PROBABILITY
AND
MATHEMATICAL STATISTICS

Vol. 44, Fasc. 1 (2024), pp. 15–28
Published online 10.6.2024
doi:10.37190/0208-4147.00139

A PREDICTIVE APPROACH TO QUANTILES: APPLICATION TO VALUE AT RISK AND TAIL VALUE AT RISK

BY

HENRYK GZYL (CARACAS)

Abstract. We prove that quantiles are best predictors in a special metric. The best predictor turns out to coincide with the notions of generalized arithmetic mean, exponential barycenter and certainty equivalent. We also show that the computation of tail value at risk (TVaR) reduces to the computation of a quantile with a higher level of confidence. This point of view makes the analysis of the statistical properties of TVaR easier.

2020 Mathematics Subject Classification: Primary 60A99; Secondary: 60D99, 62G08.

Key words and phrases: quantiles as best predictors, value at risk, tail value at risk, prediction in non-Euclidean metrics.

1. INTRODUCTION AND PRELIMINARIES

It is well known that the population mean is the best predictor of a square integrable random variable in the Euclidean (L_2) norm when there is no other information available. When there is information available, provided by a collection (σ-algebra) of events, the best predictor of a random variable is its conditional expectation given the σ-algebra. When the distance between random variables is measured by the L_1 norm, the best predictor is the median. The question that comes up is: What about the other quantiles? Are they best predictors in some norm?

Here we present a systematic approach to describing quantiles as best predictors. Our proposal consists in defining an appropriate metric on the range of the random variable and then finding the best predictor of the random variable in that metric. The basic idea behind the change of variables was proposed by Bernoulli in the 18th century. It is a key concept in economics where it is known as the utility function, and the best predictor is the certainty value. It reappeared under the name of generalized or extended arithmetic mean in the first quarter of the 20th century (see [2] or [4] for example). It was considered in the statistical literature in [1], where it is called a generalized mean, and its applications to maximum likelihood estimation are studied. We should also point out that this work is a distant

© Probability and Mathematical Statistics, 2024

relative to the work in [9] or [7], although the approach and point of view differ considerably.

We also consider the conditional expectation in the new metric, and apply it to compute a quantity that is essential in risk management, where it is known as Expected Shortfall (ES) or Tail Value at Risk (TVaR). These notions coincide for continuous random variables, which is the case that we consider here. For more on this matter, see [10] for example.

In the remainder of this section, we establish necessary notations and recall some preliminary results. In Section 2 we characterize quantiles and tail conditional expectations as best predictors in a non-Euclidean metric. In Section 3 we provide two examples, corresponding to two possible metrics. In the second example, we explicitly display the result of the computation of the TVaR as a quantile at a higher level of confidence for a few typical distributions used in risk management and reliability.

1.1. Background notations and results. The standard setup includes some complete probability space $(\Omega, \mathcal{F}, P)$ on which our random variables are defined. As usual, $E[X]$ and $E[X|\mathcal{G}]$ denote, respectively, the expected value of X and the conditional expectation of X given $\mathcal{G} \subset \mathcal{F}$. We consider only continuous random variables that have strictly increasing, continuous cumulative distribution functions. We use $F(x)$ to denote the distribution function of X. Denote by $\mathcal{I} = (a, b)$ the range of X, where $-\infty \leqslant a$ and $b \leqslant +\infty$. Let $u : \mathcal{I} \to \mathbb{R}$ be a strictly monotone, continuous function, with range $u(\mathcal{I})$.

If X, Y are random variables such that $u(X), u(Y)$ are square integrable, the d_u*-distance* between them is defined by

$$d_u(X,Y) = \big(E[(u(X) - u(Y))^2]\big)^{1/2}. \tag{1.1}$$

At this point, we mention that this distance is a geodesic distance, in a Hessian metric obtained from Bregman divergence. The connection is established in [8], where some comparison results are obtained.

DEFINITION 1.1. The number $E_u[X]$ that minimizes the distance (1.1) will be called the d_u*-mean* of X or the generalized u-mean of X, or simply the generalized mean of X.

The existence of $E_u[X]$ was discussed in [1] for finitely valued variables, but the proof carries over to the general case.

THEOREM 1.1. *Let X be a random variable such that $u(X)$ is square integrable. The best predictor of X in d_u-distance is the number $E_u[X]$ closest to X in the d_u-distance, and is given by the following generalized arithmetic mean:*

$$E_u[X] = u^{-1}\big(E[u(X)]\big). \tag{1.2}$$

Here we add that this quantity coincides with the exponential barycenter studied in [6] and the generalized mean studied in [12]. It also coincides with the certainty equivalent, which plays a key role in economics and decision theory (see [5] for example).

