covar}(\bar{Q}, Y).$$

In particular, therefore, $\operatorname{covar}(Q, Y) < 0$ (unless $\bar{Z}_0$ completely vanishes). Likewise, the expectation relations in (4.3) can be expressed equivalently by

$$\operatorname{covar}(\bar{Q}, X_j) = 0 \quad \text{for } j = 1, \ldots, n.$$

Descriptions of the risk envelopes $\mathcal{Q}$ specific to the examples we have considered are available in Rockafellar et al. [16, 17] and elsewhere. These descriptions can be applied to get further details from Theorem 4.1. We limit ourselves here to a single, but rich, example.

Example 4.1 (Optimal Coefficients in Mixed Quantile Regression). For probability levels $\alpha_k$ with $0 < \alpha_1 < \cdots < \alpha_r < 1$, and weights $\lambda_k > 0$ with $\lambda_1 + \cdots + \lambda_r = 1$, consider the deviation measure

$$\mathcal{D}(X) = \lambda_1 \mathrm{CVaR}_{\alpha_1}(X) + \cdots + \lambda_r \mathrm{CVaR}_{\alpha_r}(X)$$

and the error measure $\mathcal{E}$ in (2.23), which projects to $\mathcal{D}$. In terms of $\bar{Z} = Y - [\bar{c}_0 + \bar{c}_1 X_1 + \cdots + \bar{c}_n X_n]$, and under the simplifying assumption that the distribution of $\bar{Z}$ contains no probability atoms (i.e., the distribution function $F_{\bar{Z}}$ has no jumps), the following conditions are necessary and sufficient for $(\bar{c}_0, \bar{c}_1, \ldots, \bar{c}_n)$ to belong to the solution set $\mathcal{C}(Y)$ in problem $\mathcal{LR}(Y)$:

$$\begin{aligned} &\lambda_1 q_{\alpha_1}(\bar{Z}) + \cdots + \lambda_r q_{\alpha_r}(\bar{Z}) = 0, \\ &\lambda_1 E[X_j \mid \bar{Z} \leq q_{\alpha_1}(\bar{Z})] + \cdots + \lambda_r E[X_j \mid \bar{Z} \leq q_{\alpha_r}(\bar{Z})] = EX_j \quad \text{for } j = 1, \ldots, n. \end{aligned} \tag{4.4}$$

DETAIL. Our assumption about no atoms ensures that the quantiles for $\bar{Z}$ reduce to single values and we can avoid dealing with them as intervals. The first condition in (4.4) then merely reflects two things we already know, namely that optimality fixes $\bar{c}_0$ by requiring $\mathcal{S}(\bar{Z}) \ni 0$ (Theorem 3.1) for the statistic corresponding to $\mathcal{E}$, and that this is $\mathcal{S}(\bar{Z}) = \lambda_1 q_{\alpha_1}(\bar{Z}) + \cdots + \lambda_r q_{\alpha_r}(\bar{Z})$ (from Example 2.6). We appeal then to the description in Rockafellar et al. [16, Example 4] of the risk envelope $\mathcal{Q}$ for $\mathcal{D}$ as consisting of all $Q = \lambda_1 Q_1 + \cdots + \lambda_r Q_r$ in which $Q_k$ belongs to the risk envelope $\mathcal{Q}_k$ for $\mathcal{D}_k = \mathrm{CVaR}_{\alpha_k}$. Then $E[QX_j] = \sum_{k=1}^r \lambda_k E[Q_k X_j]$, Moreover, $\mathcal{Q}_k(\bar{Z})$ consists of the single function $Q_k$ which has $Q_k(\omega) = \alpha_k^{-1}$ for $\omega \in \Omega$ such that $\bar{Z}(\omega) \le q_{\alpha_k}(\bar{Z})$ but $Q_k(\omega) = 0$ otherwise (up to equivalence with a set of probability zero). Then $E[Q_k X_j]$ is the conditional expectation of $X_j$ subject to having $\bar{Z}$ in its lower $\alpha_k$-tail. This yields the stated result. □

Of course, the special case where there is only one quantile, not a mixture, can immediately be extracted from Example 4.1.