Intuitively, the best predictor of X, in the d_u-distance, given the information provided by a sub-σ-algebra $\mathcal{G}$ should be $u^{-1}\big(E[u(X)|\mathcal{G}]\big)$. Let us verify it.

THEOREM 1.2. *Let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$, and let X be such that $u(X)$ is square integrable. There exists a unique (up to P-null-sets) square integrable, $\mathcal{G}$-measurable random variable, denoted by $E_u[X|\mathcal{G}]$, that realizes*

$$\inf\{d_u(X,Y) \mid Y \in \mathcal{G},\ E[u(Y)^2] < \infty\}.$$

It is given by

$$E_u[X|\mathcal{G}] = u^{-1}\big(E[u(X)|\mathcal{G}]\big). \tag{1.3}$$

Proof. The proof is essentially as in [11]. Note that any $\mathcal{G}$-measurable, $u(\mathcal{I})$-valued random variable can be written as $u(Y)$ with Y being an $\mathcal{I}$-valued random variable. Let $\xi = u^{-1}(E[u(X)|\mathcal{G}])$. Then, for any $\mathcal{G}$-measurable Y we have

$$d_u^2(X,\xi) = E[(u(X)-u(Y))^2] = E[(u(X)-u(\xi))^2] + E[(u(\xi)-u(Y))^2],$$

which is minimized when $u(Y) = u(\xi) \Leftrightarrow Y = \xi$ almost surely. ■

To take care of the case where $u(X)$ is integrable but not square integrable, we follow a modification of the steps of the proof, say of [11, Theorem 23.4], plus the following comments to obtain the corresponding version of Theorem 1.2. Note that we want to predict $u(X)$ by random variables that are $u(\mathcal{I})$-valued, because $E[u(X)H] = 0$ when the support of H does not intersect $u(\mathcal{I})$.

THEOREM 1.3. *Let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$, and let X be such that $u(X)$ is integrable, but not necessarily square integrable. Then there exists a unique (up to P-null-sets) $\mathcal{G}$-measurable random variable, denoted by $E_u[X|\mathcal{G}]$, such that for any bounded, $\mathcal{G}$-measurable random variable H we have*

$$E[u(X)H] = E[u(E_u[X|\mathcal{G}])H]. \tag{1.4}$$

Proof. The proof follows the standard pattern. On the one hand, from [11, Theorem 23.4] it follows that when $u(X)$ is integrable, the conditional expectation $E[u(X)|\mathcal{G}]$ satisfies

$$E[u(X)H] = E[E[u(X)|\mathcal{G}]H]$$

for any bounded, $\mathcal{G}$-measurable random variable H. On the other hand, considering $u(\mathcal{I})$-valued random variables, there exists a $\mathcal{G}$-measurable random variable V such that $u(V)$ is integrable and

$$E[u(X)H] = E[u(V)H]$$

for any bounded, $\mathcal{G}$-measurable random variable H. Therefore, from the uniqueness part it follows that $V = u^{-1}(E[u(V)|\mathcal{G}])$. ■

Except for linearity in X, this notion satisfies the important properties of conditional expectation; in particular, if the conditioning σ-algebra is the trivial σ-algebra $\mathcal{F}_0 = \{\emptyset, \Omega\}$ then $E_u[X|\mathcal{F}_0] = E[X]$.

The following remark will play a role in the description of the confidence intervals of the best predictor. The prediction error in the transformed coordinates (or in the d_u-distance) is given by

$$\sigma_u^2(X) = \sigma^2(u(X)). \tag{1.5}$$

The following is standard and it is recalled here to establish notations.

DEFINITION 1.2. With the notations introduced above, for $0 < \alpha < 1$, we put

$$F^{-1}(\alpha) = q(\alpha). \tag{1.6}$$

Finally, we state and examine an essential assumption. Denote by $h_p(s)$ a strictly monotone function. Since h_p can be either increasing or decreasing, we require that either $h_p(0) = 0$ and $h_p(1) = 1$, or $h_p(0) = 1$ and $h_p(1) = 0$. This choice allows us to sweep many integrability issues under the carpet. With (1.2) in mind, the assumption is the following.

ASSUMPTION 1.1. Let $\mathcal{J} \subset \mathbb{R}$ be some open interval. For each $p \in \mathcal{J}$ let $h_p : [0,1] \to [0,1]$ be a continuous, strictly monotone function that maps end-points to end-points. We suppose that the mapping

$$p \mapsto h_p^{-1}\Big(\int_0^1 h_p(t)\,dt\Big)$$

is a bijection from $\mathcal{J}$ to $(0,1)$.

Note that if we consider coordinates h_p on $[0,1]$, and the metric defined by d_{h_p} as in (1.1), then, according to Theorem 1.1, the assumption asserts that for any $0 < \alpha < 1$, there is a $p \in \mathbb{R}$ such that $E_{h_p}[U] = \alpha$, where U is uniformly distributed in $[0,1]$. So, the import of the assumption will be that any quantile can be realized as a generalized expectation.

The rest of the paper is organized as follows. In Section 2 we present the basic characterization of quantiles as best predictors, that is, we show that for an appropriate choice of a parametric change of variables, the best predictor of X in the new distance is a quantile. Then we relate the tail conditional expectation to a quantile at a large confidence level. In Section 3 we consider two examples: first a change of variables inspired by the interpretation of h_p as a utility function, and then a different change of variables plus a numerical computation. In both cases we examine in detail the computation of the tail conditional expectation, and we include a short numerical example.

2. QUANTILES AND TAIL EXPECTED VALUES AS BEST PREDICTORS

The main result here is the following.

THEOREM 2.1. *Let X be a random variable with a strictly positive density, and put $F(x) = P(X \leqslant x)$. Let $h_p(s)$ be as in Assumption* 1.1 *and define $u_p(x) = h_p(F(x))$. Let $0 < \alpha < 1$ and let $p \in \mathcal{I}$ be such that*

$$h_p^{-1}\Big(\int_0^1 h_p(t)\,dt\Big) = \alpha.$$

Then the $q(\alpha)$ quantile is the best predictor of X in the d_{u_p}-distance given by (1.1)*, that is,*

$$E_{u_p}[X] = u_p^{-1}\big(E[u_p(X)]\big) = q(\alpha). \tag{2.1}$$

Proof. Theorem 1.1 asserts that

$$E_{u_p}[X] = u_p^{-1}\big(E[u_p(X)]\big) = F^{-1}\big(h_p^{-1}(E[u_p(X)])\big).$$

Since

$$E[u_p(X)] = \int_{\mathbb{R}} h_p(F(x))\,dF(x) = \int_0^1 h_p(s)\,ds,$$

invoking Assumption 1.1 we obtain $E_{u_p}[X] = q(\alpha)$, as claimed. ■

In other words, for each $0 < \alpha < 1$, there is a $p \in \mathbb{R}$ such that the α-quantile of X is the best predictor of X in the d_{u_p}-distance (1.1).

2.1. Two consistency issues. A natural question in the context of Theorem 1.1 is: How does the best predictor depend on the choice of the family h_p? Consider two possible parameterized coordinate systems defined by $h_p, g_q : \mathbb{R} \to \mathbb{R}$, both satisfying Assumption 1.1. The result is the following.

THEOREM 2.2. *For $p \in \mathcal{J}_1 \subset \mathbb{R}$ and $q \in \mathcal{J}_2 \subset \mathbb{R}$, let $h_p, g_q : \mathbb{R} \to (0,1)$ be two families of continuous bijections, both satisfying Assumption* 1.1*. Then, for each $p \in \mathcal{J}_1$ there is a $q \in \mathcal{J}_2$ (and conversely, for each $q \in \mathcal{J}_2$ there is a $p \in \mathcal{J}_1$) such that*

$$g_p \circ h_p^{-1}\big(E[h_p(U)]\big) = h_p \circ g_q^{-1}\big(E[g_p(U)]\big).$$

For $p \in \mathcal{J}_1$ set $\alpha = h_p^{-1}(E[h_p(U)])$. Then according to Assumption 1.1, there is a $q \in \mathcal{J}_2$ such that $g_p \circ h_p^{-1}(E[h_p(U)]) = h_p \circ g_q^{-1}(E[g_p(U)])$.

Sometimes it is of interest to express random variables in a scale, or a system of units, different from their natural units. Think of decibels or the Richter scale for example. In this case, we have to relate the distribution functions of the random variable in the different units to each other, as well as to other quantities of interest.

Let $K : \mathbb{R} \to \mathbb{R}$ be a bijection and let $Y = K(X)$, where X is a continuous random variable with (strictly increasing) distribution function $F(x)$. Then clearly the distribution $G(y)$ of Y satisfies $G = F \circ K^{-1}$. Similarly, for $0 < \alpha < 1$, the α quantiles $q_X(\alpha)$ and $q_Y(\alpha)$ are related by

$$q_Y(\alpha) = K(q_X(\alpha)). \tag{2.2}$$

Let us verify that this is consistent with the predictive approach.