**5. Application to parametric risk expressions.** We now take up the topic featured in the introduction, which concerns functions $f(x_1, \ldots, x_m)$ defined on $\mathbb{R}^m$ as in (1.1) with respect to a weighted combination $x_1 r_1 + \cdots + x_m r_m$ of given random variables. We suppose that each random variable $r_i$ is approximated by a linear combination

$$\hat{r}_i = c_{i0} + c_{i1} F_1 + \cdots + c_{in} F_n \tag{5.1}$$

of some other collection of random variables $F_1, \ldots, F_n$ (so-called *factors*), so that

$$r_i = \hat{r}_i + Z_i(c_{i0}, c_{i1}, \ldots, c_{in}) = \hat{r}_i + Z_{i0}(c_{i1}, \ldots, c_{in}) - c_{i0}$$

for the error expressions

$$Z_i(c_{i0}, c_{i1}, \ldots, c_{in}) = r_i - [c_{i0} + c_{i1} F_1 + \cdots + c_{in} F_n], \qquad Z_{i0}(c_{i1}, \ldots, c_{in}) = r_i - [c_{i1} F_1 + \cdots + c_{in} F_n].$$

Specifically, we envision producing these approximations by *generalized linear regression with respect to some error measure* $\mathcal{E}$. The question is how estimates of the difference between $f(x_1, \ldots, x_n)$ and the corresponding expression $\hat{f}(x_1, \ldots, x_n)$, obtained by substituting $\hat{r}_i$ for $r_i$, might influence the choice of $\mathcal{E}$.

Although regression is the basis for what we are investigating, we are not employing it directly. We are approximating the *parameter-dependent* random variable

$$Y(x_1, \ldots, x_m) = x_1 r_1 + \cdots + x_m r_m \tag{5.2}$$

by the similarly *parameter-dependent* random variable

$$\hat{Y}(x_1, \ldots, x_m) = x_1 \hat{r}_1 + \cdots + x_m \hat{r}_m \tag{5.3}$$

for some version of $\hat{r}_1, \ldots, \hat{r}_m$ generated in the form of (5.1). It is not enough for the approximation to be carried out for a fixed weight vector $(x_1, \ldots, x_m)$. If that were the case, we could merely revert to linear regression of $Y(x_1, \ldots, x_m)$ itself relative to $F_1, \ldots, F_n$. Instead, the approximation should be effective for a suitable *range* of weight vectors $(x_1, \ldots, x_m)$.

This issue will be explored with regard to the "risk aspects" of $Y(x_1, \ldots, x_m)$, by which we mean characteristics captured either by deviation measures $\mathcal{D}$ or by risk measures $\mathcal{R}$.

We deal first with deviation measures as applied to (5.2) and (5.3). We restrict our study to weight vectors $(x_1, \ldots, x_m) \in \mathbb{R}^m_+$, i.e., with $x_i \ge 0$ for $i = 1, \ldots, m$. The deviation axioms (D2) and (D3) allow us to calculate that

$$\mathcal{D}\left(\sum_{i=1}^m x_i r_i\right) = \mathcal{D}\left(\sum_{i=1}^m (x_i \hat{r}_i + x_i[r_i - \hat{r}_i])\right) \le \mathcal{D}\left(\sum_{i=1}^m x_i \hat{r}_i\right) + \sum_{i=1}^m x_i \mathcal{D}(r_i - \hat{r}_i).$$

Since $r_i - \hat{r}_i = Z_i(c_{i0}, c_{i1}, \ldots, c_{in}) = Z_{i0}(c_{i1}, \ldots, c_{in})$ and $\mathcal{D}$ is insensitive to constants, cf. (2.1), this provides the upper bound

$$\mathcal{D}\left(\sum_{i=1}^m x_i r_i\right) - \mathcal{D}\left(\sum_{i=1}^m x_i \hat{r}_i\right) \le \sum_{i=1}^m x_i \mathcal{D}(Z_{i0}(c_{i1}, \ldots, c_{in})). \tag{5.4}$$