THEOREM 2.3. *Let X be a random variable with a strictly positive density, and put $F(x) = P(X \leqslant x)$. Let $K : \mathbb{R} \to \mathbb{R}$ be a continuous bijection and let $Y = K(X)$.*

Let h_p be as in Assumption 1.1 and define $u_p(x) = h_p(F(x))$. Let $0 < \alpha < 1$ and $p \in \mathcal{I}$ be such that

$$h_p^{-1}\Big(\int_0^1 h_p(t)\,dt\Big) = \alpha \;\Leftrightarrow\; E_{u_p}[X] = q_X(\alpha).$$

Let $v_p(y) = h_p(G(y))$. Then

$$q_Y(\alpha) = K(q_X(\alpha)) = E_{v_p}[Y]. \tag{2.3}$$

Proof. Combining Theorem 2.1 and Assumption 1.1 we obtain $E_{u_p}[X] = q_X(\alpha)$. Note that from (2.2) we obtain

$$q_Y(\alpha) = K \circ u_p^{-1}(E[u_p(X)]) = K \circ F^{-1} \circ h_p^{-1} E[h_p \circ F \circ K^{-1} \circ K(X)] = E_{v_p}[Y]$$

after bringing in the definition of v_p. ■

This asserts that the predictive characterization of the quantiles of a random variable is consistent with the possible choice of units to measure it. An interesting variation on this theme goes as follows.

2.2. Tail conditional expectations. Suppose that X is a positive, continuous random variable, and that a confidence level $0 < \alpha < 1$ has been selected. Let $p \in \mathbb{R}$ be such that Assumption 1.1 holds. Then according to Theorem 2.1, we know that $E_{u_p}[X] = q(\alpha)$. Setting $F(X) = U$, we have

$$\{X > q(\alpha)\} = \{U > \alpha\}.$$

Therefore, invoking Theorem 1.2, we can establish the following result.

THEOREM 2.4. *Suppose that X is a positive random variable with a strictly increasing distribution function F. For $p \in \mathcal{J}$ related to α as in Assumption 1.1, and if $\mathcal{G}$ is the σ-algebra generated by the event $\{X > q(\alpha)\}$, then on the set $\{X > q(\alpha)\}$ we have*

$$\begin{aligned} E_{u_p}[X \mid X > q(\alpha)] &= F^{-1} \circ h_p^{-1}\Big(\frac{1}{1-\alpha} E[h_p(U); U > \alpha]\Big) \\ &= F^{-1} \circ h_p^{-1}\Big(\frac{1}{1-\alpha} \int_\alpha^1 h_p(t)\,dt\Big). \end{aligned} \tag{2.4}$$

The following approximation is interesting.

COROLLARY 2.1. *Under the assumptions and notations of Theorem* 2.4 *we have*

$$E_{u_p}[X \mid X > q(\alpha)] > q(\alpha). \tag{2.5}$$

Furthermore, if h_p is strictly increasing and both h_p and F have two continuous derivatives, then, for α close to 1,

$$E_{u_p}[X \mid X > q(\alpha)] = q(\alpha) + \frac{1}{2}K(\alpha)(1-\alpha) + O((1-\alpha)^2). \tag{2.6}$$

Here $K(\alpha) = 1/f(q(\alpha))$ and $f(x) = F'(x)$ is the probability density of X.

Proof. Since $E_{u_p}[X \mid X > q(\alpha)] = u_p^{-1}(E[u_p(X) \mid X > q(\alpha)])$, we have

$$\begin{aligned} E[u_p(X) \mid X > q(\alpha)] &= \frac{1}{1-\alpha}\int_{q(\alpha)}^{\infty} h_p(F(X))\,dF(x) = \frac{1}{1-\alpha}\int_{\alpha}^{1} h_p(t)\,dt \\ &= h_p(\alpha) + \frac{1}{1-\alpha}\int_{\alpha}^{1}(h_p(t) - h_p(\alpha))\,dt. \end{aligned}$$

To bound the second term, denote by B a bound for h_p' on $[0,1]$ and note that

$$\begin{aligned} \frac{1}{1-\alpha}\int_{\alpha}^{1}(h_p(t) - h_p(\alpha))\,dt &+ \frac{1}{1-\alpha}\int_{\alpha}^{1}\Big(\int_{\alpha}^{t} h_p'(s)\,ds\Big)\,dt \\ &= \frac{1}{1-\alpha}\int_{\alpha}^{1} h_p'(s)\Big(\int_{s}^{1} dt\Big)\,ds \\ &= \frac{1}{2}h_p'(\alpha)(1-\alpha) + \frac{1}{1-\alpha}\int_{\alpha}^{1}(h_p'(s) - h_p'(\alpha))(t-s)\,ds. \end{aligned}$$

The last term is bounded by $2B(1-\alpha)$.

Now, bring in the fact that $u_p^{-1}(h_p(\alpha)) = q(\alpha)$ and make use of the Taylor expansion up to the second order to obtain

$$E_{u_p}[X \mid X > q(\alpha)] = q(\alpha) + \frac{1}{2f(q(\alpha))}(1-\alpha) + O((1-\alpha)^2),$$

as asserted. ∎

The approximate calculation in (2.6) is particularly interesting for risk management. A quantity of interest there is the expected shortfall or TVaR, which is defined as the expected loss beyond the value at risk, or VaR, which is given by $q(\alpha)$. For the use of this concept in risk management, consider [10].

3. TWO WORKED OUT EXAMPLES

3.1. First example. Consider first the parametric family of functions $h_p : (0,1) \to (0,1)$, $p \in (-\infty, +\infty)$, given by

$$h_p(s) = \begin{cases} \dfrac{1-e^{-ps}}{1-e^{-p}} & \text{for } p \neq 0, \\ s & \text{for } p = 0. \end{cases} \tag{3.1}$$

The typical cases look as in Figure 1

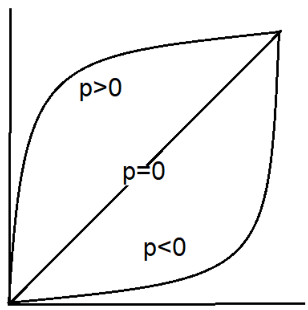

FIGURE 1. Typical $h_p(u)$ with $u \in [0,1]$

To begin with, we have:

PROPOSITION 3.1. *Let* $\{h_p(s) : p \in \mathbb{R}\}$ *be the family of functions defined in* (3.1). *Then*

(1) $h_p \to 1$ *pointwise and boundedly as* $p \to \infty$,

(2) $h_p \to 0$ *pointwise and boundedly as* $p \to -\infty$.

If we put $K(p) = (1-e^{-p})^{-1}$ *for* $p \in \mathbb{R}\backslash\{0\}$, *then*

$$I(p) = \int_0^1 h_p(s)\,ds = K(p) - \frac{1}{p} \quad \textit{for } p \neq 0 \quad \textit{and} \quad I(0) = \frac{1}{2}.$$

Then

$$\begin{aligned} I(p) &\to 1 && \textit{as } p \to \infty, \\ I(p) &\to \frac{1}{2} && \textit{as } p \to 0, \\ I(p) &\to 0 && \textit{as } p \to -\infty. \end{aligned} \tag{3.2}$$

Clearly, $p \mapsto I(p)$ *is continuous and increasing.*

Note that even though $K(p) - \frac{1}{p}$ cannot be evaluated at $p = 0$, nevertheless $K(p) - \frac{1}{p} \to 1/2$ as $p \to 0$, therefore, $I(p) = K(p) - \frac{1}{p}$ is indeed continuous at $p = 0$. This can be easily seen, for example, by writing it as $((1-e^{-p})^{-1} - 1)/p$ and using L'Hôpital's rule. Next, differentiate the numerator and the denominator with respect to p, multiply and divide by p^2 to obtain

$$\frac{p^2}{(1-e^{-p})^2}\,\frac{1-e^{-p}-pe^{-p}}{p^2}.$$

Now, expand the exponential up to order 2 to complete the verification.

It is easy to verify that $h_p^{-1}(t) = -\frac{1}{p}\ln(1 - t/K(p))$ for $t \in (0,1)$. Clearly, $t \mapsto h_p^{-1}(t)$ increases continuously from 0 to 1. Therefore, for $p \neq 0$,

$$h_p^{-1}\Big(\int_0^1 h_p(s)\,ds\Big) = -\frac{1}{p}\ln\Big(\frac{1-e^{-p}}{p}\Big)$$

also increases continuously from 0 to 1 as p ranges over $\mathbb{R}$.