A parallel calculation with the roles of $r_i$ and $\hat{r}_i$ reversed leads to the lower bound

$$\mathcal{D}\left(\sum_{i=1}^m x_i r_i\right) - \mathcal{D}\left(\sum_{i=1}^m x_i \hat{r}_i\right) \ge -\sum_{i=1}^m x_i \mathcal{D}(-Z_{i0}(c_{i1}, \ldots, c_{in})).$$

By combining the upper and lower bounds, making essential use once more of the nonnegativity of each $x_i$, we arrive at the absolute bound

$$\left|\mathcal{D}\left(\sum_{i=1}^m x_i r_i\right) - \mathcal{D}\left(\sum_{i=1}^m x_i \hat{r}_i\right)\right| \le \sum_{i=1}^m x_i \max\{\mathcal{D}(Z_{i0}(c_{i1},\dots,c_{in})), \mathcal{D}(-Z_{i0}(c_{i1},\dots,c_{in}))\}. \tag{5.5}$$

We can then explore how these bounds may be made as tight as possible.

THEOREM 5.1 (APPROXIMATION ESTIMATES FOR DEVIATION MEASURES). *Let $\mathcal{D}$ be any deviation measure and let*

$$f_{\mathcal{D}}(x_1,\dots,x_n) = \mathcal{D}(x_1 r_1 + \cdots + x_m r_m), \qquad \hat{f}_{\mathcal{D}}(x_1,\dots,x_n) = \mathcal{D}(x_1 \hat{r}_1 + \cdots + x_m \hat{r}_m),$$

*with $\hat{r}_i$ as in (5.1). Then, as an estimate of how much $\hat{f}_{\mathcal{D}}$ may differ from $f_{\mathcal{D}}$, the best upper bound coming out of (5.4) in the form*

$$f_{\mathcal{D}}(x_1,\dots,x_n) - \hat{f}_{\mathcal{D}}(x_1,\dots,x_n) \le \sum_{i=1}^m a_i x_i \quad \text{for all } (x_1,\dots,x_m) \in \mathbb{R}^m_+, \quad \text{where } a_i \ge 0,$$

*is obtained by employing an error measure $\mathcal{E}$ that projects to $\mathcal{D}$ to carry out, for each random variable $r_i$, linear regression with respect to the factors $F_1,\dots,F_n$, and then, from the regression coefficients $\bar{c}_{ij}$ thereby determined, taking*

$$a_i = \mathcal{D}(Z_{i0}(\bar{c}_{i1},\dots,\bar{c}_{in})) \quad \text{for } i = 1,\dots,m.$$

*On the other hand, the best symmetric bound of the form*

$$|f_{\mathcal{D}}(x_1,\dots,x_n) - \hat{f}_{\mathcal{D}}(x_1,\dots,x_n)| \le \sum_{i=1}^m \bar{a}_i x_i \quad \text{for all } (x_1,\dots,x_m) \in \mathbb{R}^m_+, \quad \text{where } \bar{a}_i \ge 0,$$

*is obtained in the same way, but with regression based instead on a symmetric error measure $\overline{\mathcal{E}}$ which projects to the symmetric deviation measure*

$$\overline{\mathcal{D}}(X) = \max\{\mathcal{D}(X), \mathcal{D}(-X)\}, \tag{5.6}$$

*and then taking*

$$\bar{a}_i = \overline{\mathcal{D}}(Z_{i0}(\bar{c}_{i1},\dots,\bar{c}_{in})) \quad \text{for } i = 1,\dots,m. \tag{5.7}$$