Thus, the assumptions in Theorem 2.1 are met and we gather the comments made above as:

THEOREM 3.1. *With the notations introduced above, for every* $\alpha \in (0,1) \setminus \{1/2\}$ *there is a* $p \in \mathbb{R} \setminus \{0\}$ *such that*

$$\frac{1}{p}\ln\Big(\frac{p}{1-e^{-p}}\Big) = \alpha. \tag{3.3}$$

Let $u_p(x) = h_p(F(x))$ *with* h_p *as in* (3.1). *Then*

$$E_{u_p}[X \mid X > q(\alpha)] = F^{-1}\bigg(\alpha + \ln\bigg[\Big(\frac{(1-\alpha)p}{1-e^{-(1-\alpha)p}}\Big)^{1/p}\bigg]\bigg). \tag{3.4}$$

The first assertion follows from a simple computaion. To verify the second, let $u_p(x) = h_p(F(x))$ with h_p be as in (3.1). Then, according to Theorem 2.1, the best predictor of X in the d_{u_p}-distance is given by

$$E_{u_p}[X] = F^{-1}\Big(\frac{1}{p}\ln\Big(\frac{p}{1-e^{-p}}\Big)\Big) = F^{-1}(\alpha) = q(\alpha).$$

Now, from (2.4) and some arithmetic manipulations, using (3.3) along the way, we also have

$$\begin{aligned} E_{u_p}[X \mid X > q(\alpha)] &= F^{-1}\Big(\frac{1}{p}\ln\Big(\frac{(1-\alpha)p}{e^{-\alpha p}-e^{-p}}\Big)\Big) \\ &= F^{-1}\bigg(\ln\bigg[e^{\alpha}\Big(\frac{(1-\alpha)p}{1-e^{-(1-\alpha)p}}\Big)^{1/p}\bigg]\bigg) \\ &= F^{-1}\bigg(\alpha + \ln\bigg[\Big(\frac{(1-\alpha)p}{1-e^{-(1-\alpha)p}}\Big)^{1/p}\bigg]\bigg). \end{aligned} \tag{3.5}$$

We mention as well that if $p = 0$ (i.e. $h_0(s) = s$), then $u_0(x) = F(x)$, therefore $E_{u_0}[X] = F^{-1}(1/2) = q(1/2)$. Thus, the median is a generalized best predictor in the distance associated to the change of variables defined by the distribution function.

Since each h_p maps $[0, 1]$ onto $[0, 1]$ bijectively, instead of h_p we might have started with $g_q(t) = N(q)\ln(1-qt)$, where $N(q) = (\ln(1-q))^{-1}$, the relationship between the parameters being $q = 1 - e^{-p}$.

3.2. Second example. For this example we consider $h_p(t) = t^p$ with $p \in (0, \infty)$. In this case, Assumption 1.1 amounts to saying that given $0 < \alpha < 1$, the equation $(p + 1)^{-1/p} = \alpha$ has a solution in p. That such a p can be determined is clear from the fact that the map $p \mapsto (p + 1)^{-1/p}$ from $(-1, \infty) \to (0, 1)$ is continuous and strictly increasing, tends to 0 as $p \to -1$ and to 1 as $p \to \infty$. Even though this map is defined for every p in that range, the function $u(x) = F^p(x)$ is just a constant for $p = 0$. Then $(p + 1)^{-1/p} \to e^{-1}$ as $p \to 0$. The corresponding quantile has to be obtained as a limit. When p crosses level 0, the function u goes from increasing to decreasing (or vice versa), a kind of bifurcation in the function space.

In the current example, the quantile as expected value in the d_{u_p}-metric is

$$E_{u_p}[X] = F^{-1}\big((E[F^p(X)])^{1/p}\big) = F^{-1}\left(\left(\frac{1}{p+1}\right)^{1/p}\right) = F^{-1}(\alpha) = q(\alpha).$$

Again, to study tail conditional values, there are two separate cases of interest in risk management: The variable X may denote returns, which are usually assumed to range in $(-\infty, \infty)$. In this case, the interest is in the values in the left tail of the distribution. Sometimes, e.g. in insurance or in reliability theory, X is supposed to be strictly positive and denotes pure losses or costs. In this case, the interest is in values in the right tail of the distribution. The two definitions of interest go as follows.

DEFINITION 3.1. When X ranges in $(-\infty, \infty)$ the *value at risk* (VaR) *at confidence level* $0 < \alpha < 1$ is defined to be

$$\text{VaR}_\alpha(X) = -q(1-\alpha) = -\inf\{x : F(x) \geqslant 1-\alpha\}.$$

It is interpreted as the largest loss with probability $1-\alpha$. The sign is due to the fact that one thinks of losses as positive numbers. In this case, $P(X \leqslant q(1-\alpha)) = 1-\alpha$.

When X is strictly positive, the corresponding definition is

$$\text{VaR}_\alpha(X) = q(\alpha) = \inf\{x \mid F(x) \geqslant \alpha\}.$$

Here the probability of a loss larger than $q(\alpha)$ is $P(X > q(\alpha)) = 1 - \alpha$.

We now use these definitions in conjunction with Theorem 2.4 when $h_p(s) = s^p$. We have to consider two cases separately. When X assumes any value in $(-\infty, \infty)$, for the purpose of risk management one is interested in $-E_{u_p}[X \mid X < q(1-\alpha)]$, and when X is strictly positive one is interested in $E_{u_p}[X \mid X > q(\alpha)]$. The first step in the computation of these quantities is as follows:

$$\begin{aligned} E[F^p(X) \mid X < q(1-\alpha)] &= \frac{1}{1-\alpha} \int_{-\infty}^{q(1-\alpha)} F^p(x)\, dF(x) \\ &= \frac{1}{1+p} \frac{F^{p+1}(q(1-\alpha))}{(1-\alpha)}, \\ E[F^p(X) \mid X > q(\alpha)] &= \frac{1}{1-\alpha} \int_{q(\alpha)}^{\infty} F^p(x)\, dF(x) \\ &= \frac{1}{1+p} \frac{1 - F^{p+1}(q(\alpha))}{(1-\alpha)}. \end{aligned}$$

Using the fact that $F(q(\alpha)) = \alpha$, $F(q(1-\alpha)) = 1-\alpha$ and $(p+1)^{-1} = \alpha^p$, the above identities can be written as

$$\begin{aligned} E[F^p(X) \mid X < q(1-\alpha)] &= \alpha^p \frac{(1-\alpha)^{p+1}}{(1-\alpha)} = \alpha^p (1-\alpha)^p, \\ E[F^p(X) \mid X > q(\alpha)] &= \alpha^p \frac{1-\alpha^{p+1}}{(1-\alpha)}. \end{aligned}$$

Using Theorem 2.4, the expected losses beyond the VaR computed above are characterized as a quantile with a larger level of confidence. The formal result is stated as follows.

THEOREM 3.2. *With the notations introduced above, the expected loss given that it is larger than the VaR is given by*

$$\begin{aligned} E_{u_p}[X \mid X < q(1-\alpha)] &= F^{-1}\big(E[F^p(X) \mid X < q(1-\alpha)]\big)^{1/p} \\ &= F^{-1}(\alpha(1-\alpha)), \end{aligned} \tag{3.6}$$

$$\begin{aligned} E_u[X \mid X > q(\alpha)] &= F^{-1}\big(E[F^p(X) \mid X > q(1-\alpha)]\big)^{1/p} \\ &= F^{-1}\left(\alpha\left(\frac{1-\alpha^{p+1}}{1-\alpha}\right)^{1/p}\right). \end{aligned} \tag{3.7}$$

3.2.1. *Some numerical computations of VaR and TVaR.* A short list of (α, p) values near the end of the range is presented in Table 1. The case $p = 1$ corresponds to $\alpha = 1/2$, and we obtain a characterization of the median as the best predictor in a different metric.

In the next table we list the arguments of F^{-1} for the right-hand side of (3.6) and (3.7) for the values of α considered above. To compute the TvaR we need new

TABLE 1. Pairs of a and p values up to three decimal places

α	0.01	0.05	0.1	0.9	0.95	0.99
p	−0.989	−0.940	−0.863	33.520	85.578	605.730

TABLE 2. Values of the shifted levels of confidence

α	RHS of (3.6)	α	RHS of (3.7)
0.01	0.099	0.90	0.963
0.05	0.0475	0.95	0.984
0.10	0.099	0.99	0.998

confidence levels. A list of such modified levels of confidence, corresponding to confidence levels typical in risk analysis, is collected in Table 2.

Next we present an explicit numerical example. We have computed quantiles of a few densities at the original level of confidence (α), and at the shifted level of confidence (denoted below by α^*); see Table 2. From these, the values of VaR and TVaR can be read off by changing the signs whenever appropriate. To compute the quantiles listed below, a sample of size 100,000 was generated. The parameters of several densities are displayed in Table 3:

TABLE 3. List of densities and their parameters

Density	Parameters
Normal density	$\mu = 10.000$, $\sigma = 5\mu$
Student-t:	3 degrees of freedom
Lognormal	meanlog = 1.05, sdlog= 3×meanlog
Gamma	shape= 500, scale= 100
Weibull	shape= 500, scale= 8

In Table 4 we list the α-quantile for several levels of confidence typically used in risk management.

Recall that if a random value takes negative values, and the lower quantile is negative, VaR is obtained by changing its sign, to report losses as positive quantities.

We now know that to calculate the expected loss given that it is larger than VaR, it suffices to compute a quantile at a shifted level of confidence. The new level of confidence for a few standard levels of confidence used in risk analysis is given in Table 2, and the corresponding quantiles for the same densities are listed in Table 5.