PROOF. Clearly, the upper bound in (5.4) is optimized by choosing $(\bar{c}_{i1},\dots,\bar{c}_{in})$ in order to minimize $\mathcal{D}(c_{i1},\dots,c_{in})$ for each $i$. This corresponds to linear regression as described (but with the constants $\bar{c}_{i0}$ open to any convenient choice). For the absolute bound, the argument is similar except that the max expression indicates replacing $\mathcal{D}$ by the deviation measure $\overline{\mathcal{D}}$ in (5.6). □

For a risk measure $\mathcal{R}$ instead of a deviation measure $\mathcal{D}$, bounds corresponding to (5.4) and (5.5) can be obtained in much the same manner from axioms (R2), (R3), and the nonnegativity of $x_i$. The upper bound is

$$\mathcal{R}\left(\sum_{i=1}^m x_i r_i\right) - \mathcal{R}\left(\sum_{i=1}^m x_i \hat{r}_i\right) \le \sum_{i=1}^m x_i \mathcal{R}(Z_i(c_{i0}, c_{i1},\dots,c_{in})), \tag{5.8}$$

but in the absolute bound parallel to (5.5) the right side is initially in the form

$$\sum_{i=1}^m x_i \max\{\mathcal{R}(Z_i(c_{i0}, c_{i1},\dots,c_{in})), \mathcal{R}(-Z_i(c_{i0}, c_{i1},\dots,c_{in}))\}$$

and needs to be reworked to get more out of it. A tool is available in the relations in (2.6), which lead to the inequalities

$$\begin{aligned}
\mathcal{R}(Z_i(c_{i0}, c_{i1},\dots,c_{in})) &= \mathcal{D}(Z_i(c_{i0}, c_{i1},\dots,c_{in})) + E[-Z_i(c_{i0}, c_{i1},\dots,c_{in})] \\
&\le \mathcal{D}(Z_{i0}(c_{i1},\dots,c_{in})) + |EZ_i(c_{i0}, c_{i1},\dots,c_{in})|, \\
\mathcal{R}(-Z_i(c_{i0}, c_{i1},\dots,c_{in})) &= \mathcal{D}(-Z_{i0}(c_{i0}, c_{i1},\dots,c_{in})) + EZ_i(c_{i0}, c_{i1},\dots,c_{in}) \\
&\le \mathcal{D}(Z_{i0}(c_{i1},\dots,c_{in})) + |EZ_i(c_{i0}, c_{i1},\dots,c_{in})|.
\end{aligned}$$

That way, we get an absolute bound in the form

$$\begin{aligned}
\left|\mathcal{R}\left(\sum_{i=1}^m x_i r_i\right) - \mathcal{R}\left(\sum_{i=1}^m x_i \hat{r}_i\right)\right| \le \sum_{i=1}^m x_i [&\max\{\mathcal{D}(Z_{i0}(c_{i1},\dots,c_{in})), \mathcal{D}(-Z_{i0}(c_{i1},\dots,c_{in}))\} \\
&+ |EZ_i(c_{i0}, c_{i1},\dots,c_{in})|].
\end{aligned} \tag{5.9}$$

THEOREM 5.2 (APPROXIMATION ESTIMATES FOR RISK MEASURES). *Let $\mathcal{R}$ be any averse measure of risk, and let*

$$f_{\mathcal{R}}(x_1,\dots,x_n)=\mathcal{R}(x_1r_1+\cdots+x_mr_m),\qquad \hat{f}_{\mathcal{R}}(x_1,\dots,x_n)=\mathcal{R}(x_1\hat{r}_1+\cdots+x_m\hat{r}_m),$$

*with $\hat{r}_i$ as in (5.1). Then, as an estimate of how much $\hat{f}_{\mathcal{R}}$ may differ from $f_{\mathcal{R}}$, the best upper bound coming out of (5.8) in the form*

$$f_{\mathcal{R}}(x_1,\dots,x_n)-\hat{f}_{\mathcal{R}}(x_1,\dots,x_n)\le\sum_{i=1}^m a_ix_i\quad \text{for all } (x_1,\dots,x_m)\in\mathbb{R}^m_+,\quad \text{where } a_i\ge 0,$$