As commented above, when the random variable takes negative values, VaR and TVaR are the negatives of the corresponding quantiles. Clearly, TVaR $>$ VaR in the examples.

TABLE 4. Quantiles for the list of densities

α	Normal	Lognormal	Gamma	Student-T	Weibull
1%	−106441.67	0.0019878	4.492268	−4.571527	7.926742
5%	−72430.71	0.0162146	4.637273	−2.361576	7.952440
10%	−54716.53	0.0505959	4.715295	−1.647299	7.963973
90%	73463.80	161.7767644	5.288336	1.640524	8.013412
95%	91669.15	522.0316801	5.371384	2.386815	8.017688
99%	125341.32	4523.9659797	5.531109	4.588648	8.024608

TABLE 5. Shifted confidence levels and their quantiles

α^*	Normal	Lognormal	Gamma	Student-t	Weibull
0.9%	−108006.81	1.559200e-03	4.487121	−4.744213	7.924012
4.75%	−72885.06	1.437560e-02	4.631312	−2.401811	7.951761
9.8%	−54392.05	4.717850e-02	4.711479	−1.653442	7.963614
96.3%	98841.60	8.123035e+02	5.407141	2.686339	8.019081
98.4%	117429.44	2.430261e+03	5.490509	3.773217	8.022600
99.8%	151822.82	2.730788e+04	5.664836	8.041184	8.028933

4. CONCLUDING REMARKS

There exist many ways of presenting quantiles as best predictors in metrics induced by coordinate transformations on the range of random variables. The change of variables is up to the modeler. The only restriction in our approach is that Assumption 1.1 holds. In particular, the two examples considered point to an interesting conceptual issue. In our setup, given a number ξ in the range of the variable, there is an $\alpha(\xi)$ such that $F^{-1}(\alpha(\xi)) = \xi$. Under Assumption 1.1, there is a $p(\xi)$ such that $E_{u_{p(\xi)}}[X] = \xi$.

It is worth stressing that the fact that conditional expectation given some tail events can be expressed as a quantile at a larger confidence level, makes it interesting for risk analysts, especially because the error in estimating quantiles may be easier to estimate than the error in estimating tail conditional expectations.

Acknowledgments. I thank Joel Cohen for having posed the question leading to this work. I also thank Érika Gomes-Gonçalves for computing the entries of the tables in Section 3. Also, I want to thank the referees and the editorial team for their thorough reading of the manuscript and multiple suggestions for improving it.

REFERENCES

[1] R. L. Berger and G. Casella, *Deriving generalized means as least squares and maximum likelihood estimates*, Amer. Statist. 46 (1992), 279–282.

[2] C. E. Bonferroni, *La media esponenziale in matematica finanziaria*, Annuario del Regio Istituto Superiore di Scienze Economiche e Commerciali di Bari 1923–1924, 1–14.
[3] M. de Carvalho, *Mean, what do you mean?*, Amer. Statist. 70 (2016), 270–274.
[4] B. de Finetti, *Sul concetto di media*, Giornale dell'Istituto Italiano degli Attuari 2 (1931), 369–396.
[5] L. Eeckhoudt, C. Gollir, and H. Schlesinger, *Economic and Financial Decisions under Risk*, Princeton Univ. Press, Princeton, NJ, 2005.
[6] M. Emery and G. Mokobodzki, *Sur le barycentre d'une probabilité dans une variété*, in: Séminaire de Probabilités XXV, Lecture Notes in Math. 1485, Springer, Berlin, 1991, 220–233.
[7] T. Gneiting, *Quantiles as optimal point predictors*, Int. J. Forecasting 27 (2009), 197–207.
[8] H. Gzyl, *Prediction in Riemannian metrics derived from divergence functions*, Comm. Statist. Theory Methods 51 (2022), 552–568.
[9] R. Koenker and G. Basset Jr., *Regression quantiles*, Econometrica 46 (1978), 33–50.
[10] A. J. McNeil, R. Frey, and P. Embrechts, *Quantitative Risk Analysis*, Princeton Univ. Press, Princeton, NJ, 2005.
[11] J. Jacod and P. Protter, *Probability Essentials*, Springer, Berlin, 2000.
[12] J. M. Oller and J. M. Corcuera, *Intrinsic analysis of statistical estimation*, Ann. Statist. 23 (1995), 1562–1581.

Henryk Gzyl
IESA School of Business
Caracas, Venezuela
E-mail: henryk.gzyl@iesa.edu.ve

Received 6.9.2023;
accepted 7.5.2024