*is actually achieved with*

$$a_i=0\quad \text{for } i=1,\dots,m.$$

*It is obtained through generalized linear regression of each $r_i$ with respect to the factors $F_1,\dots,F_n$, using in all cases an error measure $\mathcal{E}$ that projects to the deviation measure $\mathcal{D}$ paired with $\mathcal{R}$, and then, as in Example 3.3, replacing each $\bar{c}_{i0}$ by*

$$\bar{c}^*_{i0}=\mathcal{R}(Z_{i0}(\bar{c}_{i1},\dots,\bar{c}_{in}))\quad \text{for } i=1,\dots,m.$$

*On the other hand, the best symmetric bound coming out of (5.9) in the form*

$$\left|f_{\mathcal{R}}(x_1,\dots,x_n)-\hat{f}_{\mathcal{R}}(x_1,\dots,x_n)\right|\le\sum_{i=1}^m \bar{a}_ix_i\quad \text{for all } (x_1,\dots,x_m)\in\mathbb{R}^m_+,\quad \text{where } \bar{a}_i\ge 0,$$

*is obtained with regression based instead on a symmetric error measure $\overline{\mathcal{E}}$ which projects to the symmetric deviation measure $\overline{\mathcal{D}}$ in (5.6), taking $\bar{a}_i$ as in (5.7) and then, as in Example 3.2, replacing each $\bar{c}_{i0}$ by*

$$\bar{c}'_{i0}=EZ_{i0}(\bar{c}_{i1},\dots,\bar{c}_{in})\quad \text{for } i=1,\dots,m.$$

PROOF. Much as in Theorem 4.1, this is evident from the bounds in (5.8) and (5.9) and the regression modifications in the cited examples. □

The result in the first part of Theorem 5.2 has the following interesting interpretation. *If the regression in approximating $r_i$ by $\hat{r}_i$ with respect to the factors $F_j$ is carried out relative to the risk measure $\mathcal{R}$ as in Example 3.3, then*

$$\hat{f}_{\mathcal{R}}(x_1,\dots,x_n)\le c\quad \text{ensures}\quad f_{\mathcal{R}}(x_1,\dots,x_n)\le c\quad \text{for any } c.$$

Recall that in this setting the approximation error $r_i-\hat{r}_i$ has $\mathcal{R}(r_i-\hat{r}_i)=0$, which means that, with respect to the standard set by $\mathcal{R}$, *the risk of $r_i$ being overapproximated by $\hat{r}_i$ is acceptable*, in fact just at the limit of acceptability; cf. (2.3).

This observation has powerful consequences for the role of generalized linear regression in dealing with risk constraints in factor models in finance or elsewhere.

COROLLARY 5.1 (SAFEGUARDED APPROXIMATION OF RISK CONSTRAINTS). *For $k=1,\dots,r$, let $\mathcal{R}_k$ be an averse measure of risk and let*

$$S=\{x\in\mathbb{R}^m_+\mid \mathcal{R}_k(x_1r_1+\cdots+x_mr_m)\le c_k \text{ for } k=1,\dots,r\}.$$

*Let each $r_i$ be approximated by $\hat{r}_i$ relative to factors $F_1,\dots,F_n$ through $\mathcal{R}_i$-risk-acceptable linear regression as laid out in Example 3.3, and let*

$$\hat{S}=\{x\in\mathbb{R}^m_+\mid \mathcal{R}_k(x_1\hat{r}_1+\cdots+x_m\hat{r}_m)\le c_k \text{ for } k=1,\dots,r\}.$$

*Then $\hat{S}$ serves surely to approximate $S$ from within: one has $\hat{S}\subset S$, and in a definite sense this is the best approximation of such type.*

The surprising feature of this way of dealing with the feasibility of $x=(x_1,\dots,x_m)$ is that, to obtain the best results, *regression must be carried out in a different way for each constraint*!

Other recent work on factor models can be found, for example, in Cherny and Madan [7], where, however, different issues than these are at the center, and linear regression as presented here is not really involved.

**Acknowledgment.** Research of R. Tyrrell Rockafellar was supported by NSF Grant DMS-0104055. Research of Stan Uryasev was supported by NSF Grant 045473 "Percentile-Based Risk Management Approaches in Discrete Decision-Making Problems." The authors are grateful to the anonymous referees for their valuable comments and suggestions, which greatly helped to improve the quality of the paper.

## References

[1] Acerbi, C. 2002. Spectral measures of risk: A coherent representation of subjective risk aversion. *J. Banking Finance* **26** 1505–1518.
[2] Acerbi, C., D. Tasche. 2002. On the coherence of expected shortfall. *J. Banking Finance* **26** 1487–1503.
[3] Artzner, P., F. Delbaen, J.-M. Eber, D. Heath. 1997. Thinking coherently. *Risk* **10** 68–91.
[4] Artzner, P., F. Delbaen, J.-M. Eber, D. Heath. 1999. Coherent measures of risk. *Math. Finance* **9** 203–227.
[5] Bassett, G. W., R. Koenker, G. Kordas. 2004. Pessimistic portfolio allocation and choquet expected utility. *J. Financial Econometrics* **2** 477–492.
[6] Ben Tal, A., M. Teboulle. 2007. An old-new concept of convex risk measures: The optimized certainty equivalent. *Math. Finance* **17** 449–476.
[7] Cherny, A. S., D. P. Madan. 2006. Coherent measurement of factor risks. Available at http://ssrn.com/abstract=904543.
[8] Föllmer, H., A. Schied. 2002. *Stochastic Finance*. DeGruyter, Berlin.
[9] Koenker, R. 2005. *Quantile Regression*. Econometric Society Monograph Series, Cambridge University Press, Cambridge.
[10] Koenker, R., G. W. Bassett. 1978. Regression quantiles. *Econometrica* **46** 33–50.
[11] Rockafellar, R. T. 1970. *Convex Analysis*. Princeton University Press.
[12] Rockafellar, R. T. 2007. Coherent approaches to risk in optimization. *Tutorials Operations Research*. INFORMS, Hanover, MD, 38–61.
[13] Rockafellar, R. T., S. P. Uryasev. 2000. Optimization of conditional value-at-risk. *J. Risk* **2** 21–42.
[14] Rockafellar, R. T., S. P. Uryasev. 2002. Conditional value-at-risk for general loss distributions. *J. Banking Finance* **26** 1443–1471.
[15] Rockafellar, R. T., S. Uryasev, M. Zabarankin. 2002. Deviation measures in risk analysis and optimization. Research Report 2002-7, Department of Industrial and Systems Engineering, University of Florida, Gainesville.
[16] Rockafellar, R. T., S. Uryasev, M. Zabarankin. 2006. Generalized deviations in risk analysis. *Finance Stochastics* **10** 51–74.
[17] Rockafellar, R. T., S. Uryasev, M. Zabarankin. 2006. Optimality conditions in portfolio analysis with general deviation measures. *Math. Programming, Ser. B* **108** 515–540.
[18] Rockafellar, R. T., S. Uryasev, M. Zabarankin. 2006. Master funds in portfolio analysis with general deviation measures. *J. Banking Finance* **30** 743–778.
[19] Ruszczyński, A., A. Shapiro. 2006. Optimization of convex risk functions. *Math. Oper. Res.* **31** 433–452.
[20] Ruszczyński, A., A. Shapiro. 2006. Conditional risk mappings. *Math. Oper. Res.* **31** 544–561.
[21] Trindade, A. A., S. Uryasev, A. Shapiro, G. Zrazhevsky. 2007. Financial prediction with constrained tail risk. *J. Banking Finance* **31** 3524–3538.